PennIUR
Penn Institute for Urban Research

Federal Reserve Bank of Philadelphia

WORKING PAPER PREPARED FOR **REINVENTING OUR COMMUNITIES: TRANSFORMING OUR ECONOMIES**
FEDERAL RESERVE BANK OF PHILADELPHIA BIENNIAL CONFERENCE, SEPTEMBER 21-23, 2016

Education Reform in the Post-NCLB Era:

Lessons Learned for Transforming Urban Public Education

BY MATTHEW P. STEINBERG AND RAND QUINN
Matthew P. Steinberg, Assistant Professor, Graduate School of Education, University of Pennsylvania
Rand Quinn, Assistant Professor, Graduate School of Education, University of Pennsylvania.

SEPTEMBER 2016

Introduction

In the previous 15 years, federal policy reforms have played an increasingly important role in shaping the landscape of local public education in the United States. Indeed, the 2001 federal No Child Left Behind Act (NCLB) codified for the first time accountability standards for all U.S. public schools, requiring that all students perform at academic proficiency levels by 2014. NCLB mandated that all students in grades three through eight (and once in high school) be tested in mathematics and reading; though, states were responsible for writing both the learning standards and creating the accountability exams to assess student proficiency based on state-determined standards. For those schools that consistently failed to make adequate yearly progress toward meeting these proficiency standards, NCLB included accountability provisions that sanctioned schools in a variety of ways. Among these sanctions were the provision of school choice—granting students in persistently failing schools the option to transfer to higher-performing schools—and the option to receive free after-school tutoring (funded by federal Title I aid) through NCLB's Supplemental Educational Services provision. Schools that were chronically underperforming were threatened with turnaround or closure. More recently, the 2009 federal Race to the Top (RTTT) grant program incentivized a number of policy reforms at the state and district levels. Notable among these reforms was the implementation of common state education standards, referred to as the Common Core State Standards, and revisions to state and local human capital policies around educator (i.e., teacher and principal) evaluation.

December 2015 marked a turning point in the federal role in education policy. The reauthorization of the Elementary and Secondary Education Act, now referred to as the Every Student Succeeds Act (ESSA), grants states (and districts) newfound autonomy to revise accountability, testing, and educator evaluation policies. Given new discretion over education reform, it is imperative that state and local school leaders be empowered with evidence-based policies and practices that have been shown to improve the educational lives of students. This is particularly important for school leaders in our largest districts located in urban metropolitan centers, which serve among the most economically disadvantaged students. To provide information to policymakers and practitioners serving students in urban school districts, we document examples of policy reforms in urban school settings that show promise as effective strategies for improving school, teacher, and/or student outcomes. We attend to four topics that have received much policy and research attention in the previous 15 years; including: (a) investments in early childhood education; (b) human capital policies; (c) accountability, standards, and assessment; and (d) market-based reforms and school choice.

We focus our discussion on empirical evidence produced through rigorous analysis that lends itself to causal conclusions about the impact of education practices and policy reforms. Such evidence is critical for guiding the decisions of policymakers and urban leaders. While carefully conducted randomized control trials remain the gold standard for social policy analysis and inference about the efficacy of policy, we recognize that many education policies and practices of interest do not lend themselves to such experimental designs. Therefore, we also discuss evidence produced through rigorous quasi-experimental studies (such as regression discontinuity designs and studies relying on panel micro-data). We omit evidence from the many studies that provide correlational evidence on the efficacy of policy reforms in education.

Can education policy reforms improve the lives of students in the nation's urban schools? In the following pages, we document evidence of policies and practices that have and have not worked to improve educational outcomes for teachers and students. Our aim is that this evidence will inform the ongoing work of policymakers and school leaders to enhance the quality of schooling for all urban school students.

Early Childhood Education

Racial and ethnic differences in academic performance have been identified as far back as the Coleman Report five decades ago (Coleman et al, 1966). More recent evidence on student achievement has revealed a persistent gap in the academic performance of minority students compared to their white counterparts, a gap that emerges early in the educational careers of students (Fryer and Levitt, 2004). Indeed, minority students enter kindergarten performing below their white peers, and the achievement gap grows as students progress through the early elementary years (Fryer and Levitt, 2006). Moreover, as income inequality in the United States has increased in recent years, so too has the achievement gap between students that come from high- and low-income families (Reardon, 2011).[1] The persistent race gap in student achievement coupled with a growing achievement gap based on family income poses a particular concern for minority families in urban communities. Indeed, residential and economic segregation exist in most urban communities, and, left unaddressed, will continue to disadvantage the educational prospects of students from very early ages.

Recent attention, however, in both academic and policy circles has been given to the role early childhood education can play in preparing economically disadvantaged students for school and narrowing the achievement gap. For example, economist James Heckman has written widely on the economic benefits of investments in early childhood education as a means for reducing educational inequality (Heckman, 2011). Federal policy efforts such as the Obama administration's Race to the Top-Early Learning Challenge has provided grants to states to improve the quality of early education programs provided to among the most economically disadvantaged children. At the local level, cities such as New York and Philadelphia have recently made dramatic investments to provide greater access to high-quality pre-Kindergarten (pre-K) programs.[2] Given these recent efforts to expand access to pre-K, what effect might the provision of universal pre-K have on students' academic success prior to entering elementary school? The experience of urban districts in Tulsa and Boston provide insight into this question.

EVIDENCE FROM TULSA, OKLAHOMA

In 1998, the state of Oklahoma established a universal pre-K program for four-year-old children, under which each of the state's 543 districts could choose to participate. A key feature of the state's program is the requirement that all pre-K teachers have a college degree and a certificate in early childhood education. Moreover, these pre-K teachers received compensation equal to teachers in the state's public elementary schools. Class sizes are capped at 20 students with a student/staff ratio of 10/1 (Gormley and Gayer, 2004). In a series of studies, Gormley and colleagues examined the impact of pre-K in the state's largest urban school district—Tulsa Public Schools—on students' cognitive development and school readiness, with a particular focus on the performance of racial and ethnic minority students.

In a series of studies, Gormley and Gayer (2004) rely on a regression discontinuity design, which leverages the strict age cutoff required to qualify for the pre-K program. The authors finds that Tulsa's pre-K program provides substantive benefits to eligible four-year-old children, including significant improvements in cognitive, language and motor skill development.[3] The authors further find that the positive benefits of pre-K are greatest for Hispanic and black children, with more limited benefits for white children. Economically disadvantaged children—those who qualify for free lunch—gained more from Tulsa's pre-K program than their more

1 Reardon (2011) defines high- and low-income families as families at the 90th and 10th percentiles of the income distribution, respectively. Reardon finds that the achievement gap between children from high- and low-income families is approximately 30-40 percent greater among children born in 2001 than among children born 25 years earlier.

2 In New York City, nearly 70,000 children were enrolled in free full-day pre-K as of December 2015, an increase of approximately 50,000 enrollees in less than two years (source: http://www1.nyc.gov/office-of-the-mayor/news/954-15/mayor-de-blasio-over-68-500-students-enrolled-pre-k-all). In Philadelphia, a tax on sugar-sweetened beverages—the first such tax in any major U.S. city—was passed in June 2016, the proceeds from which will be used to expand access to early childhood education.

3 Gormley and Gayer (2004) do not find evidence that Tulsa's pre-K program impacted children's socioemotional development.

economically advantaged counterparts. In a second study, Gormley and colleagues (2005) find that the program positively and significantly impacted children's performance on cognitive tests of prereading and reading skills, prewriting and spelling skills, and math reasoning and problem-solving abilities. In a third study, Gormley (2008) specifically examines the impact of Tulsa's pre-K program among Hispanic students, and finds that Hispanic children realized substantial improvements in prereading, prewriting, and premath skills, and that children whose parents spoke Spanish at home or whose parents were born in Mexico realized the greatest benefit of Tulsa's program.

EVIDENCE FROM BOSTON

In the 2008-09 school year, Boston Public Schools (BPS) universal pre-K program provided services to approximately 2,000 children, about 34 percent of the city's four-year-old children. Unlike many pre-K programs in other districts and states, access to Boston's program was not limited by family income (Weiland and Yoshikawa, 2013). Like Tulsa's program, all BPS pre-K teachers are subject to the same educational requirements and compensation scale as K-12 teachers. All pre-K teachers must have at least a bachelor's degree and must obtain a master's degree within five years of teaching in the program.[4] Of particular note is the extent of support given to BPS pre-K teachers, including curriculum-specific training and weekly to biweekly on-site support from an experienced early childhood coach trained in the program's curricula.

To study the impact of BPS' pre-K program, Weiland and Yoshikawa (2013) employ a regression discontinuity approach that relied on the age cutoff for eligibility into Boston's pre-K program (a strategy used by Gormley and colleagues in their evaluation of Tulsa's program). The authors find large positive effects on children's language, literacy, numeracy and math skills, with smaller impacts on children's executive functioning and emotional development. Like Tulsa, economically disadvantaged students benefitted the most from BPS' program.

The pre-K experiences in Tulsa and Boston offer important insights into critical success factors for early educational interventions. Both Tulsa and Boston set teacher qualification standards for pre-K teachers, requiring teachers to hold a bachelor's degree and, in the case of Boston, mandating that teachers actively seek additional education. In doing so, pre-K teachers in both locations are treated as education professionals, compensated at scales equivalent to their district's K-12 teaching counterparts. This is in comparison to many pre-K alternatives for children—such as daycare and homecare—where the educational and training requirements of caregivers are far more limited. While there is no direct evidence on the impact that professional supports may have in these settings, Boston provides ongoing professional development to its pre-K teachers; we later discuss the efficacy of professional support and instructional feedback in K-12 education settings. Likewise, there is no evidence in the Tulsa case on the impact of limiting class sizes; though, experimental evidence from early elementary school settings finds that maintaining modest student-teacher ratios can lead to large achievement gains for students (Krueger, 1999). Taken together, these features of the pre-K programs in Tulsa and Boston provide important guidance for policymakers who are considering offering high-quality pre-K experiences for their district's children.

Teacher Human Capital

Teachers play a critical role in the life of schools. Indeed, research has long identified that teachers are the most important within-school input to the production of education—that is, improving the educational lives of students. Under NCLB, the importance of placing qualified teachers in every public school in the United

4 According to Weiland and Yoshikawa (2013), in the 2008-09 school year, 78 percent of pre-K program teachers held a master's degree and 75 percent had at least five years of teaching experience.

States was codified into law for the first time. NCLB's "highly effective teacher" provision represented the first national legislative effort to set teacher quality benchmarks. In doing so, NCLB required states to ensure that all teachers were "highly qualified," defining teacher quality by a teacher's credentials—receipt of a bachelor's degree, state certification or licensure, and proof of content-area expertise. Notably, NCLB characterized teacher quality based on inputs, or pre-service teacher qualifications. In the following sections, we examine aspects of the teacher labor supply – teacher qualifications; alternative pathways into the teaching profession; teacher induction; teacher performance evaluation; and performance pay schemes. We assess the extent to which policies and program initiatives generated improvements in both teacher performance and student achievement.

TEACHER QUALIFICATIONS

The urgency of identifying and placing highly-qualified teachers in every classroom is no greater than in our urban public schools. But do observable characteristics of teachers at the time of hire provide information to predict a teacher's subsequent effectiveness in urban classrooms? A recent study in Los Angeles Unified School District (LAUSD) examined whether teacher effectiveness—a teacher's contribution to student achievement growth—depends on either of two teacher inputs: a teacher's success on licensure exams or advanced degrees beyond a bachelor's degree (Buddin and Zamarro, 2009). Evidence from LAUSD indicates that a teacher's performance on licensure tests do not lead to success in the classroom; in particular, student achievement is not significantly affected by the basic skills, subject matter, or reading pedagogy skills of their teachers as measured on current California licensure tests. Moreover, student achievement is unaffected by whether classroom teachers have advanced degrees (Buddin and Zamarro, 2009).

If a teacher's performance on licensure exams and receipt of advanced degrees do little to improve student achievement, can other measures predict teacher effectiveness in urban schools? That is, to what extent might additional information about teachers, beyond traditional pre-service qualifications, improve urban districts' capacity to identify effective teachers at time of hire? During the 2006-07 school year, Rockoff et al. (2011) administered a survey to new math teachers in New York City teaching in grades four through eight. The authors, along with district officials, collected information about these new teachers not typically included in teacher personnel records at the time of hire, including teaching-specific content knowledge, cognitive ability, personality traits, feelings of self-efficacy, and scores on a commercially available teacher selection instrument. Combining information on teachers typically found in district personnel records (e.g., degree attainment, licensure scores, college major) with survey responses, Rockoff and colleagues construct two aggregate measures—one capturing a teacher's cognitive skills and a second capturing a teacher's noncognitive skills[5]—and find that both are highly predictive of student math achievement. These results from New York City suggest that districts should consider incorporating additional information about teacher applicants when making hiring decisions. Of course, more work needs to be done to identify which teacher quality traits, beyond those traditionally captured in personnel records, may provide better quality signals about a teacher's effectiveness. Moreover, to more carefully estimate the economic benefits of collecting such nontraditional teacher quality information, a closer accounting of the costs associated with collecting new information about teacher candidates must be weighed against the potential benefits of hiring more qualified teachers who are better able to improve student outcomes. Doing so will provide insight into whether investments in information gathering are truly worthwhile.

5 The six variables contributing to their measure of cognitive skills include a teacher's Teach for America (TFA) corps status, attending a more selective college, SAT math and verbal scores, cognitive ability as measured by the Raven's test (an intelligence test that requires no linguistic or mathematics skills), and a measure of mathematics knowledge for teaching. The five variables contributing to their measure of noncognitive skills include extraversion, conscientiousness, personal efficacy, general efficacy, and the Haberman total score. Notably, majoring in education in college was negatively associated with the aggregate cognitive skills measure.

ALTERNATIVE PATHWAYS INTO TEACHING

In addition to the inclusion of nontraditional measures of teacher quality into teacher hiring decisions, do nontraditional pathways into the teaching profession offer promise as an alternative approach to improving the teacher labor supply? Likely the most prominent among the many alternative pathways into the teaching profession is Teach For America (TFA), a national teacher recruitment program that places new college graduates into hard-to-staff schools (i.e., those with teacher supply shortages) in among the most economically disadvantaged urban (and rural) communities in the United States. TFA offers pre-service training to all corps members via a summer institute.[6] Once TFA teachers are placed in their schools, they receive ongoing support from TFA staff and faculty in the form of classroom observations, connection to resources that meet their particular professional development needs, and connection to TFA colleagues as a support network.

Decker et al. (2004) have conducted, to date, the most rigorous evaluation of the effectiveness of TFA teachers located in urban schools. Between 2001 and 2003, the authors studied the impact of TFA teachers on student achievement in Baltimore, Chicago, Los Angeles, Houston, and New Orleans. To do so, the authors randomly assigned students to classrooms in order to ensure that the TFA and control teachers (any non-TFA teacher) taught classes with, on average, identical student composition. Control teachers included any non-TFA teacher, whether traditionally certified, alternatively certified, or uncertified. TFA teachers included any teacher who entered the profession through TFA, whether current corps members in their first two years or alumni still in the profession. Notably, while the TFA teachers had strong academic backgrounds (i.e., selective colleges and high GPAs), on average they had less pre-service teacher training than the control teachers in the same schools. Moreover, TFA teachers were less likely to have education degrees, to be fully certified, or to have substantial student teaching experience prior to entering the classroom.

Results from Decker et al. (2004) indicate that the math achievement of students to whom TFA teachers were randomly assigned improved significantly more than students randomly assigned a control teacher (an effect size of 0.15). However, there was no evidence that reading achievement improved differently for students of TFA teachers compared to control teachers.[7] Notably, these estimates do not disentangle the effect of TFA recruitment on the type of teacher who enters the profession from the effect of TFA training. Moreover, given that TFA teachers tend to exit the teaching profession at much higher rates than teachers entering through more traditional pathways, the benefits to student achievement of a TFA teacher must be considered in light of the costs of replacing a TFA teacher due to atypically high turnover.

TEACHER INDUCTION

While certain recruitment and hiring strategies appear to offer promise in terms of improving the teacher labor supply, what approaches may exist to improve teacher human capital after teachers have entered the classroom? For example, does on-the-job training and support for novice teachers in the form of induction reduce teacher turnover and improve student performance? Glazerman et al. (2010) studied the impact of a comprehensive teacher induction program in 17 urban school districts. Schools within districts were randomly assigned to treatment or control conditions. The aim of this study was to examine whether additional induction services beyond the services districts usually provide to support beginning teachers (i.e., status quo teacher induction) improves teacher and student outcomes. Districts selected one of two vendors—Educational Testing

6 The summer institute incorporates four activities, including: six formal education courses; full teaching responsibility for a class of summer school students; weekly meetings of institute learning teams focused on teaching methods (organized according to subject and grade level; led by institute staff, with a focus on content and grade-specific teaching methods); and content- and grade-specific workshops. The total weekly workload in the summer institute is roughly 70 hours per week (Decker et al., 2004).

7 The positive im

pact on math outcomes for students, but not for reading test scores, is a consistent finding among many education interventions. This result, in general, can be attributed to differences in teaching math versus reading skills, the latter of which are more closely correlated with factors outside of the classroom than the former.

Service (ETS) or the New Teacher Center (NTC)—to provide comprehensive induction in schools assigned to the treatment group. The comprehensive induction included carefully selected and trained full-time mentors; a curriculum of intensive and structured support for beginning teachers, including orientation, professional development opportunities, and weekly meetings with mentors; a focus on instruction, with opportunities for novice teachers to observe experienced teachers; formative assessment tools used to evaluate teacher practice on an ongoing basis and used to observe and provide feedback to teachers; and outreach to district- and school-based administrators to educate them about program goals and to garner their systemic support for the program.

At the end of the first year of comprehensive induction, the additional support for treatment teachers did not translate into impacts on classroom practices.[8] For teachers who received one year of comprehensive induction, receipt of comprehensive induction did not impact student achievement. For teachers who received two consecutive years of comprehensive induction, there was no impact on student achievement in the first two years; however, by the end of the third year, teachers receiving comprehensive induction had a positive and statistically significant impact on student achievement.[9] Finally, neither one nor two years of comprehensive induction improved teacher retention or other teacher workforce outcomes, including teacher satisfaction and preparedness.

TEACHER PERFORMANCE EVALUATION

Given that comprehensive induction focuses attention on early career teachers, can policies designed to improve the human capital of all teachers across the experience distribution generate improvements in teacher practice, student achievement, and the teacher labor supply? Recent policy initiatives at the federal level have aimed to improve teacher human capital through more rigorous evaluation of teacher performance. Indeed, the 2009 federal Race to the Top (RTTT) grant competition was designed to incentivize states and districts to revise their approaches to educator evaluation. Such federal efforts aimed to inject greater accountability into the evaluation of teachers and to improve teacher practice through more rigorous on-the-job (i.e., in-service) evaluation. This is in light of the persistent concern that existing evaluation regimes at the state and local levels did little to differentiate teacher effectiveness. Little evidence currently exists on the impact of newly implemented teacher evaluation systems that were incentivized by RTTT grants. However, evidence from three urban districts offer promise that more rigorous evaluation of teachers can improve teacher effectiveness and the quality of the teacher labor supply.

In the 2000–01 school year, Cincinnati Public Schools launched the Teacher Evaluation System (TES) in which teacher performance was evaluated via classroom observations. While teachers in the U.S. have traditionally been evaluated through observations of classroom instruction (typically conducted by school principals), historically there has been little (if any) professional development provided to those responsible for conducting the classroom observation. Moreover, classroom observations occurred infrequently. Under Cincinnati's TES evaluation program, teachers were observed presenting a classroom lesson four times—three times by a peer evaluator (a high-performing, experienced teacher external to the school) and once by the principal or another school administrator. In addition to the classroom observation, teachers received written feedback about their performance and discussed the results of their classroom observation during a post-observation conference with their evaluator. Further, the district implemented and oversaw an intensive TES evaluator training course to prepare both peer evaluators and administrators for observing and rating teacher performance.

Evidence from Taylor and Tyler (2012) indicates that Cincinnati's TES evaluation initiative successfully improved student (math) achievement. While teacher productivity (i.e., a teacher's contribution to student achievement

8 The researchers observed teachers giving a literacy lesson in the spring of their first year and found no impacts on teachers' implementation of the literacy lesson, content of the literacy lesson, or classroom culture, relative to teachers in the control schools (Glazerman et al., 2010).

9 The third-year impacts are equivalent to effect sizes of 0.11 in reading and 0.20 in math.

growth) is positive during the year in which they are evaluated, teachers contribute even more to their students' math achievement in the years after evaluation.[10] Evidence further suggests that more rigorous evaluation of teacher practice benefits the lowest-performing teachers, as teachers who generated relatively little value-added to student test scores prior to evaluation saw the largest productivity gains in the years following evaluation. Given that TES provided detailed information to teachers about their instructional practice, the authors attribute student achievement gains to the information teachers received about their own performance and to any development of skills informed by more detailed feedback about their practice.

Chicago Public Schools (CPS) initiated a teacher evaluation pilot program among nearly 100 elementary schools in the 2008-09 school year. The pilot program—the Excellence in Teaching Project (EITP)—was designed to address many of the shortcomings of the district's traditional teacher evaluation system, which was based on a checklist of observed teacher practices. According to high-performing CPS teachers, the traditional evaluation approach did not provide meaningful feedback to improve their instruction. And, further, more than 60 percent of CPS principals believed that the checklist did not provide them with the tools necessary to adequately address teacher underperformance. Indeed, while 66 percent of CPS schools failed to meet state proficiency standards in Illinois in 2007, nearly all CPS teachers (93 percent) were deemed professionally proficient based on the district's traditional evaluation system (Steinberg and Sartain, 2015).

Similar to TES in Cincinnati, EITP in Chicago was based entirely on more rigorous observation of a teacher's classroom practice, coupled with the development of principal human capital to observe and rate teacher performance and provide detailed feedback to teachers on their practice during post-observation conferences. Indeed, principals in pilot schools randomly assigned to implement EITP in the 2008-09 school year received nearly 50 hours of training and support during the school year, including three days of initial training during the summer before implementation and follow-up sessions throughout the school year.[11] Steinberg and Sartain (2015) find that treatment schools implementing EITP in 2008-09 significantly improved student reading achievement (though positive, the math achievement effects were not statistically different from zero). The impact of EITP on school achievement persisted in the years after the pilot, though the impact was concentrated among schools that served more advantaged student populations (i.e., higher-achieving and lower-poverty schools). Given that EITP was designed to develop principal and teacher human capital through more rigorous professional development, Steinberg and Sartain (2015) note that the observed impacts of the teacher evaluation pilot on student achievement and school performance likely operated through two mechanisms: increased principal capacity as instructional leaders and improvements in teacher instructional quality through more detailed observation and feedback.

To what extent might evaluation systems designed to improve teacher practice through more intensive observation of classroom instruction serve the accountability objective of personnel evaluation? This is particularly salient given the limited role traditional evaluation systems have played in identifying and, if necessary, removing low-performing teachers from the classroom. In a follow-up study of the EITP initiative in Chicago, Sartain and Steinberg (2016) find that EITP increased the exit of low-rated and non-tenured teachers from the district. Moreover, the teachers who exited the district were lower-performing than both those who remained as well as the teachers who replaced them. These findings suggest that evaluation systems that rigorously observe and rate teacher classroom practice can induce low-performing teachers to exit the district and may improve the overall quality of the teacher workforce.

10 Student math achievement improves by approximately 0.10 standard deviations for students taught by a teacher in the year after that teacher was evaluated under TES, relative to students taught by the same teacher in the years before the teacher was evaluated under TES.

11 Principal training focused on using Charlotte Danielson's Framework for Teaching observation rubric to rate teaching practice, methods for collecting evidence, and best practices for conducting classroom observations. The training also included support for principals in coaching teachers. Principals also participated in seven monthly meetings to discuss a variety of implementation issues in the context of professional learning communities consisting of other EITP principals. During the professional learning community time, principals brought materials from classroom observations which they had conducted, and engaged in small group discussion with their colleagues, providing a rich set of supports for principals as they implemented EITP for the first time. Principals also received four half-day trainings during the school year, which provided an opportunity for them to update their understanding and use of the rubric for evaluating teachers.

The federal RTTT grant program incentivized states and districts to implement evaluation reforms that incorporated multiple measures of teacher performance, both those used in traditional evaluation systems (such as observations of classroom practice) and those tied to student achievement (so-called value-added measures, or VAMs). Washington, D.C. Public Schools (DCPS) was among the few district recipients of federal grant aid to support the development of and revisions to its existing teacher evaluation system. The resulting system— IMPACT—is among the first evaluation systems supported by RTTT aid which incorporate multiple measures of teacher performance, and ties teacher ratings to explicit personnel decisions (i.e., retention, tenure and compensation). The IMPACT system linked financial incentives as well as threat of dismissal to teacher performance on multiple measures: observation of classroom practice and student test performance. Specifically, teachers designated as Highly Effective earned a substantial increase in one-time and base compensation. In contrast, teachers designated as Minimally Effective in two consecutive years are subject to removal from teaching in DCPS.

Dee and Wyckoff (2015) evaluate IMPACT's effect on teacher performance and teacher retention. Using regression discontinuity methods, the authors find that voluntary exit from DCPS increased by 11 percentage points (more than 50 percent) among low-performing teachers (i.e., those at the threshold of the Minimally Effective rating); among low-performing teachers who remained in DCPS, teacher performance improved by 0.27 of a teacher-level standard deviation. For the highest-performing teachers (i.e., those at the threshold of the Highly Effective rating), the financial incentives improved teacher performance by 0.24 teacher-level standard deviations. Taken together, these results point to the role that incentives—dismissal threat or financial bonuses—linked to multiple measures of teacher performance can play in improving teacher performance and removing among the lowest-performing teachers in DCPS.

TEACHER MERIT PAY PROGRAMS

Can financial incentives alone—apart from a rigorously implemented teacher evaluation system—improve teacher human capital and student achievement? A recent pilot teacher incentive program in New York City provides insight into this question. In the 2007-08 through the 2009-10 school years, the United Federation of Teachers (UFT) and the New York City Department of Education (DOE) implemented a teacher incentive program in over 200 randomly selected schools, distributing approximately $75 million in teacher bonuses to over 20,000 teachers (Fryer, 2011). Each treatment school could earn up to $3,000 for every staff member (who was a UFT union member) if the school met the annual performance target set by the DOE, a target that was based on school report card scores.[12] Treatment schools were then each responsible for determining how to distribute any earned bonus money to individual teachers within their schools.[13]

Fryer (2011) estimates the causal effect of the teacher incentive program, comparing student and teacher outcomes in schools randomly selected for treatment to those schools randomly selected as controls (i.e., non-participants in the incentive program). He finds no evidence that teacher incentives increased student performance, student attendance, or graduation rates. Further, Fryer (2011) finds no evidence that the incentives changed teacher behavior—absences and retention (in school and in district). Some evidence suggests that teacher incentives may have decreased student achievement in schools with greater student enrollment. Fryer (2011) posits four potential explanations for the null results, including: the dollar value of the incentives may have been too modest in size; the incentive scheme itself may have been too complex, thereby

12 Each participating school was given $1,500 per UFT staff member if it met at least 75 percent of the performance target but not the full target. NYC's school report card scores were based on a composite measure of school environment (15 percent based on student attendance and survey scores), student achievement (25 percent based on the percentage of students meeting grade-level performance standards and graduation rates) and student academic progress (60 percent based on the change in proficiency ratings and high school exit exams).

13 The lump sum for meeting 100 percent of the annual performance target was awarded at the school level, and the school had the power to decide how to disburse the funds to individual teachers. The only restriction was that awards could not be distributed based on seniority and to participate, schools had to create a compensation committee (including the principal, a designee of the principal, two UFT staff members) to decide on the distribution of incentives. The committee had to agree unanimously on the design of the distribution of bonuses to teachers.

limiting teachers' ability to know how much effort to exert; group-based awards, rather than teacher-specific awards, may be less effective; or teachers may not know how to improve student achievement. Fryer (2011) suggests that the most reasonable explanation for the observed effects is that the pilot was too complex and limited teacher agency.

During the 2007-08 through 2010-11 school years, Chicago implemented a teacher incentive program: the Teacher Advancement Program (TAP). The TAP model was designed to improve teacher retention and student achievement by providing teachers with incentive bonuses. The TAP model included weekly meetings of teachers and mentors and regular classroom observations by a school leadership team to help teachers meet their performance goals. Teachers earned bonuses based on two performance measures: their value-added contribution to student achievement and their observed instructional performance in the classroom. In the first three years (2007-08 through 2009-10), teachers received an average bonus of approximately $1,100; this average increased to $1,400 for new Chicago TAP teachers in the fourth year (2010–11).

Relying on the random assignment of schools to TAP, Glazerman and Seifullah (2012) estimate the impact of TAP on student achievement and teacher retention. First, as might be expected, TAP increased the amount of mentoring, promotion opportunity, and compensation provided to teachers in TAP schools relative to non-TAP schools. However, the authors do not find evidence that TAP improved student achievement, with no detectable impacts on math, reading or science achievement. However, the authors did find evidence that TAP improved teacher retention, though the effect of TAP on teacher retention was not consistent across years or subgroups of students. Notably, even with the inclusion of teacher mentors in treatment schools as part of the TAP program, neither teacher attitudes about their school nor school climate improved.

Accountability, Standards and Assessment

In this section we review scholarship assessing the effects of the most prominent contemporary reform efforts based on high-stakes accountability, test-based assessments, and set standards for student performance. We begin by discussing efforts based on NCLB provisions designed to offer parents an expanded set of options. We then describe measures designed to create stronger results-based accountability in public schools.

PROVIDING SCHOOL PERFORMANCE DATA TO PARENTS

NCLB required schools to make adequate yearly progress (AYP), improving school-level and subgroup student proficiency in math and English Language Arts (ELA) every year until 100 percent proficiency was attained.[14] The law included provisions that allowed students to transfer out of Title I schools that failed to make AYP two years in a row.[15] The implied goal of these provisions was to increase academic outcomes for disadvantaged students (1) indirectly by introducing market-based competition into the public education system, thereby prompting all schools to improve, and (2) directly by providing students an immediate path to a higher performing school (Hastings and Weinstein, 2008). These mechanisms rely on the ability of parents to adequately acquire and assess information on their school options.

Justine Hastings and Jeffrey Weinstein (2008) examine the role of accessible and transparent information in the school choice decisions of parents of Charlotte-Mecklenburg Schools (CMS). Beginning in 2002, the district assigned students to schools through a choice system that required parents of students in rising grades (kindergarten, 6th, 9th) to submit their top three choices every spring.[16] In the summer of 2004, when the

14 States established timelines for adequate yearly progress such that all students in all recognized subgroups meet or exceed the State's proficiency level for academic achievement no later than 2014. No Child Left Behind Act of 2001, §1111(b)(2).

15 No Child Left Behind Act of 2001, §1116(b).

16 When the system was first implemented, ahead of the 2002-03 academic year, all parents were required to submit a choice form. Under the CMS system, students were guaranteed a seat a "home school," usually a nearby school. Admission to a non-guaranteed school was based on a lottery (Hastings and

first set of "failing" schools under NCLB was identified, parents were provided information on the academic achievement of alternate schools they had the option of transferring to in the fall. The authors compare the spring choice decisions (when school information was not readily available) with the summer decisions (after school information became available) and determine that information has a positive impact on choice. Namely, with access to information, parents choose higher performing schools.[17] In 2006, the authors worked with CMS to provide two types of simplified school information sheets to parents, one with statistics on academic achievement and another with both achievement statistics and estimated odds of admission. Both sheets were in a more readable, user-friendly format than the 2004 materials. However, despite the improved design, the effect on school choice was similar to 2004, suggesting that "the degree of simplification is not as important as just providing information on school test scores to parents at the time of choice" (p. 18). As one might expect, attending a stronger school (in this case, due to better information) was found to produce improved academic outcomes (Hastings and Weinstein, 2008).[18]

SUPPLEMENTAL EDUCATIONAL SERVICES

In addition to NCLB's transfer option, eligible students at Title I schools identified for improvement have the option of receiving supplemental educational services (SES), free of charge.[19] These services include remedial and enrichment programs in math and English Language Arts (ELA) that are aligned with state standards and are based on research. Several recent studies have examined SES (e.g., Steinberg, 2011; Springer et al., 2009; Muñoz, Potter, and Ross, 2008). For instance, a RAND study assessed the early implementation of SES in seven large urban districts (Zimmer et al, 2007).[20] By comparing the differences in achievement gains for SES participants and nonparticipants, the authors conclude that participation in SES has a positive, statistically significant effect on math and ELA scores for students in five of the districts.[21] In a subsequent meta-analysis that estimated average effects across nine districts, the authors determine that participating in SES results in achievement gains that accumulate over multiple years of program participation. Black and Latino students realize gains in both math and ELA, while students with disabilities experience gains in ELA, but not math.[22] A study by Heinrich, Meyer, and Whitten (2010) similarly estimates the effect of SES on student outcomes in Milwaukee. However, the authors are unable to conclude that participation results in a statistically significant effect on test scores.[23] While these results generally suggest that SES participation can lead to a positive and significant effect on students in math and reading, for certain districts and for certain populations no discernable gains were identified. This is likely due to the fact that features of SES—for instance, program quality, class size, and instructor experience—vary across providers. Research has yet to disentangle SES program features and their particular effects.

Weinstein, 2008).

17 Parents selected as their top choice schools with 0.62σ higher test scores (student level) than their current NCLB-sanctioned school and 0.49σ higher than their spring choice (Hastings and Weinstein, 2008).

18 The authors determined that increasing the score of the school attended by 1σ resulted in own test score gain of 0.40σ, on average (Hastings and Weinstein, 2008).

19 "The term supplemental educational services means tutoring and other supplemental academic enrichment services that are in addition to instruction provided during the school day; and are of high quality, research–based, and specifically designed to increase the academic achievement of eligible children" No Child Left Behind Act of 2001, P.L. 107-110, §1116(e).

20 The study included nine large urban districts: Baltimore City, Chicago, Denver, Long Beach, Los Angeles, Palm Beach, Philadelphia, San Diego, and Washington, D.C. Two (unidentified) districts were dropped from the SES impact analysis due to low student participation rates.

21 The student level effects of participating in SES ranged from 0.03σ to 0.38σ (math) and 0.03σ to 0.58σ (ELA) achievement gains above the district mean (at the 5 percent level).

22 The meta-analysis determined that participating in SES led to 0.09σ (math, overall), 0.8σ (ELA, overall), 0.17σ (math, 2+ years), 0.15σ (ELA, 2+ years), 0.10σ (math, African American students), 0.12σ (ELA, African American students), 0.10σ (math, Hispanic students), 0.09σ (ELA, Hispanic students), and 0.17σ (ELA, students with disabilities) achievement gains above the district mean (at the 5 percent level).

23 The authors use propensity score matching and fixed effects models and find no statistically significant effects on test scores for students who participate in SES; when taking into account total hours of SES participation, the authors find only one statistically significant relationship: for the 2005-06 year, each additional hour attended increased high school reading scores by just 0.09 of a test unit—approximately 0.10σ for 25 hours of participation (Heinrich et al., 2010).

TEST-BASED ACCOUNTABILITY

The accountability policies ushered in by NCLB have led to several test-based mechanisms that assist in the identification of underperforming students and schools, and incentivize school communities to work harder. Test-based accountability is thought to improve outcomes by "raising motivation, increasing parent involvement and improving curriculum and pedagogy" (Jacob, 2005, p. 761). However, in recent years, parents and advocates have raised concerns over the prevalent use of high-stakes standardized tests in public education. Some worry that schools, particularly under-resourced schools in urban districts, focus too much on test preparation. Others point to recent cheating scandals in Atlanta and elsewhere as an unintended consequence. While others argue that formative assessments and performance-based evaluations are of greater immediate benefit to students and teachers. In a handful of urban school districts, these concerns have led parents to "opt out" by preventing their children from taking part in state tests. With this in mind, a baseline question might be whether annual student assessments have an effect on student achievement. Thomas Dee and Brian Jacob (2011) examine the impact of NCLB and its accountability requirements on student outcomes by comparing state-level trends in achievement between states that were most affected by NCLB (treatment) and states that were less affected (control).[24] The authors find that NCLB's effect on achievement was "decidedly mixed"—the federal law led to gains in math, but not for reading.[25]

Brian Jacob (2005) examines a precursor accountability policy governing Chicago Public Schools (CPS). Introduced in 1996, the policy ended the practice of "social promotion"—advancing students to the next grade regardless of achievement, thus holding students accountable for their learning. In order to advance from 3rd, 6th, and 8th grade, students were required to meet minimum proficiency standards in reading and math. Students who fell below the cutoff were to enroll in a six-week summer school program and retake the test. If they passed, they advanced; if they failed once again, they had to repeat the grade. In addition to social promotion, schools with a proficiency rate below 15 percent were placed on probation, from which they faced reconstitution if they failed to improve. By examining changes in student achievement over time, and by comparing CPS with other large districts in the Midwest that did not implement accountability policies, Jacob (2005) finds that the CPS accountability policy improved student achievement for older students but only improved test-specific (rather than general) skills for younger students.[26] Furthermore, teachers responded "strategically" by increasing special education placements, preemptively retaining students, and substituting away from low-stakes subjects like science and social studies.

Neal and Schanzenbach (2010) examine CPS under both the 1996 reform and the post-NCLB period in order to understand how test-based accountability affected students at different points on the achievement distribution. Because rewards and sanctions are tied to the numbers of students scoring above proficiency cutoff points, the allocation of resources may vary across different types of students. Teachers and administrators may conduct "educational triage" by making "conscious and deliberate decisions to shift resources away from low-performing students and toward students who had more realistic chances of exceeding key threshold scores" (Neal and Schanzenbach, 2010, p. 265). In other words, the most cost-effective strategy for increasing proficiency counts could reasonably be to focus on students near the proficiency cut points. Furthermore, low-performing schools with few students near the cut point may not be able to avoid sanctions just by focusing on those students. The authors indeed did not find that students with the lowest levels of prior achievement—likely those with little realistic change of meeting the proficiency standard—improved following the introduction of a test-based accountability system. However, students at higher levels of achievement showed improvement.

24 Several states had accountability laws prior to the implementation of NCLB. For these states, NCLB's requirements were less consequential than for those states without prior accountability laws.

25 The authors find that by 2007, NCLB generated a 0.23σ increase in fourth grade math. Estimated effects for fourth grade reading and eighth grade math were indistinguishable from zero (Dee and Jacob, 2011).

26 Jacob (2005) estimates the effect of the policy on math and reading achievement as approximately 0.30σ and 0.20σ, respectively.

The authors also find evidence to suggest that low achieving students would fare better if they attend schools that could not meet proficiency standards through educational triage.

DATA-DRIVEN REFORM

The accountability requirements of NCLB, along with the increased technical capability districts have to accumulate and analyze data, have spurred data-driven reform efforts in recent years. One prominent example is the Data-Driven District (3D) initiative developed by the Center for Data-Driven Reform in Education (CDDRE) at Johns Hopkins University. The initiative trained school and district leaders to understand and more effectively use student performance data.[27] Through a randomized trial, Carlson, Borman, and Robinson (2011) assess whether the initiative brought about districtwide improvements in student performance. Over three cohort waves, participating districts were randomly assigned to receive CDDRE support beginning either in the first initiative year (treatment) or in the second (control). Treatment districts received a range of supports and technical assistance based on the 3D model. These included "quarterly benchmark assessments, data reviews, training in leadership and data interpretation, provision of reviews of research on effective programs and practices, and assistance in selecting and implementing proven programs" (Carlson et al., 2011, p. 383). The 3D districts exhibited modest gains in math achievement.[28] The authors speculate that the effects may be due to some combination of three mechanisms: (1) The practice that students had through multiple benchmark assessments may have served to better prepare them for the end-of-year state achievement test; (2) Benchmark assessments may have helped familiarize teachers with the state assessments, and consequently altered their instruction during the school year; and/or, (3) Benchmark assessments may have helped teachers identify "areas of weakness" for targeted instruction (Carlson et al., 2011).

SCHOOL CLOSURES

In recent years, in order to contend with declining enrollment, fiscal constraints, low student achievement, and increased competition with charters, dozens of urban districts have closed underutilized and underperforming schools (Sunderman and Payne, 2009; De La Torre and Gwynne, 2009). In one recent study, researchers examined an (unnamed) mid-sized urban district that explicitly sought to use school closures "in hopes of addressing the dual problems of low achievement and financial distress, consistent with the demands of [NCLB] and the Obama administration's *Race to the Top* initiative" (Engberg, Gill, Zamarro, and Zimmer, 2012, p. 189). Using a metric of performance based on several value-added measures, the district closed or reconstituted schools in the bottom quartile.[29] High-performing schools were kept open unless they created resource inequities. The district's policy required that displaced students be reassigned to a school at least as high-performing as the one they left. During the summer of 2006, 22 schools in the unnamed district were closed and eight were reconstituted. The authors conclude that students moving from a closed school to a new school experience adverse effects on attendance and achievement.[30] Although the effect on attendance disappears within a year of closure, the effect on achievement persists for multiple years. However, the authors find that these effects can be minimized by assigning students to higher performing schools.[31] While there are negative

27 "The goal of CDDRE is to solve the problem of scale in educational reform by working with entire school districts to help district and school leaders understand existing data on student performance, generate additional data to help guide school improvement efforts, identify root causes underlying important problems, and then select and effectively implement evidence-based programs directed toward solving those problems" (Carlson, Borman, and Robinson, 2011, p. 379).

28 Math scores for treatment districts were 0.06σ higher than control districts, on average (at the 5 percent level). For reading, the difference between treatment and control districts was positive but not statistically significant at conventional levels. The authors argue that these student-level gains in average district-level achievement is substantively meaningful. However, they note an absence of a recognized benchmark against which their results can be compared.

29 Closed middle schools were replaced by expanding the gradespan of thirteen elementary schools from K-5 to K-8.

30 In order to estimate the effect of school closures on student test scores and attendance rates, researchers use the district's residentially-based student assignment scheme as an instrument for the proportion of students arriving from closed schools.

31 The authors find that absenteeism increases by 13 percent in the first year of closure; for students moving to an equivalently performing school, the standardized effect sizes (z-score transformation) for math and reading achievement is -0.19 and -0.20, respectively (at the 5 percent level). Students moving to a higher performing school do better. For a student with an average change of 0.88 in school performance, the math and reading effect sizes are -0.11 and -0.13, respectively.

effects on attendance and achievement for students from a closed school, the authors do not find effects for students in receiving schools. The authors conclude that a policy of school closures as a means of generating increased academic outcomes is unlikely to be effective: "The evidence from this school district suggests that producing higher levels of achievement would require moving students to schools that are dramatically higher achieving than the schools they left" (Engberg et al., 2012, p. 198).

Market-Based Reforms and School Choice

SMALL SCHOOLS

During the early 2000s, large foundations (most notably, the Bill & Melinda Gates Foundation) and the U.S. Department of Education awarded grants to support the establishment of smaller learning communities ("small schools").[32] Proponents argued that by breaking up large public high schools into small schools, students would have a more personalized educational experience that would lead to improved academic outcomes. New York City was home to a system-wide effort that led to the establishment of small nonselective high schools of choice, primarily in under-resourced communities. New York's program authorized small schools "through a demanding competitive proposal process designed to stimulate innovative ideas" (Bloom, Thompson, and Unterman, 2010, p. iii). Schwartz, Stiefel, and Wiswall (2012) find that students enrolled in small schools established during this period (but not small schools established in prior years) had a 17.5 percent gain in the probability of high school graduation, compared to attending a large high school (see also, Bloom, Thompson, and Unterman, 2010).[33]

The authors identify several features of the "new" (i.e., early 2000s) small schools program that distinguished it from prior small schools efforts as well as from traditional large schools. The competitive process required applicants to describe how they would implement rigorous curricula and partner with community-based organizations. The new program exempted small schools in their start-up years from serving special needs students and from certain hiring rules. Although small schools under the new program tended to have more inexperienced teachers, they received a substantially larger per-pupil allocation. New small schools also had significantly fewer special needs and English learner students. Finally, the new program provided small schools specialized technical assistance and principal leadership training.

SCHOOL VOUCHERS

In publicly-funded school voucher systems, parents directly control taxpayer funds allocated for public education, most often for use toward private school tuition. Typically, vouchers worth several thousand dollars are issued to eligible families on behalf of their children. Families can then apply the voucher to a qualified school of their choosing. The school exchanges the voucher for payment directly from the government. Currently, there are publicly-funded voucher plans in operation in Milwaukee, Cleveland, and Washington, D.C. In addition, there are a handful of statewide programs, including in Indiana, Arizona, and Utah. In theory, vouchers can be either *universal* (available to all students within a jurisdiction) or *targeted* (available only to a particular subset of the student population). In practice, all publicly-funded voucher programs currently operating are targeted to low-income families, students in failing schools, and special needs students. Proponents claim

32 The U.S. Department of Education described smaller learning communities as structures such as "freshman academies, multi-grade academies organized around career interests or other themes, "houses" in which small groups of students remain together throughout high school, and autonomous schools-within-a-school, as well as personalization strategies, such as student advisories, family advocate systems, and mentoring programs" (U.S. Department of Education, 2015).

33 Since there are likely unobserved factors that both lead a family to select into a small school and affect student outcomes (e.g., student motivation, parent involvement), the authors instrument for enrollment in a small school with variables based on distance.

that achievement gaps will narrow under targeted voucher systems because disadvantaged students will be afforded an expanded set of educational options.

Washington D.C.'s Opportunity Scholarship Program (OSP) provides up to $7,500 to cover tuition, transportation, and other costs associated with private school attendance. If there are more applicants than available vouchers in a given year, offers are provided through a random lottery in which students attending a public school in need of improvement under NCLB receive the highest priority. A recent impact study conducted by the National Center for Educational Evaluation and Regional Assistance was unable to conclude that OSP had an effect on student achievement. In other words, eligible students who won the lottery had statistically equivalent math and ELA scores as those who lost the lottery (Wolf et al., 2010).[34] While vouchers provide an opportunity for families to access private schools, these schools may not necessarily be better at producing student outcomes. Furthermore, families may base their decision to apply for a voucher to a private school on reasons other than academic achievement. School climate, for instance, is a central concern for many. In fact, Wolf and colleagues find that the OSP increased the satisfaction and sense of school safety among parents.[35]

OSP was also found to improve high school graduation rates. The graduation rate for lottery winners was 82 percent versus 70 percent for lottery losers, a 12 percentage point difference, while the impact on graduation of actually using a voucher was 21 percentage points (Wolf et al., 2010; see also, Rouse and Barrow, 2009). To explain these results, Wolf and colleagues point to research arguing that private schools provide students with school environments and motivated peers that support improvements in student graduation rates (Evans and Schwab, 1995; Grogger and Neal, 2000; Neal, 1997; Warren, 2010).

CHARTER SCHOOLS

Charter schools are publicly funded, privately managed schools of choice. They are formally established through a legal agreement ("charter") between school leaders and an oversight authority, typically the local school board. The charter will often describe school components including mission, curricula, pedagogical approach, and organizational structure. Charter schools are authorized to operate for a pre-determined number of years and must petition for renewal at regular intervals. In exchange for this increased level of accountability, charter schools are granted autonomy from portions of state education law.

Laws authorizing charter schools have been adopted in 43 states and the District of Columbia. Charter schools now enroll over 2.6 million students in 6,600 schools (National Alliance for Public Charter Schools, 2016). A handful of large urban districts enroll a high percentage of charter students. Today, the charter enrollment share exceeds 90 percent in New Orleans, 50 percent in Detroit, 40 percent in the District of Columbia, and 30 percent in Philadelphia (National Alliance for Public Charter Schools, 2015). There are over 151,000 charter students in Los Angeles, 84,000 students in New York City, and 57,000 students in Chicago (National Alliance for Public Charter Schools, 2015). Despite pockets of active opposition, further growth is likely—particularly through the scale-up of Charter Management Organizations—as this method of reform has garnered significant private philanthropic support, mobilized educational entrepreneurs, and attracted lawmakers from both sides of the aisle (Quinn et al., 2016).

Over the past 15 years, scholars have examined the effect of charter school attendance on academic achievement (e.g., Hoxby, 2004; Buddin and Zimmer, 2003; Hanushek, Kain, Rivkin, and Branch, 2005; Center for Research on Education Outcomes, 2013). Among the recent scholarship addressing this issue is Atila

34 Evaluations of a **privately** funded voucher system in New York City that made use of a random selection design similarly were unable to report statistically significant effects on student outcomes (Mayer et al., 2002; Krueger and Zhu, 2004).

35 Wolf et al. (2010) report an effect of 0.14σ for lottery winners and 0.17σ for voucher users in parent perceptions of safety and an orderly school climate. Interestingly, the authors did not find statistically significant effects for student perceptions of climate and safety.

Abdulkadiroğlu and colleagues' (2011) study of charter schools in Boston, Massachusetts. The authors exploit the random assignment of enrollment spots produced by the legal requirement that over-subscribed charter schools must select students through admissions lotteries. This generates lottery winners who go on to enroll in the school (treatment) and lottery losers who enroll elsewhere (control).[36] In the absence of a lottery, attendance at a charter school would likely be biased due to factors like motivation or family background. The charter schools included in this study—as lottery-admission schools—are the most popular, and are likely perceived by parents to be the most effective. The authors find that each year spent in an over-subscribed charter school yields large and significant gains across tested subjects for middle and high school students (see also, Angrist et al., 2010).[37] What are the components of charter schools that make them effective? Abdulkadiroğlu and colleagues (2011) surmise that the "No Excuses" model adopted by most of the charter schools in their study may play an important role in generating test score gains. We next turn to empirical scholarship examining this model in particular.

The "No Excuses" school has attracted attention in recent years as a model of school reform that can improve student academic outcomes and close the black-white achievement gap. The model is built upon the notion that "children of all races and income levels can meet high academic standards," and thus there should be no excuse for academic failure (Carter, 2000, p. 7). No Excuses schools are marked by an extensive use of measurable and "unyielding" goals, frequent testing, expectations of parental involvement (often through parent "contracts"), attention to discipline through a rigid system of escalating rewards and punishments, and principals who are empowered to make hiring, curricular, and budgeting decisions (Carr, 2013; Carter, 2000; Seider, 2012). While No Excuses schools are most commonly associated with the charter school sector, district-operated schools throughout the nation have incorporated some features of the model.

One prominent adopter of several components of the No Excuses model is the Promise Academy of the Harlem Children's Zone (HCZ), a renowned multi-service nonprofit providing a coordinated suite of programs, from a "Baby College" for new parents to a "College Success Office" for first-generation students (Harlem Children's Zone, 2015). By virtue of being part of HCZ, Promise Academy students and their families have access to a wide variety of activities and programs. The schools provide medical, dental, and mental health services free of charge and offer several other amenities like food baskets and bus fare for families, Model United Nations Summits, and comprehensive nutrition education, to name a few. HCZ raises considerable revenue for their schools, beyond the per pupil allocation provided by the New York Department of Education.[38] Following the No Excuses model, Promise Academy schools have a longer day and year, offer extensive remediation opportunities for math and ELA, and incentivize both students and teachers for achievement. Unlike some other No Excuses schools, Promise Academy schools do not require signed behavioral contracts for students or parents (Harlem Children's Zone, 2015; Dobbie and Fryer, 2011).

A recent study by Will Dobbie and Roland Fryer (2011) examines the effect of this "ambitious social experiment" on educational outcomes. Similar to the Abdulkadiroğlu and colleagues' (2011) study, Dobbie and Fryer take advantage of the random assignment generated through Promise academy elementary and middle school admissions lotteries. In addition, the authors employ a complementary strategy that relies on the open enrollment nature of charter schools (i.e., any parent in the district, regardless of home address, can enroll in a Promise Academy), the active recruitment of parents within a 24-block area covered by the HCZ, and the age requirements for enrollment at the time of the schools' openings.[39] The assumption here is that home

36 Some student win the lottery and enroll elsewhere, some students lose the lottery yet end up enrolling through waitlists and sibling preferences. Researchers typically identify two causal estimands: the intent-to-treat (ITT) effect of being assigned to a treatment group (e.g., being offered a spot through an admissions lottery), and the effect of treatment-on-treated (TOT), in other words, of actually receiving the treatment (e.g., enrolling in a charter school).

37 The authors find effects of 0.40σ for middle school math, 0.20σ for middle school ELA, 0.36σ for high school math, 0.30σ for high school ELA, 0.35σ for writing topic, and 0.21σ for writing composition.

38 Dobbie and Fryer (2011) report the per pupil allocation from the NYDOE as $12,443 and the additional revenue from HCZ as $6,829 (2008-09 academic year).

39 HCZ has since expanded its coverage area to 97 blocks.

address and cohort year are exogenous variables that affect academic achievement only through their effect on the likelihood of enrollment in a Promise Academy. Taking these instrumental variables into account allows for causal inferences to be drawn. Both strategies led to similar conclusions: the effect of attending a Promise Academy was enough to close the black-white achievement gap in elementary and middle school math and elementary ELA.[40] The authors do not find a statistically significant effect on ELA in the middle grades; but they conclude that it is likely positive, albeit less dramatic. By comparing the effects of attendance for students who live inside and outside of the Zone, as well as comparing students with siblings who did not attend a Promise Academy, the authors determine that the observed effect is due to the quality of the schools rather than the dense network of non-school community supports available through HCZ. The authors note that the elementary and middle school math effects and the elementary ELA effects of attending a Promise Academy are larger than the reported effects of such well-known reforms as class size reduction, early childhood programs, and teacher bonuses. Although there are no rigorous studies estimating the effect of particular features of the No Excuses model, nor studies that examine the effect of the model outside of the charter school context, Dobbie and Fryer's results examining the impact of the Promise Academy program provide an encouraging path forward for urban district school improvement.

Conclusion

In this article, we described the new educational landscape in the United States by examining four domains of reform. We were particularly concerned with how these efforts played out in urban school systems, and focused our discussion on empirical evidence that provided causal insights into the efficacy of policy reforms.[41]

We close by summarizing actionable takeaways suggested by existing research for policymakers and school district leaders:

1. Early childhood education programs need qualified professionals. Programs, such as those in Tulsa and Oklahoma, that require a bachelor's degree and compensate pre-K teachers in line with their K-12 colleagues may generate positive results. In addition, providing ongoing professional development for pre-K teachers and limiting class sizes can be important program features.

2. Teacher human capital policies must be comprehensive and multifaceted. Teacher induction programs that provide rigorous supports can have a positive impact over time. Effective evaluation systems in Cincinnati and Chicago entailed comparatively frequent classroom observations and provided teachers with detailed feedback during post-observation conferences. High-stakes evaluation systems such as DCPS's IMPACT may improve teacher performance and lead low-performing teachers to exit.

3. High-stakes testing may generate improvements in academic achievement, although unintended consequences—focusing just on those students near threshold scores or focusing just on test-specific skills—may occur. Furthermore, accountability hinges on the effective use of relevant data. Urban school districts can harness data to drive improvements in achievement, and data on the academic performance of schools can be a valuable resource for parents when making school choice decisions.

4. Certain choice-based reforms have been effective in improving academic outcomes. Small schools in New York, where applicants seeking to establish a school underwent a rigorous selection process and schools

40 The effect of attending a Promise Academy in elementary school is as much as 0.32σ in math and 0.42σ in ELA for every year a student is enrolled. If we consider a student enrolled in a Promise Academy for six years (kindergarten through 5th grade), we would expect a 1.94σ increase in math and a 2.52σ increase in ELA. The effect of attending a Promise Academy in middle school is an increase of as much as 0.23σ in math achievement. Accordingly, after three years, 6th to 8th grade, the effect is 0.69σ.

41 This naturally limits the scholarship we cover and the outcomes assessed. Although much of the evidence we present is based on student performance on standardized exams in math and reading, we recognize that there are a variety of additional student measures and outcomes that policymakers, school leaders, and parents care about.

received substantially larger per-pupil funding enjoyed high graduation rates. Similarly, No Excuses charter schools, like the Promise Academy in Harlem, can produce impressive results for their students. These schools are unique, however, in that they have large per-pupil allocations, require longer school days and a longer school year, and provide various achievement incentives for both students and teachers.

Our aim in this article was to inform policymakers, school leaders, and the public on critical issues in contemporary school reform, and the extent to which these efforts have improved the educational conditions in our major urban districts. Ultimately, we believe that future reform efforts should be informed by past experiences in order to create and expand opportunities for upward mobility in urban America.

ACKNOWLEDGEMENTS

The authors thank Amanda Jones-Layman for excellent research assistance. Authors contributed equally to this article. Steinberg (corresponding author) can be contacted at steima@upenn.edu; Quinn can be contacted at raq@upenn.edu.

REFERENCES

Abdulkadiroğlu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., and Pathak, P. A. 2011. Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *The Quarterly Journal of Economics* 126(2): 699–748.

Angrist, J. D., Dynarski, S. M., Kane, T. J., Pathak, P. A., and Walters, C. R. (2010). Inputs and impacts in charter schools: KIPP Lynn. *American Economic Review* 100(2): 239-243.

Bloom, H. S., Thompson, S. L., and Unterman, R. 2010. Transforming the high school experience: how New York City's new small schools are boosting student achievement and graduation rates. *MDRC,* June.

Buddin, R., and Zamarro, G. 2009. Teacher qualifications and student achievement in urban elementary schools. *Journal of Urban Economics* 66(2): 103–115.

Buddin, R., and Zimmer, R. 2005. Student achievement in charter schools: A complex picture. *Journal of Policy Analysis and Management* 24(2): 351-371.

Carlson, D., Borman, G. D., and Robinson, M. 2011. A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis* 33(3): 378-398.

Carr, S. 2013. *Hope against hope: Three schools, one city, and the struggle to educate America's children*. New York, NY: Bloomsbury.

Carter, S. C. 2000. No Excuses: Lessons from 21 High-Performing, High-Poverty Schools. Heritage Foundation, 214 Massachusetts Avenue, NE, Washington, DC 20002.

Center for Research on Education Outcomes. 2013. *National charter school study*. Stanford, CA.

Coleman, J., E. Campbell, C. Hobson, J. McPartland, A. Mood, F. D. Weinfeld, and P. York. 1966. *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office.

De la Torre, M., and Gwynne, J. 2009. *When schools close: Effects on displaced students in Chicago Public Schools. Research Report.*

Decker, P. T., Mayer, D. P., Glazerman, S., and others. 2004. The effects of Teach for America on students: Findings from a national evaluation. University of Wisconsin–Madison, Institute for Research on Poverty. Retrieved from: http://www.ssc.wisc.edu/irpweb/publications/dps/pdfs/dp128504.pdf.

Dee, T. S., and Jacob, B. 2011. The impact of No Child Left Behind on student achievement. Journal of Policy Analysis and Management 30(3): 418–446.

Dee, T. S., and Wyckoff, J. 2015. Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management* 34(2): 267–297.

Dobbie, W., and Fryer Jr, R. G. 2009. *Are high quality schools enough to close the achievement gap? Evidence from a social experiment in Harlem*. National Bureau of Economic Research.

Engberg, J., Gill, B., Zamarro, G., and Zimmer, R. 2012. Closing schools in a shrinking district: Do student outcomes depend on which schools are closed? Journal of Urban Economics 71(2): 189–203.

Evans, W. N., and Schwab, R. M. 1995. Finishing high school and starting college: Do Catholic schools make a difference? *The Quarterly Journal of Economics 110*: 941-974.

Fryer, R. G. 2011. Teacher incentives and student achievement: Evidence from New York City public schools. National Bureau of Economic Research. Retrieved from: http://www.nber.org/papers/w16850.

Fryer, R. and Levitt, S. 2004. Understanding the black-white test score gap in the first two years of school. The Review of Economics and Statistics 86(2): 447-464.

Fryer, R. and Levitt, S. 2006. The black-white test score gap through third grade. American Law and Economics 8(2): 249-281.

Glazerman, S., E. Isenberg, S. Dolfin, M. Bleeker, A. Johnson, M. Grider, and M. Jacobus. 2010. Impacts of Comprehensive Teacher Induction: Final Results From a Randomized Controlled Study (NCEE 2010-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Glazerman, S., and Seifullah, A. 2012. An evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after four years. Final Report. Mathematica Policy Research, Inc. Retrieved from http://eric.ed.gov/?id=ED530098.

Gormley, W. T. 2008. The effects of Oklahoma's pre-K program on Hispanic children. Social Science Quarterly 89(4): 916–936.

Gormley, W. T., Gayer, T., Phillips, D., and Dawson, B. 2005. The effects of universal pre-K on cognitive development. Developmental Psychology 41(6): 872–884.

Gormley, W. T., and Gayer, T. 2004. Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program. *Journal of Human Resources 40*(3): 533–558.

Grogger, J., Neal, D., Hanushek, E. A., and Schwab, R. M. 2000. Further Evidence on the Effects of Catholic Secondary Schooling [with Comments]. *Brookings-Wharton Papers on Urban Affairs*, 151-201.

Hanushek, E. A., Kain, J. F., Rivkin, R. G., and Branch, G. F. 2005. *Charter school quality and parental decision making with school choice* (NBER Working Paper No. 11252). Cambridge, MA: National Bureau of Economic Research.

Harlem Children's Zone. 2015. Harlem Children's Zone 2014-2015 Biennial Report: A Community of Opportunity. Retrieved from: http://hcz.org/documents-and-publications/.

Hastings, J. S., and Weinstein, J. M. 2007. *Information, school choice, and academic achievement: Evidence from two experiments*. National Bureau of Economic Research.

Heckman, J.J. 2011. The economics of inequality: The value of early childhood education. *American Educator* 35(1): 31-35.

Heinrich, C. J., Meyer, R. H., and Whitten, G. 2010. Supplemental Education Services Under No Child Left Behind Who Signs Up, and What Do They Gain?. *Educational Evaluation and Policy Analysis 32*(2): 273-298.

Hoxby, C. M. 2004. A straightforward comparison of charter schools and regular public schools in the United States. Working paper, Harvard Institute of Economic Research.

Jacob, B. A. 2005. Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics 89*(5-6): 761-796.

Kane, T. J., and Staiger, D. O. 2008. Estimating teacher impacts on student achievement: An experimental

evaluation. National Bureau of Economic Research. Retrieved from: http://www.nber.org/papers/w14607

Krueger, A. 1999. Experimental estimates of education production functions. *The Quarterly Journal of Economics 114*(2): 497-532.

Krueger, A. B., and Zhu, P. 2004. Another look at the New York City school voucher experiment. *American Behavioral Scientist 47*(5): 658-698.

Mayer, D. P., Peterson, P. E., Myers, D. E., Tuttle, C. C., and Howell, W. G. 2002. *School Choice in New York City after Three Years: An Evaluation of the School Choice Scholarships Program.* Final Report.

Muñoz, M. A., Potter, A. P., and Ross, S. M. 2008. Supplemental educational services as a consequence of the NCLB legislation: Evaluating its impact on student achievement in a large urban district. *Journal of Education for Students Placed at Risk 13*(1): 1-25.

National Alliance for Public Charter Schools. 2015. *A growing movement: America's largest charter school communities. Tenth Annual Edition*. Washington, DC.

National Alliance for Public Charter Schools. 2016. Data dashboard. Retrieved from: http://www.publiccharters.org/dashboard/home.

Neal, D. 1997. The effects of Catholic secondary schooling on educational achievement. *Journal of Labor Economics*, 98-123.

Neal, D., and Schanzenbach, D. W. 2010. Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics* 92(2): 263–283.

Quinn, R., Oelberger, C., and Meyerson, D. in press. Getting to scale: Ideas, opportunities, and resources in the early diffusion of the charter management organization, 1999-2006. *Teachers College Record.*

Reardon, S.F. 2011. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In R. Murnane and G. Duncan (Eds.), *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low-Income Children.* New York: Russell Sage Foundation Press .

Rockoff, J., Jacob, B., Kane, T., and Staiger, D. 2011. Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, 43-74.

Rouse, C. E., and Barrow, L. 2009. School vouchers and student achievement: Recent evidence and remaining questions. *Annual Review of Economics* 1(1): 17–42.

Sartain, L., and Steinberg, M.P. 2016. Teachers' labor market responses to performance evaluation reform: Experimental evidence from Chicago public schools. *Journal of Human Resources* 51(3).

Schwartz, A. E., Stiefel, L., and Wiswall, M. 2013. Do small schools improve performance in large, urban districts? Causal evidence from New York City. *Journal of Urban Economics,* 77, 27–40.

Seider, S. 2012. *Character Compass. How Powerful School Culture Can Point Students Toward Success.* Harvard Education Press.

Springer, M. G., Pepper, M. J., and Ghosh-Dastidar, B. 2009. Supplemental educational services and student test score gains: Evidence from a large, urban school district. In *NCLB: Emerging Findings Research Conference at the Urban Institute*, Washington, DC on August, 12: 2009.

Steinberg, M. P. 2011. Educational Choice and Student Participation: The Case of the Supplemental Educational

Services Provision in Chicago Public Schools. *Educational Evaluation and Policy Analysis* 33(2): 159–182.

Steinberg, M. P., and Sartain, L. 2015. Does Teacher Evaluation Improve School Performance? Experimental Evidence from Chicago's Excellence in Teaching Project. *Education Finance and Policy* 10(4): 535–572.

Sunderman, G. L., and Payne, A. 2009. Does Closing Schools Cause Educational Harm? A Review of the Research. Information Brief. Mid-Atlantic Equity Center.

Taylor, E. S., and Tyler, J. H. 2012. The Effect of Evaluation on Teacher Performance. *American Economic Review* 102(7): 3628–3651.

U.S. Department of Education. 2015. Smaller Learning Communities Program. Washington, D.C. Retrieved from: http://www2.ed.gov/programs/slcp/.

Warren, J. R. 2010. *Graduation Rates for Choice and Public School Students in Milwaukee, 2003-2008.* Milwaukee, WI: School Choice Wisconsin.

Weiland, C., and Yoshikawa, H. 2013. Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development* 84(6): 2112–2130.

Wolf, P., Gutmann, B., Puma, M., Kisida, B., Rizzo, L., Eissa, N., and Carr, M. 2010. Evaluation of the DC Opportunity Scholarship Program: Final Report. NCEE 2010-4018. National Center for Education Evaluation and Regional Assistance.

Zimmer, R., Gill, B., Razquin, P., Booker, K., and Lockwood III, J. R. 2007. State and local implementation of the No Child Left Behind Act (Volume I- Title I School Choice, Supplemental Educational Services, and student achievement). RAND.