Article

Urban Education
2015, Vol. 50(3) 346–376

DOI: 10.1177/0042085913507060
uex.sagepub.com
SAGE

The Students in Front of Us: Reform for the Current Generation of Urban High School Students

Joe Burks[1] and Craig Hochbein[2]

Abstract

The implementation of education policies requiring the turnaround of persistently low-achieving schools has demanded reforms that will not only improve achievement, but also deliver results in a short period of time. To meet such demands, Jefferson County Public Schools educators implemented Project Proficiency (PP). Results from state-administered mathematics tests demonstrated that all participating schools reported substantial increases in student proficiency. We examined the impact of PP on the performance of students, who met dropout predictive criteria established by Balfanz, Herzog, and MacIver. Study results suggested that PP students at risk of dropout realized meaningful and statistically significant achievement gains.

Keywords

high school programs, academic achievement, urban education, school effectiveness, school reform

[1]Jefferson County Public Schools, Louisville, KY, USA
[2]Lehigh University, Bethlehem, PA

Corresponding Author:
Craig Hochbein, Lehigh University, College of Education, 111 Research Drive, Bethlehem, PA 18015.
Email: craig.hochbein@lehigh.edu

In 2009, the guidelines for the Race to the Top grant competition sponsored by the United States Department of Education implicitly stated the educational priorities and objectives of the newly appointed administration. Included among these goals was the turnaround of the 5,000 lowest performing schools in the country. In support of the priority and the means to achieve the turnaround of persistently low-achieving (PLA) schools Secretary of Education Arne Duncan (2009) wrote, "If we don't take aggressive action to fix the problems of these schools, we are putting the children in them on track for failure" (p. 36). To increase the likelihood of earning one of the Race to the Top grant awards, Kentucky state legislators aligned state statutes with federal turnaround models for schools identified as PLA. Adoption of these aggressive actions, which included the possible removal of building principals and 50% of the school faculty, generated additional layers of school accountability for schools scoring in the bottom five or 5% of the state in reading and math (Persistently Low Achieving School, 2010).

With test scores that ranked among the lowest in the state, state education officials identified and sanctioned 10 of the 21 comprehensive high schools in Jefferson County Public Schools (JCPS), and placed two additional schools on probation. Navigating these unprecedented sanctions, high school educators in JCPS, a large urban district in Louisville, Kentucky, had 1 year to dramatically improve students' scores on statewide tests. The sanctioned schools collectively averaged a student composition of 59% minority and 80% qualifying for free or reduced-price lunch (FRL). In the fall of 2010, JCPS educators developed and implemented a plan intended to rapidly improve the academic performance of high school students in PLA schools.

Nearly a year after implementation of this initiative, state test results demonstrated the largest gains in high school reading and math proficiency rates in JCPS history. Every JCPS high school gained in reading and math proficiency, with the PLA schools performing particularly well. The JCPS PLA schools averaged gains of 14% in reading and 17% in math, which outpaced mean state improvements of 5% and 6%, respectively. Yet, even after improvement, five high schools remained among the lowest performing schools in the state. To escape PLA status, JCPS educators needed to maintain the momentum of its high school reform and strategically address critical student learning gaps. Increasing overall averages in student proficiency can mask the need to improve individual student achievement. Particularly in PLA schools, genuine and sustainable reform efforts must reach low-achieving students, which generally consist of low-income, minority, and transient students, to proficient performers (Zavadsky, 2009). The purpose of this study was to examine the influence of the JCPS high school reform model on the academic achievement of a segment of at-risk students. To achieve this purpose, the researchers used a comparative study to answer the question:

When compared with similar at-risk students, do statistically significant differences exist in state administered math scale scores and gains for students participating in the JCPS high school reform model?

Background

Concerned that declining educational standards produced American mediocrity and an ill-equipped 21st-century workforce, Americans have demanded high school reform (Labaree, 2012; National Commission on Excellence in Education, 1983; Shealey, 2006). To counter this perceived lack of rising educational standards, policymakers established high-stakes accountability measures for public schools (Ravitch, 2010) and business and civic leaders encouraged school districts to emulate international education practices (National Center on Education and the Economy, 2008). Despite the implementation of a variety of school reforms, Lasky et al. (2005) affirmed "the nation's best measure of school-age academic achievement, the National Assessment of Educational Progress (NAEP), has, over its three decades of existence, documented almost no progress toward demonstrable improvement in student performance" (p. 28). Even though recent high school results demonstrated an increase in proficiency in reading and math from 2005 to 2009, reading scores fell below 1992 results and achievement gaps among subgroups of students remained unchanged (National Assessment of Education Progress [NAEP], 2011). Graduation rates have trended slightly upward this past decade (Bruce, Bridgeland, Fox, & Balfanz, 2011), yet little evidence exists that indicates school districts have moved high school reform to scale and obtained significant gains in student achievement (Balfanz & Legters, 2004a; Earl, Torrance, & Sutherland, 2006; Stringfield & Datnow, 1998).

Targeting Reforms Toward Potential Dropouts

If future American prosperity depends on elevating each student's preparedness for a rapidly changing and globally competitive environment, then urban districts' reforms must impact their high schools' large numbers of students who drop out (America's Promise, 2010). Dropouts tend "to be unemployed, living in poverty, receiving public assistance, in prison, on death row, unhealthy, divorced and ultimately single parents with children who drop out of high school themselves" (Bridgeland, Dilulio, & Morrison, 2006, p. 2). With low-skill work increasingly outsourced internationally and commanding low wages intranationally, dropouts have affected the general economy by limiting their wage-earning power, resulting in a loss of billions of dollars in family incomes and American tax revenue (Harlow, 2003; Land & Legters, 2002; Levin, Belfield, Muennig, & Rouse, 2007; Wald & Losen, 2003).

Table 1. Students' End-of-6th Grade Measures Predictive of 60% Dropout Probability.

Below 80% attendance
End-of-course failure in math
End-of-course failure in English
Suspended or low end-of-course conduct grade

Note. Students meeting at least one of these four criteria at the end of sixth grade have a 60% chance of dropping out of school.
Source: Adapted from Balfanz, Herzog, and MacIver (2007).

America's large numbers of dropouts have exacerbated such problems. Balfanz and Legters (2004b) reported, "between 1993 and 2002, the number of high schools with the lowest levels of success in promoting freshmen to senior status on time, a strong correlate of high dropout and low graduation rates, increased by 75%" (p. 4). To preserve and strengthen America's competitiveness, urban high school reform must convert potential dropouts into well-educated workers in a knowledge-based economy (Toch, 2003).

Middle school prevention:Too late. Recent studies recommended that urban districts attack the dropout epidemic through early identification and prevention. For instance, in a study of on-time graduation characteristics of freshmen in the Chicago Public Schools, Allensworth and Easton (2005) found students who entered high school from the bottom quarter of their eighth grade were more than 40% off track to graduate by the end of freshman year. Balfanz, Herzog, and MacIver (2007) examined longitudinal data of 12,972 Philadelphia public school students and concluded that urban districts could use sixth-grade individual factors of low attendance, failure of math or English, or suspensions to identify and prevent 60% of potential high school dropouts (Table 1).

Additional studies of large urban districts have affirmed that students entering ninth grade overaged, chronically absent, from low-income families, presenting poor behavior and low-achievement scores exhibit a higher dropout rate in high school and require middle school interventions (Alexander, Entwisle, & Kabbini, 2001; De Wit, Karioja, & Rye, 2010; MacIver, 2010; MacIver, Durham, Plank, Farley-Ripple, & Balfanz, 2007; Neild & Balfanz, 2006; Zvoch, 2006). Questioning the impact of high school adult advocacy on student engagement, attendance, on-time promotion, and graduation, MacIver (2011) found no significant effect on dropouts and concluded "relatively well-implemented strategies that are research-based to prevent dropout will not necessarily yield positive effects unless systematically linked to a complete framework that begins at least in the middle schools" (p. 181).

These studies suggested successful transition from middle school to high school as the key to significant reduction in high school dropout rates.

Unfortunately, middle school prevention strategies for at-risk students provide no remedy for the students currently in PLA high schools or in districts without early warning systems. In large urban districts, lagging indicators arrive too late for high school educators to determine the effectiveness of previously implemented strategies and target struggling students for specific remediation (Mishook, Foley, Thompson, & Kubiak, 2008). This lack of prevention compels educators in large urban districts to own, transform, and prepare the students who are in the desks in front of them for proficient performance on annual state assessments.

Whole school reform: Too slow and sporadic. Lacking the luxury of middle school interventions to improve at-risk student performance, urban district leaders have struggled to implement and sustain reform at scale. Educational leaders concerned by the performance of public schools have attempted a multitude of initiatives to increase student achievement. Charter schools (Bifulco & Ladd, 2006; Lubienski, 2003), trimester schedules (Lybbert, 1998; Winn, Menlove, & Zsiray, 1997), career academies (Kemple & Willner, 2008; McLaughlin, 1990), and school choice (Hoxby, 2003; Levin, 2001) have been attempted, but none has proven dramatically successful. Organizational bureaucracy has compounded the problem, stifling the accelerated change required to reform the high percentage of low-performing urban high schools (Bryan, Klein, & Elias, 2007; Payne, 2008).

Despite pockets of possibility providing hope for scale up, Earl et al. (2006) argued, "there are no examples anywhere of successful whole district high school reform. There are a few high schools, here and there, that have improved significantly, but none as a group" (p. 126). For instance, after examining the impact of Chicago Public Schools' reforms, Bryk, Sebring, Allensworth, Luppescu, and Easton (2010) emphasized the incredible difficulty of expanding a school's successful program for disadvantaged students into an organization-wide reform. The Baltimore City Public School System (BCPSS) has replicated some successful reform elements such as freshman academy and targeted support of at-risk students, yet has not expanded the Talent Development Model to scale across the district despite significant increases in performance found at Patterson High School (McPartland, Balfanz, Jordan, & Legters, 1998). In contrast, Gambone, Klem, Summers, Akey, and Sipe (2004) demonstrated the successful turnaround of all high schools after district implementation of First Things First reform model. However, Steve Gering, deputy superintendent for the Kansas City Schools, acknowledged that system-wide changes did not occur for 3 to 4 years (Viadero, 2005).

Whole-school coordination of high school staff and supports that have generated successful outcomes have typically required time, organization, and resources that were valuable for long-term reform, but impractical for producing short-term, district-wide increases in student achievement. In a longitudinal, mixed-method case study, Stringfield and Yakimowski-Srebnick (2005) found that in 6 years, the effects of accountability-driven reforms in BCPSS of testing, governance, and federal No Child Left Behind (NCLB) legislation generated a 13.1% gain in the high school graduation rate. In a study of the San Diego Unified School district, Darling-Hammond et al. (2005) found over 5 years that due to systemic improvements in principal instructional leadership, higher level course offerings to all high school students, and extended learning opportunities, the percentage of high schools that met state and subgroup NCLB targets increased from 19% to 56%. The Chicago Public Schools' reforms targeted students on track for graduation and decreased the dropout rate by more than 4% over 8 years (Allensworth & Easton, 2005, 2007; Bryk, Sebring, Kerbow, Rollow, & Easton, 1998; Hess, 2003).

In addition, the North Carolina New Schools Project launched by the governor and the North Carolina Education Cabinet in 2003 created schools that reported an average 1.3% smaller dropout rate than comparison schools across the state. Almeida, Steinberg, Santos, and Le (2010) reported that the New York City Public schools established 42 high schools that graduated above average numbers of "over-age and undercredited" students. Useem, Offenberg, and Farley (2007) studied the School District of Philadelphia's attempt to affect student outcomes through improving teacher certification and quality. Despite multiple improvements in the hiring process, they concluded the school placement process continued to make it difficult to "move fast in hiring the best and brightest in a timely way" (Useem et al., 2007, p. 20). Unfortunately, these programs and projects improved only a percentage of the schools and took several years to demonstrate gains. Although some high school reforms have shown promise, Balfanz, Legters, West, and Weber (2007) advised that even leading-edge reforms could take 4 years to move struggling students in PLA schools to proficient performance.

Improving student perceptions:Too little for dramatic gains. Interventions specifically designed to improve academic skills might require improvement of academic self-concept, or the belief in one's academic ability. In a study of 1,211 secondary students in Australia, researchers found that student levels of academic self-concept predicted measures of school disengagement (Bodkin-Andrews, O'Rourke, Dillon, Craven, & Yeung, 2009). Marsh and Craven (2006) reported that self-concept of academic abilities yielded stronger student outcomes than self-esteem. Based on a longitudinal data, Marsh and

O'Mara (2008) developed a "reciprocal effects model" (p. 549) and established that increased academic self-concept led to improved performance, which cyclically led to further increases in academic self-concept.

Evidence has indicated that students' perceptions of improved climate and a sense of belonging influenced their effort and performance, which should bolster high school reform (De Wit et al., 2010; Rumberger & Lim, 2008). Yet, findings also have demonstrated that student perception of peer, teacher, and emotional support typically decreases at the secondary level (Barber & Olsen, 2004; Furrer & Skinner, 2003; Marks, 2000; Reschly, Huebner, Appleton, & Antaramian, 2008). The contradictory findings highlight the need to improve high school students' self-confidence and cultivate caring learning environment. However, according to Bryk, Sebring, et al. (2010), without simultaneous improvements in teaching, improved student confidence and perceptions are unlikely to generate "substantial improvements in student learning" (p. 17). In fact, Rivera-McCutchen (2012) found that a variety of reforms targeting affective needs of low-performing students failed to sufficiently impact their academic performance.

National reform recommendations for our most disadvantaged students include middle school intervention (Balfanz et al., 2007), whole high school reform (Nunnery, 1998), and student-perception adjustments (Baker, 2006). Although necessary, for the students sitting in the desks in front of us, who are just months away from the next high-stakes assessment, none of these reforms offers a sufficient solution to help them. Future middle school prevention programs would arrive too late for current high school students. Similarly, whole school reform efforts have take years to demonstrate gains, which would not help currently enrolled and low-achieving high school students. Adjusting student attitudes alone delivers too little impact on reading and math test scores.

JCPS High School Reform: Project Proficiency (PP)

In 2010, the Kentucky Department of Education (KDE) identified 10 PLA high schools in JCPS, and noted that several others operated on the verge of PLA status. To address the performance deficiencies of these schools, JCPS district leaders developed a high school reform strategy known as PP. The objectives of PP included three challenging goals (JCPS, 2011). The first goal was to generate substantial gains in reading and math proficiency. Previously implemented high school reforms in JCPS, such as leadership development, school choice, smaller learning communities, inquiry-based curriculum, data-tracking systems, and equitable funding, had yielded positive, but incremental gains. From 2007 to 2010, many of the JCPS PLA and near-PLA high

Table 2. Reading and Math Student Proficiency Rates (%), 2007–2011.

	Proficient reading students				
	2007	2008	2009	2010	2011
JCPS PLA	46.1	46.5	44.8	43.2	57.2
JCPS	63.0	63.6	62.0	62.5	70.2
State	60.2	60.0	61.8	61.3	65.9
	Proficient math students				
	2007	2008	2009	2010	2011
JCPS PLA	26.8	23.7	25.8	20.0	36.7
JCPS	43.6	41.8	43.0	40.4	54.5
State	39.3	38.5	41.2	40.3	46.0

schools demonstrated sporadic increases in average proficiency in reading and math, but their scores still ranked in the bottom 5% of state scores (Table 2).

The second goal of PP was to quickly improve achievement. In addition to substantial gains, the new accountability requirements expected JCPS leaders and educators to demonstrate improvements of PLA high schools in a short period of time. Results from annual state testing potentially triggered sanctions of staff removal, charter take-over, and school closing for the "lowest-achieving schools in improvement, corrective action, or restructuring in a state" (School Improvement Fund, 2010). To avoid such sanctions, JCPS leaders sought an initiative that prepared students to perform proficiently on the forthcoming round of state assessments.

The third goal of PP was to propagate the reform throughout all PLA high schools. The volume, as well as the student composition of the PLA schools forced JCPS district leaders to move PP to scale (Coburn, 2003). The large numbers of at-risk high school students in these schools, typical of large urban districts (Balfanz et al., 2007; Darling-Hammond et al., 2005; Earl et al., 2006; Stringfield & Yakimowski-Srebnick, 2005), compelled JCPS leaders to move beyond targeted reform for students with near-proficient scores at selected schools. Instead, JCPS leaders developed PP as a broad reform for every high school math classroom in PLA schools.

Combining effective practices gleaned from previous JCPS high school initiatives and current school turnaround strategies, JCPS administrators and teacher leaders developed and launched PP across PLA high schools for the 2010-2011 school year (Appendix A). To accomplish the three objectives of PP, district leaders and high school educators developed four fundamental

and interrelated strategies. First, for each fixed grading-period, district officials, curriculum specialists, and teacher leaders collaboratively reduced the focus of core math courses to three key standards and a corresponding summative assessment. Second, operating within professional learning communities (PLCs; R. DuFour & Eaker, 1998; Eilers & Camacho, 2007), teachers of common courses "co-constructed" (Datnow & Stringfield, 2000, p. 188) lessons, tasks, and interventions through which students could demonstrate competency of each key standard. Third, PP relied on standards-based teaching and assessment of student work with the intent of guaranteeing competency for each key standard regardless of a student's starting point or background knowledge. Fourth, although PP promoted complete remediation and competency "before" rather than "after" each summative assessment, educators in PP schools guided students who scored below 80% on the summative assessment to recover missed content until students earned 80% or higher.

Although schools shared the same three objectives and four strategies, district administrators empowered building principals to effectively implement PP. This flexible implementation design enabled building leaders and educators to capitalize on the local contexts of their schools to maximize organizational strengths to attend to individual deficiencies. Despite the local differences, district officials reported a number of common elements of the PP implementation. For instance, through reteaching, differentiated instruction (Tomlinson & McTighe, 2006), redesign of student tasks to assess student competency (Stiggins, 2008), responsive interventions (Fuchs & Fuchs, 2006), and standards-based grading (Guskey, 2009; Lekholm & Cliffordson, 2008; Marzano, 2010), teachers collectively sought to ensure student competency for each key standard and acquired a shared knowledge base of effective instructional practices for subsequent grading periods (Allen & Blythe, 2004; R. P. DuFour, DuFour, Eaker, & Karhanek, 2004).

In addition, district administrators provided PP educators with a system for tracking student demonstration of competency, which utilized web-based technology, titled Classroom Assessment System and Community Access Dashboard (CASCADE). Results of scanned multiple-choice diagnostic assessments provided early indications of competency or misunderstanding of each key standard. Throughout the grading period, teachers posted online when students demonstrated competency for a standard through daily tasks, contributing to a school dashboard of progress by standard, student, and teacher. Teachers also manually entered a grade for student reflection. At the end of the grading period, teachers scanned into CASCADE a summative proficiency assessment for each student, which produced reports for an initial test score and an item analysis to guide student recovery of missed standards.

Although by definition, standards-based evaluation of student understanding should not result in grades, the landscape of tradition, Carnegie units, and grade-point averages required by competitive colleges demanded final grades. Consequently, CASCADE converted students' posted data of their demonstrations of competency, or assessments for learning, into a daily grade, formerly derived from an average of homework, quiz, and project scores. Technology provided PP educators the ability to genuinely exercise standards-based grading, allowing them to determine grades by evaluating understanding of standards rather than measuring each student's cumulative activity through an average of scores. CASCADE generated a recommended set of final 6 weeks grades based on 20% student reflection (effort), 40% standards-based assessment for learning (competency), and 40% summative assessment (proficiency).

Despite the existence of PP systems, structures, and expectations, the design did not mandate standard operating procedures (SOP). Unlike other high school reform initiatives, PP did not prescribe a set of tightly coupled SOP (Gambone et al., 2004; McPartland et al., 1998). Instead, PP used guiding principles similar to those of Highly Reliable Organizations (HROs) and utilized by the successful Neath-Port Talbot (NTE) Local Education Authority in southern Wales, Great Britain (Datnow & Stringfield, 2000; Stringfield, Reynolds, & Schaffer, 2008). Leaders and educators in NTE developed a finite set of shared goals, powerful databases, a balance of tight and loose SOP, and collegial decision making (Dee, Henkin, & Singleton, 2006; Jarrett & Stenhouse, 2011; Stringfield et al., 2008; Stringfield, Reynolds, & Schaffer, 2010). PP administrators facilitated innovation rather than manage prescription (Bryk, 2009; Henkin & Holliman, 2009). Tight expectations of PP included usage of key standards, standards-based grading, and ensured learning, but balanced these with loose expectations for processes of local implementation. Rather than lock-step compliance, PP-enabled educators to communicate emerging successful instructional strategies and add to the collective instructional knowledge base of their schools and district.

The positive results from 2011 state testing reinvigorated JCPS staff, inspired many students, and provided the community with the hope that effective reform had turned around PLA schools. However, JCPS district officials questioned whether PP could generate continuous improvement beyond 1 year of implementation. Stringfield and Datnow (1998) asserted that reform efforts aimed at urban districts have lacked systemic sustainability. Payne (2008) concluded that while urban districts tinkered with a variety strategies for improving performance of disadvantaged students, "we don't know how to implement these things with fidelity at scale" (p. 94). Significant increases in proficiency by JCPS at-risk students could provide supporting

evidence that PP represented a potentially scalable and effective high school reform. The purpose of this study is to examine if PP was associated with achievement gains of a segment of at-risk students.

Method

Participants

JCPS enrolled approximately 100,000 students in 90 elementary schools, 25 middle schools, 21 high schools, and 20 alternative settings. The alternative settings included schools for pregnant teens, zero-tolerance offenders, adjudicated students, dropout candidates, and state agency students. The composition of the district's student body included 56.5% White, 36% Black, and 7.5% Other students. More than half of JCPS students resided in single-parent homes and approximately 63% qualified for FRL. In addition, schools in JCPS served 14% of Kentucky's total student population and nearly 50% of its African American students.

The initial sampling frame included all JCPS students who attended the district's 11 PLA or near-PLA high schools during the 2009-2010 (2010) and 2010-2011 (2011) school years. We divided the sample into two cohorts. The comparison group, the non-PP cohort, took the 11th-grade math and social studies Kentucky Core Contest Tests (KCCT) in 2010 without participating in PP. The treatment group, the PP cohort, took 11th-grade math and social studies KCCT in 2011 after participating in PP.

Through purposive, nonprobability sampling (Trochim & Donnelly, 2008), we further reduced both cohorts to focus on a segment of at-risk students. We narrowed each cohort to include only students with corresponding 8th- and 11th-grade math and social studies KCCT scores. The non-PP cohort students had corresponding scores from 2007 and 2010 and the PP cohort students had corresponding scores from 2008 and 2011. We further winnowed the cohorts by selecting only students who finished their sixth-grade year with at least one of four dropout-predictive criteria researched by Balfanz et al. (2007). The final non-PP cohort included 241 6th-grade students of 2005, and the final PP cohort included 264 6th-grade students of 2006 who met the same criteria (Table 3).

Although the study did not examine dropout-prevention initiatives, the Balfanz et al. (2007) criteria provided an independent and valid operational definition of at-risk students. The additional requirements utilized in the sampling strategy resulted in a group of students who persisted with their education, when others with similar predictive criteria had dropped out of school. This restricted sample demonstrates weak external validity with the greater

Table 3. Cohort Characteristics of Students At-Risk of Dropping Out.

	Non-PP cohort		PP cohort	
Characteristic	*n*	%	*n*	%
Race/ethnicity				
White	64	26.56	79	29.92
Black	170	70.54	173	65.53
Other	7	3.90	12	4.55
Gender				
Male	166	68.88	172	65.15
Female	75	31.12	92	34.85
FRL	189	78.42	212	80.30
ECE	58	24.07	67	25.38
ESL	5	2.07	13	4.92

Note. FRL = Free or reduced price lunch. ECE = Exceptional Child Education (special education eligible). ESL = English as Second Language.

population of students who drop out of school. However, the identified sample consisted of students who demonstrated academic difficulty for multiple years yet remained diligent in their pursuit of a high school diploma. We later discuss potential limitations resulting from this sampling strategy.

Measures

To assess student achievement, we analyzed math and social studies scores from the state-administered KCCT. To promote test security, KDE officials administer multiple versions of the KCCT in each subject area. The KCCT Technical Guide identifies Cronbach's alpha measures to report internal consistency. Each of the 8th-grade mathematics and social studies tests consisted of six test versions, with $\alpha = .89$. The 11th-grade mathematics and social studies tests also consisted of six test versions, with $\alpha = .90$. Item and description indices were identified for each test version and converted to mean scale scores from 0 to 80. For state and NCLB reporting purposes, KDE officials divided mean scale scores into four nominal performance-level descriptors: novice, apprentice, proficient, and distinguished (Table 4).

We analyzed mean scale scores from 8th- and 11th-grade students in math and social studies to compare growth within the cohorts, as well as performance between cohorts. However, the different subject areas and years produced differing score distributions. Mean differences between math and social studies differed by no more than 2 points, but initial 8th-grade means

Table 4. KCCT Math and Social Studies Mean Scale Score Range and Performance Descriptors.

	Performance level description range			
	Novice	Apprentice	Proficient	Distinguished
8th grade	0-19	20-39	40-62	63-80
11th grade	0-19	20-39	40-63	64-80

Note. KCCT = Kentucky Core Contest Tests.

ranged from 3 to 6 points higher than corresponding 11th-grade means. To ameliorate the differences in mean scale scores from different years, we group mean-centered the scale scores for each student around the appropriate annual state mean scale score. To calculate these group-centered scores, we subtracted the state mean scale score from each student's scale score. The adjusted set of scores resulted in values that communicated the performance of the students relative to the mean performance of all students in that given subject, grade, and year. (Table 5)

Design and Procedures

We designed the study to test the assumption that PP was positively associated with KCCT math performance for a segment of students at-risk of dropping out of high school. We first collected descriptive statistics for each cohort to determine whether PP was associated with increased student achievement between the 8th and 11th grades. We then conducted paired-sample *t* tests to determine the strength of the relationships between the 8th-grade and corresponding 11th-grade test scores and compared the mean differences. Finally, because the study lacked random assignment, we used a quasi-experimental, nonequivalent groups design (Shadish, Cook, & Campbell, 2002; Trochim & Donnelly, 2008) for a pretest–posttest comparison of KCCT math and a pretest–posttest comparison of KCCT social studies mean-centered scale scores for the non-PP and the PP cohorts. The resulting contrasts compared the performances of similar students from the 11 PLA schools in different years, as well as the performance of the same students in the 11 PLA schools in different subject areas.

The PP treatment was administered only to 11th-grade students in the subject area of math 2011. Therefore, we compared social studies scores with math scores for PP students to expose possible historical validity threats during the PP year. Equally improved scores in math and social studies for the PP cohort would reduce the validity of claims about the strength of PP's influence

Table 5. KCCT Mean Scale Scores for 8th and 11th Grade Math and Social Studies.

	Mean math score			Mean social studies score		
	Cohort	State	Centered cohort	Cohort	State	Centered cohort
Non-PP cohort						
2007 8th grade	27.44	39	−11.56	29.54	41	−11.46
2010 11th grade	22.54	36	−13.46	24.75	35	−10.25
Mean change	−4.90	−3	−1.90	−4.79	−6	1.21
PP cohort						
2008 8th grade	27.47	41	−13.53	28.01	42	−13.99
2011 11th grade	28.87	37	−8.13	24.39	36	−11.61
Mean change	1.40	−4	5.40	−3.62	−6	2.38

Note. KCCT = Kentucky Core Contest Tests; PP = project proficiency.

on student performance. Similarly, significant differences between math and social studies scores would suggest an association between PP and student math achievement.

For data analyses, we performed independent-samples *t* tests for the equality of means to determine the demographic comparability of the non-PP and PP cohorts on the following variables: minority membership, gender, FRL status, Exceptional Child Education (ECE; special education), and English as a second language (ESL) designation. We then conducted a second independent-samples *t* test to compare the respective 8th-grade pretest means of the KCCT math and social studies mean-centered scale scores. In addition to the validity threats of differences in demographics and initial academic performance inherent in nonequivalent group design, we considered historical and maturation factors (Trochim & Donnelly, 2008). Finally, we conducted a one-way ANOVA to determine whether statistically significant differences existed between cohorts in the mean KCCT math gains and social studies gains from the 8th to 11th grades. These analyses used mean math or social studies gains as a dependent variable and PP as the independent variable.

Results

Cohort Comparability

We did not randomly select the non-PP and PP cohorts, and therefore began with tests to examine demographic and academic cohort comparability. A

Table 6. Independent-Samples *t* Test Comparing Demographic and Pretest Variables.

	Non-PP Cohort (n = 241)		PP cohort (n = 264)		Equality of means		
Demographic	*M*	*SD*	*M*	*SD*	*t*	*df*	*p*
Minority	.73	.44	.70	.46	.84	503	.40
Gender	.31	.47	.35	.48	−.89	503	.38
FRL	.78	.41	.80	.40	−.52	503	.60
ECE	.24	.43	.25	.44	−.34	503	.73
ESL	.02	.14	.05	.22	−1.76	458	.08
8th grade KCCT	*M*	*SD*	*M*	*SD*	*t*	*df*	*p*
Math	27.33	17.26	25.66	18.63	1.04	503	.30
Social studies	29.53	14.55	26.27	16.26	2.38	503	.02

Note. PP = project proficiency; FRL = Free or reduced price lunch; ECE = Exceptional Child Education (special education eligible); ESL = English as Second Language; KCCT = Kentucky Core Contest Tests.

MANOVA compared the student composition of the cohorts. No significant effect was found on the collective demographic variables, which prompted us to conduct subsequent univariate tests. Independent-samples *t* tests of non-PP and PP cohorts on the same factors also yielded no significant differences, indicating demographic comparability of the cohorts (Table 6).

To control for initial academic differences between cohorts in math and social studies, we established the KCCT 8th-grade scale scores as pretest covariates using an ANCOVA (Hinkle, Wiersma, & Jurs, 2003). After controlling for differences in the 8th-grade pretest scores, the ANCOVA indicated no significant differences in 11th-grade math means. Corroborating the ANCOVA results for math, an independent-samples *t* test comparing the means of 8th-grade KCCT math tests between non-PP and PP cohorts yielded no significant difference. In contrast, ANCOVA results indicated a significant difference between 8th-grade social studies pretest scores of the PP and non-PP cohorts. However, after controlling for these initial differences, further analyses revealed no statistically significant differences in the variance in 11th-grade social studies scores between the two cohorts, $F(1, 502) = 2.87$, $p > .05$. The absence of significant differences between initial math and social studies pretests affirmed academic comparability of the two cohorts.

Table 7. Performance Distribution of 11th Grade KCCT Math and Social Studies Proficiency Status.

			11th grade math status		
8th grade math status	Cohort	*n*	% Novice	% Apprentice	% Prof/Dist
Novice	PP	112	50.0	44.6	5.4
	Non-PP	94	70.2	26.6	3.2
Apprentice	PP	96	17.7	58.3	24.0
	Non-PP	86	34.9	57.0	8.1
Prof/Dist	PP	56	3.6	19.6	76.8
	Non-PP	61	11.5	55.7	32.8
			11th grade social studies status		
8th grade social studies status	Cohort	*n*	% Novice	% Apprentice	% Prof/Dist
Novice	PP	100	71.0	28.0	1.0
	Non-PP	61	68.9	29.5	1.6
Apprentice	PP	102	29.4	59.8	10.8
	Non-PP	128	39.1	53.9	7.0
Prof/Dist	PP	62	6.4	46.8	46.8
	Non-PP	52	5.8	42.3	51.9

Note. KCCT = Kentucky Core Contest Tests; PP = project proficiency.

Changes in Proficiency Levels

Examination of the student-level NCLB proficiency designations in the 8th and 11th grades revealed positive results for the PP math cohort (Table 7). Compared with the non-PP cohort performance in math, PP students earned higher proficiency designations. For instance, of the 112 students in the PP cohort who earned an initial 8th-grade novice designation, 50% earned apprentice or higher on the 11th-grade test. Similarly, of the 96 students in the PP cohort who earned an apprentice designation in the 8th grade, 24% subsequently earned a proficient or distinguished mark in the 11th grade. In contrast, approximately 30% of the non-PP novice students exceeded this designation in the 11th grade, and only 8% of the non-PP apprentice students earned higher marks in the 11th grade.

In addition, a lower percentage of students from the PP cohort experienced a reversion in their proficiency designations. Among the 56 PP students who initially earned a proficient or distinguished designation, 77% remained at that level on the 11th-grade exam. Of the 96 PP cohort apprentice students,

Table 8. Comparison of Mean-centered Scale Scores for KCCT Math and Social Studies.

	8th grade		11th grade		Equality of means		
Cohort/subject	*M*	*SD*	*M*	*SD*	*t*	*p*	*D*
Non-PP cohort							
Math	−11.56	17.25	−13.46	14.87	2.19	.03	.14
Social studies	−11.46	14.56	−10.25	15.06	−1.48	.14	.10
PP cohort							
Math	−13.53	18.24	−8.13	17.81	−5.12	.00	.32
Social studies	−13.99	16.32	−11.61	15.83	−2.55	.01	.16

Note. KCCT = Kentucky Core Contest Tests; PP = project proficiency.

17% reverted to a novice level on the 11th-grade exam. In contrast, 33% of the non-PP proficient or distinguished students continued their success on the 11th-grade exam, and 35% of the non-PP apprentice students reverted to a novice designation.

Similar comparisons did not reveal similar patterns of student performance in social studies. On the social studies tests, the PP and non-PP revealed comparable patterns of increase and decrease at all three proficiency designations. Unlike performance on the math test, the PP cohort did not demonstrate greater performance than the non-PP cohort on the social studies test. The PP and non-PP cohorts on the social studies test did not exhibit the positive designation increases, or the reduction in designation reversions. Comparison of the performance of these nonequivalent control groups reduced the historical threat to validity, and strengthened claims about the association between PP and the academic achievement of some at-risk students.

Comparisons of Mean-Centered Scale Scores

For both cohorts, a Pearson correlation coefficient revealed a moderately strong positive relationship between each pair of math and social studies assessments (Hinkle et al., 2003). These correlations confirmed that students with higher scores on 8th-grade KCCT tended also to have higher scores on their 11th-grade tests, regardless of subject area, which suggested comparability of the 8th- and 11th-grade tests. We performed paired-sample *t* tests to compare mean KCCT math and social studies scores from 8th to 11th grade (Table 8). The most prominent finding was the statistically significant increase by the PP cohort from 8th- to 11th-grade math complemented by the

practical significance of a small to medium effect size of .32 ($d = .32$; Cohen, 1988). Results also indicated a significant increase for social studies for the PP cohort from 8th to 11th grade. However, the effect size of .16 ($d = .16$) was in a range that Cohen (1988) identified as below small, and after controlling for a pretest covariate, the previously reported ANCOVA revealed no statistically significant differences in the variance in 11th-grade social studies scores between the two cohorts.

In addition to a statistically significant increase in KCCT math scores from 8th to 11th grade, the average mean-centered score on the math test of the PP cohort shifted from 13.53 points below the state mean in the 8th grade to 8.13 points below the state mean in the 11th grade. In contrast, for the non-PP cohort, the distance in the average mean-centered score from the state mean widened in math by approximately 2 points from 11.56 points to 13.46 points below the state mean. PP was also associated with the relationship between prior achievement and future performance, as Pearson correlations revealed a decrease in the relationship between 8th-grade and 11th-grade mathematics achievements, with values of $r = .55$ and .66 in PP and non-PP cohorts, respectively. Since prior achievement generally predicts future performance (V. E. Lee & Bryk, 1989; Raudenbush, 2004; Raudenbush & Willms, 1995), the reduced correlation strength between prior achievement and posttest math scores, combined with the increased mean, indicated that lower performing students on the 8th-grade math test scored at higher levels on the 11th-grade math test. These results suggested that PP was associated with increased math proficiency among initially low-achieving students.

Finally, we conducted a one-way ANOVA to compare the mean gains in KCCT math mean-centered scale scores of the two cohorts from 8th grade to 11th grades. Previous ANCOVA results revealed no significant differences in social studies scores, and thus eliminated the need for conducting an ANOVA for social studies. Corroborating the previous evidence of the influence of PP on 11th-grade KCCT math scores, we found a significant difference between cohort gain scores, $F(1, 503) = 49.42$, $p < .01$. Results indicated that students in the non-PP cohort decreased in their KCCT math gain score by nearly 2 points ($M = -1.79$, $SD = 13.4$), whereas the PP cohort increased by approximately 7 points ($M = 7.05$, $SD = 14.7$). Due to the use of an ANOVA to compare two cohorts, we applied omega squared (Hinkle et al., 2003) to estimate the effect size, with $\omega^2 = .09$, which complemented statistical significance with practical significance. Borman, Hewes, Overman, and Brown (2003) concluded from a meta-analysis of comprehensive school reform that researchers can expect between a .09 and .15 effect size for district-wide samples for school reforms that "go beyond the effect of Title I" (p. 35).

Discussion

This study indicated that when compared with a previous cohort of students with similar at-risk factors predictive of dropping out (Balfanz et al., 2007), PP students achieved statistically significant greater gains in the KCCT math scale scores between the 8th and 11th grades. This study found that the centered mean math gain for at-risk students in the non-PP cohort actually decreased by nearly 2 points from the 8th to the 11th grades, whereas the PP cohort increased by nearly 5 points, a statistically significant and educationally meaningful reversal. Students from the PP cohort also improved their proficiency designations from 8th to 11th grade at a greater rate than non-PP students. In addition, students from the PP cohort demonstrating math proficiency in the 8th grade maintained that designation at a greater rate than their peers from the non-PP cohort.

To further strengthen the nonequivalent group design, social studies scores were examined alongside math scores for PP cohort students. Given that social studies was not connected with PP and had the least similarities with math content, similar patterns among results of math and social studies could exclude a relationship between PP and math gains. However, within the PP cohort, for the same students taking a different test we found no significant difference in social studies gains, even after controlling for an 8th-grade pretest covariate. Therefore, statistically significant achievement gains, considerable increases in proficient performance of at-risk students across the district, and a noteworthy effect size strengthened the credibility of PP as an effective reform for a segment of at-risk students currently enrolled in urban high schools.

Design Considerations

To complement the established external validity of the KCCT state assessments, the internal validity of this study was strengthened by demonstrating no statistically significant demographic or academic differences between the non-PP and PP cohorts of at-risk students. However, we only included students who had not dropped out, had arrived without delay to the 11th grade, and had corresponding 8th- and 11th-grade test scores, arguably excluding some of each cohort's most struggling students. Although we excluded a segment of at-risk students from the non-PP and PP cohorts, the identified sample represented an important component of any educational reform. The sampled students had demonstrated persistence toward graduation, yet maintained poor academic achievement. To quickly improve school performance

metrics, improvement initiatives must address the current generation of students who sit in front of us.

As with all research and implications, this study contains limitations that potentially temper the findings. First, the study examined the effects of a single-year treatment. A multiyear study would provide not only trend data, but also opportunities for additional types of analyses. Second, although the non-PP and PP cohorts proved statistically comparable, PP was implemented amid the threat of state sanctions and staff removal, which could have influenced students' and teachers' performances. Yet, social studies scores did not increase. Third, all but 1 of the 11 schools in the study received some level of state assistance that may have influenced achievement gains. Four schools received state-provided resource teachers, substantial financial resources for professional development, and stipends for extended staff time, and six schools received moderate supplementary funding. However, given the amount of evidence and results, we assert that such confounds slightly mitigate rather than negate PP as an effective and scalable high school reform.

A fourth limitation, related to the data collected about the initiative, also potentially tempers the findings. Aside from anecdotal evidence provided by one of the authors who assisted with the development and implementation of PP, we had no evidence regarding fidelity of implementation. As outlined by the PP guidelines, the study assumed that, on average, the PLA schools' math teachers implemented district curriculum, evaluated student competency through standards-based evaluation of work, sought to guarantee student competency of key standards each grading period, and ensured a fail-safe score on each grading-period summative assessment (Appendix B). Although these guidelines existed, PP did not include a list of SOP like other school reform models. Instead, akin to the HRO reform studied by Stringfield et al. (2008, 2010; Stringfield, Reynolds, & Schaffer, 2012) and suggestions by Bryk, Gomez, and Grunow (2010), district leaders empowered principals to manage innovation rather than ensure compliance. From the disparate results of the PP and non-PP cohorts we inferred that under the PP reform, teachers' instruction and students' learning changed. Yet, we have no valid or reliable data that observed altered actions of principals, teachers, or students.

Finally, the sampling strategy, which relied on dropout-predictive criteria for students attending PLA schools, identified a specific sample of students that included only those who persisted to the 11th grade. As measured by Averaged Freshmen Graduation Rate, which was the calculation utilized by state officials during the studied time period, all 11 schools in the study demonstrated low graduation rates during the time that sampled students would have attended them (Table 9). Efforts to increase graduation rates could have

Table 9. Averaged Freshman Graduation Rate Metrics for Schools in the PP Sample.

	2008	2009	2010	2011
9th grade membership	368.6	390.3	368.4	368.6
10th grade membership	280.2	302.5	292.0	296.7
AFGR average membership	324.4	346.4	330.2	332.7
Graduates	198.5	207.6	191.2	198.7
AFGR	60.7	59.1	57.2	59.6

Note. Average Membership = 9th + 10th Membership/2; PP = project proficiency; AFGR = graduates/average membership.

influenced the results. However, the cohorts consisted of the same PLA schools. This longitudinal comparison increased the reliability of the results, but potentially decreased the generalizability to not only student dropouts or less-persistent students, but also schools with higher graduation rates.

Implications for Practice, Policy, and Research

Establishing effective high school reform that rapidly moves substantial numbers of students to levels of proficiency across an urban district challenges educational practice, policy, and research. Darling-Hammond et al. (2005) affirmed, "high schools have presented a perennial challenge to school reform efforts" (p. 169). Although JCPS high schools realized significant gains for at-risk students in math after implementing PP in 2011, most of the PLA schools remained ranked among the lowest performing schools in Kentucky. In 2012, JCPS high school teachers in PLA and near-PLA high schools entered the second year of PP implementation facing new standards, completely different state assessments, six principal changes, and considerable restaffing of faculties. However, results from the previous year and productive instructional practices compelled schools to maintain the momentum of PP and the goal to guarantee competency of key standards for every student.

Guided by the overarching PP goal to guarantee key-standard competency for each student, districts should adopt the most reproducible elements of PP: to create conditions of urgency (Stringfield et al., 2008), instructional coherence of curriculum, instruction, and assessment (Cooper, 2000; P. Lee, 1999; Oxley, 2008), and "co-construction" (Stringfield & Datnow, 2002, p. 269) opportunities with teachers for implementation and decision making. Teachers should claim the power of collectively ensuring student learning by

collaborating to evaluate student understanding of standards, instead of settling for the averaging grades (Guskey, 2009; Lekholm & Cliffordson, 2008; Marzano, 2010). Teachers should also create common formative assessments to measure individual student progress, engage learners in self-reflection, and seek instructional implications (Stiggins & DuFour, 2009). Finally, teachers should adjust instruction and interventions to guide each student to demonstrate an acceptable level of competency in key standards.

The results of our study also indicated that legislators and school boards might provide fewer prioritized goals, invest in existing principals and teachers, and support systems and processes. Amid mandates to teach a growing number of new standards, PP provided practitioners with only three prioritized standards per 6-weeks grading period. Schools were provided common diagnostic and summative assessments on fewer standards, a goal each grading period to guarantee a level of individual student competence that a teacher respected, and the balance of supervision and support for general implementation. Rather than the replacement of faculty, policymakers should consider these PP practices and "enable schools to enhance the stability and professional capacity of staff members and the academic performance and active engagement of students" (Berman & Camins, 2011, p. 28).

Further study of the effectiveness of PP provides a challenge for researchers. Hargreaves and Fink (2006) proposed that after effectively implementing change, "the biggest challenge of all is to make it durable and sustainable" (p. 2). As suggested, researchers should monitor the performance of the schools for multiple years to determine the longitudinal success of the reform. Researchers might focus specifically on limitations of the current study by investigating implementation of the reform, as well as the influence of PP on other samples of students, such as dropouts or even college-going students. Similarly, focused on limitations, researchers might utilize different methodological and analytical techniques to assess the influence of PP, such as propensity score matching, interrupted time series, event history analysis, and others.

Evidence from the study supports that through PP, JCPS educators in 11 high schools moved high school reform to scale at the district level for students most at risk of dropping out. Although results were statistically and meaningfully significant, these schools have only begun their journey to move each student to proficient performance. To sustain and build from these gains, they must depend on additional expertise from educational peers, policymakers, and researchers to help them maintain their momentum of urban high school reform. Through PP, JCPS educators achieved significant gains with a segment of at-risk high school students, and this study provides evidence, traction, and hope for understanding elusive urban high school reform.

Appendix A

Design, Training, and Implementation Activity Timeline for PP

Timeline	Activity	Presenters	Recipients
April and May 2010	Designed PP structures and determined necessary support	HS Superintendent	HS Liaison PSMs Select Principals
June 2010	Discussed input that principals collected from their administrative staffs and department chairs	HS Superintendent HS Liaison	School Principals
July 2010	Refined PP structures and resources with critical stakeholders.	HS Superintendent HS Liaison	C/O Math Specialists Resource Teachers Teacher Leaders
	Commenced strategic and logistic planning of data tracking system.	HS Liaison C/O Math Specialists Select Principals	C/O Research Staff C/O Technology Staff
August 2010	Conducted input session and idea exchange with teachers' union representatives	HS Superintendent Resource Teachers	Union Representatives
	Visited PLA schools to provide training about key standards, standards-based teaching, learning, and grading, and logistics of web-based tracking of competency	HS Superintendent C/O Research Staff C/O Math Specialists Resource Teachers	School Principals Assistant Principals Math Teachers
September 2010 through May 2011	Regularly visited teacher learning teams to facilitate protocols of examining data, adjusting teaching, redesigning assessments, and revising intervention strategies.	HS Superintendent HS Liaison PSMs C/O Math Specialists Resource Teachers	Math Teachers
	During quarterly meetings principals presented competency data, shared tasks that generated student competency, and sought feedback about instructional leadership issues.	HS Superintendent HS Liaison	School Principals

Note. HS = High School; PSM = Priority School Manager; C/O = Central Office.

Appendix B

Evolution of Classroom Operational Procedures Expectations

	Before PP	After PP
Grading practices	Quantified student activity: Assigned points and averaged grades for, correct responses on tests/quizzes, submission of projects and notebooks, completion of daily class work and homework, pursuit of extra credit and extra effort.	Evaluated student understanding: Holistically judged student competency of standards from daily task performance. Converted evaluation of student conceptual understanding into a daily work grade. Required student self-assessment of progress toward demonstrating competency of each key standard. Guided each student to recover missed content on a final unit test until student scored 80% or higher.
Data usage	Summative testing of total content: Monitored comprehensive curriculum maps and pacing guides of all state-mandated content. Administered quarterly summative tests district-wide to measure growth in student proficiency. Examined previous year's state assessment scores to determine student performance gaps.	Formative assessing of key topics: Identified three key standards per six-week grading period. Administered district pre- and post-tests aligned with grading-period key standards to assess student growth. Through a web-based dashboard, teachers recorded and tracked students' daily progress toward competencies. Teacher teams weekly examined student work and reflections to assess learning gaps.
Student interventions	Remediated after summative tests: Provided remediation opportunities outside of class time for students with lower test scores. Remediation instruction included reteaching, self-paced worksheets with a teacher circulating to provide assistance, or independent work on computer program modules.	Remediated before summative test: Based on student pretest data, weekly competency totals, and student work samples, teacher teams collectively adjusted and reinvented student tasks, instructional strategies, formative assessments, and intervention time, space, and methods.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Alexander, K. L., Entwisle, D. R., & Kabbini, N. S. (2001). The dropout process in life course perspective: Early risk factors at home and school. *Teachers College Record*, *103*, 760-882.

Allen, D., & Blythe, T. (2004). *The facilitator's book of questions: Tools for looking together at student and teacher work*. New York, NY: Teacher's College Press.

Allensworth, E. M., & Easton, J. Q. (2005). *The on-track indicator as a predictor of high school graduation*. Chicago, IL: Consortium on Chicago School Research at the University of Chicago.

Allensworth, E. M., & Easton, J. Q. (2007). *What matters for staying on-track indicator and graduating in Chicago public high schools*. Chicago, IL: Consortium on Chicago School Research at the University of Chicago.

Almeida, C., Steinberg, A., Santos, J., & Le, C. (2010). Six pillars of effective dropout prevention and recovery: An assessment of current state policy and how to improve it. *Jobs for the Future*. Retrieved from http://www.jff.org/sites/default/files/DropoutBrief-090810.pdf

America's Promise Alliance. (2010). Available from http://www.americaspromise.org

Baker, J. A. (2006). Contributions of teacher–child relationships to positive school adjustment during elementary school. *Journal of School Psychology*, *44*, 211-229.

Balfanz, R., Herzog, L., & MacIver, D. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, *42*, 223-235.

Balfanz, R., & Legters, N. (2004a). Locating the dropout crisis: Which high schools produce the nation's dropouts? In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 57-84). Cambridge, MA: Harvard Education Press.

Balfanz, R., & Legters, N. (2004b). *Locating the dropout crisis: Which high schools produce the Nation's dropouts? Where are they located? Who attends them?* (Research Report No. 70). Washington, DC: Center for Social Organization of Schools, Johns Hopkins University.

Balfanz, R., Legters, N., West, T. C., & Weber, L. M. (2007). Are NCLB's measures, incentives, and improvement strategies the right ones for the nation's low-performing high schools? *American Educational Research Journal*, *44*, 559-593.

Barber, B. K., & Olsen, J. A. (2004). Assessing the transitions to middle and high school. *Journal of Adolescent Research*, *19*, 3-30.

Berman, S. H., & Camins, A. (2011, November). Investing in turnaround that endures. *Education Week*, *31*, 28.

Bifulco, R., & Ladd, H. F. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy*, *1*, 50-90.

Bodkin-Andrews, G. H., O'Rourke, V., Dillon, A., Craven, R. G., & Yeung, A. S. (2009, January). *Explaining away aboriginality: Causal modeling of academic self-concept and disengagement for indigenous and non-indigenous Australian students*. Paper presented at the Fifth Global SELF International Biennial Conference, United Arab Emirates University, Al Ain.

Borman, G. D., Hewes, G., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *73*, 125-230.

Bridgeland, J., Dilulio, J., & Morrison, K. (2006). *The silent epidemic: Perspectives of high school dropouts*. Washington, DC: Civic Enterprises, LLC.

Bruce, M., Bridgeland, J. M., Fox, J. H., & Balfanz, R. (2011). *On track for success: The use of early warning indicator and intervention systems to build a grad nation*. Washington, DC: Civic Enterprises, Johns Hopkins University.

Bryan, K., Klein, D., & Elias, M. J. (2007). Applying organizational theories to action research in community settings: A case study in urban schools. *Journal of Community Psychology*, *35*, 383-398.

Bryk, A. S. (2009). Support a science of performance improvement. *Phi Delta Kappan*, *90*, 597-600.

Bryk, A. S., Gomez, L. M., & Grunow, A. (2010). *Getting ideas into action: Building networked improvement communities in education*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.

Bryk, A. S., Sebring, P. B., Kerbow, D., Rollow, S., & Easton, Q. (1998). *Charting Chicago school reform*. Boulder, CO: Westview Press.

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*, 3-12.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cooper, R. (2000). Urban school reform from a student-of-color perspective. *Urban Education*, *34*, 597-622.

Darling-Hammond, L., Hightower, A., Husbands, J. L., LaFors, J. R., Young, V. M., & Christopher, C. (2005). *Instructional leadership for systemic change: The story of San Diego's reform* (Vol. 3). Lanham, MD: Scarecrow Education.

Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk*, *5*, 183-204.

Dee, J., Henkin, A., & Singleton, C. (2006). Organizational commitment of teachers in urban schools: Examining the effects of team structures. *Urban Education*, *41*, 603-627.

De Wit, D. J., Karioja, K., & Rye, B. J. (2010). Student perceptions of diminished teacher and classmate support following the transition to high school: Are they related to declining attendance? *School Effectiveness and School Improvement*, *21*, 451-472.

DuFour, R., & Eaker, R. E. (1998). *Professional learning communities at work: Best practices for enhancing student achievement*. Bloomington, IN: National Education Service.

DuFour, R. P., DuFour, R. B., Eaker, R. E., & Karhanek, G. (2004). *Whatever it takes: How professional learning communities respond when kids don't learn*. Bloomington, IN: National Educational Service.

Duncan, A. (2009, June 17). Start over. *Education Week*, *28*, 36.

Earl, L., Torrance, N., & Sutherland, S. (2006). Changing secondary schools is hard: Lessons from 10 years of school improvement in the Manitoba school improvement program. In A. Harris, & J. Chrispeels (Eds.), *Improving schools and educational systems* (pp. 109-128). London, England: Routledge.

Eilers, A., & Camacho, A. (2007). School culture change in the making: Leadership factors that matter. *Urban Education*, *42*, 616-637.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*, 93-99.

Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, *95*, 148-162.

Gambone, M. A., Klem, A. M., Summers, J. A., Akey, T. A., & Sipe, C. L. (2004). *Turning the tide: The achievements of the first things first education reform in the Kansas City, Kansas public school district*. Philadelphia, PA: Youth Development Strategies.

Guskey, T. R. (2009). *Practical solutions for serious problems in standards based grading*. Thousand Oaks, CA: Corwin Press.

Hargreaves, A., & Fink, D. (2006). *Sustainable leadership*. San Francisco, CA: Jossey-Bass.

Harlow, C. (2003). *Education and correctional populations*. Washington, DC: US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Henkin, A., & Holliman, S. (2009). Urban teacher commitment: Exploring associations with organizational conflict, support for innovation, and participation. *Urban Education*, *44*, 160-180.

Hess, A. G. (2003). *Changes in student achievement in Illinois and Chicago, 1990-2000*. Chicago, IL: Northwestern University.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston, MA: Houghton Mifflin.

Hoxby, C. (2003). School choice and school productivity: Could school choice be a tide that lifts all boats? In C. Hoxby (Ed.), *The economics of school choice* (pp. 287-341). Chicago, IL: University of Chicago Press.

Jarrett, O., & Stenhouse, V. (2011). Problem solution project: Transforming curriculum and empowering urban students and teachers. *Urban Education*, *46*, 1461-1495.

Jefferson County Public Schools. (2011). *Project proficiency guide*. Unpublished manuscript: Author.

Kemple, J. J., & Willner, C. J. (2008). *Career academies: Long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood*. New York, NY: MDRC.

Labaree, D. F. (2012). *Someone has to fail: The zero-sum game of public schooling*. Cambridge, MA: Harvard University Press.

Land, D., & Legters, N. (2002). The extent and consequences of risk in U.S. education. *Yearbook of the National Society for the Study of Education*, *101*, 1-288.

Lasky, S., Stringfield, S., Teddlie, C., Kennedy, E., Schaffer, E., Chrispeels, J., . . .McDonald, D. (2005). Designing and conducting a gold standard effective schools study. *Journal for Effective Schools*, *4*, 27-45.

Lee, P. (1999). In their own voices: An ethnographic study of low-achieving students within the context of school reform. *Urban Education*, *34*, 214-244.

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, *62*, 172-192.

Lekholm, A. K., & Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school levels: Effects of gender and family background. *Educational Research and Evaluation*, *14*, 181-199.

Levin, H. M. (2001). *Privatizing education: Can the marketplace deliver choice, efficiency, equity, and social cohesion?* Boulder, CO: Westview Press.

Levin, H. M., Belfield, C., Muennig, P., & Rouse, C. (2007). *The costs and benefits of an excellent education for all of America's children*. New York, NY: Columbia University, Teachers College.

Lubienski, C. (2003). Innovation in education markets: Theory and evidence on the impact of competition and choice in charter schools. *American Educational Research Journal*, *40*, 395-443.

Lybbert, B. (1998). *Transforming learning with block scheduling: A guide for principals*. Thousand Oaks, CA: Corwin Press.

MacIver, M. A. (2010). *Gradual disengagement: A portrait of the 2008-09 dropouts in the Baltimore city schools*. Baltimore, MD: Baltimore Education Research Consortium.

MacIver, M. A. (2011). The challenge of improving urban high school graduation outcomes: Findings from a randomized study of dropout prevention efforts. *Journal of Education for Students Placed at Risk*, *16*, 167-184.

MacIver, M. A., Durham, R. E., Plank, S. B., Farley-Ripple, E. F., & Balfanz, R. (2007). *The challenge of on-time arrival: The seven-year flight paths of Baltimore's sixth graders of 1999-2000*. Baltimore, MD: Baltimore Education Research Consortium.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal*, *37*, 153-184.

Marsh, H., & Craven, R. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science*, *1*, 133-163.

Marsh, H., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept, personality, and social psychology. *Personality and Social Psychology Bulletin*, *34*, 542-552.

Marzano, R. J. (2010). *Formative assessment and standards-based grading*. Bloomington, IN: Marzano Research Laboratory.

McLaughlin, M. W. (1990). The RAND change agent study revisited: Macro perspectives and micro realities. *Educational Researcher*, *19*, 11-16.

McPartland, J., Balfanz, R., Jordan, W., & Legters, N. (1998). Improving climate and achievement in a troubled urban high school through the talent development model. *Journal of Education for Students Placed at Risk*, *3*, 337-361.

Mishook, J., Foley, E., Thompson, J., & Kubiak, M. (2008, Winter). Beyond test scores: Leading indicators for education. *Voices in Urban Education*. Retrieved from http://www.annenberginstitute.org/VUE/winter08/Mishook.php

National Assessment of Education Progress. (2011). Retrieved from http://nces.ed.gov/nationsreportcard/pdf/about/2011471.pdf

National Center on Education and the Economy. (2008). *Tough choices or tough times: The report on the new commission on the skills of the American work force, revised and expanded*. San Francisco, CA: Jossey-Bass.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the secretary of education*. Washington, DC: Superintendent of Documents, Government Printing Office.

Neild, R. C., & Balfanz, R. (2006). *Unfulfilled promise: The dimensions and characteristics of Philadelphia's dropout crisis, 2000-2005*. Philadelphia, PA: Philadelphia Youth Network, Johns Hopkins University, and University of Pennsylvania.

Nunnery, J. A. (1998). Reform ideology and the locus of development problem in educational restructuring. *Education and Urban Society*, *30*, 277-295.

Oxley, D. (2008). Creating instructional program coherence. *Principal's Research Review*, *3*, 1-7.

Payne, C. M. (2008). *So much reform so little change: The persistence of failure in urban schools*. Cambridge, MA: Harvard Press.

Persistently Low-Achieving School and School Intervention Defined. Kentucky Revised Statutes 160.346 (2010).

Raudenbush, S. W. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?* Princeton, NJ: Educational Testing Service.

Raudenbush, S. W., & Willms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Sciences*, *20*, 307-355.

Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York, NY: Basic Books.

Reschly, A. L., Huebner, E. S., Appleton, J. J., & Antaramian, S. (2008). Engagement as flourishing: The contribution of positive emotions and coping to adolescents' engagement at school and with learning. *Psychology in the Schools*, *45*, 419-431.

Rivera-McCutchen, R. L. (2012). Caring in a small urban high school: A complicated success. *Urban Education*, *47*, 653-680.

Rumberger, R. W., & Lim, S. (2008). *Why students drop out of school: A review of 25 years of research*. Santa Barbara: California Dropout Research Project.

School Improvement Fund. (2010). *School Improvement Grants; American Recovery and Reinvestment Act of 2009 (ARRA); Title I of the Elementary and Secondary Education Act of 1965, as Amended (ESEA)*. Retrieved from http://www2.ed.gov/programs/sif/legislation.html

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

Shealey, M. W. (2006). The promises and perils of "scientifically based" research for urban schools. *Urban Education*, *41*, 5-19.

Stiggins, R. J. (2008). *Assessment manifesto: A call for the development of balanced assessment systems*. Princeton, NJ: Educational Testing Service.

Stiggins, R. J., & DuFour, R. (2009). Maximizing the power of formative assessments. *Phi Delta Kappan*, *90*, 640-644.

Stringfield, S., & Datnow, A. (1998). Scaling up school restructuring designs in urban schools. *Education and Urban Society*, *30*, 269-276.

Stringfield, S., & Datnow, A. (2002). Systemic supports for schools serving students placed at risk. In S. Stringfield, & D. Land (Eds.), *Educating at-risk students* (pp. 269-285). Chicago, IL: National Society for the Study of Education.

Stringfield, S., Reynolds, D., & Schaffer, E. C. (2008). Improving secondary students' academic achievement through a focus on reform reliability: 4- and 9-year findings from the high reliability schools project. *School Effectiveness and School Improvement*, *19*, 409-428.

Stringfield, S., Reynolds, D., & Schaffer, E. C. (2010, October). *Toward highly reliable, high-quality public schooling*. Paper presented at the McREL Best in the World Consortium Meeting, Denver, CO.

Stringfield, S., Reynolds, D., & Schaffer, E. C. (2012). Making best practice standard- and lasting. *Phi Delta Kappan*, *94*, 45-50.

Stringfield, S., & Yakimowski-Srebnick, M. E. (2005). Promise, progress, problems, and paradoxes of three phases of accountability: A longitudinal case study of the Baltimore City public schools. *American Educational Research Journal*, *42*, 43-75.

Toch, T. (2003). *High schools on a human scale: How small schools can transform American education*. Boston, MA: Beacon Press.

Tomlinson, C. A., & McTighe, J. (2006). *Integrating differentiated instruction & understanding by design: Connecting content and kids*. Alexandria, VA: Association for Supervision and Curriculum Development.

Trochim, W. M., & Donnelly, J. P. (2008). *The research methods knowledge base* (3rd ed.). Mason, OH: Atomic Dog Publishing.

Useem, E., Offenberg, R., & Farley, E. (2007). *Closing the teacher quality gap in Philadelphia: New hope and old hurdles*. Philadelphia, PA: Research in Action.

Viadero, D. (2005, July). Reform model found to spur gains in Kansas. *Education Week*, *24*, 10.

Wald, J., & Losen, D. (Eds.). (2003). *Deconstructing the school to prison pipeline. New directions for youth development (Number 99)*. San Francisco, CA: Jossey-Bass.

Winn, D., Menlove, R., & Zsiray, S. W. (1997). An invitation to innovation: Rethinking the high school day. *NASSP Bulletin*, *81*, 10-18.

Zavadsky, H. (2009). *Bringing school reform to scale: Five award-winning urban districts*. Cambridge, MA: Harvard Education Press.

Zvoch, K. (2006). Freshman year dropouts: Interactions between student and school characteristics and student dropout status. *Journal of Education for Students Placed at Risk*, *11*, 97-117.

Author Biographies

Joe Burks served 37 years in the Jefferson County Public Schools (JCPS), Louisville, Kentucky, as a secondary school teacher, principal, and assistant superintendent. He earned his EdD and MEd at the University of Louisville and his undergraduate BS at Centre College of Kentucky. In 2010-2011, he guided the development and implementation of Project Proficiency, a JCPS proficiency-based teaching, learning, and grading initiative that increased proficient performance in all 21 JCPS high schools, with 11 of the district's lowest performing high schools doubling the state's gain in reading and tripling the state's gain in math.

Craig Hochbein is an assistant professor of educational leadership at Lehigh University. His research utilizes quantitative analyses to examine the longitudinal development of school performance. Hochbein has specifically investigated factors associated with declining academic achievement and the effectiveness of policies intended to improve school performance.