008:15:LP:cm:LH:LP

DC Public Education Reform Amendment Act (PERAA) Report No. 3 Supplemental

December 3, 2014

The Education Consortium for Research and Evaluation (EdCORE)

EdCORE at The George Washington University

DC Public Education Reform Amendment Act (PERAA)
Report No. 3 Supplemental:

The Impact of Replacing Principals on Student Achievement in DC Public Schools (DCPS)

Submitted to the

ODCA Office of the District of Columbia Auditor

December 3, 2014

In collaboration with

MATHEMATICA Policy Research

The Education Consortium for Research and Evaluation (EdCORE) is led by the Graduate School of Education and Human Development at the George Washington University, in partnership with American Institutes for Research, Mathematica Policy Research, Policy Studies Associates, Quill Research Associates, RAND, and SRI.

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Graduate School of Education & Human Development
THE GEORGE WASHINGTON UNIVERSITY

Foreword

This is a supplemental to the third in a series of interim reports on the 2007 Public Education Reform Amendment Act (PERAA) prepared by EdCORE at the George Washington University under contract with the National Academy of Sciences (NAS). The reports provide analysis to inform the work of the NAS Committee on the Five Year Summative Evaluation of the District of Columbia Public Schools in response to the mandate for independent evaluation included in PERAA. The interim reports provide updates by school year, highlight trends across years in selected evaluation topic areas, and provide additional data collected for longer-term analysis.

With guidance from the Office of the District of Columbia Auditor (ODCA), and in adherence to the PERAA legislation, four broad topics have been the focus of EdCORE's inquiry:

- Business practices and strategies, including organizational structure and roles, financial management, operations management, facilities and maintenance; resource allocations; public accountability; interagency collaboration; and stakeholder engagement and responsiveness.

- Human resources operations and human capital strategies, including the number (and percentage) of highly qualified teachers; retention rates for effective teachers; schools and wards served by effective teachers; length of time principals and administrators serve; types of leadership strategies used; and responsibilities of central office versus school level leadership.

- Academic plans, including integration of curriculum and program specific focus into schools and grade progression and credit accumulation.

- Student achievement, including a description of student achievement that includes academic growth; proficiency; and other (non-academic) educational outcomes.

The previous EdCORE reports provided annual snapshots of each topic and preliminary trend analysis in the areas of business practices and academic planning. This third report addresses human resources and includes two parts and a supplemental addendum. Part I (originally released June 30, 2014), focuses on teacher effectiveness (as measured by IMPACT) and retention in District of Columbia Public Schools (DCPS). Part II provides information on trends in teacher effectiveness by ward and socioeconomic status (SES) of schools in DCPS (released August 22, 2014). The supplemental consists of a report on principal retention prepared by Mathematica Policy Research (MPR) and funded by the Walton Family Foundation.

As noted in our previous reports, we caution readers about findings and implications of these analyses. Though informative, the data presented here are not sufficient to fully describe PERAA implementation across the entire citywide education landscape or infer causes of observed trends. It would be imprudent to attribute observed differences to the enactment and implementation of PERAA. Causal inferences of this sort cannot be established without substantially more information and analysis.

In addition, to the support provided by the DC government through NAS, the work reported here was supported indirectly by funds from a combination of public and private organizations that have helped create and build EdCORE. Acknowledgments go to the National Science Foundation (NSF), American Institutes for Research (AIR), CityBridge Foundation, and Judy and Peter Kovler for their generous support of EdCORE. We also wish to thank GW Vice President for Research Leo Chalupa and Provost Steven Lerman for their institutional support, without which it would not have been possible to fulfill the demands of the PERAA evaluation and related activities.

As always, we are grateful to our colleagues at the NAS – Robert Hauser, Stuart Elliott, Natalie Nielsen, Alix Beatty, and committee co-chairs Lorraine McDonnell and Carl Cohn along with the members of their committee – for their confidence in EdCORE and for their wise counsel on matters technical and stylistic. We wish also to thank Acting District of Columbia Auditor Lawrence Perry for his consummate professionalism and gracious management of the budget and contracting process.

Dr. Heather Harding and Sahar D. Sattarzadeh provided conceptual, technical, and logistical support at all phases of the work. We want to acknowledge Dr. Elias Walsh for his flexibility and gracious handling of this supplemental addendum. The deft management of contractual matters by Viola Horek (of NAS) and Maxine Freund, Vinnie Mitchell, and, Christine Tomasik (of GW) was also greatly appreciated.

Heather Harding, Ed.D.
Executive Director, EdCORE, Graduate School of Education and Human Development
The George Washington University

Michael J Feuer, Ph.D.
Dean, Graduate School of Education and Human Development
The George Washington University

MATHEMATICA
Policy Research

REPORT

FINAL REPORT

The Impact of Replacing Principals on Student Achievement in DC Public Schools

December 3, 2014

Elias Walsh
Dallas Dotter

Submitted to:
The Walton Family Foundation
125 West Central, Suite 314
Bentonville, AR 72712

Project Officer: Marc Holley
Contract Number: 2012-12-04

Submitted by:
Mathematica Policy Research
1100 1st Street, NE
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director: Elias Walsh
Reference Number: 40196.100

CONTENTS

TABLES

FIGURES

ACKNOWLEDGEMENTS

We would like to thank the Walton Family Foundation for funding the work. We would also like to thank Kelly Linker and Alden Wells at the District of Columbia Public Schools and Jeffrey Noel at the Office of the State Superintendent of Education of the District of Columbia for providing the data used in this study. At Mathematica Policy Research, John Hotchkiss and Lisa McCusker, assisted by Mark Timms, Adele Costigan, and Carolyn Chuong, processed the data and provided expert programming. Eric Isenberg and Steven Glazerman provided valuable comments. Betty Teller edited the report and Jackie McGee provided word processing and production support. Any opinions expressed herein are those of the authors and do not necessarily represent the views of the Walton Family Foundation or any other organization.

ABSTRACT

In 2007, the District of Columbia (DC) passed the Public Education Reform Amendment Act (PERAA), which established mayoral control of DC Public Schools (DCPS) and led to the appointment of Michelle Rhee as school chancellor. In an effort to boost student achievement, Chancellor Rhee replaced many school principals as one of her first reforms. For the 2008–2009 school year, 39 percent of the principals in the school district—51 individuals—did not return, and more were replaced in the following years. We measured whether students in a school with a new principal performed better on standardized tests than they would have if the original principal had been retained. To do so, we analyzed the changes in student achievement that occurred when principals who left at the end of each of the school years from 2007–2008 through 2010–2011 were replaced. We compared the achievement of students in DCPS schools before and after a change in school leadership, and then compared this change to the change in the achievement of students from a sample of comparison schools within DCPS that kept the same principal. We found that after three years with a new principal, the average student's reading achievement increased by 4 percentile points (0.09 standard deviations) compared to how the student would have achieved had DCPS not replaced the previous principal. For students in grades 6 to 8, the gains were larger and statistically significant in both math and reading.

I. INTRODUCTION

A. Overview

In 2007, the District of Columbia (DC) began a process of school reform with the Public Education Reform Amendment Act (PERAA). PERAA led to numerous reforms that changed nearly every aspect of the DC Public Schools (DCPS), including school governance structures, human capital policies, and leadership. PERAA placed DCPS under mayoral control; Adrian Fenty, the mayor of DC, used his authority to appoint Michelle Rhee as the first chancellor of DCPS. In an effort to boost student achievement, Chancellor Rhee replaced many school principals as one of her first reforms.

Although DCPS has made annual renewal decisions about school principals prior to and since PERAA, Chancellor Rhee changed the implementation of the principal retention policy to engage in a conscious strategy of replacing poor-performing principals. For the 2008–2009 school year, 39 percent of the principals in the school district—51 individuals—did not return. Less than 30 percent of principals left DCPS in any other school year between 2003–2004 and 2010–2011. Because as many as half of the exits that occurred at the end of the 2007–2008 school year were intentional dismissals by Rhee (Turque 2008), these circumstances provide a unique opportunity to understand the impact of targeted principal dismissals on student achievement. Whereas most previous research on the impact of principal transitions has focused on typical principal turnover or rotations across schools, the replacements in DCPS provide evidence on the effectiveness of a policy of principal dismissals.

In this report, we measure whether students in a school with a new principal performed better on standardized tests than they would have if the original principal had been retained. To do so, we analyze the changes in student achievement that occurred when exiting DCPS principals were replaced. We examine principal exits that occurred at the end of each of the school years from 2007–2008 through 2010–2011, and in particular we examine the 2008 replacements following the enactment of PERAA.

The primary challenge in this analysis is to distinguish between changes in school-wide student achievement caused by the new principal from those that might have occurred even if the dismissed principal had continued to lead the school. Achievement gains could have occurred if other factors besides the principal also changed in DCPS schools and affected student achievement. For example, PERAA also led to changes in human capital policies for teachers, which may have affected achievement trends in all DCPS schools. To address this issue, our analysis uses a comparison group of DCPS schools that did not experience transitions in school leadership. Doing so allows us to focus on how achievement trends differ in schools with and without replacements.

We implemented this strategy using a “difference-in-differences” design, so called because the design involves making two comparisons. The first comparison was between the achievement of students in DCPS schools before and after a principal’s replacement. We then compared this change to the change in the achievement of students from a sample of comparison schools within DCPS in which the principal was not replaced. In doing so, we estimated how the achievement of students in the schools with new principals would have performed in the absence of a leadership transition. Our analysis also accounts for students' prior achievement and background characteristics to address changes in the composition of students in a school over time.

We also address other challenges in our analysis. For example, some of the 51 principal exits following the 2007–2008 school year coincided with school closures and combinations. As a result, changes in achievement after replacements may reflect not only the impact of the change in leadership, but also the impact of combining schools. In addition, we address challenges that arise from the possibility that the schools in our comparison sample differ importantly from schools with exiting principals. For example, DCPS may have selected principals for dismissal from schools with declining achievement trends. If so, then achievement trends in comparison schools may not reflect how achievement would have evolved in schools with replacements had DCPS not replaced any principals, and our difference-in-differences design would produce estimates that are confounded by the differences between the two groups of schools.

We found that new principals led to significantly higher achievement for students in reading. The average student's reading achievement in schools led by new principals increased by 4 percentile points compared to how the student would have achieved had DCPS not replaced their previous principals. New principals did not have immediate impacts on achievement—we found statistically significant impacts on reading achievement following new principals' third year—but we found no evidence that student achievement declined after replacements, even temporarily. Although not as strong, the pattern was similar for math. For students in grades 6 to 8, the gains were larger and statistically significant in both subjects after two years; new principals improved achievement of the average 6th- to 8th-grade student by 9 percentile points in math and 8 percentiles in reading.

B. Previous research

Several previous studies have attempted to measure the contributions of principals to student achievement. For this review, we focus on studies that, like ours, use individual student data and account for students' prior test scores and other background characteristics when measuring these contributions. We omit studies that focus on school-wide average achievement, because these studies can misattribute to a new principal a change in student achievement that is caused by a change in student composition.

Most recent studies of principals' impact on achievement do account for student background, and many do so by calculating school "value added." School value added isolates the school's contribution to student achievement from the contributions of factors that are outside the control of the school, including the background characteristics of students. In addition to principal effectiveness, school value added may also measure the effectiveness of teachers in the school, contributions of school resources and facilities to achievement, and other school-level factors. By comparing a school's value added before and after a principal was replaced, this approach can isolate a principal's impact on achievement from other school-level factors.

Recent studies that compare a school's value added in the years before and after a principal transition have found that principals account for well under half of the differences in the level of student achievement across schools, with other school-level factors responsible for the remaining differences. Using data from Pennsylvania, Chiang et al. (2012) found that principals are responsible for at most 25 percent of the school's contribution to student achievement. Results from studies in Miami-Dade County Public Schools (Grissom et al. 2012), and Texas (Branch et al. 2012) are consistent with a figure that is less than 15 percent. [1] If DCPS principals were responsible for 15 percent of the school's contribution to student achievement, then replacing a principal who is at the 16th percentile of effectiveness with an average principal—an improvement of one standard deviation—would improve the average student's achievement by 1 percentile point.[2]

Several studies have found that it may take a few years for a new principal to make full impact in a new school. Using data from New York City, Clark et al. (2009) found that new principals' contributions to student achievement improve by approximately 0.01 standard deviations between the principals' first and third year of experience. Coelli and Green (2012) studied principal transitions in British Columbia, Canada, and found that it may take three or more years for a new principal to reach full impact in a school—and this impact can be much larger than the average impact over the first few years.[3] Two studies examined cumulative impacts of new principals on achievement over time. Dhuey and Smith (2013b), who also studied British Columbia principals, found that cumulative exposure for three years to a new principal who is one standard deviation more effective can boost student scores by 0.4 standard deviations. Gates et al. (2013) studied outcomes of students in 10 districts that recruited principals from New Leaders—a program designed to recruit, train, and support highly effective principals—and found that cumulative exposure to these principals over three years improved scores by approximately 0.03 standard deviations.[4]

Finally, Miller (2013) suggests that too much credit may be given to new principals if they were hired after a drop in the school's achievement under the previous principal. Using data from North Carolina, she found that although new principals improve over their first few years in a new school, after five years the new principal is only as effective as the previous principal's highest level of performance. Miller (2013) warns against attributing all of the post-transition gains to the new principal. Had the original principal instead been retained, the pre-transition

[1] However, Dhuey and Smith (2013a), studying principals in British Columbia, found that the same principal may have a larger contribution to student achievement if he or she is placed in a different school where the principal is a better "match" for the specific challenges that school faces. Also, Branch et al. (2012) found that the impact of individual principals may vary more in schools with more low-income students, suggesting larger variation in match quality in these schools—although this finding could instead result from differences in the principals who lead these schools compared to schools with higher-income students.

[2] We estimate that a DCPS school that is one standard deviation more effective improves student achievement by 0.20 student-level standard deviations, equivalent to improving the average student's achievement by 8 percentiles. If principals are responsible for 15 percent of that improvement, a principal who is one standard deviation more effective would contribute 0.03 standard deviations, or a 1 percentile point improvement.

[3] Branch et al. (2012) also examine estimates of principals' impacts on student achievement that are allowed to change with tenure in the school, but conclude the measures are too imprecise to be useful in their data.

[4] Estimates in Gates et al. (2013) are not directly comparable to those in the other studies. Whereas the other studies presented impacts from improving principal effectiveness by one standard deviation, the New Leaders principals in Gates et al. (2013) may differ by more or less than one standard deviation of principal effectiveness from the principals they replaced.

drop in performance may have proven to be only temporary. Thus, some or all of the gains associated with the new principal might also have been achieved had no transition occurred.[5]

C. Our contribution

This study makes two contributions to the previous research. First, all principal transitions in DCPS were precipitated by principals who left the district, many of whom DCPS targeted for replacement. In contrast, previous studies have focused on rotations between schools or other typical nonretention. Thus, the exiting principals may be more likely to be low performers than in previous studies. The DCPS replacements are more likely to be new hires or promotions, although some replacements were transferred from other schools that closed or were combined. Consequently, the impact of the new DCPS principals may differ from the impact of transitions in previously studied states and districts. Furthermore, this is the first study to examine the impact of a strategy of replacing poor-performing principals similar to the one precipitated by PERAA in DCPS.

Second, we provide evidence of the impact of a new principal on student achievement in each year up to four years following the previous principal's exit. Similarly, we are also able to observe possible trends in achievement *prior* to the replacements, such as the declines that Miller (2013) warns could lead to overstating the impact of a new principal. Our eight-year panel of student achievement data allows us to investigate these patterns to understand whether post-transition impacts can be fully attributed to the impact of the new principal. We are not aware of any previous study using longitudinal student-level data that obtains such rich information about the timing of student achievement impacts from new principals.

[5] Although Miller's study uses school-level rather than student-level data, leaving the possibility that the impact estimates may be partly caused by changes in school composition and not only principal transitions, she attempts to address concerns related to changes in student composition that arise from using school-level data.

II. EMPIRICAL APPROACH

A. Difference-in-differences design

Our approach to measuring the impact of a new principal on student achievement is to compare the trend in achievement observed in schools with replacements to the same trend in schools that did not experience a transition in school leadership. The change in achievement before and after the change in school leadership is the first difference in our "difference-in-differences" design; the second difference is between this achievement trend and the trend over the same time period in a set of comparison schools that kept the same principal. In doing so, we also account for changes in the composition of students in schools with and without new principals.

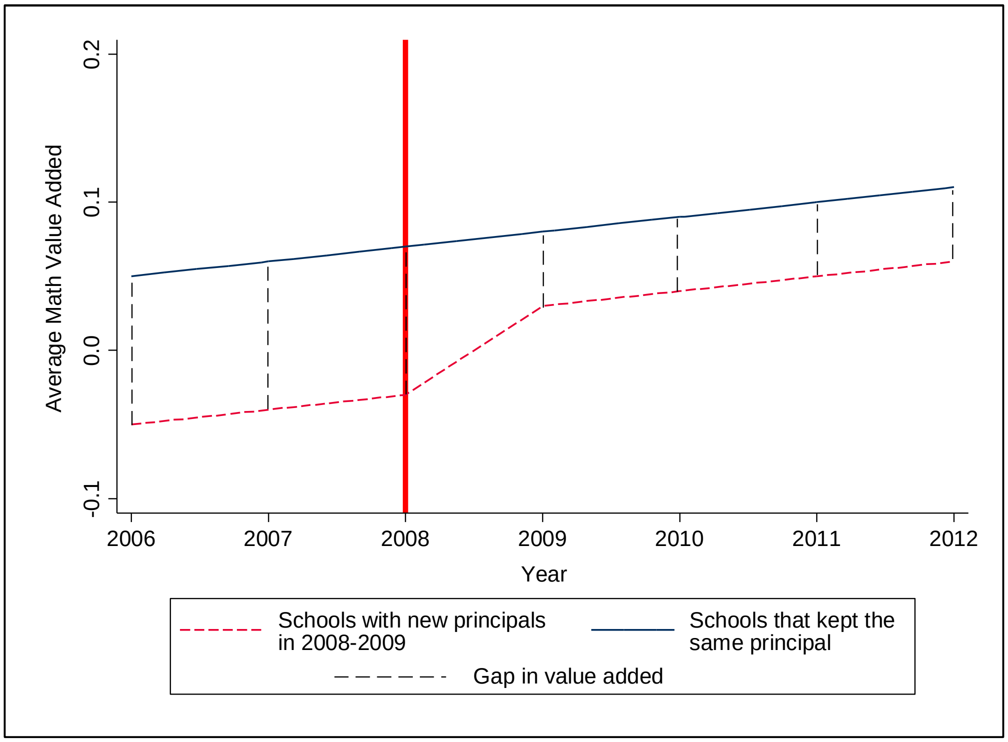
In Figure II.1, we use hypothetical data to illustrate the difference-in-differences approach to obtaining estimates of the impact of the changes in school leadership that occurred at the end of the 2007–2008 school year. The horizontal axis indicates the spring of a school year so that 2008 is the last year that an exiting principal led the school, indicated by the vertical solid red line. The outcome in Figure II.1 is school value added—a measure of the contribution of school-level factors (including but not limited to principals) to student achievement. In this hypothetical example, schools that kept the same principal have higher value added compared to schools with new principals in the 2008–2009 school year, so the solid blue line for schools that kept the same principals is above the dashed red line for schools with new principals.

In the hypothetical example, the new principals led to positive impacts on achievement. The gap between the two trends in value added, one of the two differences in the difference-in-differences research design, is represented by the vertical dashed black lines. These gaps are constant from 2006 through 2008, but narrow starting in 2009. The narrower gap in 2009 indicates that student achievement in schools with new principals improved after one year with the new principal compared to student achievement in comparison schools for the same years. The gap remains constant after 2009, indicating that achievement gain was sustained through the 2011–2012 school year in this hypothetical example. The change in the gap between schools with and without changes in school leadership encapsulates the two comparisons in our difference-in-differences research design.

We use regression analysis to estimate the impacts of the new principals on achievement. The regressions model trends in math and reading achievement for schools with and without transitions. In addition to prior achievement and other characteristics of students, the regression accounts for differences in achievement between schools that do not change over time, such as those that may be caused by differences in school resources or other school-level factors. The regression also accounts for changes in the overall average student achievement levels over time and across grades that may have arisen from other district-wide changes or DCPS policies.

Our impact estimates give the change in the gap in achievement between schools with and without new principals from a baseline year—the last school year the exiting principal led the school. We estimate the change in the gap for each of the four years following and the five years prior to a change in school leadership. We formally describe our regression specification in Appendix A.

Figure II.1. Hypothetical achievement trends for schools with and without new principals in the 2008–2009 school year



Source: Hypothetical data.

B. Limitations

Although our study makes important contributions to understanding the impact of new principals on student achievement that ultimately resulted from PERAA reforms, we acknowledge three limitations. First, although we examine turnover that results in many cases from an intentional policy of attempting to replace ineffective principals with highly effective new principals, we cannot distinguish principals who left voluntarily from those who left DCPS involuntarily. Thus, we examine the impact of replacing principals who left DCPS voluntarily or otherwise. However, even if our data did distinguish between voluntary and involuntary exits, some voluntary exits may actually be more similar to involuntary exits. For example DCPS could offer incentives to retain highly effective principals and, in so doing, make it less likely that less-effective principals will return. Rather than formal performance incentives, these incentives could be intangible, such as more cordial relationships with DCPS leadership. This may have occurred for teachers in DCPS. Dee and Wyckoff (2013) found that the DCPS IMPACT evaluation system led more lower-performing teachers to exit the district even though they were not subject to dismissal under IMPACT. If DCPS provided incentives for principals to voluntarily exit, then estimating the impact of replacements for all exits combined may be preferred rather than attempting to distinguish voluntary from involuntary exits. As a consequence, our results provide an estimate of the impact on student achievement that DCPS achieved from replacements that occurred as a result of targeted dismissals combined with the impact of more typical nonretention.

Second, our estimates of the impact of new principals could be confounded with other changes over time within schools that are not caused by the exits.[6] Although we account for changes in the composition of students within schools, DCPS may have implemented other changes in schools at the same time they were replacing principals. Our analysis accounts for these changes if they have the same impact on achievement in schools with and without new principals. However, some changes may have differentially affected achievement in these groups. Most notably, DCPS closed or combined many schools, with some of these school closings and combinations occurring simultaneously with the changes in school leadership. Although we conduct analyses to address simultaneous school closures and exits, changes to school resources that coincided with a change in school leadership may have also occurred and are more difficult to measure. For example, DCPS may have provided new principals with additional resources to support the transition. If so, our impact estimates would be too large, as they would conflate the impact of the new principal with the impact of the additional resources.

Finally, our approach to estimating the impact of new principals requires that schools in the comparison group are unaffected by the policy of selectively replacing principals, but this may not be the case. Some new principals were drawn from comparison schools that were closed or combined. Movement of students out of closed comparison schools will necessarily lead to changes in the composition of students in other comparison and treatment schools. To help address concerns that the composition of students in comparison schools may change over time, we account for student background characteristics when measuring trends in the contributions of comparison school principals, just as we do for schools with changes in school leadership. However, there may be other ways in which comparison schools are affected by this human capital policy that we cannot address. For example, the threat of dismissals in DCPS may have incentivized principals to bring about higher achievement in their schools. If so, principals in comparison schools may have been retained in part because they responded to those incentives by improving student achievement. In this case, our impact estimates would be lower than they would be in the absence of any incentives. Alternatively, if the incentives affected both groups of principals similarly, our impact estimates would reflect only the effects of replacing principals and not the full impact of the policy including incentive effects. Consequently, the full impact of the principal dismissal strategy could be larger than our estimates suggest.

[6] In other words, our difference-in-differences strategy requires that we assume that any unobserved determinants of student achievement that vary across schools do not also vary over time in a way that is related to whether schools did or did not have a principal replaced.

III. DATA

We use administrative data provided by DCPS and the Office of the State Superintendent of Education of DC (OSSE). The data include (1) a list of DCPS principals' school assignments for each school year from 2000–2001 through 2011–2012, (2) student background characteristics including math and reading test scores in grades 3 through 8 and 10 for the 2002–2003 through 2011–2012 school years, and (3) information on students' school enrollment. Although our main analysis focuses on student outcomes in the seven school years from 2005–2006 through 2011–2012, we use data from 2000–2001 through 2004–2005 to construct a measure of principal experience, and in some sensitivity analyses.

Our analysis divides schools into those in which DCPS replaced principals and those in which it did not between the 2007–2008 and 2010–2011 school years. The percentage of principals who left DCPS varied substantially over time, and some principals were forced out of jobs due to school combinations or closings. The annual turnover rate of DCPS principals varied between 14 to 39 percent from 2003–2004 through 2010–2011 (last row of Table III.1). The largest percentage of principals leaving DCPS occurred at the end of the 2007–2008 school year, Michelle Rhee's first year as chancellor, when 39 percent of principals—51 individuals—did not return to DCPS.

School restructuring in DCPS creates challenges for tracking student achievement over time in schools with and without changes in school leadership. For example, schools that closed do not have new principals. In most cases, the school of a departing principal remained open; however, as shown in rows 2 and 3 of Table III.1, these schools were sometimes closed or combined. The principal exits from 2007–2008 coincided with substantial restructuring of the schools; nine of the schools with a departing principal combined with another school, and nine other schools with departing principals closed. The next rows of Table III.1 show that school closures and combinations also affected some returning principals. Again taking the 2007–2008 school year as an example, the principals in six schools that closed transferred to a different school in DCPS and three principals continued leading their school after it was combined with one of the nine combined schools led by a departing principal.[7]

We measure the impacts of replacing principals in DCPS using student test scores in math and reading. The test scores are from SAT-9 tests from spring 2003 through spring 2005 and DC Comprehensive Assessment System (DC CAS) in spring of the subsequent seven years. We standardized the test scores to have a mean of zero and standard deviation of one within each combination of grade, year, and subject. This step translated math and reading test scores in every grade and year into a common metric; the DC CAS scores otherwise would not be comparable across these groups (that is, they are not "vertically aligned"). Standardizing the scores means that we cannot track DC-wide changes in achievement levels over time; however,

[7] The 12 principals in the 2007–2008 schools that were combined were involved in six combinations of two schools each, so that these 12 schools were led by six of the principals in the 2008–2009 school year.

that is not a goal of our analyses. Instead, we compare trends in achievement between students in schools with and without principal transitions.[8]

Table III.1. Principal transitions in DCPS by school year and status

Principal and school status	2003–2004	2004–2005	2005–2006	2006–2007	2007–2008	2008–2009	2009–2010	2010–2011
Left DCPS								
School remained open	17	32	24	18	33	22	29	21
School combined	0	0	1	0	9	0	0	0
School closed	0	0	1	0	9	2	1	1
Stayed in DCPS								
School remained open	102	87	95	112	71	85	77	84
School combined	0	0	1	2	3	0	0	0
School closed	0	0	0	0	6	0	0	0
Total	119	119	122	132	131	109	107	106
Left DCPS (%)	14	27	21	14	39	22	28	21

Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: The table includes principals in schools with at least 50 tested students in grades 4 through 8 or in grade 10.

The table describes transitions that occurred at the end of each school year.

Because of concerns with the accuracy and completeness of the SAT-9 test scores, we did not use these scores as outcomes for our main analyses. For analysis that did include the SAT-9 test scores, we excluded scores from grades 4, 6, and 7 in the 2004–2005 school year because we obtained relatively few test scores for students in those grades. Because we account for pre-test scores from the previous year in our analysis, excluding these SAT-9 scores also meant that we excluded all students in grades 5, 7, and 8 in the 2005–2006 school year from our analysis.

To account for student background, we used indicators for race/ethnicity categories, subsidized meals eligibility, English language learner status, receipt of special education services, gender, and whether a student transferred between schools during the year. Individual student data on subsidized meals eligibility is lacking for students attending a community-eligible school because these schools do not collect annual information about individual student poverty status.[9] Beginning in the 2005–2006 school year, for students who attended community-

[8] In standardizing across years, we also assumed that the dispersion in student ability is the same in each year. This would not be the case if the reforms following PERAA had a larger impact on the achievement of low-performing compared to high-performing students so that the gap in achievement between these two groups of students narrowed. Although we cannot rule out this possibility, in a sensitivity analysis we examine impacts for lower- and higher-achieving students separately.

[9] Schools are eligible to become community eligible if they have a student population composed of at least 40 percent with an identified need for free lunch based on direct certification, where students qualify based on their families' participation in state welfare or food stamp programs. These schools provide free breakfasts and lunches to all enrolled students and save on administrative costs by forgoing the collection of individual student subsidized meals applications.

eligible schools, we used a subsidized meals status for the student from another school or year, when available, and otherwise marked students in those schools as eligible for free lunch. We marked 3.5 percent of students in these years as eligible for free lunch for this reason.

We make several restrictions to the students and schools included in the analysis. We include in our main analysis students in each of the seven school years from 2005–2006 through 2011–2012 who are linked to at least one DCPS school for which we have identified a principal in the year. We also require that students have both a post-test and a pre-test from the same subject. For students in grades 4 through 8, the pre-test was from the previous grade and year. For students in grade 10, the pre-test was from grade 8 two years prior to the post-test. We then excluded 12 schools with new principals and 6 schools without changes in school leadership because of possibly compromised test scores in those schools. These schools were identified in a *USA Today* report as ones where at least half of tested classrooms showed evidence of cheating in at least one of the 2007–2008, 2008–2009, or 2009–2010 school years. Tests were flagged by the DC test score publisher if they had high rates of incorrect answers that were erased and replaced with correct answers (USA Today 2011).[10] As a final step, we excluded schools that were missing from any of the seven years in the panel.[11,12,13] Although we include closed schools in some of our analyses, this restriction to our primary analysis sample removes all 20 schools that closed before the 2011–2012 school year. The final analysis sample retains 88 percent of students who have post-tests from one or more of the seven school years.

Our analysis focuses on the principals who left DCPS in the years following the enactment of the PERAA school reform legislation and their replacements. In Table III.2, we provide counts of the new principals included in our analysis in each of the three school years following PERAA and identify whether the new principals had previously led a DCPS school. Of the 32 new principals who replaced a 2007–2008 principal, 23 were not leading a DCPS school in the previous year. The remaining nine new principals either led a different DCPS school during the 2007–2008 school year or their previous school was combined with an exiting principal's school. Prior to assuming leadership of a school, new principals may have been teaching or in administration within DCPS, or may have been recruited from outside DCPS. The counts of new principals in the last row of Table III.2 are lower than the total number of transitions in Table III.1 because of the restrictions we made to the analysis sample and because some schools have had multiple post-PERAA changes in school leadership but are counted only once in Table III.2. For the 21 schools with multiple transitions, we include only the first new principal following PERAA. In doing so, we treat the subsequent replacements as a consequence of the first post-PERAA replacement.

---

[10] We present results that instead include these 18 schools with compromised test scores in Appendix E.

[11] For a sensitivity analysis, we also included students in the previous two school years for a nine-year panel. The nine-year panel excludes grade 10 because we do not have test scores from the 2001–2002 school year.

[12] For this step, we treated two schools that were combined at some point into a single school as having been the same school in all years to avoid excluding these schools from the analysis.

[13] Prior to restricting to the seven-year panel, we excluded from our analysis school-year combinations with fewer than 50 remaining students in any grade and subject.

Table III.2. New principals after PERAA by school year and status

	Last school year departing principal led school			
New principal status	**2007–2008**	**2008–2009**	**2009–2010**	**2010–2011**
New principal is:				
Not previously a DCPS principal	23	9	8	3
From another DCPS school	9	0	1	0
New principals in analysis sample	32	9	9	3

Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: The table includes schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined, so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Counts of new principals include only the first replacement year for the 21 schools with multiple post-PERAA replacements. Of the 32 schools with replacements for departing principals from the 2007–2008 school year, 17 have had at least one subsequent replacement.

One new principal from the 2010–2011 school year began leading two schools, so there are 53 new principals in our analysis but the total number of schools with new principals is 54.

The schools in our analysis sample have slightly higher average achievement compared to the averages for DCPS schools generally. Table III.3 provides averages and standard deviations of school characteristics. The average levels of student achievement in the analysis sample can be different from zero because we standardized test scores using all students with test scores, not only those used in our analysis. Schools in our analysis sample have slightly higher achievement than average by 0.04 standard deviations in math and reading, shown on rows 1 and 2.

Although students in included schools have slightly higher test scores than students in schools overall, the included schools were no more effective at raising student achievement than excluded schools. To measure school effectiveness, we used value added to student achievement, a measure of the contribution of school-level factors that includes but is not limited to principals.[14] The average math and reading value-added estimates for schools in DCPS is zero by definition. Thus, rows 3 and 4 of Table III.3 indicate that that the schools included in the analysis sample are representative of the average value added of all schools in DCPS. Finally, rows 5 though 9 of Table III.3 show average student characteristics. For example, in the average school, 68 percent of students in the sample are eligible for free or reduced-price lunch and 80 percent are black.

Compared to principals who did not leave, those who left DCPS at the end of the 2007–2008 school year had lower-achieving students in math and reading in the year of the exit and had lower school value-added estimates, indicating that, on balance, the schools in which principals exited (voluntarily or otherwise) were not as effective at raising student achievement as schools

[14] We use school value added for these descriptive statistics and for a preliminary examination of the impact of replacing principals on student achievement, but our main analysis is conducted using student-level achievement data. Results based on the school value-added estimates are similar. We estimated value added for schools in each school year using data on student test scores and background. We describe our approach to estimating value added in Appendix B.

in which principals were retained. Additionally, returning principals led schools with fewer students who were eligible for free or reduced-price lunch compared to exiting principals. Other characteristics of principals' students did not significantly differ for principals returning versus exiting after the 2007–2008 school year, and principals who left did not have significantly more or less experience leading schools in DCPS compared to returning principals. When pooling the 2008–2009, 2009–2010, and 2010–2011 school years, we found no statistically significant differences between student characteristics in schools with principals returning versus exiting after these years, but exiting principals were less likely to have two to five years of experience. We present these differences in Table III.4.

Table III.3. Characteristics of DCPS schools

	School characteristic	Average	Standard deviation
(1)	Average math achievement (standard deviations of student achievement)	0.04	0.52
(2)	Average reading achievement (standard deviations of student achievement)	0.04	0.50
(3)	Math value added (standard deviations of student achievement)	0.00	0.21
(4)	Reading value added (standard deviations of student achievement)	0.00	0.18
(5)	Fraction of students eligible for free or reduced-price lunch	0.68	0.25
(6)	Fraction of students that are English language learners	0.08	0.14
(7)	Fraction of students that receive special education services	0.17	0.08
(8)	Fraction of students that are black	0.80	0.27
(9)	Fraction of students that are Hispanic	0.11	0.18

Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: The table includes schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined, so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Averages and standard deviations were calculated with one observation per school-year combination and are not weighted.

Math and reading achievement is standardized to have an average of zero and a standard deviation of one within each grade, year, and subject among all DCPS students. Value added is measured in student-level standard deviations of math or reading achievement and has an average of zero within each year and subject among all DCPS schools.

The table treats schools that are combined as distinct schools prior to being combined, and as a single school after being combined. The sample includes 82 schools, of which 12 were involved in six combinations. The averages include a total of 543 school-year records from the seven school years.

Although the schools with returning principals—the 22 comparison schools in the analysis—are lower achieving compared to schools with new principals, this is not necessarily a concern. The difference-in-differences research design accounts for differences in average characteristics between the two groups of schools so long as a key assumption holds. We assume that the difference in achievement between schools with and without new principals before the replacements occur would be the same as the difference in the years following the replacements in the hypothetical case that no replacements actually occurred. In other words, we allow a pre-transition gap in achievement between schools with and without new principals as in Figure II.1, but we assume that the trend in schools without new principals represents how achievement would have evolved in schools with changes in school leadership had DCPS not replaced any principals. As with similar assumptions for all research designs that rely on non-experimental methods, this assumption is not directly testable because we do not observe the hypothetical case of no changes in school leadership. However, we provide some important evidence in support of this assumption by testing for differences in achievement trends in the two groups of schools prior to the changes in school leadership.

Table III.4. Average characteristics of returning and exiting principals by time period

School or principal characteristic	Principals from 2007–2008		Principals from 2008–2009 through 2010–2011	
	Returning	Exiting	Returning	Exiting
Average achievement (standard deviations of student achievement)				
Math	0.31	-0.12*	0.22	0.02
Reading	0.28	-0.12*	0.17	0.05
Value added (standard deviations of student achievement)				
Math	0.08	-0.04*	0.06	-0.04
Reading	0.07	-0.06*	0.02	-0.01
Experience leading a DCPS school				
One year	0.18	0.19	0.02	0.09
Two to five years	0.55	0.56	0.56	0.23*
Six or more years	0.27	0.25	0.42	0.68
Fraction of students who are:				
Eligible for free or reduced-price lunch	0.58	0.71*	0.66	0.67
English-language learners	0.06	0.09	0.07	0.07
Special education	0.14	0.18	0.15	0.16
Black	0.85	0.80	0.77	0.79
Hispanic	0.08	0.12	0.10	0.09
Number of schools	22	32	22	22

Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: The table includes schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined, so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Averages and standard deviations were calculated with one observation per school-year combination and are not weighted.

Math and reading achievement is standardized to have an average of zero and a standard deviation of one within each grade, year, and subject among all DCPS students. Value added is measured in student-level standard deviations of math or reading achievement and has an average of zero within each year and subject among all DCPS schools.

Principals in schools with multiple replacements between the 2007–2008 and 2010–2011 school years are only counted as exiting for the first of these replacements, and are not included in the averages in the subsequent years. Returning principals include only principals who were not replaced between the 2007–2008 and 2010–2011 school years.

* = statistically significant at the 5 percent level

IV. RESULTS

A. Preliminary examination of trends in student achievement

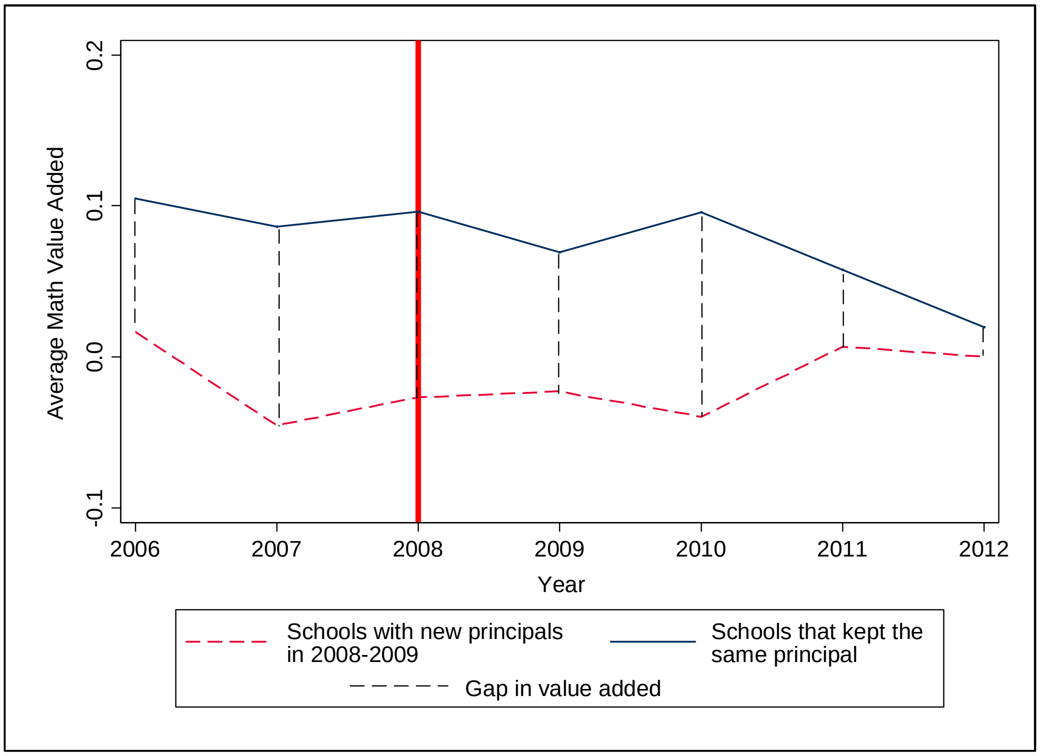
To implement our difference-in-differences research design, we analyze changes in the gap in student achievement between schools with and without new principals. We illustrate this approach using the changes in school leadership that occurred at the end of the 2007–2008 school year. Figure IV.1 plots the average math value added in schools with and without new principals, and Figure IV.2 plots the same for reading. The vertical solid red line at 2008 indicates the year in which changes in leadership occurred for schools with new principals; value added in this year is the schools' contribution in the last year of the exiting principals' tenure. The solid blue line for schools that kept the same principals is above the dashed red line for schools with new principals because schools without changes in leadership tend to have higher value added (Table III.4). An increase in value added for one group of schools may not reflect an actual year-to-year increase in the math skills of students because we standardized the value-added estimates within each year. Consequently, we focus on the gap between the two groups, rather than the trend for either group alone. The gap is represented by the vertical dashed black lines.[15] As in the hypothetical example in Figure II.1, the change in the gap between schools with and without transitions encapsulates the two comparisons in our "difference-in-differences" research design.

The pre-transition gaps in Figures IV.1 and IV.2 support a key assumption underlying the difference-and-differences research design—that outcomes for the two groups of schools would have trended similarly had DCPS not replaced any principals. This assumption cannot be tested directly because we cannot know how the trend in value added for schools with new principals would have evolved had the original principal remained in the school. However, differences in the trends before the transition—such as a widening or narrowing of the gap leading up to the transition—would be evidence that trends for the two groups of schools would also have appeared different after 2008 had DCPS not replaced any principals. Although the pre-transition trends are not perfectly parallel leading up to 2008 (as they are for the hypothetical data in Figure II.1), there is no evidence of a systematic widening or narrowing of the gaps in math or reading. For example, the gap in math widens between 2006 and 2007, but narrows between 2007 and 2008 (Figure IV.1).

The figures also provide evidence that schools with new principals in the 2008–2009 school year improved relative to comparison schools. Between 2008 and 2012 (the four years following the change in school leadership), the gap between schools led by new principals and comparison schools narrowed for both math and reading value added. For the main results, we use regression analysis to estimate the size of the gaps in each year relative to the gap in the year of the transition, and pool estimates of the impact of changes in school leadership from 2007–2008 through 2010–2011.

[15] We plot value added in this figure rather than achievement because value-added estimates account for student background characteristics similarly to the approach we use in our analysis. Replacing value-added estimates with average achievement in this figure leads to larger gaps between the two groups of schools, but the patterns are otherwise similar.

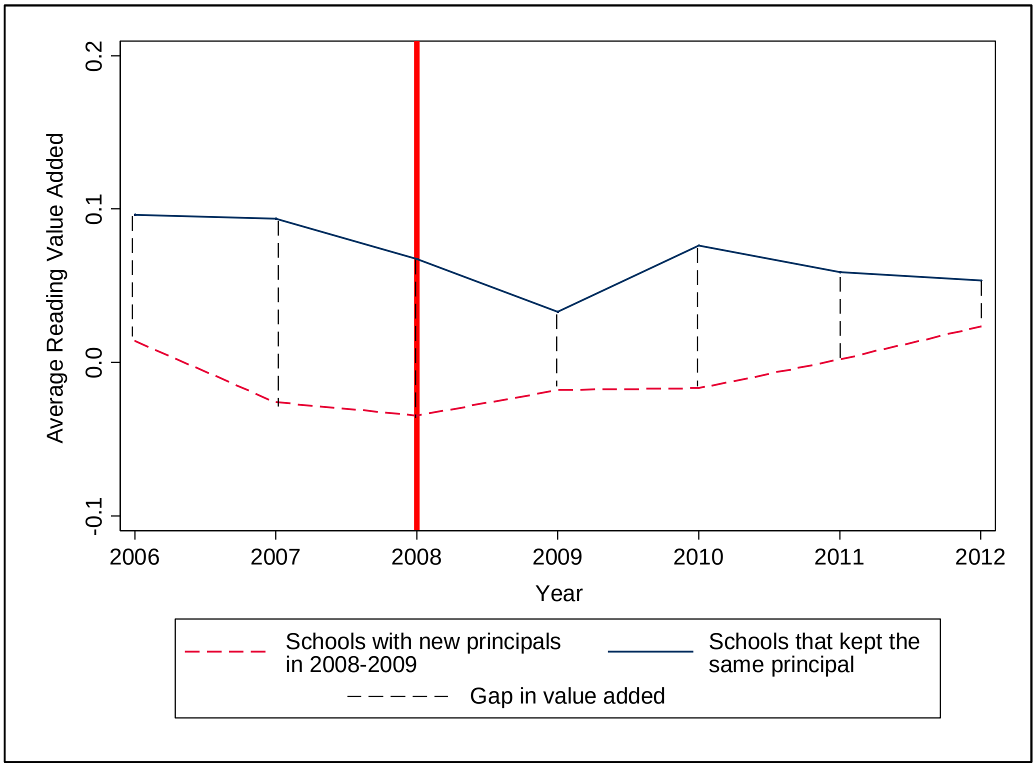
Figure IV.1. Trends in math value added for schools with and without new principals



Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: The figure includes 54 schools, of which 32 had new principals in the 2008–2009 school year. The figure includes schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Figure IV.2. Trends in reading value added for schools with and without new principals



Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: The figure includes 54 schools, of which 32 had new principals in the 2008–2009 school year. The figure includes schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

B. Impact of post-PERAA new principals

1. Impact on math and reading achievement

New principals produced higher reading achievement after three years in the school compared to the level of achievement prior to the change in school leadership. We found positive but insignificant impacts on math achievement. Figure IV.3 shows the estimate of impacts on math achievement for replacements that occurred between the 2007–2008 and 2011–2012 school years. All impact estimates in the figure are measured as changes in the gap in math achievement between schools with and without replacements from the gap that was present during the year of the replacement. This baseline gap in achievement is shown as a single dot at 0.0 standard deviations in the final school year before the transition occurred (called "year 0"). We measure changes in the gap using student-level standard deviations of student achievement. The gap in the year immediately following the replacement (year 1) is nearly identical to the baseline gap, indicating that new principals had no impact on math achievement after one year. However, the point estimate is larger in year 2, indicating that math achievement in schools with new principals improved relative to schools that kept the same principal after two years with the new principal. This positive impact in year 2 of 0.05 standard deviations is not statistically significant (the confidence interval crosses the dashed line at 0.0). Although also not statistically significant in years 3 and 4, the impact estimate is 0.07 standard deviations in both of these years, suggesting that a higher level of achievement may have been sustained in these schools through the fourth year with the new principal. An impact of 0.07 standard deviations is equivalent to improving the average student's performance by 3 percentiles.

Figure IV.4 shows the same impact estimates for reading. As with math, we find no impact of the new principals on reading achievement in the first two years after a change in school leadership. The impact on reading achievement is 0.09 standard deviations in year 3 and 0.10 standard deviations in year 4, and both estimates are statistically significant.[16] An impact of 0.10 standard deviations is equivalent to improving the average student's performance by 4 percentiles. The impact estimates for math and reading are also shown in Table IV.1.

Impacts of 0.07 to 0.10 standard deviations of student-level achievement are equivalent to an increase in student achievement of between 3 and 4 percentiles for an average student. Impacts of this magnitude are consistent with new principals who are about two to three standard deviations more effective than the principals they replaced (Chiang et al. 2012; Grissom et al. 2012; Branch et al. 2012).[17] Gains of these magnitudes would be expected when replacing a principal in the bottom 5 percent of the distribution of DCPS principals with one who is in the middle of the distribution. Furthermore, the improvement in reading was large enough to have increased the proficiency rate in affected schools during the 2006–2007 school year from 36 to 43 percent.[18]

2. Pre-transition gaps in achievement

We do not find strong evidence of pre-transition gaps in achievement between schools with and without replacements, suggesting that our analysis is adequately accounting for the selection of principals for replacements. None of the pre-transition impacts in math or reading are statistically significant—the confidence intervals for pre-transition years in Figures IV.3 and IV.4 all overlap 0.0. Furthermore, there is no evidence that the achievement of students at schools with new principals compared to students in comparison schools declined steadily leading up to the transitions, in contrast to Miller's (2013) findings for principal transitions in North Carolina. The gap is nearly identical in the year immediately prior to the dismissal compared to the baseline gap. Although the "impacts" on achievement in the second and third year prior to the transitions are larger than those in the subsequent years (but not statistically

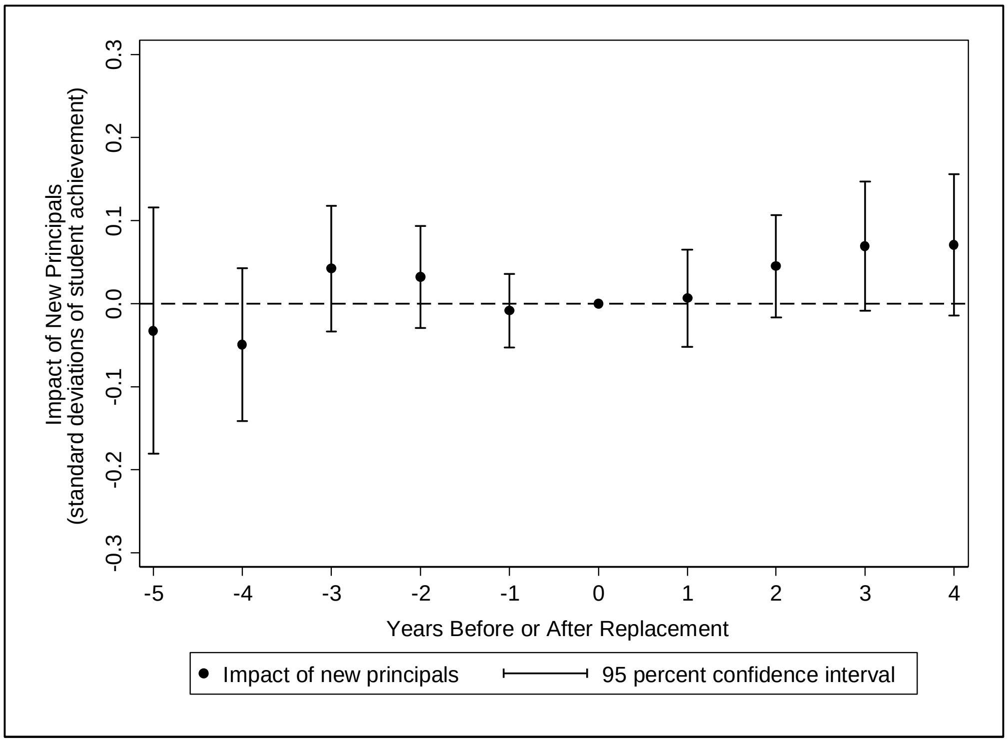
[16] Because we observe outcomes only through the 2011–2012 school year, the year 4 impact estimates are based only on principal replacements from the 2007–2008 school year. In Appendix C, we show that impact estimates only for this first cohort of replacement principals are slightly larger than those based on all principals. Thus, the year 4 estimate may slightly overstate the impact of the later three cohorts of new principals.

[17] For this calculation, we assume principals are responsible for 15 percent of schools' contributions to student achievement, and that a one standard deviation improvement in school effectiveness leads to an improvement in student achievement of 0.20 student-level standard deviations.

[18] For this calculation, we applied the impact of new principals in the third year after the replacements to the 2007 test scores of students who were enrolled in the schools with post-PERAA new principals during the 2006–2007 school year. We obtained the proficiency levels for each grade and subject from technical documentation of the 2007 DC CAS (CTB/McGraw Hill 2008).

significantly so), these are not part of an overall trend downward; the point estimates in the earlier pre-transition years are lower. Even so, because the pre-intervention impacts are imprecise, it is possible that achievement declined in schools leading up to a change in school leadership, so we consider the possibility that principals who were replaced were simply unlucky in riding a downward trend in test scores in the years leading up to their departure. In a conservative analysis to account for this possibility, impact estimates after three and four years with the new principal are reduced to 1 to 3 percentiles for an average student and are not statistically significant. See Appendix D for the results of this analysis.

Figure IV.3. Impact of new principals on math achievement by year relative to replacement



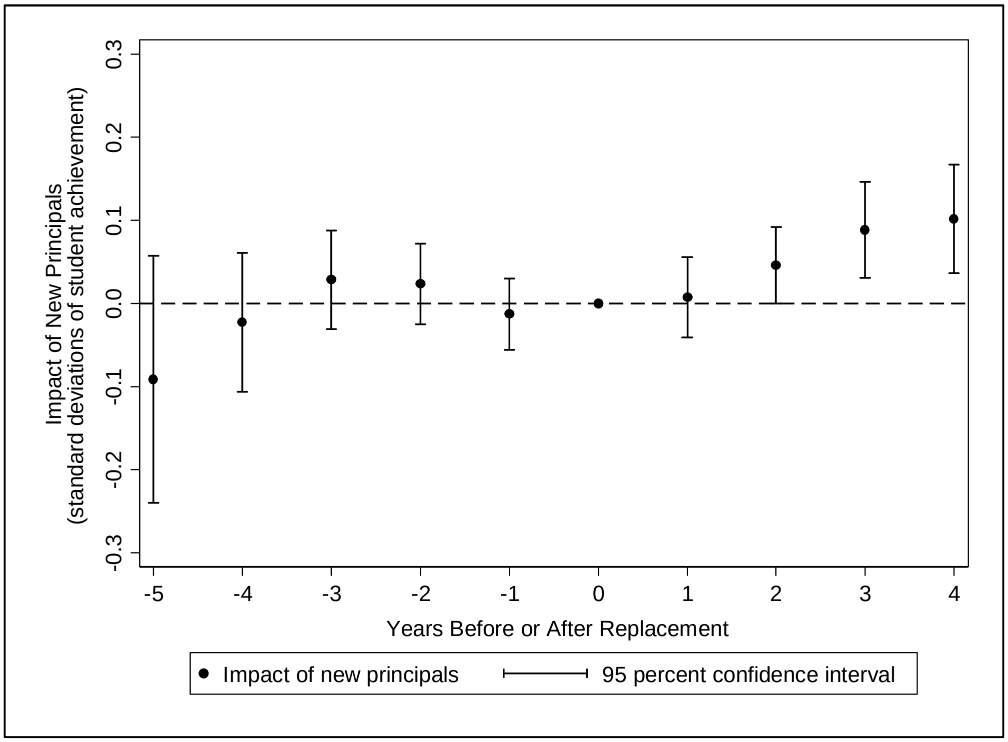
Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Notes: The figure includes 76 schools, of which 54 had post-PERAA new principals. The figure includes schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes 18 schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Impacts are measured relative to outcomes in year zero, the last year exiting principals led their schools. Outcomes for the seven school years are DC CAS scores.

Confidence intervals are based on standard errors that are clustered at the school-year level.

Figure IV.4. Impact of new principals on reading achievement by year relative to replacement



Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Notes: The figure includes 76 schools, of which 54 had post-PERAA new principals. The figure includes schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes 18 schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Impacts are measured relative to outcomes in year zero, the last year exiting principals led their schools. Outcomes for the seven school years are DC CAS scores.

Confidence intervals are based on standard errors that are clustered at the school-year level.

Table IV.1. Impact of new principals on math and reading achievement

	Impact by year since replacement (standard deviations of student achievement)			
Subject	Year 1	Year 2	Year 3	Year 4
Math	0.01	0.05	0.07	0.07
	(0.03)	(0.03)	(0.04)	(0.04)
Reading	0.01	0.05	0.09*	0.10*
	(0.02)	(0.02)	(0.03)	(0.03)

Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: The table includes 76 schools, of which 54 had post-PERAA new principals. The table includes schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes 18 schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Impacts are measured relative to outcomes in year zero, the last year exiting principals led their schools. Outcomes for the seven school years are DC CAS scores.

Standard errors are clustered at the school-year level.

* = statistically significant at the 5 percent level

C. Impact of new principals for subgroups of students and schools

We also estimate the impact of new principals for subgroups of students and schools, including for grade spans, higher- and lower-achieving students, and more- and less-experienced principals. Impacts could be larger for students in higher grades if discipline and school culture policies allow principals more influence on student achievement in those grades. Results for lower-achieving students could be larger if these students are more sensitive to changes in leadership. Principals with less experience may be less effective, which would lead to larger impact estimates for that subgroup. However, there are many reasons besides these that results could differ between these subgroups. Furthermore, because of the number of subgroups we examine, it is possible that we might obtain different results for one or more subgroups based only on chance. Consequently, the differences should only be considered suggestive of which groups might benefit more from new principals. Furthermore, none of the differences in impact estimates between these groups are statistically significant.

1. Results by grade span

We found larger impacts of new principals for students in grades 6 to 8, compared to students in grades 4 and 5.[19] In the third year after a replacement—the first year in which we found significant impacts for the full sample—we found no impact on math achievement for students in grades 4 or 5 in math and a statistically insignificant impact of 0.06 standard deviations in reading (Panel A of Table IV.2). For students in grades 6 to 8, the impact after three years with the new principal was 0.24 standard deviations in math and 0.19 standard deviations in reading (Panel B of Table IV.2). These impacts for students in grades 6 to 8 are equivalent to an increase in student achievement of between 8 and 9 percentiles for an average

[19] Although we include grade 10 in the full sample results, we do not report separate results for grade 10 because they were very imprecise; only 11 schools had grade 10 students.

student. The improvements after three years were large enough to have increased the proficiency rate in affected middle schools during the 2006–2007 school year from 36 to 47 percent in reading, and from 28 to 42 percent in math.

2. Results for higher- and lower-achieving students

We did not find strong evidence that the impact of new principals differed for higher- and lower-achieving students. We define higher-achieving students as those who scored above the district-wide average score within a grade and year, and lower-achieving students as those who scored below that same average. The impact three years after a replacement for higher-achieving students was 0.07 standard deviations for math and 0.08 for reading (Panel C of Table IV.2). For lower-achieving students, these impacts after three years with a new principal were 0.08 standard deviations for math and 0.12 for reading (Panel D of Table IV.2). For both groups, the estimate was statistically significant for reading but not for math.

3. Results for principals with more and less experience

Finally, we estimated the impact of replacing more- and less-experienced principals. We define less-experienced principals as those with three or fewer years of experience leading schools in DCPS at the time that they were replaced.[20] By this definition, schools with less-experienced principals recently experienced a previous transition. Consequently, a higher impact of replacing a less-experienced principal compared to a more-experienced principal could result from the less-experienced principal having been less effective, from lower achievement in the school as a result of a recent transition in leadership, or from both. Our analysis cannot distinguish these possibilities.

Three years after a replacement, we found larger impact estimates from replacing the less-experienced principals, although the differences are not statistically significant. The impact three years after a replacement for less-experienced principals was 0.12 standard deviations for math and 0.14 for reading (Panel E of Table IV.2). For more-experienced principals, this year 3 impact was a statistically insignificant 0.05 standard deviations for both math and reading (Panel F of Table IV.2). However, four years after a replacement, the impacts were more similar for the two principal experience subgroups.

[20] We limited the sample of schools with replacements to those with more- or less-experienced principals, but did not similarly limit the sample of comparison schools. Doing so was necessary to obtain precise results. Consequently, the comparison group of schools for both principal experience subgroups includes the same 22 schools.

Table IV.2. Impact of new principals on math and reading achievement by subgroup

	Impact by year since replacement (standard deviations of student achievement)			
Subject	**Year 1**	**Year 2**	**Year 3**	**Year 4**
Panel A: Grades 4 and 5 (41 schools with new principals, 17 comparison schools)				
Math	0.01	0.02	0.00	0.04
	(0.04)	(0.05)	(0.05)	(0.06)
Reading	0.04	0.04	0.06	0.09*
	(0.03)	(0.04)	(0.04)	(0.05)
Panel B: Grades 6 to 8 (45 schools with new principals, 13 comparison schools)				
Math	0.04	0.14*	0.24*	0.22*
	(0.04)	(0.04)	(0.07)	(0.07)
Reading	-0.01	0.10*	0.19*	0.19*
	(0.03)	(0.03)	(0.05)	(0.05)
Panel C: Higher-achieving students (54 schools with new principals, 22 comparison schools)				
Math	-0.01	0.04	0.07	0.10*
	(0.03)	(0.03)	(0.04)	(0.05)
Reading	0.00	0.02	0.08*	0.08*
	(0.02)	(0.02)	(0.03)	(0.03)
Panel D: Lower-achieving students (54 schools with new principals, 22 comparison schools)				
Math	0.03	0.06	0.08	0.04
	(0.04)	(0.04)	(0.05)	(0.05)
Reading	0.03	0.10*	0.12*	0.13*
	(0.03)	(0.03)	(0.04)	(0.05)
Panel E: Three or fewer years of experience in DCPS (20 schools with new principals, 22 comparison schools)				
Math	0.09	0.12*	0.12*	0.11
	(0.05)	(0.04)	(0.05)	(0.06)
Reading	0.08*	0.11*	0.14*	0.13*
	(0.03)	(0.03)	(0.04)	(0.05)
Panel F: More than four years of experience in DCPS (34 schools with new principals, 22 comparison schools)				
Math	-0.02	0.02	0.05	0.06
	(0.04)	(0.04)	(0.04)	(0.05)
Reading	-0.01	0.03	0.05	0.09*
	(0.03)	(0.03)	(0.03)	(0.04)

Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: The table includes only schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes 18 schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined, so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Impacts are measured relative to outcomes in year zero, the last year exiting principals led their schools. Outcomes for the seven school years are DC CAS scores.

Standard errors are clustered at the school-year level.

* = statistically significant at the 5 percent level

D. Accounting for school closures and combinations

Some principal transitions coincided with school restructuring, especially in the 2007–2008 school year, when 12 principals were in schools that were combined the following year (Table III.1). In addition, 15 schools closed in that same school year, and 4 more closed over the next three years. Our main analysis above includes combined schools but excludes the 19 closed schools.

School restructuring could have an impact on student achievement besides through a principal. For example, Ozek et al. (2012) show that school closures in DC led to a temporary decline in achievement for affected students. If so, then the impact of new principals could be confounded with the impact of restructuring. To address this concern, we also examined results that exclude combined schools, and results that include closed schools.

1. Results when excluding schools that combined

Impacts based on our main analysis measure the combined effect of school combinations and transitions when they occurred simultaneously. This could understate the impact of the policy of replacing principals if principals in newly combined schools faced extra challenges improving student achievement that we did not account for. To understand whether school combinations affect our results, we conducted a sensitivity analysis excluding six schools with simultaneous combinations and transitions, all from the 2007–2008 school year.[21]

Compared to our main results, impact estimates from the sensitivity analysis that excludes combined schools are identical three years after the replacements, and slightly larger after four years. Our main results indicated a 0.07 standard deviation impact on math achievement by year 4 and a 0.10 standard deviation impact in reading (Panel A of Table IV.3). The alternative results show year 4 impacts of 0.10 standard deviations in math and 0.11 standard deviations in reading when excluding combined schools (Panel B of Table IV.3).

2. Results when including schools that closed

Because simultaneous school closures and principal exits were part of the strategy employed by DCPS after PERAA, we also conducted analysis that attributed outcomes for students who previously attended closed schools to the closed school after it closed. In effect, we considered principals in schools that students enrolled in after their school closed as replacements for the principal of the closed school who exited DCPS. In contrast, our main analysis attributed outcomes for these students to their new school. We excluded schools that closed from our main analysis to prevent our results from fluctuating based on changes in the composition of schools included in our analysis across years.

[21] For this sensitivity analysis we count distinct schools prior to their being combined, so the 6 schools we excluded were combined into 3 distinct schools for the 2008–2009 school year. Of the 12 2007–2008 schools that were combined into 6 distinct schools, 9 had simultaneous combinations and replacements (Table III.1). We excluded only 6 of these 9 schools from the sensitivity analysis, because 3 had already been excluded from our main analysis sample for other reasons. Similarly, we had also previously excluded the 3 2007–2008 schools that were combined with another school but were led by principals that were not replaced.

In the sensitivity analysis, we included schools that closed by tracking the achievement of students who attended these schools and treating these students as if they were still attending the closed school. In practice, this meant that approximately 900 students in seven comparison group schools from our main analysis were "moved" to closed schools with new principals for the sensitivity analysis.[22] Therefore the number of schools with new principals in this sensitivity analysis increased to 61 from 54. For example, for grade 4 students in a school that closed at the end of the 2007–2008 school year, we would include their grade 5 through 8 achievement in the subsequent four years in the analysis, linked to the closed school. Doing so was not possible for all students in closed schools because we required that some students from the schools be observed in all seven years used in the analysis. For example, for schools that closed at the end of the 2007–2008 school year, this was only possible for students in grades 4. Grade 5 students in the 2007–2008 school year would have been in grade 9 during the 2011–2012 school year, and we did not observe grade 9 achievement.

Considering schools that closed as additional schools with new principals could lead to lower impact estimates, if student achievement is negatively affected when students transfer to a new school after their previous school closes; however, we do not find that this is the case. Impact estimates from analysis that excludes combined schools are nearly identical to those from our main analysis that included closed schools (Panel C of Table IV.3). The negligible role of including closed schools could be due to the relatively small number of additional students who are included in the sensitivity analysis—the 900 new students represent a 3.5 percent increase in the number of students in schools with new principals.

[22] Although some school closures occurred in schools without replacements (that is, the schools closed but the principal stayed in DCPS to lead a new school), most were in schools that had transitions (Table III.1). The number of comparison group schools in the sensitivity analysis for closed schools remains 22 because for no closed comparison group school did we observe students who previously attended that school in each school year through 2011–2012.

Table IV.3. Impact of new principals on math and reading achievement including or excluding schools that combined or closed

	Impact by year since replacement (standard deviations of student achievement)			
Subject	**Year 1**	**Year 2**	**Year 3**	**Year 4**
Panel A: Main results including schools that combined and excluding closed schools (54 schools with new principals, 22 comparison schools)				
Math	0.01	0.05	0.07	0.07
	(0.03)	(0.03)	(0.04)	(0.04)
Reading	0.01	0.05	0.09*	0.10*
	(0.02)	(0.02)	(0.03)	(0.03)
Panel B: Excluding schools that combined (48 schools with new principals, 22 comparison schools)				
Math	0.01	0.05	0.07	0.10*
	(0.03)	(0.03)	(0.04)	(0.05)
Reading	0.00	0.04	0.09*	0.11*
	(0.03)	(0.03)	(0.03)	(0.03)
Panel C: Including closed schools (61 schools with new principals, 22 comparison schools)				
Math	0.01	0.05	0.07	0.07
	(0.03)	(0.03)	(0.04)	(0.04)
Reading	0.01	0.05*	0.09*	0.10*
	(0.02)	(0.02)	(0.03)	(0.04)

Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: The table includes schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes schools where likely cheating occurred. In the case of closed schools, students from the school must have been tested in each of these school years. Schools that combined are treated as the same school before and after they combined, so they can be included in this sample.

Impacts are measured relative to outcomes in year zero, the last year exiting principals led their schools. Outcomes for the seven school years are DC CAS scores.

Standard errors are clustered at the school-year level.

* = statistically significant at the 5 percent level

V. CONCLUSION

Our analysis suggests that the DCPS principal replacement strategy led to higher reading achievement in schools and positive but statistically insignificant impacts in math. We followed schools with new principals for at most four years. We found that in the first year, new principals had no impact on achievement. Statistically significant achievement gains began in the third year the new principal led the school. The gains persisted through the fourth year with the new principal, the last year for which we were able to estimate impacts. The impact estimates are consistent with new principals that are two to three standard deviations more effective compared to the principals they replaced, or an increase in student achievement of 3 to 4 percentiles for an average student. The gains for students in grades 6 to 8 were consistent with an increase in student achievement of between 8 and 9 percentiles for an average student.

VI. REFERENCES

Ashenfelter, Orley. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics*, vol. 60, no. 1, 1978, pp. 47–57.

Branch, Gregory F., Eric A. Hanushek, and Steven G. Rivkin. "Estimating the Effect of Leaders on Public Sector Productivity: The Case of School Principals." Working paper #17803. Cambridge, MA: National Bureau of Economic Research, 2012.

Buonaccorsi, John P. *Measurement Error: Models, Methods, and Applications*. Boca Raton, FL: Chapman & Hall/CRC, 2010.

Chiang, Hanley, Stephen Lipscomb, and Brian Gill. "Is School Value Added Indicative of Principal Quality?" Working paper. Cambridge, MA: Mathematica Policy Research, November, 2012.

Clark, Damon, Paco Martorell, and Jonah Rockoff. "School Principals and School Performance." Working paper 38. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, December, 2009.

Coelli, Michael, and David A. Green. "Leadership Effects: School Principals and Student Outcomes." *Economics of Education Review*, vol. 31, 2012, pp. 92–109.

CTB/McGraw-Hill. *Technical Report for the Washington, DC Comprehensive Assessment System (DC-CAS) Spring 2007*. Monterey, CA: CTB/McGraw-Hill, 2008.

Dee, Thomas and James Wyckoff. "Incentives, Selection, and Teacher Performance: Evidence from IMPACT." Working paper #19529. Cambridge, MA: National Bureau of Economic Research, 2013.

Dhuey, Elizabeth, and Justin Smith. "How School Principals Influence Student Learning." Unpublished manuscript, 2013a.

Dhuey, Elizabeth, and Justin Smith. "How Important Are School Principals in the Production of Student Achievement?" Unpublished manuscript, 2013b.

Gates, Susan M., Laura S. Hamilton, Paco Martorell, Susan Burkhauser, Paul Heaton, Ashley Pierson, Matthew Baird, Mirka Vuollo, Jennifer J. Li, Diana Catherine Lavery, Melody Harvey, and Kun Gu. "Preparing Principals to Raise Student Achievement: Implementation and Effects of the New Leaders Program in Ten Districts." Santa Monica, CA: RAND Corporation, 2014.

Grissom, Jason A., Demetra Kalogrides, and Susanna Loeb. "Using Student Test Scores to Measure Principal Performance." Working paper #18568. Cambridge, MA: National Bureau of Economic Research, 2012.

Isenberg, Eric, and Heinrich Hock. "Measuring School and Teacher Value Added in DC, 2011–2012 School Year." Washington, DC: Mathematica Policy Research, 2012.

Isenberg, Eric, and Elias Walsh. “Measuring Teacher Value Added in DC, 2013–2014 School Year.” Washington, DC: Mathematica Policy Research, 2014.

Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan. “Earnings Losses of Displaced Workers.” *American Economic Review*, vol. 83, 1993, pp. 685–709.

Miller, Ashley. “Principal Turnover and Student Achievement.” *Economics of Education Review*, vol. 36, 2013, pp. 60–72.

Ozek, Umut, Michael Hansen, and Thomas Gonzalez. “A Leg Up or a Boot Out? Student Achievement and Mobility Under School Restructuring” Working paper 78. Washington, DC: National Center for the Analysis of Longitudinal Data in Education Research, June, 2012.

Turque, Bill. “Rhee Has Dismissed 24 Principals.” *Washington Post*, Friday, May 16, 2008.

USA Today. “Hundreds of Classes in D.C. Public Schools Flagged for Erasures.” *USA Today*, April 8, 2011.

APPENDIX A
REGRESSION MODEL

REGRESSION MODEL

We estimated the following regression of students' post-test scores on student background characteristics and variables that identify achievement trends for schools with new principals and comparison schools:

$$(1)\quad Y_{igst} = \left(\sum_{j=-5}^{-1} \delta_j R_{stj}\right) + \left(\sum_{j=1}^{4} \delta_j R_{stj}\right) + \mu_{gt} + \lambda_{1gt} S_{igt} + \lambda_{2gt} O_{igt} + \boldsymbol{\beta}'\mathbf{X}_{it} + \upsilon_{gs} + \varepsilon_{igst}\,.$$

In this regression, *Y* is the post-test for student *i* in grade *g*, school *s*, and year *t*. The summation terms represent a set of "relative year" indicators *R* and coefficients δ for each year from five years previous to a replacement to four years after a replacement. Whereas year *t* is a school year from 2005–2006 through 2011–2012, the index *j* represents a year relative to a replacement that may have occurred after any one of the 2007–2008 through 2010–2011 school years. The two summations exclude year $j = 0$ because our primary specification excludes the relative year indicator for year 0—the last school year the exiting principal led the school—so that the coefficients on the remaining indicators measure changes relative to achievement in exiting principals' final year in DCPS.[23]

The next term, μ_{gt}, is a set of indicators for each grade-year combination to account for differences in achievement levels over time and across grades that arise because of which students are included in the analysis sample. The variables *S* and *O* represent the same- and opposite-subject pre-tests with associated coefficient vectors λ_1 and λ_2. We estimated a separate pre-test coefficient for each grade and year. The vector **X** includes the other individual student background characteristics and the coefficient vector $\boldsymbol{\beta}$ provides relationships between each characteristic and achievement that are constrained to be the same in every grade and year.

To account for characteristics of schools that do not change over time, including fixed differences between schools with new principals and comparison schools, we included indicators for each school-grade combination in υ_{gs}.[24] The error term ε represents any other student-, school-, or year-specific factors.[25] Finally, we weighted each record in the regression based on a

[23] We also estimated a version of regression (1) using propensity score weights to construct a comparison group that was more similar to the group of schools with replacements based on value added from the 2005–2006 and 2006–2007 school years and demographics of students in the schools. Results from these matched specifications are reported in Appendix E.

[24] Our main results treat combined schools as the same school when constructing fixed effects, but we also obtain similar results based on treating combined schools and pre-combined schools as distinct units. The latter approach implicitly excludes simultaneous transitions and school combinations from the group of schools with replacements.

[25] We account for heteroskedasticity and correlation of regression errors within school-year combinations. We use the more conservative approach and account for clustering using a school variable that groups each combined school with its pre-combined school units rather than treating these as three distinct schools.

dosage variable that gives less weight to each record for students who attended multiple schools during the year.[26]

[26] In addition to the student-level regression approach in equation (1), as a sensitivity analysis we also estimated a school-level regression based on school value-added estimates as the outcome:

$V_{gst} = \left(\sum_{j=-5}^{4} \delta_j R_{stj} \right) + \mu_{gt} + \upsilon_{gs} + \varepsilon_{gst}$. We assigned each record in this regression a weight based on the estimated standard errors of school value added to reduce the influence of imprecise value-added estimates. Specifically, we used the inverse of the squared standard error. Results based on this specification were similar to those from the student-level specification.

APPENDIX B
OVERVIEW OF THE VALUE-ADDED MODEL

OVERVIEW OF THE VALUE-ADDED MODEL

In this appendix, we provide a brief description of the school value-added model that we used for descriptive statistics in Section III and for our preliminary investigation of the impact of new principals in Section IV.A. We also used these estimates to conduct a sensitivity analysis described in Appendix A. The features of the approach are described in more detail in the description of the school value-added model in Isenberg and Hock (2012); some of the methods from that description have been updated as in the description of the teacher value-added model in Isenberg and Walsh (2014).

We produced math and reading value-added estimates for each school-year-grade combination and then obtained a single math and reading estimate for each school-year combination by combining the estimates across grades. To do so, we estimated separate regression models in math and reading for each grade using data at the student-year level. Each regression related achievement on the post-test to achievement on the same- and opposite-subject pre-tests, other student background characteristics, and variables for each school-year combination. The pre-test relationships were allowed to vary for each grade-year combination, but we estimated one coefficient per grade (pooling across years) on each of the other student characteristics. For a student who attended a single school in a year, we assigned a 1 to the school-year variable for the student's school in that year and a 0 to all other school-year variables. For students in multiple schools, we divided by the number of schools and then assigned a fractional amount of "dosage" to each school-year variable where the student attended and a 0 to all other school-year variables.[27] We assigned dosage for students who attended schools that are not included in the analysis file to a catch-all school for the year.

The coefficients on the school-year variables provided initial estimates of school value added in the grade and subject. We standardized the initial estimates so that the mean and standard deviation of the distribution of teacher estimates is the same across grades and then averaged the estimates across grades within each school-year combination.[28] The resulting estimates are measured in standard deviations of student-level test scores.

We ran the regressions in two stages to allow an errors-in-variables correction for measurement error in the pre-tests. As a measure of true student ability, standardized tests contain measurement error, causing standard regression techniques to produce estimates of teacher effectiveness with systematic error. By netting out the known amount of measurement

[27] Although we obtained data on time enrolled in each school, we chose not use them because they are not of equal quality across all the years in the panel.

[28] We used an adjusted standard deviation that removes estimation error to reflect the dispersion of underlying teacher effectiveness. Using the unadjusted standard deviations to scale estimates for combining across grades could lead to over- or underweighting one or more grades when the extent of estimation error differs across grades. This is because doing so would result in estimates with the same amount of total dispersion—the true variability of teacher effectiveness and the estimation error combined—in each grade, but the amounts of true variability in each grade would not be equal. Instead, we scaled the estimates so that estimates of teacher effectiveness in each grade have the same true standard deviation, by spreading out the distribution of effectiveness in grades with relatively imprecise estimates. Estimates based more on estimates from imprecise grades will have larger standard errors; we account for these standard errors in the analysis.

error, the errors-in-variables correction eliminates this source of error (Buonaccorsi 2010). In applying the errors-in-variables approach, we used grade-specific reliability data available from the test publisher (for example, CTB/McGraw Hill 2008). Reliability data were not available for the SAT-9 test scores or for DC CAS scores from 2006; we instead used the average DC CAS reliability in the same grade from 2007 through 2012.

Correcting for measurement error required a second regression step because of computational limitations with the measurement error-correction method related to producing measures of precision. After running the errors-in-variables regression, we used the measurement error-corrected values of the pre-test coefficients to calculate an adjusted post-test that nets out the contribution of the pre-tests. We estimated a second regression step that excluded the pre-test variables and replaced the post-test with the new adjusted post-test. This step was necessary to obtain standard errors that are consistent in the presence of both heteroskedasticity and clustering at the student level, because the regression includes multiple observations for the same student.[29]

[29] Because we accounted for the precision of the estimates in the analysis, we did not apply empirical Bayes shrinkage, as is often done for value-added estimates. Shrinkage is most important when the goal is to interpret individual value-added estimates, which is not a goal in our analysis.

APPENDIX C
IMPACT OF REPLACING PRINCIPALS IN THE FIRST YEAR OF PERAA

IMPACT OF REPLACING PRINCIPALS IN THE FIRST YEAR OF PERAA

In addition to examining the impact of all replacements between the 2007–2008 and 2010–2011 school years, we also examined the impact of only those replacements that occurred at the end of the first year of PERAA. To do this, we excluded the 22 schools with transitions in the 2008–2009, 2009–2010, and 2010–2011 school years from the analysis, leaving the 32 schools with transitions that occurred at the end of the 2007–2008 school year. We compared outcomes in these schools to the same 22 comparison schools used in the full analysis.

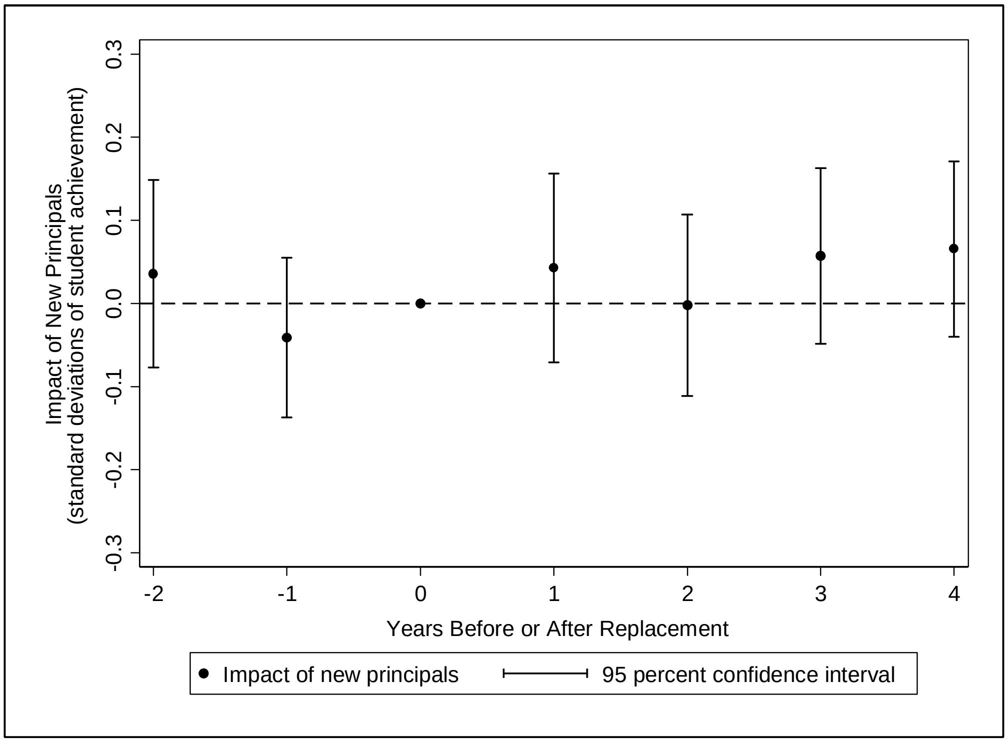
In addition to paying special interest to this single cohort of new principals because they represent the first replacements to occur following PERAA, we also examined them to investigate whether differences in impacts across cohorts could confound the timing of impacts that pool across the four cohorts of new principals. Because we did not observe outcomes four years after a change in school leadership for all schools with post-PERAA transitions, trends in the impact estimates that pool across the four cohorts may reflect differences in the impacts across cohorts of new principals rather than effects of additional time a new principal has in the school after a replacement. For example, the impact of new principals in the first year of PERAA may be larger than those in the later years if the first round of replacing principals removed the least effective of the group. If so, the year 4 impact could be larger than in years 1, 2, and 3 even if the additional year the new principal was at the school had no impact on outcomes.[30]

Because of the smaller number of transitions, results based only on these schools are less precise. We found no significant impact on math achievement for this first cohort of new principals, although the impact estimates for the first through fourth years with the new principal range from 0.00 to 0.07 standard deviations and are similar to those based on all four cohorts (Figure C.1). The impact estimates in reading for the first through fourth years with the new principal are slightly larger than those based on all four cohorts; they range from 0.07 to 0.12 and, like the full sample results, are statistically significant in the third and fourth year the new principals led schools (Figure C.2).

Because the impact estimates for the first cohort of new principals are slightly larger than those for all four cohorts, the impact estimates for the fourth year with the new principal in Table IV.1 could overstate the impact that would be obtained if we had included outcomes from after the 2011–2012 school year that would allow us to measure the impact of the later cohorts of new principals four years after replacement.

[30] Differences in impacts between cohorts would affect the results in this way, but the results would not be similarly affected if there were only differences in the composition of the schools, such as different levels of achievement. Because the analysis includes school fixed effects, we account for differences in composition across schools.

Figure C.1. Impact of new principals in the 2008–2009 school year on math achievement by year relative to replacement



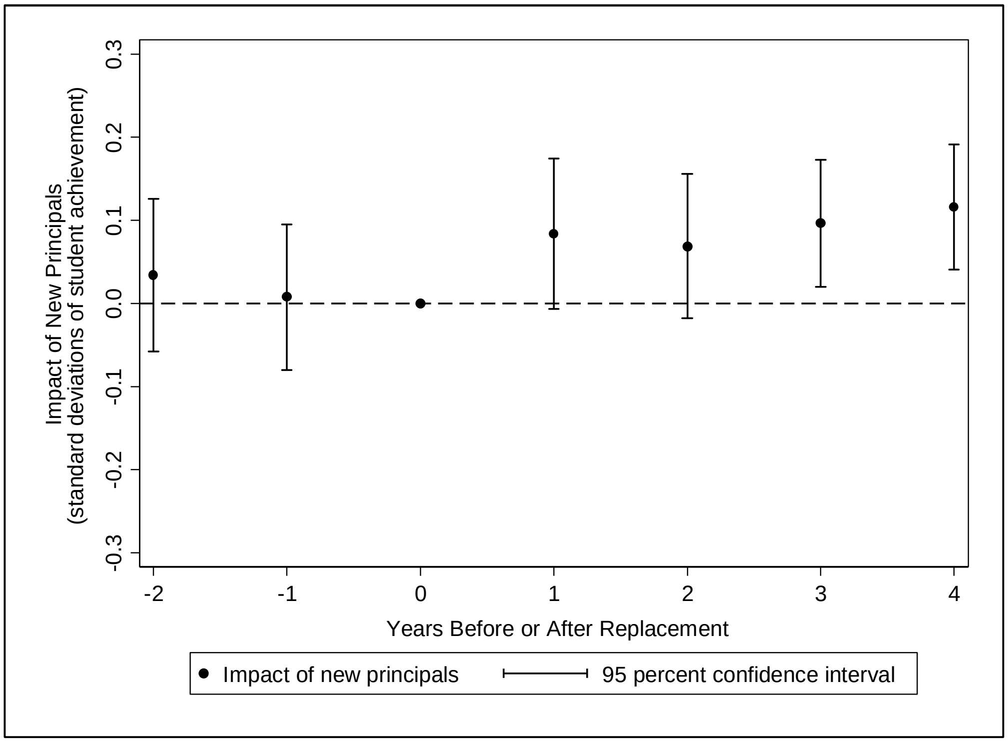
Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Notes: The figure includes 54 schools, of which 32 had new principals in the 2008–2009 school year. The figure includes only schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Impacts are measured relative to outcomes in year zero, the last year exiting principals led their schools. Outcomes for the seven school years are DC CAS scores.

Confidence intervals are based on standard errors that are clustered at the school-year level.

Figure C.2. Impact of new principals in the 2008–2009 school year on reading achievement by year relative to replacement



Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Notes: The figure includes 54 schools, of which 32 had new principals in the 2008–2009 school year. The figure includes only schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Impacts are measured relative to outcomes in year zero, the last year exiting principals led their schools. Outcomes for the seven school years are DC CAS scores.

Confidence intervals are based on standard errors that are clustered at the school-year level.

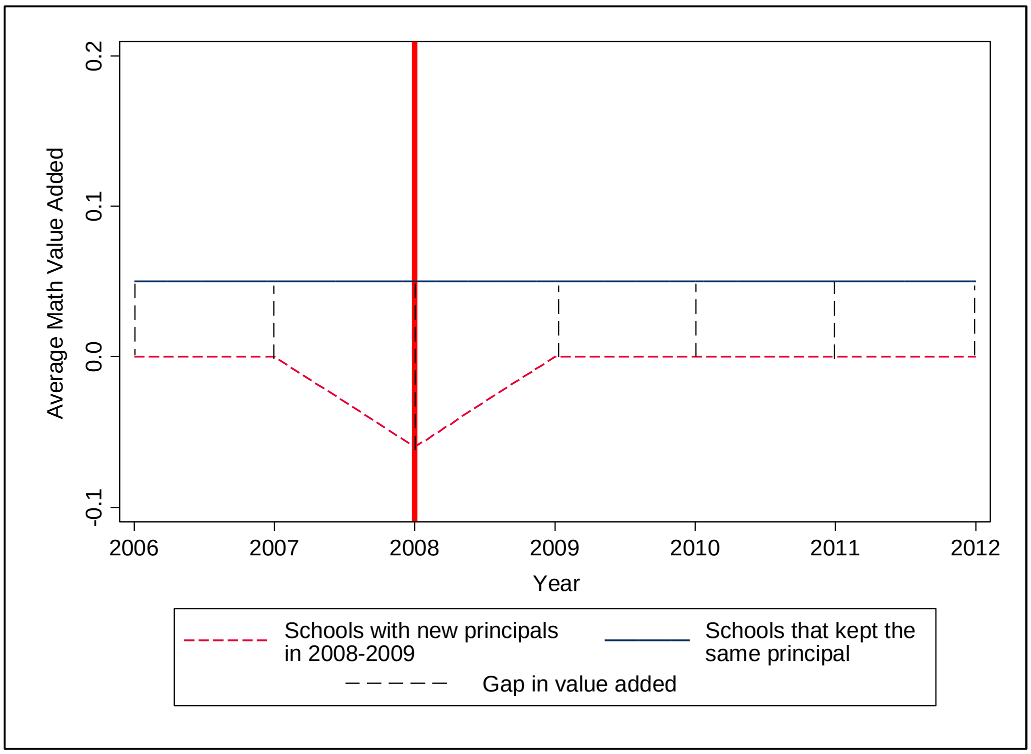
APPENDIX D
ACCOUNTING FOR A POSSIBLE PRE-TRANSITION DECLINE

ACCOUNTING FOR A POSSIBLE PRE-TRANSITION DECLINE

Using pre-transition data, our analyses can address the possibility that some of the change in the gap between schools with and without new principals following the transition may have occurred even if the exiting principal had remained in the school (Miller 2013).

It is possible for our difference-in-differences estimates to incorrectly attribute student achievement gains to a new principal. Similar to the hypothetical example in Figure II.1, we show hypothetical data in Figure D.1 to illustrate how this might occur. The hypothetical gap is the same in 2006 and 2007, becomes larger in 2008, and then returns to its previous level for 2009 through 2012. Because we compare the gap in each year to the gap from the last year exiting principals' led their schools—which is 2008 in this example—our difference-in-differences impact estimates from Section IV would attribute the entire decline in the gap between 2008 and 2009 to the new principal. However, the gaps in 2009 through 2012 do not represent an improvement compared to the earlier gaps in 2006 and 2007.

Figure D.1. Hypothetical achievement trends for schools with and without new principals in the 2008–2009 school year with pre-transition decline



Source: Hypothetical data.

In this case, we would not be able to distinguish between two possible explanations: (1) the original principal became worse over time, leading up to the replacement or (2) the original principal was "unlucky" to experience a temporary downward trend in performance that caused DCPS to remove the principal, but this trend is not related to the principal's effectiveness. The second possibility might occur because achievement based on standardized tests includes some measurement error. Consequently, standardized test scores can fluctuate even in a school with no actual changes in the skills of students over time. Under this explanation, the downward pre-transition trend in performance is not due to the quality of the original principal and would have rebounded had the principal remained in the school. Unlike the first explanation, the post-transition impacts should not be attributed to the new principal, since the exiting principal would have achieved the same result had he or she remained in the school.[31]

Our panel of student achievement data allows us to investigate these patterns to understand whether post-transition impacts can be fully attributed to the impact of the new principal. We address this concern by comparing post-transition impacts to the baseline level of achievement from before a possible decline in achievement. Table D.1 contrasts the results based on using the last year exiting principals' led their schools (year 0) as the baseline, as in our main analysis (Panel A), and the alternative approach that uses achievement from two to five years prior to the exiting principals' last year in DCPS (Panel B). The results in the year 0 column of Panel B indicate that the gap in achievement between schools with and without new principals declined by 0.03 standard deviations in math and reading before the transition occurred. This decline is not statistically significant. As a consequence of this pre-transition decline in achievement from the baseline years, the post-transition impacts that account for the decline are smaller and lose statistical significance.

These lower alternative impact estimates likely understate the true impact. The evidence of a pre-transition decline is based largely on positive gaps observed in two pre-transition years (the second and third year prior to the exiting principals' last year), rather than a systematic trend downwards (Figures IV.3 and IV.4), as in the hypothetical example in Figure D.1. If, instead, we were to estimate the impact using achievement from three to five years prior to the exiting principals' last year in DCPS, the impact estimates would have changed less from those in Panel A of Table D.1 and might even have *increased*. Indeed, when we include achievement from additional pre-transition years as we do in a sensitivity analysis in Appendix E, accounting for a pre-transition decline leads to results that are more similar to those that do not account for the possible pre-transition decline. However, we caution that the quality of test scores from these additional pre-transition years may not be as high as those from the later years.

[31] This is an example of what economists frequently call an "Ashenfelter Dip," which referred originally to falsely attributing wage gains to a training program that may only have returned participants to the wage rate they would have obtained without the program. The apparent gains arose because workers who had experienced a dip in wages were the ones who chose to participate in the program (Ashenfelter 1978; Jacobson et al. 1993).

Table D.1. Impact of new principals with and without accounting for possible pre-transition decline in achievement

	Impact by year since replacement (standard deviations of student achievement)				
Subject	**Year 0**	**Year 1**	**Year 2**	**Year 3**	**Year 4**
Panel A: Not accounting for pre-transition decline (impact relative to exiting principals' last year in DCPS)					
Math	n.a.	0.01	0.05	0.07	0.07
		(0.03)	(0.03)	(0.04)	(0.04)
Reading	n.a.	0.01	0.05	0.09*	0.10*
		(0.02)	(0.02)	(0.03)	(0.03)
Panel B: Accounting for pre-transition decline (impact relative to 2 to 5 years prior to exiting principals' last year in DCPS)					
Math	-0.03	-0.03	0.01	0.03	0.03
	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Reading	-0.03	-0.02	0.02	0.06	0.07
	(0.02)	(0.03)	(0.03)	(0.04)	(0.04)

Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: The tables include 76 schools, of which 54 had post-PERAA principal transitions. The table includes only schools observed in each of the seven school years from 2005–2006 through 2011–2012, but excludes 18 schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined, so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Standard errors are clustered at the school-year level.

* = statistically significant at the 5 percent level

n.a. = not applicable

APPENDIX E
ALTERNATIVE ESTIMATES OF THE IMPACT OF NEW PRINCIPALS

ALTERNATIVE ESTIMATES OF THE IMPACT OF NEW PRINCIPALS

In this appendix, we provide estimates of the impact of new principals based on (1) a nine-year panel of schools instead of the seven-year panel used for the estimates in Section IV, (2) weights to make the schools with and without new principals in our analysis more similar, and (3) including schools that were identified by a report in *USA Today* as having incidences of cheating on assessments (USA Today 2011).

1. Results based on a nine-year panel

We excluded outcomes from the 2003–2004 and 2004–2005 school years from our main analysis because our data on the SAT-9 scores from these years are incomplete. In the nine-year panel, we included scores from these years. Doing so could provide a more complete picture of how impacts evolved over time, including possible trends in pre-transition impacts. One main difference between the samples used in the seven- and nine-year panels is that the latter excludes grade 10 students because we did not observe a pre-test for these students in the 2003–2004 school year.

The results using the nine-year panel are presented in Figure E.1 for math and Figure E.2 for reading, and are similar to those based on the seven-year panel in Figures IV.3 and IV.4. Due to excluding grade 10, the magnitudes of the post-transition impact estimates are slightly larger than in the seven-year panel results. Only one of the pre-transition "impact" estimates is statistically significant; the impact from seven years prior to exiting principals' last year in DCPS in math is significantly smaller than zero. As with our main results from the seven-year panel (shown in Panel A of Table E.1), these results are generally consistent with the key assumption of our analysis that outcomes for the two groups of schools would have progressed similarly in the absence of transitions. The post-transition impact estimates based on the nine-year panel are also shown in Panel B of Table E.1.

Accounting for a possible pre-transition decline in achievement for schools with new principals has less impact when using the nine-year panel compared to the seven-year panel. The nine-year panel results in Panels A and B of Table E.2 that respectively do and do not account for the possible decline are very similar. Although these nine-year panel estimates do suggest that the seven-year panel estimates in Panel B of Table D.1 that account for a possible pre-transition decline are too conservative, we caution that this conclusion is based on SAT-9 test score data that may be incomplete and lack the level of quality that may come with the additional scrutiny paid to the DC CAS scores by DCPS and other stakeholders.

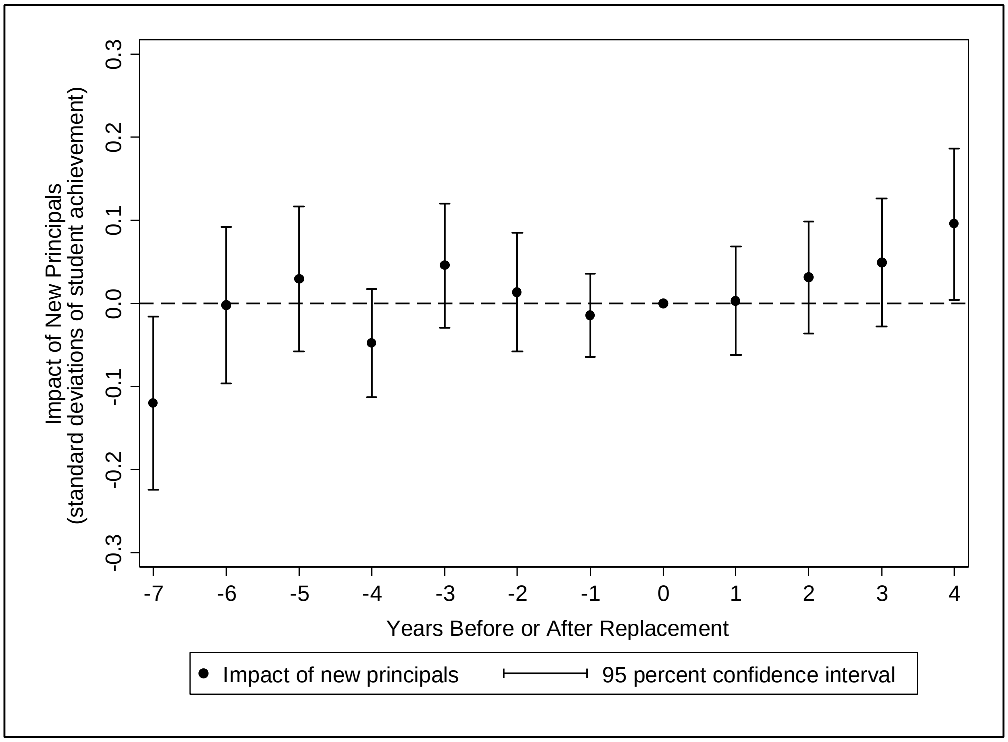
2. Results based on propensity score weights

We also estimated impacts using propensity score weights to construct a comparison group that was more similar to the group of schools with new principals based on value added from the 2005–2006 and 2006–2007 school years and demographics of students in the schools. Schools that kept the same principal that were more similar to those with new principals received more weight in the analysis and schools with new principals that were more similar to those that kept the same principal also received more weight. Results from these matched specifications are similar to our main results for reading, and smaller and not significant in math (Panel C of Table E.1).

3. Results based on including schools with potentially compromised test scores

Finally, we estimated impacts that included 12 schools with new principals and 6 schools that kept the same principal that were identified in a *USA Today* report as showing evidence of cheating in at least half of tested classrooms in at least one of the 2007–2008, 2008–2009, or 2009–2010 school years. Tests were flagged by the DC test score publisher if they had high rates of incorrect answers that were erased and replaced with correct answers (USA Today 2011). Although similar in magnitude to results that exclude these schools, including these schools leads to statistically significant impacts in math and reading two years after the replacements (Panel D of Table E.1).

Figure E.1. Impact of post-PERAA new principals on math achievement by year relative to replacement, nine-year panel



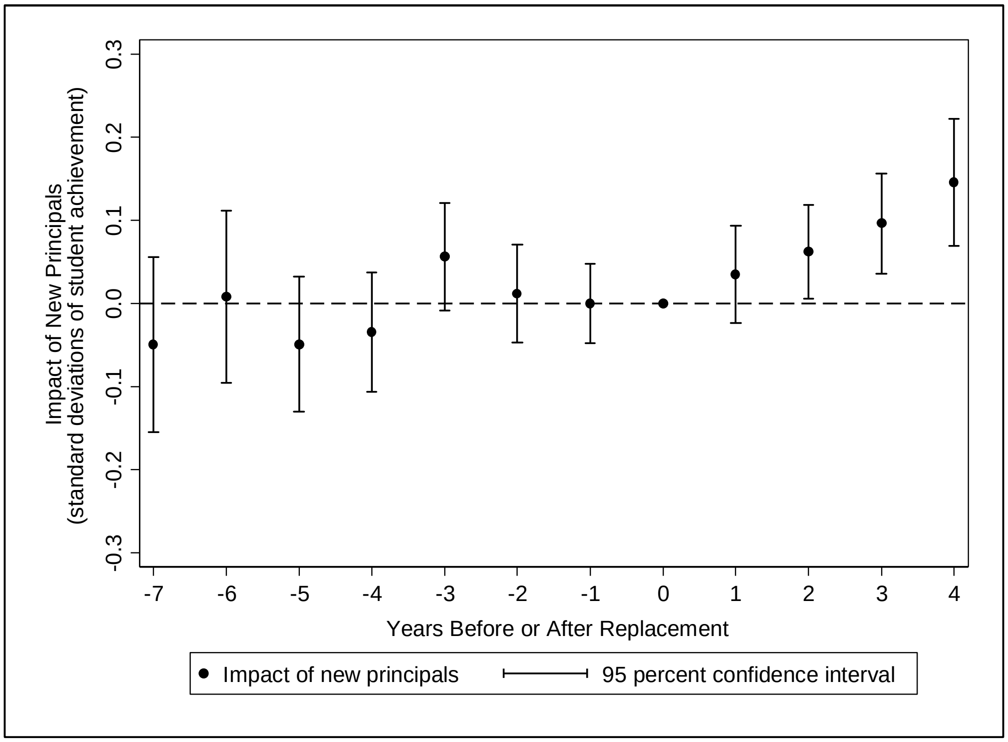
Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Notes: The figure includes 57 schools, of which 41 had post-PERAA new principals. The figure includes schools observed in each of the nine school years from 2003–2004 through 2011–2012, but excludes schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Impacts are measured relative to outcomes in year zero, the last year exiting principals led their schools. Outcomes are SAT-9 scores for the 2003–2004 and 2004–2005 school years and DC CAS scores for 2005–2006 through 2011–2012.

Confidence intervals are based on standard errors that are clustered at the school-year level.

Figure E.2. Impact of post-PERAA new principals on reading achievement by year relative to replacement, nine-year panel



Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Notes: The figure includes 57 schools, of which 41 had post-PERAA new principals. The figure includes schools observed in each of the nine school years from 2003–2004 through 2011–2012, but excludes schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Impacts are measured relative to outcomes in year zero, the last year exiting principals led their schools. Outcomes are SAT-9 scores for the 2003–2004 and 2004–2005 school years and DC CAS scores for 2005–2006 through 2011–2012.

Confidence intervals are based on standard errors that are clustered at the school-year level.

Table E.1. Alternative estimates of the impact of new principals on math and reading achievement

	Impact by year since replacement (standard deviations of student achievement)			
Subject	**Year 1**	**Year 2**	**Year 3**	**Year 4**
Panel A: Main results using the seven-year panel (54 schools with new principals, 22 comparison schools)				
Math	0.01	0.05	0.07	0.07
	(0.03)	(0.03)	(0.04)	(0.04)
Reading	0.01	0.05	0.09*	0.10*
	(0.02)	(0.02)	(0.03)	(0.03)
Panel B: Nine-year panel (41 schools with new principals, 16 comparison schools)				
Math	0.00	0.03	0.05	0.10*
	(0.03)	(0.03)	(0.04)	(0.05)
Reading	0.04	0.06*	0.10*	0.15*
	(0.03)	(0.03)	(0.03)	(0.04)
Panel C: Propensity score weights (54 schools with new principals, 22 comparison schools)				
Math	0.02	0.04	0.05	0.04
	(0.03)	(0.03)	(0.04)	(0.04)
Reading	0.02	0.05	0.09*	0.11*
	(0.02)	(0.02)	(0.03)	(0.03)
Panel D: Including schools with possible cheating (66 schools with new principals, 28 comparison schools)				
Math	0.00	0.07*	0.06	0.07
	(0.03)	(0.03)	(0.04)	(0.04)
Reading	0.00	0.05*	0.07*	0.09*
	(0.02)	(0.02)	(0.03)	(0.03)

Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: Panel A includes schools observed in each of the seven school years from 2005–2006 through 2011–2012. Panel B includes schools observed in each of the nine school years from 2003–2004 through 2011–2012. Schools that combined are treated as the same school before and after they combined so they can be included in these samples. Schools where likely cheating occurred are excluded except in Panel D.

The propensity score specification places more weight on comparison schools that are more similar to those with new principals based on value added from the 2005–2006 and 2006–2007 school years and demographics of students in the schools.

Impacts are measured relative to outcomes in year zero, the last year exiting principals led their schools. Outcomes for the nine-year panel are SAT-9 scores for the 2003–2004 and 2004–2005 school years and DC CAS scores for 2005–2006 through 2011–2012.

Standard errors are clustered at the school-year level.

* = statistically significant at the 5 percent level

Table E.2. Impact of new principals with and without accounting for possible pre-transition decline in achievement, nine-year panel

	Impact by year since replacement (standard deviations of student achievement)				
Subject	**Year 0**	**Year 1**	**Year 2**	**Year 3**	**Year 4**
Panel A: Not accounting for pre-transition decline (impact relative to exiting principals' last year in DCPS)					
Math	n.a.	0.00	0.03	0.05	0.10*
		(0.03)	(0.03)	(0.04)	(0.05)
Reading	n.a.	0.04	0.06*	0.10*	0.15*
		(0.03)	(0.03)	(0.03)	(0.04)
Panel B: Accounting for pre-transition decline (impact relative to two to seven years prior to exiting principals' last year in DCPS)					
Math	-0.01	-0.01	0.00	0.05	0.10
	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)
Reading	-0.01	0.03	0.05	0.09*	0.14*
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)

Source: Mathematica calculations based on administrative data from DCPS and OSSE.

Note: The tables include 57 schools, of which 41 had post-PERAA new principals. The table includes only schools observed in each of the nine school years from 2003–2004 through 2011–2012, but excludes schools where likely cheating occurred. Schools that combined are treated as the same school before and after they combined so they can be included in this sample. Schools that closed prior to the 2011–2012 school year are not included.

Outcomes for are SAT-9 scores for the 2003–2004 and 2004–2005 school years and DC CAS scores for 2005–2006 through 2011–2012.

Standard errors are clustered at the school-year level.

* = statistically significant at the 5 percent level

n.a. = not applicable