A Report on the Denver Public Schools Highest Priority Incentive Program: Descriptive Results and Estimates of Causal Impacts

October 19, 2020

Authored by:

Dr. Allison Atteberry

Dr. Mimi Engel

Michelle Doughty

Dan Mangan

This report is a product of the Teacher Workforce Collaborative (TWC), a Research-Practice Partnership (RPP) between Denver Public Schools (DPS) and University of Colorado Boulder's School of Education (CU). All results and writing here are the product of analyses by the CU Boulder TWC Team. The research reported here is supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305H190060 to University of Colorado Boulder (CU Boulder). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. This partnership is also generously supported through grants from the CU Boulder SoE (Place-Based Partnership Seed Grant), the William T. Grant Foundation's Institutional Challenge Grant, and a CU Boulder Research & Innovation Office (RIO) grant. We thank TWC RPP members Sarah Almy and Heather George for reviewing report drafts and for their extensive support and collaboration in producing this report. We also thank the DPS Impact Office Research & Evaluation Team for sharing data, answering questions, and Brooks Rosenquist for reviewing earlier drafts of this report.

School of Education
UNIVERSITY OF COLORADO BOULDER

DENVER PUBLIC SCHOOLS

DCTA
Denver Classroom Teachers Association

Table of Contents 2
Purpose of this Report 4
Solicitation of Feedback and Review 4
Executive Summary 5
Overview of Research Objectives 6
Research Objective 1: Describe Longitudinal Patterns in Teacher Retention in DPS 6
Research Objective 2: Describe HPI Schools and Teachers 6
Research Objective 3: Estimate Impacts of HPI on Teacher, Principal, and Student Outcomes 6
HPI Program Description and Overview 7
Background on HPI School Selection 8
Estimating HPI Impacts: CITS with Matching on Pre-HPI Trends 9
The Causal Challenge 9
Methodological Approach: CITS with Matching on Pre-HPI Outcome Trends 10
Other Approaches Considered: CITS with PSM-Based Matching & Regression Discontinuity 11
Data Sources and Key Variables 12
Sample of HPI Schools Included in CITS Impact Analysis 14
Results 14
Research Objective 1: What are longitudinal patterns in teacher retention in, and exit from, DPS? 15
1A. Teacher Retention Rates Over Time 15
1B. Retention Rates Across Schools 15
1C. What School Factors Predict Whether Teachers Leave? 15
1D. What Teacher Characteristics Predict Whether Teachers Leave Their School? 17
Research Objective 2: Describe HPI Schools and Teachers 18
2A. Who Receives HPI Incentives and Reported Job Satisfaction? 18
2B. How do HPI Schools Differ from non-HPI Schools? 19
2C. Did Retention Increase in HPI Schools once HPI Began? 19
2D. Where Do Teachers Who Leave HPI Schools Go Next? 20
Research Objective 3: Estimate Impacts of HPI Policy on Teacher, Principal, and Student Outcomes 20
3A. Did the HPI Policy Increase Teacher In-School Retention? 20
Matching Results for CITS Analysis of In-School Teacher Retention 20
CITS Results: In-School Teacher Retention 21

CITS Results: Are Certain Teachers Retained In-School at Higher Rates? 22
3B. Did the HPI Policy Increase Teacher In-District Retention? 22
Matching Results for CITS Analysis of In-District Teacher Retention 22
CITS Results: In-District Teacher Retention 23
3C. Did the HPI Policy Attract Teachers with Higher LEAP Scores to HPI Schools? 24
Matching Results for CITS Analysis of LEAP Scores of Recruited (New-to-School) Teachers 24
CITS Results: LEAP Scores of Recruited (New-to-School) Teachers 24
3D. Are principals in HPI schools more likely to be retained than their non-HPI counterparts? 24
3E. Did the HPI Policy Improve Student Achievement in HPI Schools? 25
Matching Results for CITS Analysis of HPI Impacts on Student Achievement 25
Future Work: Phase II - Explore Mechanisms for Phase I Findings 25
Summary and Conclusion 27
Contextualizing HPI Results Using Relevant Research 28
References 31
Tables, Figures, and Appendices 33

Purpose of this Report

This report is a result of a Work Scope (Appendix A) agreed upon by Denver Public Schools (DPS) and the Denver Classroom Teachers Association (DCTA) for analysis of the Highest Priority Incentive program (HPI). The report was produced by the Teacher Workforce Collaborative (TWC), a research-practice partnership (RPP) between DPS and report authors Allison Atteberry and Mimi Engel, faculty at the University of Colorado Boulder (CU). See Appendix B for a description of TWC's Mission, as well as report author biographies.

The goal of the report is to help DPS and DCTA understand the effects of the HPI program between 2016 and 2019.[1] The report describes the HPI program as well as whether and to what extent the program has affected teacher retention (the next revision of this report will estimate effects on principal retention and student achievement). The "Tentative Agreement Between School District #1 DPS and DCTA" states: "DPS and DCTA will commission a joint research project conducted during the term of this agreement to examine the root causes of educator retention and turnover throughout the district's High Priority Schools in order to identify possible solutions" (p. 5). This report is intended to meet that requirement.

We leverage existing quantitative data to estimate the overall effect of the HPI program on outcomes in DPS. Using existing data, we report results of a retrospective evaluation of the HPI program's effects from 2016-2019.

Solicitation of Feedback and Review

This report has been reviewed for completeness and factual accuracy by a small group of DPS administrators including Sarah Almy, Heather George, and Brooks Rosenquist. Following the first review for quality assurance, the report was provided to both DCTA and DPS to solicit feedback from internal stakeholders and make any requested changes. Following approval, the CU Team may disseminate results for academic audiences. DPS and DCTA leaders will determine how best to communicate results to their constituents and stakeholders.

[1] *Throughout this report, we refer to school years by the spring of the school year (e.g., 2019= SY 2018-2019).*

Executive Summary

During the 2015-2016 school year, DPS implemented the Highest Priority Incentive (HPI) program. The HPI program targets tiered financial incentives (based on performance rating) to employees who remain in one of approximately 30 DPS schools identified for HPI. To date, over 1500 teachers, school leaders, Special Service Providers, and other DPS employees have received HPI payments. Teachers are eligible for two HPI incentives: The first is a Monthly Retention Incentive ranging from $125 to $250 per month if the teacher remains in an HPI school. The second is an Annual Incentive ranging from $500 to $1000 for teachers who remain in an HPI school. Smaller incentives are paid to SSPs and other employees in HPI schools. Much larger bonuses – up to $30,000 per year in total – are paid to HPI school leaders. School leader bonuses are not dependent on evaluation scores.

Teachers working in HPI schools receive bonuses if they remain in an HPI school over time and have performance evaluations above the lowest “not meeting” designation. Over 99% of HPI teachers receive an HPI incentive payment of some sort, with bonus size increasing based on performance evaluation category. The intent of HPI payments is to increase retention of teachers with adequate performance in HPI schools. The program was designed to contribute to efforts to retain teachers in DPS schools that had the clear need for additional support. The current study estimates the effect of the program on teacher retention in HPI schools.

HPI’s goal was to target incentives to schools that struggle to recruit and retain effective teachers. Descriptive results indicate that HPI schools did indeed serve larger proportions of non-White and ELL designated students and had lower teacher retention rates in the three years before HPI began compared with non-HPI schools. Isolating the causal effects of the HPI program is a challenge addressed using a quasi-experimental method, Comparative Interrupted Time Series (CITS), which uses a comparison group of non-HPI schools with similar trends in teacher retention prior to the onset of HPI to estimate the effect of HPI on teacher retention.

Results from analyses reported here do not find evidence that HPI had effects on key outcomes of interest. Results of the CITS analyses indicate no statistically significant average effect of the HPI program on teacher retention (in-school or in-district). Though one might hypothesize that HPI effects could be concentrated among teachers receiving the largest incentives (based on LEAP), we find no evidence of HPI effects by LEAP score category. We also do not find that HPI helped attract teachers with higher LEAP scores to HPI schools. All of the school leaders in HPI schools were retained in the first year following implementation of the HPI program. However, descriptive analyses suggest that HPI school leader retention was not sustained at a higher rate than that of school leaders in non-HPI schools in subsequent years. The report concludes with a consideration of what prior research suggests might explain these null results.

It is important to note several limitations of these analyses. First, CITS can be used to estimate an average policy effect only; we cannot determine whether HPI caused changes within individual schools. Second, because there are a relatively small number of HPI schools, the fact that estimates indicate no effect of HPI must include the caveat that these analyses would be unlikely to detect small effects of the HPI program.

Overview of Research Objectives

Below are the research objectives that are addressed in this report.

Research Objective 1: Describe Longitudinal Patterns in Teacher Retention in DPS

- **1A. Retention rates over time.** What percentage of teachers leave their schools and/or the District each year, and how has this changed over time?
- **1B. Retention rates across schools.** How much do retention rates vary across schools?
- ***1C. What school factors predict whether teachers leave?*** Which school-level factors predict teacher retention/exit, *including malleable (e.g., leadership, climate, DPS initiatives)* and non-malleable factors (e.g., school location, student demographics)?
- ***1D. How do teacher characteristics predict whether teachers leave?*** Which teacher-level factors predict teacher retention/exit from a school or DPS (e.g., years of experience, demographic characteristics, subject, grade level)? Why do teachers report leaving (exit surveys)? Has this changed over time?

Research Objective 2: Describe HPI Schools and Teachers

- **2A. Who Receives HPI Incentives and Reported Job Satisfaction?** How many teachers receive HPI incentives, for how long, and at which schools? *Do teachers in HPI schools (over 99% of whom receive HPI incentives) report higher job satisfaction than similar counterparts who do not?*
- **2B. How do HPI schools differ from non-HPI schools?** How do HPI-eligible teachers/schools differ from non-HPI schools/teachers? Are there comparable non-HPI schools in DPS?
- **2C. Did teacher retention increase in HPI schools once HPI began?** How do HPI-eligible teachers/schools differ from non-HPI schools/teachers? Are there comparable non-HPI schools in DPS?
- **2D. Where do teachers who leave HPI schools go?** When teachers exit HPI schools, what do they do next (e.g., new non-HPI school, new HPI school, exit DPS)?

Research Objective 3: Estimate Impacts of HPI on Teacher, Principal, and Student Outcomes

- **3A. Did the HPI policy increase In-School teacher retention?** Once HPI began, did In-School teacher retention rates increase in HPI schools more than in similar non-HPI schools?
- **3B. Did the HPI policy increase In-District teacher retention?** Once HPI began, did In-District teacher retention rates increase in HPI schools more than in similar schools?
- **3C. Did the HPI policy attract teachers with higher LEAP scores to HPI schools?** Does the HPI program help school leaders attract teachers to high priority schools?
- ***3D. Are principals in HPI schools more likely to be retained than their non-HPI counterparts?***
- **3E. Did the HPI policy improve student achievement in HPI schools?**

HPI Program Description and Overview

During the 2015-2016 school year, DPS began providing targeted financial incentives to over 1,500 teachers and specialized service providers (SSPs) in approximately 30 schools as part of its Highest Priority Incentives program (HPI). The program was developed by the Teacher Retention Task Force (Denver Public Schools, n.d.)—a working group of district officials and teachers convened to determine how to attract and retain teachers in high-need schools. Teachers working in HPI schools receive bonuses if they remain in their school over time and have performance evaluations above the lowest "not meeting" designation (bonus size increases based on performance category). The intent of HPI payments is to increase retention of teachers with adequate performance in HPI schools. In this way, the HPI program was designed to contribute to efforts to ensure "great teachers in every classroom, especially those in [DPS's] highest-poverty schools" (HPI Overview & FAQs).

Prior to launching HPI, DPS had sought to fund similar incentives through reductions in bonus payments from *ProComp*, the District's Pay-for-Performance (PFP) program, which had been in place since 2005. Input from DCTA led DPS to choose not to reallocate ProComp funds. In Spring 2015, DPS announced the new HPI program, coinciding with Colorado Governor John Hickenlooper's signing of a bill that reduced the District's yearly contribution to the state employees' retirement fund (PERA), thus providing an estimated $20 million in savings used to fund HPI. DCTA protested the decision to implement HPI without buy-in from the union and filed a grievance (Gorski, 2015).

Incentive payments began at the start of the 2015-2016 school year and are paid, in addition to ProComp incentives, to school leaders, teachers, SSPs, and other DPS employees working in HPI schools, as detailed in Table 1. Incentive eligibility for teachers and SSPs is contingent upon performance. For teachers, this is measured using Leading Effective Academic Practice (LEAP), DPS' multiple measure teacher evaluation system. LEAP teacher evaluation includes a combination of classroom observations (conducted by school leaders and certified teacher leaders/observers), student perception surveys, and other factors related to student growth.[2] HPI school teachers and SSPs who receive evaluation ratings above the lowest performance category (0.05% -- or ½ of 1% of all DPS teachers are rated in the lowest performance category) receive incentive payments.

Teachers are eligible for two HPI incentives. The first is a Monthly Incentive ranging from $125 to $250 per month. The second is an Annual Incentive ranging from $500 to $1000 for those who remain in an HPI school. The current study estimates the effect of HPI on In-School teacher retention. We note that a very small number of teachers, ranging from 41 in 2016 (3.2% of all HPI teachers) to 20 in 2019 (1.6% of HPI teachers) transferred from one HPI school to another in each year. Teachers who transfer between HPI schools are still eligible to receive the HPI retention bonus. Smaller incentives are paid to SSPs and other employees in HPI schools. Much larger bonuses – up to $30,000 per year in total – are paid to HPI school leaders. All HPI school leaders who remain in their same school receive incentives, as they are not

---

[2] *Factors used to estimate student growth for LEAP scores include Student Learning Objectives (SLOs), School Performance Growth (SPF), and state test results for individual teachers when available.*

dependent on evaluation scores. Unlike teachers, school leaders who transfer between HPI schools do not receive a retention bonus for the transfer year.

HPI Theory of Action

Research consistently finds that teachers are the most influential malleable in-school factor associated with student outcomes (e.g., Aaronson et al., 2007; Boyd et al., 2011; Clotfelter et al., 2007; Goldhaber, 2002; Hanushek et al., 2005; Murnane & Phillips, 1981; Rockoff, 2004). Thus, attracting and retaining effective teachers in schools serving historically marginalized student populations is considered a promising means for closing achievement and opportunity gaps. The HPI program is informed by this knowledge and aims, ultimately, to improve student outcomes.

Similar to other incentive programs implemented across US districts and states during the past decade, the HPI program is implicitly grounded in a standard microeconomic framework of labor supply and demand where the available pool of teachers is a function of wages, working conditions, and other opportunities available to those who teach (Hanushek et al., 1999). Theoretically, then, the unequal distribution of teacher talent across and within schools and districts results from both the traditional uniform teacher salary schedule and varied working conditions (Springer et al., 2016). Thus, teacher incentive programs are often designed to increase the capacity of schools that are considered 'difficult-to-staff' to attract and retain effective, qualified, and experienced teachers.

Evidence on the effects of teacher incentive programs, however, is mixed. While a number of studies find positive effects of incentives aimed at attracting and retaining excellent teachers to high needs schools (e.g., Glazerman et al., 2013), other studies find such programs may not translate to improved student achievement (e.g., Cowan & Goldhaber, 2018). Evidence suggests that long-term, as opposed to one-time bonuses, may be more effective (Clotfelter et al., 2008). In addition, research has shown that incentives to induce teacher transfer must be very large: Glazerman, Protik, Teh, Bruch, and Max (2013) found in the Talent Transfer Initiative (TTI) that offers of $20,000 incentives to transfer to low-achieving schools only induced 5% of TTI teacher candidates to actually transfer. There is also evidence that inadequate communication regarding program eligibility may diminish overall effects (Clotfelter et al., 2008).

Background on HPI School Selection

Thirty HPI schools were chosen prior to the start of the 2016 school year.[3] Of those, 10 elementary and 10 secondary schools were selected based, loosely, on a combination of student demographics, school size, and teacher mobility factors. Relevant factors included the schools' percentage of students who were eligible for the federal Free and Reduced-Price Lunch Program (FRPL) or special education services, were designated as English language learners (ELL), and the volatility (turnover) of the school teaching staff.

[3] *Two middle schools were added to the HPI school list to replace 2 original HPI schools that closed in 2017. See* 'Sample of HPI Schools Included in CITS Impact Analysis' *for a full discussion of HPI school counts in this Report.*

DPS combines these 4 factors into a single measure called the School Characteristics Indicator (SCI) index.[4] Higher SCI index scores indicate that a school enrolls more historically-underserved students and/or suffers from greater teacher turnover. Indeed, at both the elementary and secondary levels, the schools with the 10 highest SCI index scores and with enrollments of at least 250 students were HPI schools.[5] See Figure 1 (elementary) and Figure 2 (secondary) for a graphic display of DPS schools by their school enrollment (Y-axis) and SCI index scores (X-axis). HPI schools are indicated by green circles.

According to interviews with DPS central office administrators involved with the HPI implementation, an additional 10 schools were selected at the discretion of DPS leadership. The exact reasons those schools were chosen appeared varied, and interviewees reported that HPI school selection occurred in direct collaboration with then-superintendent Tom Boasberg. Both the superintendent and several key decision-makers from 2015 have since left the district. Moreover, there is no definitive list of which 10 of the HPI schools were chosen based on leadership discretion, as opposed to chosen based on student/teacher data. Since 10 of the elementary HPI schools and 10 of the secondary HPI schools appear to follow a consistent pattern based on SCI index scores and school enrollment, we assume the remaining 10 HPI schools were the ones selected with discretion.

For Research Question 2B, we provide a demographic profile of HPI schools, relative to non-HPI schools. Those analyses show that HPI schools served larger proportions of non-White and ELL designated students and had lower teacher retention rates in the three years before HPI began. HPI's goal was to target incentives to schools that struggle to recruit and retain effective teachers. Research consistently documents that teacher recruitment and retention is more challenging in schools serving larger proportions of historically-underserved students populations (Engel, Jacob, and Curran 2014; Boyd, Loeb, Wyckoff 2013; Boyd, Lankford, Loeb, Ronfeldt & Wyckoff 2011). HPI schools served very large proportions of FRPL-eligible students, and also tended to have highly variable In-School[6] teacher retention rates in the years just prior to HPI's implementation. See Figure 3 for a graphic display of all DPS schools, their average In-School teacher retention rates (Y-axis) and average FRPL-eligible student percentages in the 3 years prior to HPI. See Table 2[7] for a list of all schools that were ever designated HPI, their SCI index scores, student demographics, and pre-HPI In-School teacher retention rates.

<u>Estimating HPI Impacts: CITS with Matching on Pre-HPI Trends</u>

<u>*The Causal Challenge*</u>

[4] *The SCI index formula is: SCI= .4(FRL) + .2(ELL) + .2(SPED) + .2(Volatility)*

[5] *It appears that SCI scores used to identify the 10 schools for HPI came from 2014 for elementary schools and 2015 for secondary schools.*

[6] *In this report, we discuss both "In-School" and "In-District" teacher retention. The former captures the percentage of a school's teachers that return to that school the following year. If a teacher transfers to another school in DPS, they are retained "In-District" but not "In-School". We focus on In-School retention as it is a key objective of the HPI incentive, which seeks to increase teacher retention in HPI schools. However, we also show some results for In-District retention.*

[7] The current version of this report includes school names and unit numbers. If revised for the general public, the report will use pseudonyms in all cases when the information in the report is not readily available to the public via other sources.

A central challenge in the social and observational sciences is knowing the cause of an observed outcome. For example, we might know that, on average, student outcomes in a particular school district or set of schools are improving or declining. However, it is typically the case that there are many factors that may be correlated with changes in student outcomes including the composition of both the student body and the teachers who teach those students, as well as other school, district, state, and federal programs and policies. The difficulty faced in estimating the effects of the HPI program, thus, is isolating the effect(s) of HPI from the myriad factors that occurred at the same time including, for example, other policies within DPS or Colorado, the local economy, and the composition of the DPS teacher workforce and student population as a whole.

For example, a descriptive analysis could indicate that teacher retention *declined* in HPI schools during the implementation period, despite the fact that the HPI program was designed to increase teacher retention in the designated schools. However, an analysis of teacher retention at the district level might reveal a secular trend of declining teacher retention, with higher retention rates (despite a decline) in HPI schools. Results such as these could indicate a positive effect of HPI. In contrast, positive teacher retention rates in HPI schools that might suggest an 'HPI effect' could, instead, be found to be indistinguishable from a general trend in DPS or across the state of Colorado.

The causal challenge, thus, is to apply an analytic method that allows us to isolate the effect of the HPI program from other programs, policies, and trends. The "gold standard" for estimating causal effects in the social sciences is generally considered to be randomized controlled experiments which allow for the estimation of an average treatment effect. The random assignment of units (e.g., students, teachers, or schools) to 'treatment' and 'control' conditions means that any differences found across groups can be attributed to the 'treatment', or, in the case of education, program or policy.

In reality, programs and policies in education are rarely randomly assigned,[8] as was the case with HPI, where the selection of schools for HPI designation was focused intentionally on schools considered to be "high needs". Thus, efforts to isolate the effects of HPI on outcomes of interest (i.e., teacher retention, student achievement) statistically control for differences across students, teachers, and schools through analysis of DPS administrative data collected before, at the onset, and during the implementation of the HPI program. Similar to analyses in randomized controlled experiments, however, one can only estimate an average policy effect using the approach described below. Thus, estimates will be provided of an average effect of the HPI program on outcomes of interest. These analyses cannot be done at the school-level, meaning that it is not possible to estimate a causal effect of HPI for any particular school.

Methodological Approach: CITS with Matching on Pre-HPI Outcome Trends

The approach used in this report to estimate the causal effect of HPI on teacher retention rates is called a comparative interrupted time series (CITS). This analytic approach can be used in cases where the timing of onset of a program or policy is specific and known and the program was implemented among some units but not others. HPI meets these criteria, as implementation began at a distinct point in time—Fall of

[8] *There are a number of exceptions to this, however, and the application of RCTs in education continues to grow, even in areas such as teacher incentive programs.*

the 2015-2016 school year—and there are schools that were and were not designated HPI. In addition to these requisite factors for using CITS to estimate the effects of HPI, a non-equivalent comparison group of schools from among those that were not designated HPI is required.

To identify an appropriate set of non-HPI comparison schools, we began by examining each HPI school's pre-HPI In-School teacher retention trend for the 10 years prior to the onset of HPI.[9] For each HPI school, we then attempted to identify a "matching" non-HPI school with a similar pre-HPI trend in In-School teacher retention. We refer to this as 'Pre-Trend Matching' (as opposed to 'PSM-Based matching', which we allude to briefly below).[10] The logic of Pre-Trend matching is that we compare how In-School teacher retention changed in the HPI school once HPI began relative to a non-HPI school with similar retention trends in the decade preceding the onset of HPI.

While the comparison group does not need to be equivalent to the HPI group of schools, a requirement for causal interpretation of CITS-based impact estimates is that the two sets of schools (HPI schools and comparison schools) have similar ***pre-HPI trends*** in terms of the outcome of interest prior to the implementation of HPI. In the case of HPI, this means that comparison schools need to have similar trends in In-School teacher retention rates prior HPI implementation which began in Fall 2015.

Two types of effects are estimated in CITS analyses. First, an average effect can be observed at the immediate onset of the program. Referred to as a "shock" effect, this occurs when there is a statistically significant difference in the level of the outcome of interest, compared with the comparison group that coincides with program onset. The second type of effect, a "trend" effect, occurs when the average trend in the outcome of interest is statistically significantly different than the trend for the comparison group. It is possible to have either a shock effect or a trend effect or to observe both -- a shock at program onset followed by a difference in post-implementation trends. It is also important to keep in mind that it is possible that a program will have no effects. Again, CITS and other quasi-experimental methodological approaches allow us to estimate average policy effects. Thus, we are unable to estimate school-level effects of the HPI policy.

<u>Other Approaches Considered: CITS with PSM-Based Matching & Regression Discontinuity</u>

[9] *For example, suppose a given HPI school's pre-HPI trend in In-School teacher retention was -2 percentage points (PP) per year. This would mean that, during the decade leading up to the start of HPI, that HPI school's In-School teacher retention rate was falling, on average, by 2 PP each year. An ideal match for this HPI school would be a non-HPI school that also had an In-School pre-trend of -2 PP/year.*

[10] *Specifically, our primary Pre-Trend matching method is called 'One-to-One (1:1) Nearest Neighbor (NN) matching without replacement.' Using this method, each HPI school was matched to a single non-HPI school. The school with the most similar (nearest) outcome pre-trend becomes a given HPI school's match. Without replacement means that once a non-HPI school has been selected as a match for an HPI school, it is not matched with another HPI school. Note that we also present results using 1:1 NN matching w/o replacement, as well as a secondary matching method called 'One-to-Many (1:M) caliper matching with replacement' in the appendices. We prefer 1:1 NN matching, because matching with replacement in small samples can result in the repeated use of the same very small number of non-HPI schools as matches. Results are similar across methods. See Appendix C for tables and figures detailing results of matching using each of these methods.*

We considered 2 additional approaches to estimating HPI impacts on teacher retention outcomes: (1) CITS w/ PSM-Based Matching, and (2) Regression Discontinuity. We describe why we originally pursued these options, as well as the reasons why these methods are not the preferred analytic approach for HPI.

(1) We considered an alternative approach—propensity score matching (PSM)—to identify a suitable non-HPI comparison group of schools for the CITS analysis. An appeal of conducting a CITS analysis using PSM-based matching is that, in theory, HPI schools can be matched to non-HPI schools on a much larger number of pre-HPI factors (as opposed to the Pre-Trend Matching we use in this Report, which matches schools based on pre-HPI In-School teacher retention rates alone). The goal of a PSM-Based matching approach is to identify a set of non-HPI schools that were, based on a range of characteristics, predicted to be *just as likely* to be designated HPI schools as those that were actually selected.

In practice, however, we were unable to identify a sufficient number of non-HPI schools using the PSM-Based matching approach with very similar likelihoods of becoming an HPI. This challenge can largely be explained by the fact that SCI index scores are such a strong predictor of HPI school selection (see Figure 1 and Figure 2). When we used the PSM approach to matching, the non-HPI schools were typically not very similar to the HPI schools—a key assumption of PSM methods. While the results of the PSM approach to matching indicated the method was not a good analytic approach for this study, we anticipate that some readers might suggest a PSM approach to matching. A full discussion of the CITS with PSM-based matching and results are provided in Appendix D. Again, because the PSM-based matches were not successful, this is not our preferred method.

(2) We also explored the possibility of using a regression discontinuity (RD) method to estimate HPI impacts. The appeal of RD is that, when the design works, results are convincing (i.e., the method, when successful, has a strong "causal warrant"). Recall that the SCI index scores appear to almost perfectly determine 20 of the 30 HPI schools (see Figure 1 and Figure 2). The 10 elementary and secondary schools with the 10 highest SCI index scores were designated HPI schools, while the school with 11th highest SCI index score just missed the cutoff. The logic of the RD approach is to compare the outcomes of schools that fell just above/below a cutoff (here, top 10), the logic being that there are no substantive differences between schools very near the SCI cutoff. In other words, in an RD approach, we think of schools just above/below an arbitrary cutoff to be "as good as random" and can interpret results causally.

However interviews with former central office administrators who were present at the time of HPI school selection reported that SCI index scores were not, in practice, consistently used to select HPI schools. The strong correspondence between SCI index scores and HPI selection may be more incidental than intentional. In addition, when we conducted RD analyses, we found that they were underpowered, meaning that the sample size was too small to allow for the detection of meaningful statistically significant results. Nonetheless, since an RD approach was a possibility, and, had it been successfully implemented would have been our preferred analytic method, we include a full discussion of the RD analyses and results in Appendix E.

<u>Data Sources and Key Variables</u>

This report uses DPS administrative data from years 2006 to 2020. Since the HPI program began at the start of the 2015-2016 school year, we have 10 years of pre-HPI school-level retention data and 4 years following HPI implementation. The data includes 3,500-4,500 teachers per year (14,049 distinct teachers) and 120-170 schools per year (197 distinct schools). Figure 4 presents a graphical summary of school counts over the past 19 years, including charter and alternative schools, HPI schools, and regular schools (non-charter/alternative and non-HPI). We use secondary data to answer the questions agreed upon by DPS and DCTA in the work scope. These data include student achievement outcomes, teacher satisfaction and retention outcomes (both In-School and In-District), and school outcomes (e.g., stability of the teaching staff, school climate). We also use survey data from the DPS Teacher Exit Survey and the Annual Teacher CollaboRATE Survey (these data will be added and analyzed in the next draft of this report) related to teachers' opinions about HPI, their schools, and their decisions about where to work.

HPI Impact Analysis Key Outcome Variable: Teacher Retention

The key outcome of interest in this report is teacher retention. We consider both *In-School* and *In-District* retention. The HPI policy and this Report focuses on In-School teacher retention, as that aligns with the theory of action that underlies HPI, however for Research Question 3B, we also examine In-District teacher retention. Because we do not have data on charter schools, teachers who leave DPS traditional schools for charter schools are described as leaving the district. The lack of charter school data will make In-District retention rates reported herein appear slightly lower than DPS would report overall. However, given HPI's focus on retaining teachers in specific, non-charter schools, losing teachers to DPS charter schools can be considered an unfavorable outcome for the purpose of the HPI analyses.

We measure retention for a given teacher within a given year using a variable that indicates whether that teacher was retained or not (equal to 1 if a teacher will return in the following fall, equal to 0 if they do not). When aggregated to the school-level in a given year, In-School retention captures the percent of teachers who will be retained in the same school in a given year.[11]

HPI Impact Analysis Explanatory Variables

The focus on this report is on whether or not a teacher works in an HPI school. The HPI program took effect in the 2015-2016 school year. This primary explanatory variables of interest in analyses designed to estimate HPI impacts will be (a) an indicator for whether a given school will become an HPI school, as well as (b) an indicator for whether the outcome of interest—typically In-School teacher retention—is observed before or after HPI began.

Our focus in this Report is on estimating the effects of the HPI incentives on In-School teacher retention. Our main analyses use the CITS approach using Pre-Trend matching to estimating HPI impacts, and in these analyses we do not include controls for other factors. However, in Appendix D we also present results from a CITS approach using PSM-Based matching (described above), and in those models, we do also include control variables which are listed therein.

[11] *The last year for which we can currently calculate In-School/District retention is 2018-2019 since we do not have data from Fall of 2020, we do not yet know which teachers from the 2019-2020 school year will return.*

Sample of HPI Schools Included in CITS Impact Analysis

In addition to the 30 original HPI schools, 2 middle schools were added to the HPI school list in later years (Bear Valley and Kepner Beacon). They replaced 2 of the original 30 HPI schools that closed after the 2016-17 school year (Kepner Middle and Henry World Middle). The total number of schools that were ever designated HPI is therefore 32.

Of the 30 original HPI schools, 6 were co-located.[12] When HPI schools were chosen, each co-located school was counted once (not twice). For example, Bruce Randolph School— which includes a co-located middle and high school — was counted as 1 of the original 30 HPI schools. In the current analyses, we treat each co-located unit as 2 distinct schools. We do so because there were insufficient co-located non-HPI schools to serve as viable matches for the 6 co-located HPI schools. By separating them, we could potentially match, for example, Bruce Randolph Middle School to any other non-HPI middle school, and Bruce Randolph High School to any other non-HPI high school. Therefore, when counted separately, the 6 original co-located HPI schools result in a total of 12. Once co-located HPI schools are counted individually, a total of 36 (24 + 12) schools were originally selected for HPI. Due to the addition of 2 schools, a total of 38 schools (counting the co-located HPI schools individually) were—at least for some portion of the period under study—HPI schools.

For inclusion in the CITS analyses with Pre-Trend Matching each included school is required to have at least 3 years of outcome data, before and after the start of HPI, in order to estimate pre-HPI and post-HPI trends in retention. Of the 38 HPI schools, 5 did not meet that minimum data requirement.[13] Therefore the total number of HPI schools included in the CITS analyses of HPI impacts is 33. See Table 3 for a full list of all 38 schools ever designated as an HPI school, the annual In-School teacher retention rate in each school between 2006 and 2019, and a count of the number of pre-HPI and post-HPI years of data for each school (the 5 schools with insufficient data are noted in red). Table 3 also shows which schools exhibited an increase in average In-School teacher retention rates from the 4 years before to the 4 years after HPI began (the 20 HPI schools shown in green). As we will discuss later, however, we should not interpret these increases as clear evidence that HPI was responsible for those increases.

Results

[12] *One HPI high school shared a building with another school that was not designated HPI-- they are not strictly considered "co-located" since they do not share leadership, but we refer to them as such here because they share a physical space but need to be separated to arrive at the correct HPI school sample size: Respect Academy at Lincoln (506) was not designated HPI and is an alternative school, while Lincoln High (450) was HPI.*

[13] *Oakland elementary opened in 2015—just one year before HPI began—and therefore has too few years of retention data pre-HPI. Kepner Middle School and Henry World Middle schools closed in 2017 and therefore have too few years of retention data post-HPI. Bear Valley Middle School and Kepner Beacon Middle School opened in 2017 and were added to HPI to replace the 2 middle schools that closed. Therefore, they do not have any pre-HPI data.*

<u>Research Objective 1: What are longitudinal patterns in teacher retention in, and exit from, DPS?</u>

1A. Teacher Retention Rates Over Time

What percentage of teachers leave their schools and/or DPS each year and how has this changed over time?

Figure 5 provides information on In-School and In-District DPS teacher retention between 2002 and 2019. As can be seen, retention within DPS schools has generally been increasing since 2013. Retention increased sharply at the onset of the Great Recession, reaching a 2009 peak of 88% within the District and 82% within schools, followed by a decline through 2013.

When comparing the percentages reported in Figure 5 with those reported officially by DPS, differences are explained by the fact that the analytic sample used for this report includes only teachers in regular, non-charter schools in the District, while DPS numbers include non-traditional schools. Additionally, our report of percentages retained In-District is most similar to DPS' "retained in role," as, in this report, we consider a teacher to be retained in the district if they are still a DPS classroom teacher in the schools in our analytic sample.[14]

1B. Retention Rates Across Schools

How much do retention rates vary across schools?

Figure 6 shows substantial variation in In-School retention rates among DPS schools across years. In some years, a small number of schools retain as few as 10-20% of their teachers, while a substantial number of schools approach or reach 100% teacher retention in some years. A large majority of schools retain 70-90% of their teachers each year. The most recent year, 2019, saw a marked decrease of medium-to-low retention rates in schools (significantly fewer schools retaining around 60% of their teachers), and an increase of schools retaining about 80% of their teachers. This indicates that the recent increase in In-School teacher retention is driven by schools that had lower teacher retention rates in past years.

1C. What School Factors Predict Whether Teachers Leave?

Which school-level factors predict teacher retention/exit, including non-malleable factors (e.g., school demographic composition) and malleable factors (e.g., how teachers perceive their school leaders, and school engagement)?

We begin by considering non-malleable school-level factors. Table 4 shows the teacher count and average In-School and In-District teacher retention by different school characteristics. When considering teacher retention by school level, we can see that middle schools experience lower rates of teacher retention both In-School and In-District. For the remaining descriptors, we separated schools into terciles according to some of the demographic characteristics of the students they serve, and then compared the teachers in each set of schools. For example, the third of schools with the highest enrollment (schools ranging from

[14] *When calculated In-District retention, DPS includes <u>retention in any role</u> (i.e., not just teaching positions), which accounts for the difference in numbers reported by DPS compared with what is reported here.*

596 to 2659 students) employ nearly 50% of DPS teachers and have slightly higher In-School teacher retention rates. The third of schools serving the smallest proportions of disadvantaged students (students eligible for FRPL or in historically marginalized racial/ethnic groups) have higher In-School and In-District teacher retention rates.

Figure 7 illustrates several of the key associations between school-level non-malleable factors and In-School teacher retention rates shown in Table 4. Schools with more White students tend to have higher teacher retention on average (green line), although the association between teacher retention and proportion of White students served is flatter in schools where over 50% of students are White. Schools with a greater percentage of Hispanic/Latinx students (red line) and Black students (orange line) tend to have lower teacher retention rates. However, some schools with the highest percentage of Hispanic/Latinx students (>70%) tend to have slightly higher teacher retention rates. Schools with a higher percentage of FRPL-eligible students consistently tend to have lower average teacher retention.

We next consider malleable school-level predictors of teacher retention--specifically, the extent to which teachers' perceptions of school leadership and engagement in their school are associated with retention. Figure 8 illustrates that there are some significant differences between how teachers who decide to leave (yellow bar) or remain (purple bar) in their school at the end of a given year perceive their schools' leadership, according to their CollaboRATE survey responses. These results, shown with averages represented by the height of bars and 95% confidence intervals around each peak, show that teachers who are retained In-School report significantly more positive perceptions of school leadership and engagement on all indices shown.

The largest difference is found in retained and non-retained teachers' scores on their School Engagement index. There are also clear differences in satisfaction with leaders overall and with principals, specifically, between teachers who remain in the same school and those who do not. Responses to questions in the district's Overall DPS Engagement index, as well as Teacher Leader and Assistant Principal questions, show statistically significant, though smaller differences in teacher retention. Overall, these results suggest that a teacher's engagement at the school level and relationship with school leaders are among the most meaningful predictors of whether a teacher will leave their school. Since the School Engagement Index and "All Leader" Index are the strongest predictors of retention, we focus on these two CollaboRATE measures in the next two Figures.

The relationship between perceptions of school leadership and teacher retention is not uniform across all teachers. For instance, we examine whether these associations vary based on teacher experience levels (Figure 9) or teacher race/ethnicity (Figure 10). Figure 9 suggests that—for teachers at all levels of experience—retained and non-retained teachers report very different perceptions of their school leaders. Reported engagement and perceptions of school leaders are equally predictive of retention for teachers at any level of experience. This is not the case when examining these associations by teacher race/ethnicity: Figure 10 shows the association between retention and the two key CollaboRATE indices—School Engagement and Overall Leader[15]. These associations differ across Asian, Black, Hispanic, and

---

[15] *The school engagement index and DPS engagement index are composites of five questions identified in the DPS [2020 Thrive Guide](). The questions are: (1) I enjoy my work at school/DPS. (2) I feel valued as an employee of my school/DPS. (3) I would*

White DPS teachers. The confidence intervals are larger in Figure 10 than in Figure 8 because the teacher-level sample sizes are smaller once we conduct analyses separately by teacher race/ethnicity. While we observe statistically significant differences between retained and non-retained teachers' reports on the School Engagement Index among Asian, Hispanic, and White teachers, differences among Black teachers who are and are not retained are smaller in size and are not statistically significant. In terms of perceptions of School Leaders, differences between retained and not-retained teachers who are Hispanic, White, or Asian are statistically significant, while differences among Black teachers are not. Further, while there is a very small (and not statistically significant) difference in perceptions of school leaders among Black teachers who are and are not retained, Black teachers who are not retained have slightly higher scores on the All Leader Index. This suggests a need for additional analyses of what predicts teacher retention for Black teachers in DPS.

1D. What Teacher Characteristics Predict Whether Teachers Leave Their School?

Which teacher-level factors predict teacher retention/exit from a school (e.g., years of experience, demographic characteristics, subject, grade level)? Why do teachers report leaving (exit surveys)? Has this changed over time?

Table 5 and Figures 11 and 12 present results from logistic regression models predicting teacher retention both In-School and In-District. We ran separate logistic models for In-District and In-School teacher retention as well as separate models including school fixed effects (right side columns labeled "Comparing within Schools) and without them (right side columns labeled "Comparing Across Schools).

Figure 11 illustrates whether teacher race/ethnicity predicts In-School teacher retention, and Figure 12 illustrates whether other teacher characteristics—gender, salary, having an MA or higher degree—predict In-School and In-District teacher retention. The full logistic regression results from which these Figures are made are presented in Table 5.

Figure 11 illustrates how teachers' racial/ethnic identification predicts their In-School retention in comparison to White teachers (White teachers are the reference category because they are the largest racial/ethnic category of teachers in DPS). For Figures 11 and 12, circles represent regression coefficients from Table 5 and vertical lines around circles are 95% confidence intervals. If the confidence interval line intersects the x-axis at 0, then the coefficient is statistically indistinguishable from zero (or no difference) and we can say that the variable does not predict retention. Some racial/ethnic groups, such as Asian teachers, multi-racial teachers, and other or unknown ethnicity teachers are not retained at statistically significantly higher or lower rates than White teachers. Hispanic/Latinx teachers are retained at statistically significantly higher rates than White teachers when comparisons are made within the same school. Black teachers are significantly less likely to be retained than White teachers overall, but when the comparison is made within schools, Black teachers are no longer less likely to be retained. This implies that Black teachers are more likely to work in schools with overall lower retention rates, as explored in question 1C, and that they are more likely to stay in those schools than their White counterparts. The confidence interval for Native American teachers is very wide due to their small numbers in DPS, leading

---

recommend my school/DPS to others as a good place to work. (4) My job has a positive impact on my school/DPS. (5) I am proud to tell people I work for my school/DPS. Of these, only questions 1 and 2 were included in the 2018 survey.

to imprecise estimates. However, despite this imprecision, Native American teachers are statistically significantly less likely to be retained when estimates compare teachers across schools.

Figure 12 provides information on how other teacher characteristics predict the likelihood of a teacher being retained in their school. The figure shows regression coefficients, each from a separate regression, predicting the likelihood of a teacher being retained. The confidence interval line for "female" crosses the x axis, indicating that female and male teachers both within and across schools are retained In-School at similar rates. In contrast, the coefficients for salary, age, and experience are all positive and the confidence intervals do not intersect with zero, meaning that higher values of salary are associated with greater likelihood of retention. The masters-plus variable shows that when comparing teachers across all schools, teachers with and without advanced degrees are retained at roughly the same rates.

The teacher characteristics shown in Figures 11 and 12 all have a simple, linear association with teacher retention. In contrast, teacher age and teacher experience have non-linear associations with teacher retention and are thus shown separately in Figure 13. For teacher age, we can see that retention increases with age between 20 – 50 and then begins to decrease with age thereafter. Similarly, teacher retention peaks for teachers with between 15 and 20 years of DPS experience, while teachers with less or with more experience have lower retention rates. These patterns are similar to those commonly seen in the literature and tend to indicate that very young or very new teachers tend to be in transition or uncertain of their careers, while very old or very experienced teachers are beginning to retire. Age and years of DPS experience are positively correlated (with a correlation of 0.58).

We were also interested in leveraging teacher responses to the DPS Teacher Exit Survey, administered in September of 2016, 2017, and 2018 for teachers who left before that fall. However, because the response rate for these surveys is only about 25%, we do not include analysis of these in the report.

Research Objective 2: Describe HPI Schools and Teachers

2A. Who Receives HPI Incentives and Reported Job Satisfaction[16]?

How many teachers receive HPI incentives, for how long, and at which schools? How many HPI payments (and total amount)? Do HPI teachers[17] report higher job satisfaction than similar counterparts who do not?

Table 6 presents the number of employees who receive HPI incentive payments and the mean dollar amount of those payments, across the 4 years of HPI implementation. As the table indicates, the number

---

[16]*The items and indices included here from 2018 and 2019 administrations of the CollaboRATE survey are those that provide the most information about "job satisfaction". Hence, we use these items as they provide useful information and are proxies for teacher job satisfaction. We currently do not have incentive payment data in a form that allows us to determine which teachers in HPI schools received bonuses. We note that, because virtually all DPS teachers receive LEAP scores above the bottom category, that very few teachers in HPI schools were ineligible for bonuses.*

[17]*Less than 1% of all teachers in HPI schools received a "does not meet" LEAP category rating, and those are the only teachers who would not receive an HPI incentive. The job satisfaction of all teachers in HPI schools is therefore indistinguishable from the job satisfaction of teachers receiving HPI payments. We assume that HPI incentives were distributed in accordance with the HPI policy.*

of employees receiving the HPI Retention Bonus has increased in each year, with 924 receiving it in 2019. The number of employees receiving an HPI Monthly bonus ranges between roughly 1200 to 1300 each year. As is noted in Table 1, Table 6 shows that School Leaders are paid bonuses that are much larger than those received by teachers.

Figure 14 shows CollaboRATE responses from surveys administered in 2018 and 2019 for teachers in HPI and non-HPI schools. All indices shown in the figure are constructed from responses to CollaboRATE survey items ranging on a Likert Scale from 1 (least satisfied) to 5 (most satisfied). Results indicate that teachers in HPI and non-HPI schools report similar average levels of overall engagement and perceptions of school leadership. Teachers in HPI schools report slightly higher engagement at the District level, but, on average, report slightly lower school engagement. As we explored in Q1C, the school-level engagement index is the stronger predictor of teacher retention. Lower school-level engagement ratings on CollaboRATE predict a lower likelihood of retention, on average, among teachers in HPI schools compared with teachers at non-HPI Schools.

2B. How do HPI Schools Differ from non-HPI Schools?

How do HPI-eligible teachers/schools differ from non-HPI schools/teachers? Are there comparable non-HPI schools in DPS?

Table 7 compares the demographic profile of HPI and non-HPI schools, averaged across the three years before and after HPI began, with student characteristics differing substantially across HPI and non-HPI schools without much change across the pre- and post-HPI three year averages. Results indicate that average years of teacher experience among DPS teachers decline over this time period which is related to the fact that both HPI and non-HPI schools were staffed with more teachers in their first three years of teaching in DPS in the three years after HPI implementation began. Across the three years prior to and 3 years following HPI implementation, teachers in HPI schools had lower retention rates, lower LEAP scores, and fewer years of experience teaching in DPS. For example, in the post-HPI implementation years, 44% of teachers in HPI schools had taught in DPS for three years or less, compared with about 34% of teachers in non-HPI schools.

These patterns could arise for several reasons. For example, HPI schools may struggle to recruit and/or retain more experienced and effective teachers, as is suggested by average information on teacher experience and LEAP scores. It is also possible that it is more difficult to receive high performance evaluations in harder-to-serve school settings. Both of these hypotheses are supported by evidence from other districts and states. It is beyond the scope of this Report to disentangle the causes of differential LEAP scores by HPI status. In both HPI and non-HPI schools, there were notable increases in LEAP scores and the percentage of teachers rated effective or higher across the pre-HPI and post-HPI periods.

2C. Did Retention Increase in HPI Schools once HPI Began?

Figure 15 illustrates pre-HPI and post-HPI trends in In-School teacher retention rates, separately for elementary, middle, K-08, and high schools. A pattern of results that would indicate clear effects of HPI in these figures would show a systematic shift at the "shock" point when HPI was initiated (2016), and/or a notable shift in post-HPI trends in teacher retention. As Figure 15 shows, there does not seem to be a consistent pattern on retention rates before and after HPI was implemented across HPI schools.

Figure 16 illustrates retention pre- and post-HPI by teacher LEAP scores. These data are only available starting in the 2013-2014 school year, making it difficult to see trends. As Figure 16 shows, retention is higher for higher-rated teachers in both HPI and non-HPI schools, but there is no clear pattern of differences between HPI and non-HPI.

2D. Where Do Teachers Who Leave HPI Schools Go Next?

When teachers exit HPI schools, what do they do next (e.g., new non-HPI school, new HPI school, exit DPS)?

Figure 17 shows the movement of teachers in HPI schools after the end of the school year in 2016 through 2019. The majority of teachers are retained in their schools, while the next largest number leave the DPS traditional non-charter school workforce (unfortunately, we cannot distinguish between those who leave the district entirely and those who move to charter schools within the district). Some teachers move to non-teacher positions within the district, and small numbers move to a teaching position in another DPS school. More teachers move to non-HPI schools than to HPI schools, which is in part a result of the fact that there are many more non-HPI schools than HPI schools in the district. The fact that so few HPI teachers transfer to different HPI schools means that analyses of In-School retention for HPI captures the vast majority of HPI teachers who remain in HPI schools.

We can better compare the movement to HPI and non-HPI schools if we compare teachers leaving HPI schools to teachers leaving non-HPI schools, as shown in Table 8. We can see that teachers leaving HPI schools are more likely to move to another HPI school than teachers leaving non-HPI schools.

Research Objective 3: Estimate Impacts of HPI Policy on Teacher, Principal, and Student Outcomes

3A. Did the HPI Policy Increase Teacher In-School Retention?

Matching Results for CITS Analysis of In-School Teacher Retention

Successful matching occurs when we can identify, for each HPI school, a non-HPI school that had a very similar trend in In-School teacher retention in the years leading up to the start of the HPI policy. We match schools by level (e.g., elementary HPI schools can only be matched to elementary non-HPI schools). Figure 18 provides an example of an HPI school—Goldrick Elementary (in green)—and the non-HPI elementary school to which it was matched (blue) using our primary matching method—1:1 NN Matching without replacement.[18]

---

[18] *Our preferred matching method is "1:1 NN Matching without Replacement". This means that each HPI school will be matched to exactly 1 non-HPI school (1:1), each HPI school will be matched to its "nearest neighbor" (NN) with the most similar trend in In-School teacher retention in the pre-HPI period of 2006 – 2015, and that each non-HPI school can serve as a match to only 1*

For Goldrick Elementary School (School ID 244), In-School retention rates were falling from 2006 through 2015 by -2.2 percentage points (PP) per year. The matching process for Goldrick Elementary is considered successful because its matched non-HPI school had a very similar pre-HPI trend in In-School teacher retention. The pre-trend for its matched non-HPI school was also -2.2 PP/year. Recall that only pre-trends (as opposed to levels) need to align for the CITS approach to produce unbiased estimates of causal effects.

In Figure 19, we show all DPS elementary schools, ranked from lowest (left side of X-axis) to highest (right side of X-axis) in terms of their pre-HPI trend in In-School teacher retention rates. Schools in green are HPI schools, and Goldrick Elementary is shown with its -2.2 PP/year pre-trend (Y-axis), and labeled with a "1" to indicate it was matched to 1 nearby non-HPI school (blue Xs).

Figure 19 shows which HPI schools (green) were matched using the primary matching method, with how many (green number), each school's pre-HPI trend in In-School teacher retention, which non-HPI schools were selected as matches (blue), and which non-HPI schools were not included in the analysis (red) because they were not matched to any HPI schools. The upper-left corner of Figure 19 provides a summary report of the number of elementary HPI schools that were matched (N=12) and the number that did not have a match (N=0, all elementary schools were matched). It also reports the number of elementary non-HPI schools that were used as a match (N=12) and those that were not (N=49). Please see Appendix C for a version of Figure 19 for In-School Retention matching for middle, K-08, and high schools.

In Table 9, we summarize the total number of HPI and non-HPI schools, by In-School retention match status and school level. Overall, 30 of the 33 HPI schools were successfully matched with 30 non-HPI schools. Note that there are 3 HPI high schools that had no In-School retention match using the primary matching method.[19] For these 3 HPI high schools, there were no non-HPI high schools in DPS with similarly negative pre-HPI trends in In-School teacher retention.

CITS Results: In-School Teacher Retention

Figure 20 presents the results of the CITS analysis for elementary schools using the preferred matching method. As described in the methodological section (see page 8), there are two potential effects of interest—a shock effect and a trend effect. If HPI had a positive effect, we would expect to see a positive and statistically significant shock and/or trend.

The results graphed in Figure 20 are not statistically significant (indicated by the (ns) superscript). The shock effect is -6.2 percentage points (PP). This means that the average In-School teacher retention rate dropped by 6.2 PP *more* in HPI schools than in similar non-HPI schools in the first year following HPI

HPI school (without replacement). In Appendix C, we also use a matching approach that allows many non-HPI schools within a certain range to be matched to each HPI school instead of just 1 match per HPI school (caliper matching), and non-HPI schools can be re-used as a match to multiple HPI schools (with replacement). The overall pattern of results is similar using this method.

[19] *See Appendix C for matching results using an alternative matching approach, M:1 caliper matching with replacement. However, under both matching approaches 3 HPI high schools remain unmatched.*

implementation. However, this difference between HPI and non-HPI schools is not statistically distinguishable from 0.

In both HPI schools and the matched non-HPI schools, we can see that within school teacher retention was falling prior to 2016, and in both sets of schools, these retention rates began to increase in 2016, with HPI school trends increasing by 2.0PP more than non-HPI schools. If results were statistically significant, this would indicate a positive effect of HPI, but the effect is not statistically distinguishable from zero.

Table 10 presents the CITS-estimated shock and trend effects on In-School teacher retention, separately for elementary, middle, K-08, and high schools, as well as all schools combined. The top panel presents results using the preferred matching approach (1:1 NN without replacement), for which none of the estimated shock or trend effects are statistically significant. Further, the results do not indicate a clear pattern in either a positive or negative direction for the impacts of HPI on In-School teacher retention.

Turning to results from the secondary matching approach (1:M caliper matching with replacement) shown in Table 10, all but 1 of the 10 estimated shock/trend effects are not statistically significant. We do observe a positive trend effect of +3.1 PP/year for elementary schools using the secondary matching approach. However, when we combine all schools into a single analysis, the trend effect (+1.9) is no longer significant. We interpret this single statistically significant coefficient as possibly being the result of chance variation as opposed to an effect of the HPI policy.

Taken together, these results indicate that there were no clear impacts of the HPI policy on In-School teacher retention. Here, again, we remind the reader that experimental, and quasi-experimental research methods such as the CITS approach employed here, allow for the estimation of average effects of the HPI program. It is not possible to determine whether the HPI policy caused changes in outcomes in any specific HPI school.

CITS Results: Are Certain Teachers Retained In-School at Higher Rates?

Finally, we estimated CITS models to explore whether the analyses might detect differential effects of the HPI program on In-School Retention separately for teachers receiving larger HPI incentive payments-- that is, separately by LEAP rating. In theory, it is possible that HPI could have incentivized teachers who receive the largest bonuses--teachers with "distinguished" LEAP ratings"--to remain in their schools at higher rates. In practice, there are statistical power limitations with such an analysis. The CITS analyses described above using all teachers *together* suffer from having fewer observations than would be ideal. Once we attempt to separate the CITS analyses into smaller groups of teachers by LEAP category rating, the sample sizes become even smaller. We find no evidence of any clear patterns in HPI In-School Retention effects by teacher LEAP category.

3B. Did the HPI Policy Increase Teacher In-District Retention?

Matching Results for CITS Analysis of In-District Teacher Retention

We now turn to exploring whether the HPI policy had any measurable impact on schools' In-District (as opposed to In-School) teacher retention rates. Figure 21 shows which HPI elementary schools (green) were matched using the primary matching approach (1:1 NN), as well as how many matches each HPI school received (green number), each school's pre-HPI trend in In-District teacher retention, which non-HPI schools were selected as matches (blue), and which non-HPI schools were not included in the analysis (red) because they were not matched with any HPI schools.

The upper-left corner of Figure 21 (for In-District teacher retention) provides a summary report of the number of elementary HPI schools that were matched (N=12) and the number that did not have a match (N=0, all elementary schools were matched). It also reports the number of elementary non-HPI schools that were used as a match (N=12) and those that were not (N=49). Please see Appendix C for a version of Figure 21 for In-District Retention matching for middle, K-08, and high schools.

In Table 11, we summarize the total number of HPI and non-HPI schools, by match status and school level for In-District teacher retention. Overall, 29 of the 33 HPI schools were successfully matched with 29 non-HPI schools. Table 11 also highlights a limitation of this CITS analysis. There are 4 HPI high schools with no match in terms of the pre-HPI trend in In-District teacher retention (using the primary matching approach, 1:1 NN without replacement). See Figure 22 for the high school matching results for In-District Retention.

Figure 22, which presents high school matching results for In-District retention, illustrates why there are 4 HPI schools without matches. There are 7 HPI high schools with negative pre-HPI trends in In-District retention (i.e., below the horizontal red line in Figure 22). However, there are only 3 non-HPI high schools with negative pre-HPI trends. Therefore, 4 of the 7 HPI high schools remained unmatched. This single statistically significant coefficient as a possibly being the result of chance variation as opposed to an effect of the HPI policy.

CITS Results: In-District Teacher Retention

Table 12 presents results from the CITS analysis using the primary matching method (1:1 NN without replacement). If HPI had a positive effect, we would expect to see a positive and statistically significant shock and/or trend. Table 12 presents the CITS-estimated shock and trend effects for In-District teacher retention separately for elementary, middle, K-08, and high schools, as well as all schools combined. The top panel presents results using the preferred matching approach (1:1 NN without replacement), for which only 1 of the 10 estimated shock or trend effects are statistically significant. The significant coefficient suggests the possibility that, for high schools, the HPI program may have had an immediate, positive impact on In-District teacher retention. However, this result should be interpreted with caution, given that all of the other shock and trend estimates are not statistically significant and that only 4 of the 7 HPI high schools could be matched and are included in this analysis. Further, this single statistically significant coefficient is possibly the result of chance variation as opposed to an effect of the HPI policy.

Turning to results shown in the lower panel of Table 12 from the secondary matching approach (1:M caliper matching with replacement), none of the estimated shock or trend effects are statistically significant. Taken together, these results suggest that there were no clear impacts of the HPI policy on In-District teacher retention.

3C. Did the HPI Policy Attract Teachers with Higher LEAP Scores to HPI Schools?

Does the HPI program help school leaders attract teachers with higher LEAP scores to high priority schools?

Matching Results for CITS Analysis of LEAP Scores of Recruited (New-to-School) Teachers

We now turn to exploring whether the HPI policy had any measurable impact on the LEAP scores of the teachers principals are able to attract to their schools. Here, the outcome of interest is the average LEAP score of those teachers who are new to a given school in a given year. Note that LEAP scores are only available in 2014 through 2019. We hypothesize that principals could first leverage HPI bonuses to attract new teachers to their schools as early as Fall of 2016.

In Table 13, we summarize the total number of HPI and non-HPI schools, by match status and school level for the CITS analysis of new-to-school teachers' LEAP scores. Overall, 26 of the 31 HPI schools that have at least 3 years of pre-HPI LEAP data were successfully matched with 26 non-HPI schools. Table 11 also highlights a limitation of this CITS analysis. There are 6 HPI high schools with no match in terms of the pre-HPI trend in the LEAP scores of new-to-school teachers (using the primary matching approach, 1:1 NN without replacement).

CITS Results: LEAP Scores of Recruited (New-to-School) Teachers

Table 14 first presents results from the CITS analysis using the primary matching method (1:1 NN without replacement). If HPI had a positive effect, we would expect to see a positive and statistically significant shock and/or trend. Table 14 presents the CITS-estimated shock and trend effects for new-to-school teachers' LEAP scores separately for elementary, middle, K-08, and high schools, as well as all schools combined. The top panel presents results using the preferred matching approach (1:1 NN without replacement), and the lower panel of Table 14 from the secondary matching approach (1:M caliper matching with replacement). None of the estimated shock or trend effects are statistically significant. Taken together, these results suggest that there were no clear impacts of the HPI policy on the LEAP scores of the teachers that principals were able to recruit to HPI schools.

3D. Are principals in HPI schools more likely to be retained than their non-HPI counterparts?[20]

Figure 23 shows the percentage of principals who were retained In-School between 2006 and 2019 separately for HPI Schools, all non-HPI schools, and non-HPI schools designated hard-to-serve. Note that the variable and increasing number of HPI schools over time reflects new and/or changing schools prior

[20] *We note that this question is not included in the Work Scope agreed upon by DPS and DCTA. While a causal analysis answering this question is not possible, we provide descriptive information addressing this question here.*

to the onset of HPI as well as details on the number of HPI schools following the program's onset described in detail on page 15 under "Sample of HPI Schools Included in CITS Impact Analysis**"**.

Beginning in 2014, prior to the onset of HPI, we see a small upward trend in principal retention across schools through 2016. In 2017, however, we observe a marked increase in principal retention in HPI schools, with 97% of HPI principals retained; a pattern that diverges from a slight downward trend in principal retention in non-HPI schools. However, the jump in principal retention in HPI schools is not sustained; in 2018, principal retention rates in HPI schools dropped to around 70%, bumping back up to 83% in 2019. Principal retention rates in non-HPI schools during this time were relatively stable. Therefore, in 2 of the 4 post-HPI years (2017 and 2019), principal retention rates were higher in HPI schools than in non-HPI schools. However, in both 2016 and 2018, principal retention rates were lower in HPI schools, even when compared with non-HPI schools that were designated hard-to-serve. The near-100% principal retention rates in HPI schools in 2017 suggests a possible single-year effect of the HPI principal bonus.

3E. Did the HPI Policy Improve Student Achievement in HPI Schools?

Matching Results for CITS Analysis of HPI Impacts on Student Achievement

We now turn to exploring whether the HPI policy had any measurable impact on student achievement scores on assessments given statewide. Here, the outcome of interest is the school's mean achievement score (math or ELA) in a given year, and scores have been standardized around the district mean within subject-grade-year.

In Table 15, we summarize the total number of HPI and non-HPI schools, by match status and school level for the CITS analysis of student achievement. Overall, 28 of the 29 HPI schools that have at least 3 years of pre-HPI student achievement data were successfully matched with 28 non-HPI schools. There was 1 middle school that could not be matched.

If HPI had a positive effect, we would expect to see a positive and statistically significant shock and/or trend. Table 16 presents the CITS-estimated shock and trend effects on student achievement (ELA and Math) separately for elementary, middle, K-08, and high schools, as well as all schools combined. The top panel presents results using the preferred matching approach (1:1 NN without replacement), and the lower panel of Table 16 presents results from the secondary matching approach (1:M caliper matching with replacement). Taken together, the CITS results suggest that there were no clear impacts of the HPI policy on ELA or Math achievement scores. Of the 20 estimated shock effects, only 1 was statistically significant (which is what we would expect to occur by chance alone). Likewise, of the 20 trend effect estimates, only 1 was statistically significant.

Future Work: Phase II - Explore Mechanisms for Phase I Findings

The Work Scope (Appendix A) states the following: "After the completion of Phase I [the current Report], results will be presented to both DCTA and DPS, and we will consider whether there is interest in pursuing Phase II [exploring mechanisms for Phase I Findings]. The design of any new data collection materials would be done in consultation with DPS and DCTA stakeholders." Given that we did not find effects, positive or negative, of HPI on In-School teacher retention, it is not clear at this point whether there will be a Phase II component to this work. This will be a point of discussion once the current report has been released.

Summary and Conclusion

During the 2015-2016 school year, DPS began providing targeted financial incentives to over 1,500 teachers, school leaders, Special Service Providers, and other DPS employees working in approximately 30 schools as part of its Highest Priority Incentives program (HPI). Incentive eligibility for teachers and SSPs is contingent upon performance. Teachers are eligible for two HPI incentives. The first is a Monthly Incentive ranging from $125 to $250 per month. The second is an Annual Incentive ranging from $500 to $1000 for those who remain in an HPI school. Smaller incentives are paid to SSPs and other employees in HPI schools. Much larger bonuses – up to $30,000 per year in total – are paid to HPI school leaders. School leader bonuses are not dependent on evaluation scores.

Teachers working in HPI schools receive bonuses if they remain in their school over time and have performance evaluations above the lowest "not meeting" designation. Over 99% of HPI teachers receive bonuses, with bonus size increasing based on performance evaluation category. The intent of HPI payments is to increase retention of teachers with adequate performance in HPI schools. In this way, the HPI program was designed to contribute to efforts to retain teachers in the District's schools that had the greatest need for additional support. The current study estimates the effect of the program on teacher retention in HPI schools.

HPI's goal was to target incentives to schools that struggle to recruit and retain effective teachers. Descriptive results indicate that HPI schools served larger proportions of non-White and ELL designated students and had lower teacher retention rates in the three years before HPI began. Isolating the causal effects of the HPI Program is a challenge that is addressed using a quasi-experimental method, Comparative Interrupted Time Series (CITS), which uses a comparison group of non-HPI schools with similar trends in teacher retention prior to the onset of HPI to estimate the effect of HPI on teacher retention.

We do not find evidence that HPI had measurable impacts on key outcomes of interest. Results of the CITS analyses indicate no statistically significant average effect of the HPI program on teacher retention (in-school or in-district). Though one might hypothesize that HPI effects could be concentrated among teachers receiving the largest incentives (based on LEAP), we find no evidence of differential HPI effects by LEAP score categories. We also do not find that HPI helped attract teachers with higher LEAP scores to HPI schools. All of the school leaders in HPI schools were retained in the first year following implementation of the HPI program. However descriptive analyses suggest that school leader retention in HPI schools was similar to that observed in non-HPI schools in subsequent years. We conclude the Report with a reflection on what relevant research suggests might explain the null results reported here.

It is important to note several limitations of these analyses. First, CITS can be used to estimate an average policy effect only; we cannot determine whether HPI caused changes within individual schools. Second, because there are a relatively small number of HPI schools, the fact that estimates indicate no effect of HPI must include the caveat that these analyses would be unlikely to detect small effects of the HPI program.

Contextualizing HPI Results Using Relevant Research

In considering why the HPI program did not have detectable positive effects on teacher outcomes, it is useful to consider relevant research from other contexts. We refer the interested reader to Podgursky & Springer (2007) for a comprehensive review of theories (and early evidence) that both support and challenge the basic logic of teacher financial incentives (e.g., difficulties of measuring teacher job performance, the team-based nature of the profession, the potential for incentivizing neglect of unmeasured dimensions of teachers' work, a misalignment with what intrinsically motivates teachers). At that time, little empirical evidence existed to adjudicate between these competing theories. Since then, a substantial number of incentive pay programs have been implemented, coupled with a growing body of empirical work documenting program effects.

Some studies focus on estimating policy impacts on *teacher* outcomes, and these find mixed results. For example, Glazerman & Seifullah (2012) found short-term impacts of the Chicago Teacher Advancement Program on in-school teacher retention, though not in-district retention. A $5000 retention bonus for highly-rated teachers in low-performing Tennessee schools increased retention for tested-subject teachers (Springer, Swain, & Rodriguez, 2016). Feng & Sass (2018) found that loan forgiveness incentives for Florida teachers in hard-to-staff positions reduced attrition among these teachers. The Talent Transfer Initiative (TTI) that offered $20,000 incentive payments to highly effective teachers to transfer to low-achieving schools induced 5% of TTI candidates to transfer (Glazerman, Protik, Teh, Bruch, & Max, 2013). One study indicated that pay-for-performance (PFP) did not cause short-term changes in teacher practices (Yuan et al., 2013), while another showed that teachers changed their practice in response to a threatened loss of incentives (Fryer, Levitt, List, & Sadoff, 2012).

Other studies have focused on the effects of teacher incentive payment programs on student outcomes. Using nationally representative survey data, Figlio & Kenny (2007) found higher test scores in districts with incentive pay. However, studies of individual policies have not found positive effects on student achievement (Glazerman & Seifullah, 2012; Springer et al., 2011). While many studies based in the U.S. have found little to no impact of PFP on student achievement (see also Cowan & Goldhaber, 2018; Fryer, 2011; Marsh, Springer, McCaffrey, Yuan, & Epstein, 2011), some international studies in developing countries have found significant effects of incentive pay on student outcomes (see, e.g., Atkinson et al., 2009; Duflo, Hanna, & Rya, 2012; Kremer, Ilias, & Glewwe, 2003; Muralidharan & Sundararaman, 2008).

In sum, the evidence base on teacher incentive pay is mixed for both student achievement and teacher retention. Evidence also indicates that long-term, as opposed to one-time, financial incentives may be more effective, and that communication regarding program eligibility is important (Clotfelter et al., 2008). Findings have been particularly tepid with respect to stimulating cross-school transfers or raising overall teacher retention within a district. However a recent study by Dee & Wyckoff (2015) found that Washington DC's IMPACT teacher evaluation and incentive system increased retention among high performing teachers, improved teacher performance, and raised student achievement. This suggests that strategic incentives for retention of effective teachers may be a promising avenue for further study. However, evidence indicates that these incentives likely need to be sizable, differentiated based on the characteristics of teachers the district seeks to retain, sustained over time, and communicated clearly to teachers.

Recommendations

While HPI includes both monthly and annual incentives that continue over time, those incentives are relatively small, particularly when compared with the teacher incentive programs that have been documented as successful in the research described above. Further, although HPI was designed, in principle, to focus on the strategic retention of effective teachers, in practice 99% of teachers in HPI schools receive HPI incentive pay.

1. ***Given lessons from prior research, when DPS revisits the HPI program design, it may be worth considering offering larger payments to a smaller and more intentionally-selected subset of teachers whom the district identifies as especially important to recruit and retain to HPI schools.***

In contrast to teacher incentives, the HPI incentive payments for school leaders are large. However, they are not linked to evidence of effective leadership. This could have the unintended consequence of retaining school leaders who have received negative evaluations from their teachers. If this has occurred, it could unintentionally result in the exit of teachers who are dissatisfied with their school leadership and for whom the bonus pay does not make up for perceived working conditions. HPI was designed not only to stabilize teacher/staff turnover, but also to recruit and retain the strongest teachers and school leaders. The HPI program as implemented to date may not be incentivizing this strategic retention.

2. ***When the HPI program design is revisited, it will be worth considering whether and, if so, how to include or continue the leadership incentives.***

There is also evidence from prior research that inadequate communication regarding program eligibility may diminish overall effects (Clotfelter et al., 2008). Effective communication about a new teacher compensation program is therefore essential. It is possible that a teacher incentive program might be designed in a way that could be quite effective, however, if school leaders or teachers are not *aware* of the program, do not have buy-in, or are not notified when they receive those payments, potential program effects could be undermined.

3. ***Future versions of the HPI program would benefit from a clear plan for (a) communicating with school leaders and their supervisors about how the incentive could be leveraged for recruitment and retention, (b) explaining the purpose and goals of the HPI program districtwide to increase buy-in at all levels, and (c) experimenting with how to effectively communicate with teachers about both their eligibility for and receipt of incentives.***

Finally, we reflect briefly on the challenges the authors of this report confronted in evaluating whether HPI had a causal effect on teacher outcomes. More than 30 schools in DPS struggle to retain effective teachers; therefore, there was no single, "right" way to select 30 schools for the HPI program. Given that the selection criteria remained somewhat unclear after-the-fact, there were likely missed opportunities to estimate causal effects of HPI.

4. ***When possible, it would be beneficial to identify clear school eligibility and selection criteria at the outset, and – in cases where more schools meet those criteria than can possibly be served by a program or policy – use a randomization procedure to equitably select eligible schools, leaders, or teachers for inclusion. TWC, the RPP between CU and DPS, is positioned not only to conduct evaluations of DPS programs after-the-fact, but to engage in co-design (e.g., design-based implementation research; Fishman, Penuel, Allen, Cheng, & Sabelli, 2013). of programs, policies, and practices. Engaging TWC members during the design and early implementation phase of a new program like HPI may improve the likelihood that, several years into program implementation, it is possible to ascertain whether a given policy had the intended effects. We hope that TWC, the growing partnership between CU-Boulder and DPS, along with its connections to other key Denver stakeholders such as DCTA – might provide such a resource to DPS in the future.***

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, *25*(1), 95–135.

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (n.d.). *Teacher Turnover, Teacher Quality, and Student Achievement in DCPS*. 23.

Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H., & Wilson, D. (2009). Evaluating the impact of performance-related pay for teachers in England. *Labour Economics*, *16*(3), 251–261.

Boyd, Don, Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2011). The role of teacher quality in retention and hiring: Using applications to transfer to uncover preferences of teachers and schools. *Journal of Policy Analysis and Management*, *30*(1), 88–110.

Boyd, Donald, Lankford, H., Loeb, S., & Wyckoff, J. (2013). Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Disentangling the Preferences of Teachers and Employers. *Journal of Labor Economics*, *31*(1), 83–117. JSTOR.

Clotfelter, C. T., Glennie, E. J., Ladd, H. F., & Vigdor, J. L. (2008). Teacher Bonuses and Teacher Retention in Low-Performing Schools: Evidence from the North Carolina $1,800 Teacher Bonus Program. *Public Finance Review*, *36*(1), 63–87.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review*, *26*(6), 673–682.

Cowan, J., & Goldhaber, D. (2018). Do bonuses affect teacher staffing and student achievement in high poverty schools? Evidence from an incentive for national board certified teachers in Washington State. *Economics of Education Review*, *65*, 138–152.

Dee, T. S., & Wyckoff, J. (2015). Incentives, Selection, and Teacher Performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, *34*(2), 267–297.

Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives Work: Getting Teachers to Come to School. *The American Economic Review*, *102*(4), 1241–1278. JSTOR.

Duncan, G. J., & Murnane, R. J. (2011). *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*. Russell Sage Foundation.

Engel, M., Jacob, B. A., & Curran, F. C. (2014). New Evidence on Teacher Labor Supply. *American Educational Research Journal*, *51*(1), 36–72.

Feng, L., & Sass, T. R. (2018). The Impact of Incentives to Recruit and Retain Teachers in "Hard-to-Staff" Subjects. *Journal of Policy Analysis and Management*, *37*(1), 112–135.

Figlio, D. N., & Kenny, L. W. (2007). Individual teacher incentives and student performance. *Journal of Public Economics*, *91*(5), 901–914.

Fishman, B. J., Penuel, W. R., Allen, A. R., Cheng, B. H., & Sabelli, N. O. R. A. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. *National society for the study of education*, *112*(2), 136-156.

Fryer, J., Roland G., Levitt, S. D., List, J., & Sadoff, S. (2012). *Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment* (Working Paper No. 18237; Working Paper Series). National Bureau of Economic Research.

Fryer, R. G. (2011). *Teacher Incentives and Student Achievement: Evidence from New York City Public Schools* (Working Paper No. 16850; Working Paper Series). National Bureau of Economic Research.

Glazerman, S., Protik, A., Teh, B., Bruch, J., & Max, J. (2013). Transfer Incentives for High-Performing Teachers. In *Mathematica Policy Research Published Reports; Princeton*. Mathematica Policy Research.

Glazerman, S., & Seifullah, A. (2012). An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) After Four Years. *Final Report*, 106.

Glewwe, P., Ilias, N., & Kremer, M. (2010). Teacher Incentives. *American Economic Journal: Applied Economics*, *2*(3), 205–227. JSTOR.

Goldhaber, D. (2002). Mystery of good teaching: The evidence shows that good teachers make a clear difference in student achievement. The problem is that we don’t really know what makes a good teacher. (Feature). *Education Next*, *2*(1), 50-. Gale Academic OneFile.

Gorski, E. (2015, September 3). *Denver teachers union: DPS skirted bargaining rules with new incentive pay program*. Chalkbeat Colorado. https://co.chalkbeat.org/2015/9/3/21092588/denver-teachers-union-dps-skirted-bargaining-rules-with-new-incentive-pay-program

Hanushek, E. A., Kain, J. F., O’Brien, D. M., & Rivkin, S. G. (2005). *The Market for Teacher Quality* (Working Paper No. 11154; Working Paper Series). National Bureau of Economic Research.

Hanushek, E., Kain, J., & Rivkin, S. (1999). *Do Higher Salaries Buy Better Teachers?* (No. w7082; p. w7082). National Bureau of Economic Research.

Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., Epstein, S., Koppich, J., Kalra, N., DiMartino, C., & Peng, A. (2011). *Big Apple for Educators: New York City’s Experiment with Schoolwide Performance Bonuses:Final Evaluation Report*. RAND Corporation, The.

Muralidharan, K. (2008). *Teacher Performance Pay: Experimental Evidence from India*. Program on Education Policy and Governance, Kennedy School of Government.

Murnane, R. J., & Phillips, B. R. (1981). Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance. *Economics of Education Review*, *1*(4), 453–465.

Podgursky, M., & Springer, M. G. (2007). Credentials Versus Performance: Review of the Teacher Performance Pay Research. *Peabody Journal of Education*, *82*(4), 551–573.

Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *The American Economic Review*, *94*(2), 247–252. JSTOR.

Springer, M. G., Ballou, D., Hamilton, L., Vi-Nhuan L, J. R., Lockwood, D. F., McCaffrey, M., & Stecher, B. M. (2011). *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching (POINT)*. SREE.

Springer, M. G., Swain, W. A., & Rodriguez, L. A. (2016). Effective Teacher Retention Bonuses: Evidence From Tennessee. *Educational Evaluation and Policy Analysis*, *38*(2), 199–221.

Yuan, K., Le, V.-N., McCaffrey, D. F., Marsh, J. A., Hamilton, L. S., Stecher, B. M., & Springer, M. G. (2013). Incentive Pay Programs Do Not Affect Teacher Motivation or Reported Practices: Results From Three Randomized Studies. *Educational Evaluation and Policy Analysis*, *35*(1), 3–22.

Tables, Figures, and Appendices

(see next page)

List of Tables and Figures

List of Tables and Figures 34
Tables 36
Table 1. HPI Payment Amounts, by School Role and Evaluation Ratings 36
Table 2. All HPI Schools with Information on SCI Scores, Current Student Demographics, and Pre-HPI Retention Rates 37
Table 3. Annual In-School Retention Rates in HPI Schools, with Depiction of Available Data 38
Table 4. RQ 1C – School-Level Retention Rates, Separately by School Characteristics 39
Table 5. RQ 1D – Teacher Characteristics that Predict In-School and In-District Retention 40
Table 6. RQ 2A – Number of Employees Receiving HPI Incentive Payments and Average Amounts, by School Year 41
Table 7. RQ 2B – School Characteristics of HPI and Non-HPI Schools, Averaged across the 3 Years Pre-HPI and 3 Years Post-HPI 42
Table 8. RQ 2D – HPI vs Non-HPI Schools: Annual Movements of Teachers, Separately in the 4 Years of HPI Implementation 43
Table 9. RQ 3A – Matching on Pre-HPI Trend Results for CITS Analysis of In-School Teacher Retention: Number of Matched and Unmatched Schools 44
Table 10. RQ 3A –CITS Results using Pre-Trend Matching: HPI Shock and Trend Effects on In-School Teacher Retention 45
Table 11. RQ 3B – Matching on Pre-HPI Trend Results for CITS Analysis of In-District Teacher Retention: Number of Matched and Unmatched Schools 46
Table 12. RQ 3B –CITS Results using Pre-Trend Matching: HPI Shock and Trend Effects on In-District Teacher Retention 47
Table 13. RQ 3C – Matching on Pre-HPI Trend Results for CITS Analysis of LEAP Scores of Recruited (New-to-School) Teachers: Number of Matched and Unmatched Schools 48
Table 14. RQ 3C –CITS Results using Pre-Trend Matching: HPI Shock and Trend Effects on LEAP Scores of Recruited (New-to-School) Teachers 49
Table 15. RQ 3E – Matching on Pre-HPI Trend Results for CITS Analysis of Student Achievement: Number of Matched and Unmatched Schools 50
Table 16. RQ 3E –CITS Results using Pre-Trend Matching: HPI Shock and Trend Effects on Student Achievement 51

Figures 52
Figure 1. Elementary Schools Selected for HPI as a Function of Enrollment and SCI Measures 52
Figure 2. Secondary Schools Selected for HPI as a Function of Enrollment and SCI Measures 53
Figure 3. Average In-School Retention Rate and Average % of FRPL-Eligible Students in 3 Years Prior to HPI (does not include charter schools) 54
Figure 4. Number of DPS Schools including Main Analytic, Alternative/Charter Schools 2002 - 2020 55
Figure 5. RQ 1A –Teacher Retention Over Time, 2002–2019, Separately for In-School and In-District Retention Rates 56
Figure 6. RQ 1B – Variability in In-School Retention Rates across DPS schools 2006 to 2019 57
Figure 7. RQ 1C—Relationship Between DPS Schools' Annual Retention Rate and the School's Student Demographics 58
Figure 8. RQ 1C—Relationship between Teachers' Perceptions of Their School (Leadership, Engagements) and Teacher In-School Retention 59
Figure 9. RQ 1C — Does the Relationship between CollaboRATE and Retention Differ by Years of DPS Experience? 60
Figure 10. RQ 1C—Does the Relationship between CollaboRATE and Retention Differ by Teacher Race? 61
Figure 11. RQ 1D – Does Teacher Race/Ethnicity Predict In-School Teacher Retention? 62
Figure 12. RQ 1D – Do Other Teacher Characteristics Predict In-School Teacher Retention? 63
Figure 13: RQ 1D – Percent of Teachers Retained in School at Different Teacher Age Bins (Left Panel) and Teacher Experience (Right) 64
Figure 14. RQ 2A — CollaboRATE Results in HPI and Non-HPI Schools 65
Figure 16. RQ 2C—Teacher Retention by LEAP Rating over time in HPI and non-HPI schools 67
Figure 17. RQ 2D – Retention and Mobility of Teachers in HPI School, over the 4 HPI Years 68
Figure 18. RQ 3A – Example of Pre-Trend Matching for an HPI Elementary School 69
Figure 19. RQ3A – Elementary Schools, Ranked on Trend in In-School Retention Rate in Pre-HPI Years (2006 – 2015) 70
Figure 20. RQ 3A – CITS Estimate of HPI Impacts on In-School Teacher Retention for Elementary Schools 71
Figure 21. RQ 3B – Elementary Schools Ranked by Trend in In-District Retention Rate in Pre-HPI Years (2006 – 2015) 72
Figure 22 RQ 3B – High Schools Ranked by Trend in In-District Retention Rate in Pre-HPI Years (2006 – 2015) 73
Figure 23. RQ 3D. Principal Retention In-School by HPI Status 74

Tables

Table 1. HPI Payment Amounts, by School Role and Evaluation Ratings

			Yearly Total of Monthly Incentives			Annual Retention Incentive		
Role	Eval Type	School Level	Distinguished	Effective	Approaching	Distinguished	Effective	Approaching
Teacher	LEAP	All	$3,000.00	$2,250.00	$1,500.00	$1,000.00	$750.00	$500.00
SSP	GPS	All	$3,000.00	$2,250.00	$1,500.00	$1,000.00	$750.00	$500.00
Other	CPE	All	$1,500.00	$1,500.00	$1,500.00	$500.00	$500.00	$500.00
Principal	N/A	Elementary & ECE-8 (<600 students)		$17,000.00		Yr 2 Yr 3 Yr 4 Yr 5+	$2,000.00 $4,000.00 $6,000.00 $8,000.00	
Principal	N/A	Secondary & ECE-8 (>600 students)		$20,000.00		Yr 2 Yr 3 Yr 4 Yr 5+	$2,500.00 $5,000.00 $7,500.00 $10,000.00	
AP	LEAD	Elementary & ECE-8 (<600 students)		$10,000.00			N/A	
AP	LEAD	Secondary & ECE-8 (>600 students)		$12,000.00			N/A	

Note. Incentive amounts are determined using the most recent end-of-year performance reviews. Teachers and SSPs with performance ratings of "Not Meeting" are ineligible for HPI incentive payments (0.05% of teachers are rated in this bottom category). Employees evaluated under the CPE system are ineligible if they have a rating of "unsatisfactory". School leader bonuses are not dependent on effectiveness ratings, thus, all school leaders receive bonuses. Employees must be active with benefits (not on leave) to maintain eligibility. LEAP = Leading Effective Academic Practice; GPS = Growth and Performance System; CPE = Comprehensive Performance Evaluation; SSP = Specialized Service Provider.

Table 2. All HPI Schools with Information on SCI Scores, Current Student Demographics, and Pre-HPI Retention Rates

				Relevant SCI		18/19 Student demographics				
School name	Code	Level	Years as HPI	13/14	14/15	% Hispanic	% White	% Black	% FRL	Pre-HPI Retention
Castro Elementary School	287	Elementary	15/16 - now	59.8		85.4%	3.2%	4.2%	94.6%	85.7%
Charles M. Schenck (CMS) Community School	270	Elementary	15/16 - now	61.4		91.1%	1.3%	3.5%	92.6%	77.4%
Cheltenham Elementary School	218	Elementary	15/16 - now	53.5		69.5%	6.2%	18.1%	92.3%	38.7%
Cowell Elementary School	224	Elementary	15/16 - now	57.0		88.7%	4.3%	3.8%	94.8%	65.6%
DCIS At Ford	166	Elementary	15/16 - now	53.7		76.6%	3.0%	15.5%	94.5%	42.5%
Goldrick Elementary School	244	Elementary	15/16 - now	56.4		82.7%	4.6%	5.1%	95.5%	46.5%
Harrington Elementary School	248	Elementary	15/16 - now	52.3		71.6%	6.7%	16.5%	90.9%	39.4%
Knapp Elementary School	250	Elementary	15/16 - now	57.8		90.0%	5.3%	3.3%	94.8%	74.5%
Mcglone Elementary	299	Elementary	15/16 - now	57.5		74.0%	2.1%	14.9%	92.5%	77.1%
Munroe Elementary School	260	Elementary	15/16 - now	60.5		88.9%	3.9%	4.9%	96.7%	75.0%
Oakland Elementary School	150	Elementary	15/16 - now			65.2%	5.0%	24.5%	83.3%	50.0%
Schmitt Elementary School	271	Elementary	15/16 - now	53.7		76.6%	4.1%	11.1%	90.5%	53.6%
Swansea Elementary School	280	Elementary	15/16 - now	57.7		91.1%	3.1%	4.1%	93.8%	81.6%
Marie L. Greenwood Academy	258	K-8	15/16 - now	56.7	55.9	84.6%	3.1%	9.7%	90.6%	79.1%
Place Bridge Academy	190	K-8	15/16 - now	58.8	58.7	31.8%	12.6%	28.7%	95.0%	80.0%
Trevista Ece-8 At Horace Mann	189	K-8	15/16 - now	54.1	53.5	64.9%	15.2%	16.5%	88.7%	40.5%
Bear Valley International School	338	Middle	18/19 - now			72.2%	16.9%	2.8%	74.0%	
Henry World School	418	Middle	15/16 - 17/18		47.6					39.6%
Kepner Beacon Middle School	384	Middle	17/18 - now			85.3%	6.6%	3.9%	96.2%	
Kepner Middle School	408	Middle	15/16 - 17/18		60.5					51.3%
Lake International School	448	Middle	15/16 - now		57.5	71.6%	10.3%	12.0%	94.2%	68.0%
Bruce Randolph School	423_463	Mid/High	15/16 - now		59.2	88.0%	1.7%	8.0%	91.3%	75.0%
DCIS At Montbello	447_466	Mid/High	15/16 - now		56.3	80.5%	1.7%	12.9%	88.1%	73.2%
Martin Luther King Early College	419_469	Mid/High	15/16 - now		51.7	61.4%	5.7%	22.6%	78.5%	61.3%
Noel Community Arts School	434_467	Mid/High	15/16 - now		53.4	58.6%	7.1%	28.5%	87.8%	69.4%
West Generations Academy	397_510	Mid/High	15/16 - now		55.6	80.6%	5.8%	8.9%	93.5%	65.3%
West Leadership Academy	396_511	Mid/High	15/16 - now		59.7	87.4%	4.6%	5.2%	94.0%	90.0%
Abraham Lincoln High School	450_506	High	15/16 - now		59.6	86.9%	2.2%	5.2%	88.3%	73.9%
Collegiate Preparatory Academy	468	High	15/16 - now		48.9	61.3%	3.8%	23.6%	79.7%	55.2%
Manual High School	464	High	15/16 - now		59.8	54.1%	4.0%	36.2%	87.6%	66.7%
North High School	455	High	15/16 - now		50.5	69.1%	16.8%	8.1%	69.6%	67.2%
Northeast Early College	471	High	15/16 - now		52.0	76.2%	5.9%	15.0%	82.3%	59.4%

Note. In this table, each of the 7 co-located HPI schools are combined in a single row. Note that 2 HPI schools closed after HPI began, and 2 new HPI schools were added to replace them. A total of 32 schools were designated HPI at some point during the period of study. If the co-located schools are counted separately, a total of 39 schools have been designated HPI.

Table 3. Annual In-School Retention Rates in HPI Schools, with Depiction of Available Data

	School Level	Year First Observed	School Name (DPS ID)	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	# of Pre-Years	# of Post-Years	Mean Retention Pre-HPI*	Mean Retention Post-HPI*	Pre --> Post Change in Retention	Included in CITS?
				In-School Retention Rate															
1	Elem.	2002	Castro Elementary School (287)	76%	76%	86%	68%	74%	86%	77%	50%	61%	76%	10	4	78%	66%	-12%	Yes
2	Elem.	2002	Charles M. Schenck (Cms) Community Sch	81%	91%	73%	33%	64%	77%	70%	77%	69%	75%	10	4	62%	73%	11%	Yes
3	Elem.	2002	Cheltenham Elementary School (218)	88%	91%	80%	66%	58%	38%	61%	62%	80%	82%	10	4	60%	71%	11%	Yes
4	Elem.	2002	Cowell Elementary School (224)	81%	90%	76%	65%	82%	66%	61%	69%	75%	80%	10	4	72%	71%	-1%	Yes
5	Elem.	2002	Goldrick Elementary School (244)	89%	90%	94%	84%	68%	47%	38%	56%	58%	81%	10	4	73%	58%	-15%	Yes
6	Elem.	2002	Harrington Elementary School (248)	94%	83%	78%	81%	86%	39%	26%	67%	62%	67%	10	4	71%	55%	-16%	Yes
7	Elem.	2002	Knapp Elementary School (250)	78%	89%	77%	79%	79%	74%	82%	73%	78%	88%	10	4	77%	80%	3%	Yes
8	Elem.	2002	Mcglone Elementary (299)	70%	10%	55%	47%	81%	80%	87%	80%	75%	74%	10	4	66%	79%	13%	Yes
9	Elem.	2002	Munroe Elementary School (260)	74%	83%	64%	69%	69%	75%	79%	76%	70%	81%	10	4	69%	76%	7%	Yes
10	Elem.	2002	Schmitt Elementary School (271)	85%	85%	88%	54%	78%	54%	37%	65%	48%	58%	10	4	68%	52%	-16%	Yes
11	Elem.	2002	Swansea Elementary School (280)	91%	71%	68%	73%	63%	82%	67%	77%	79%	77%	10	4	71%	75%	4%	Yes
12	Elem.	2012	Dcis At Ford (166)			78%	43%	72%	43%	65%	57%	58%	72%	4	4	59%	63%	4%	Yes
13	Elem.	2015	Oakland Elementary School (150)			.	.	.	50%	56%	48%	64%	76%	1	4	50%	61%	11%	No
1	Middle	2002	Henry World School (418)	94%	78%	69%	72%	71%	40%	42%	48%			10	2	**63%**	**23%**	**-40%**	**No**
2	Middle	2002	Kepner Middle School (408)	90%	81%	72%	66%	40%	51%	64%	35%			10	2	**57%**	**25%**	**-33%**	**No**
3	Middle	2002	Martin Luther King Early College Ms (419)	75%	73%	81%	43%	67%	63%	61%	60%	59%	74%	10	4	63%	63%	0%	Yes
4	Middle	2003	Bruce Randolph School Ms (423)	84%	68%	89%	87%	63%	74%	69%	74%	69%	77%	10	4	78%	72%	-6%	Yes
5	Middle	2011	Lake International School (448)		67%	67%	44%	40%	68%	67%	71%	41%	54%	5	4	55%	58%	3%	Yes
6	Middle	2012	Dcis At Montbello Ms (447)			100%	45%	42%	71%	54%	54%	71%	59%	4	4	64%	60%	-5%	Yes
7	Middle	2012	Noel Community Arts School Ms (434)			90%	42%	67%	70%	63%	60%	55%	50%	4	4	67%	57%	-10%	Yes
8	Middle	2013	West Generations Academy Ms (397)			.	58%	80%	56%	63%	79%	82%	79%	3	4	64%	76%	11%	Yes
9	Middle	2013	West Leadership Academy Ms (396)			.	89%	78%	95%	64%	75%	78%	85%	3	4	87%	76%	-12%	Yes
10	Middle	2017	Bear Valley International School (338)			.	.	.	.	.	60%	86%	68%		3	.	72%	.	No
11	Middle	2017	Kepner Beacon Middle School (384)			.	.	.	.	.	90%	70%	81%		3	.	80%	.	No
1	High	2002	Abraham Lincoln High School (450)	93%	82%	77%	72%	74%	74%	75%	65%	78%	81%	10	4	75%	75%	0%	Yes
2	High	2002	Manual High School (464)	69%	58%	48%	68%	54%	67%	69%	63%	76%	81%	8	4	59%	72%	13%	Yes
3	High	2002	North High School (455)	62%	58%	63%	63%	71%	71%	70%	85%	90%	75%	10	4	67%	80%	13%	Yes
4	High	2007	Bruce Randolph School Hs (463)	84%	88%	71%	76%	66%	76%	72%	81%	74%	68%	9	4	72%	74%	2%	Yes
5	High	2007	Martin Luther King Early College Hs (469)	61%	65%	74%	56%	65%	60%	80%	74%	71%	80%	9	4	64%	77%	13%	Yes
6	High	2012	Collegiate Preparatory Academy (468)			58%	27%	33%	57%	54%	63%	75%	54%	4	4	44%	61%	17%	Yes
7	High	2012	Dcis At Montbello Hs (466)			56%	63%	48%	78%	80%	74%	61%	83%	4	4	61%	74%	13%	Yes
8	High	2012	Noel Community Arts School Hs (467)			67%	38%	64%	68%	61%	63%	46%	74%	4	4	59%	61%	2%	Yes
9	High	2012	Northeast Early College (471)			67%	36%	40%	59%	71%	61%	61%	67%	4	4	50%	65%	14%	Yes
10	High	2013	West Generations Academy Hs (510)			.	36%	79%	71%	64%	54%	85%	100%	3	4	62%	76%	14%	Yes
11	High	2013	West Leadership Academy Hs (511)			.	100%	100%	84%	83%	92%	85%	87%	3	4	95%	87%	-8%	Yes
1	K_08	2002	Marie L. Greenwood Academy (258)	91%	79%	67%	82%	73%	79%	77%	66%	78%	48%	10	4	76%	67%	-8%	Yes
2	K_08	2009	Place Bridge Academy (190)	91%	88%	82%	81%	78%	80%	83%	69%	77%	69%	7	4	80%	75%	-6%	Yes
3	K_08	2009	Trevista Ece-8 At Horace Mann (189)	82%	62%	11%	50%	54%	40%	64%	56%	54%	54%	7	4	39%	57%	18%	Yes

Note. The 6 co-located HPI schools are described here separately, each in their own row. 2 HPI schools closed after HPI began, and 2 schools were added to replace them. A total of 38 schools were HPI at some point, but 5 schools do not have sufficient data to be included in CITS analyses. Asterisk indicates this is the mean for the 4 years either pre- or post-HPI

Table 4. RQ 1C – School-Level Retention Rates, Separately by School Characteristics

	Tercile range	Avg. teachers per year	**Retained in school**	Retained in DPS
By school level				
Elementary		2,225	**76%**	84.6%
K-8		485	**77%**	84.3%
Middle		615	**70%**	79.3%
High		869	**77%**	82.9%
By school enrollment				
Bottom third	(27 - 414)	916	**72%**	81.5%
Middle third	(415 - 595)	1,315	**76%**	84.2%
Top third	(596 - 2659)	1,962	**77%**	83.8%
By % FRL students				
Bottom third	(3% - 64%)	1,418	**81%**	86.0%
Middle third	(64% - 88%)	1,394	**73%**	81.9%
Top third	(89% - 99%)	1,380	**73%**	82.4%
By % Hispanic students				
Bottom third	(4% - 41%)	1,466	**80%**	85.3%
Middle third	(41% - 76%)	1,329	**73%**	81.8%
Top third	(77% - 97%)	1,398	**74%**	83.0%
By % White students				
Bottom third	(0% - 6%)	1,371	**71%**	81.1%
Middle third	(6% - 27%)	1,416	**75%**	83.4%
Top third	(27% - 89%)	1,406	**81%**	85.8%
By % Black students				
Bottom third	(0% - 4%)	1,372	**77%**	85.0%
Middle third	(4% - 17%)	1,299	**76%**	83.5%
Top third	(17% - 88%)	1,520	**74%**	81.9%

Note: this table shows the rate of teacher retention in schools of different types across the entire panel of study (2006 - 2019). School-year observations are divided into terciles (thirds) by a characteristic: for example, the school-year observations with the highest, middle, and lowest enrollments. The characteristics of the terciles are shown as tercile range: for example, that the third of schools with the smallest enrollment have between 27 and 414 students. The average in-school retention of all teachers across the panel who have worked at schools in a certain group are shown in bold. From this we can see broad trends in teacher retention by categories of schools.

Table 5. RQ 1D – Teacher Characteristics that Predict In-School and In-District Retention

	Comparing Across Schools				Comparing Within Schools			
	Retained in school		Retained in district		Retained in school		Retained in district	
	Coeff	P-R2	Coeff	P-R2	Coeff	P-R2	Coeff	P-R2
Female	0.0147	0.0000	0.0458	0.0001	0.0026	0.0364	-0.0035	0.0191
	(0.0269)		(0.0294)		(0.0268)		(0.0301)	
Salary ($1,000s)	0.0181 ***	0.0103	0.0189 ***	0.0105	0.0149 ***	0.0434	0.0178 ***	0.0282
	(0.0008)		(0.0008)		(0.0008)		(0.0009)	
Masters+	-0.0018	0.0000	0.0212	0.0000	-0.0736	0.0376	-0.0226	0.0209
	(0.0226)		(0.0247)		(0.0221)		(0.0247)	
Teacher race (vs. White)		0.0004		0.0013		0.0369		0.0206
Hispanic/Latinx	0.0472		0.2346 ***		0.1486 ***		0.2741 ***	
	(0.0315)		(0.0354)		(0.0314)		(0.0360)	
Black	-0.1520 **		-0.0957		0.0779		0.0678	
	(0.0552)		(0.0597)		(0.0566)		(0.0610)	
Asian	0.0470		-0.0169		0.0409		-0.0124	
	(0.1231)		(0.1288)		(0.1209)		(0.1248)	
Native American	-0.4031 *		-0.1848		-0.2911		-0.0997	
	(0.1804)		(0.1792)		(0.1794)		(0.1833)	
Multiracial	0.0261		0.1323 *		0.0884		0.1834 **	
	(0.0588)		(0.0626)		(0.0572)		(0.0619)	
Other/unknown ethnicity	-0.1122		0.1081		-0.0389		0.1997	
	(0.1940)		(0.2199)		(0.1784)		(0.2090)	
Teacher age		0.0076		0.0141		0.0417		0.0305
Age	0.1399 ***		0.1925 ***		0.1186 ***		0.1738 ***	
	(0.0069)		(0.0076)		(0.0069)		(0.0076)	
Age^2	-0.0016 ***		-0.0022 ***		-0.0014 ***		-0.0020 ***	
	(0.0001)		(0.0001)		(0.0001)		(0.0001)	
Teacher DPS experience		0.0134		0.0262		0.0461		0.0411
DPS Exp.	0.0966 ***		0.1470 ***		0.0807 ***		0.1330 ***	
	(0.0046)		(0.0055)		(0.0045)		(0.0054)	
DPs Exp.^2	-0.0028 ***		-0.0044 ***		-0.0025 ***		-0.0041 ***	
	(0.0002)		(0.0002)		(0.0002)		(0.0002)	

*Note. This table includes teachers in schools in our analytic sample from 2006 - 2019 (the full available panel). The table shows logistic regression coefficients from separate regressions. Coefficients listed under "Teacher race" (racial/ethnic categories), "Teacher age", and "Teacher DPS experience" were produced from three separate logistic regression models. The column "Comparing Across Schools" doesn't include a teacher fixed effect, while the Column "Comparing Within Schools" does. "P-R2" is the model pseudo R-squared. *p < .05, **p < .01, *** p < .001*

Table 6. RQ 2A – Number of Employees Receiving HPI Incentive Payments and Average Amounts, by School Year

	N Employees Receiving Bonus				Mean Payment Amount in Dollars (SD)			
	2016	2017	2018	2019	2016	2017	2017	2019
HPI Retention	-	17	838	924	- -	233 (237)	737 (177)	760 (170)
HPI Monthly	1193	1330	1313	1291	481 (144)	2,455 (972)	1,835 (595)	1,504 (512)
School Leader Continuity	-	19	23	-	- -	2,263 (256)	3,804 (1,165)	- -
School Leader HPI	94	87	92	-	13,926 (3,673)	13,931 (3,757)	14,173 (3,828)	- -

Note. School Leader HPI payments are analogous to teacher HPI Monthly payments but are distributed in a lump sum at the end of the school year. School Leader Continuity Payments are analogous to HPI Retention payments but increase in amount based upon the number of years the school leader has worked at the school.

Table 7. RQ 2B – School Characteristics of HPI and Non-HPI Schools, Averaged across the 3 Years Pre-HPI and 3 Years Post-HPI

	Pre-Program		Post-Program	
	HPI	non-HPI	HPI	non-HPI
Student Characteristics				
% FRL*	93.0%	64.6%	90.7%	59.9%
% ELL*	51.2%	29.3%	53.8%	26.5%
% SPED*	12.4%	11.0%	12.9%	10.7%
% Hispanic	79.5%	50.2%	77.5%	46.8%
% White	3.9%	28.9%	5.0%	31.8%
% Black	11.5%	13.5%	11.9%	13.0%
% Asian	2.7%	3.0%	2.8%	3.1%
Teacher Characteristics				
Avg. Years DPS Exp.	5.4	7.5	4.5	6.5
% in First 3 Years in DPS	41.1%	26.9%	44.0%	33.7%
% Retained in School	65.5%	74.2%	68.0%	75.8%
Avg. Salary	$ 46,973	$ 48,304	$ 51,046	$ 52,356
% with Masters+	50.7%	55.7%	54.6%	58.9%
Avg. LEAP Points	68.9	72.1	73.6	76.4
% Rated Effective+	70.9%	82.4%	83.1%	88.5%
School Characteristics				
Avg. Enrollment	640	536	558	519
% Elementary	42.6%	65.8%	41.9%	64.6%
% K-8	10.4%	11.0%	9.7%	9.8%
% Middle	10.4%	9.9%	12.9%	8.8%
% High	17.4%	7.1%	16.1%	8.8%
School Count	30	104	32	106

Note. Asterisk indicates that the given variable was used as part of the SCI index score formula. SCI scores strongly predict HPI school status. The table compares HPI and non-HPI schools in the three years immediately before (12/13 - 14/15) and after (15/16 - 18/19) the program.*

Table 8. RQ 2D – HPI vs Non-HPI Schools: Annual Movements of Teachers, Separately in the 4 Years of HPI Implementation

	After 2015/2016		After 2016/2017		After 2017/2018		After 2018/2019		Overall	
	HPI	Non-HPI	HPI	Non-HPI	HPI	Non-HPI	HPI	Non-HPI	HPI	Non-HPI
Moved to HPI school	10.0%	6.1%	9.5%	6.6%	10.8%	4.7%	5.8%	5.0%	9.1%	5.6%
Moved to non-HPI school	23.5%	18.0%	24.3%	20.3%	20.2%	21.9%	19.1%	20.3%	21.9%	20.1%
Non-teacher position in DPS	10.2%	10.6%	11.0%	12.2%	10.0%	11.6%	11.8%	15.3%	10.7%	12.4%
Left DPS	56.4%	65.5%	55.3%	60.9%	59.1%	62.0%	63.3%	59.5%	58.3%	62.0%
Leaving teacher count	422	840	400	898	381	860	330	838	1533	3436

Note. With the exception of the bottom row, values are column percentages of the total teacher leaving count for that year and school type (which is the value shown in the bottom row).

Table 9. RQ 3A – Matching on Pre-HPI Trend Results for CITS Analysis of In-School Teacher Retention: Number of Matched and Unmatched Schools

	HPI Schools		Non-HPI Schools		
	At Least 1 Match Found	No Matches Found	Used as a Match	Not Used as a Match	All Schools
Elementary	12	0	12	49	73
Middle	7	0	7	3	17
K-08	3	0	3	7	13
High	8	3	8	0	19
All Schools	30	3	30	59	122

Table 10. RQ 3A –CITS Results using Pre-Trend Matching: HPI Shock and Trend Effects on In-School Teacher Retention

Matching Approach	School Level	Shock Effect	Trend Effect
1:1 NN Matching wo/ Replacement	Elementary	-6.2 *(ns)*	+2.0 *(ns)*
	Middle	-2.0 *(ns)*	-0.3 *(ns)*
	K-08	+10.8 *(ns)*	-5.2 *(ns)*
	High	+12.2 *(ns)*	+1.9 *(ns)*
	All	+2.7 *(ns)*	+1.2 *(ns)*
1:M Caliper Matching w/ Replacement	Elementary	-7.5 *(ns)*	+3.1 *
	Middle	+8.1 *(ns)*	+1.3 *(ns)*
	K-08	+12.2 *(ns)*	+0.1 *(ns)*
	High	+16.0 *(ns)*	+0.8 *(ns)*
	All	+5.0 *(ns)*	+1.9 *(ns)*

*Note. (ns) indicates that an estimated effect is not significant at the 0.05 level, * for p < .05, ** for p < .01, and *** for p < .001).*

Table 11. RQ 3B – Matching on Pre-HPI Trend Results for CITS Analysis of In-District Teacher Retention: Number of Matched and Unmatched Schools

	HPI Schools		Non-HPI Schools		
	At Least 1 Match Found	No Matches Found	Used as a Match	Not Used as a Match	All Schools
Elementary	12	0	12	49	73
Middle	7	0	7	3	17
K-08	3	0	3	7	13
High	7	4	7	1	19
All Schools	29	4	29	60	122

Table 12. RQ 3B –CITS Results using Pre-Trend Matching: HPI Shock and Trend Effects on In-District Teacher Retention

Matching Approach	School Level	Shock Effect	Trend Effect
1:1 NN Matching wo/ Replacement	Elementary	-7.7 *(ns)*	+2.3 *(ns)*
	Middle	+7.2 *(ns)*	+0.7 *(ns)*
	K-08	+0.8 *(ns)*	-0.4 *(ns)*
	High	+15.8 *	-0.9 *(ns)*
	All	+3.0 *(ns)*	+1.1 *(ns)*
1:M Caliper Matching w/ Replacement	Elementary	-4.8 *(ns)*	+1.7 *(ns)*
	Middle	+10.0 *(ns)*	-1.8 *(ns)*
	K-08	+2.6 *(ns)*	-2.7 *(ns)*
	High	+16.4 *(ns)*	-1.0 *(ns)*
	All	+5.1 *(ns)*	-0.0 *(ns)*

*Note. (ns) indicates that an estimated effect is not significant at the 0.05 level, * for p < .05, ** for p < .01, and *** for p < .001).*

Table 13. RQ 3C – Matching on Pre-HPI Trend Results for CITS Analysis of LEAP Scores of Recruited (New-to-School) Teachers: Number of Matched and Unmatched Schools

	HPI Schools		Non-HPI Schools		
	Match Found	No Matches Found	Used as a Match	Not Used as a Match	All Schools
Elementary	12	0	12	46	70
Middle	6	1	6	4	17
K-08	3	0	3	7	13
High	5	6	5	3	19
All Schools	26	7	26	60	119

Table 14. RQ 3C –CITS Results using Pre-Trend Matching: HPI Shock and Trend Effects on LEAP Scores of Recruited (New-to-School) Teachers

Matching Approach	School Level	Shock Effect	Trend Effect
1:1 NN Matching wo/ Replacement	Elementary	+0.3 *(ns)*	-0.2 *(ns)*
	Middle	-0.4 *(ns)*	-1.6 *(ns)*
	K-08	-5.3 *(ns)*	+2.9 *(ns)*
	High	+5.5 *(ns)*	+3.5 *(ns)*
	All	+0.7 *(ns)*	+0.5 *(ns)*
1:M Caliper Matching w/ Replacement	Elementary	+2.9 *(ns)*	-1.4 *(ns)*
	Middle	+4.2 *(ns)*	-2.3 *(ns)*
	K-08	-6.1 *(ns)*	+2.3 *(ns)*
	High	+1.7 *(ns)*	+2.0 *(ns)*
	All	+2.4 *(ns)*	-0.5 *(ns)*

Table 15. RQ 3E – Matching on Pre-HPI Trend Results for CITS Analysis of Student Achievement: Number of Matched and Unmatched Schools

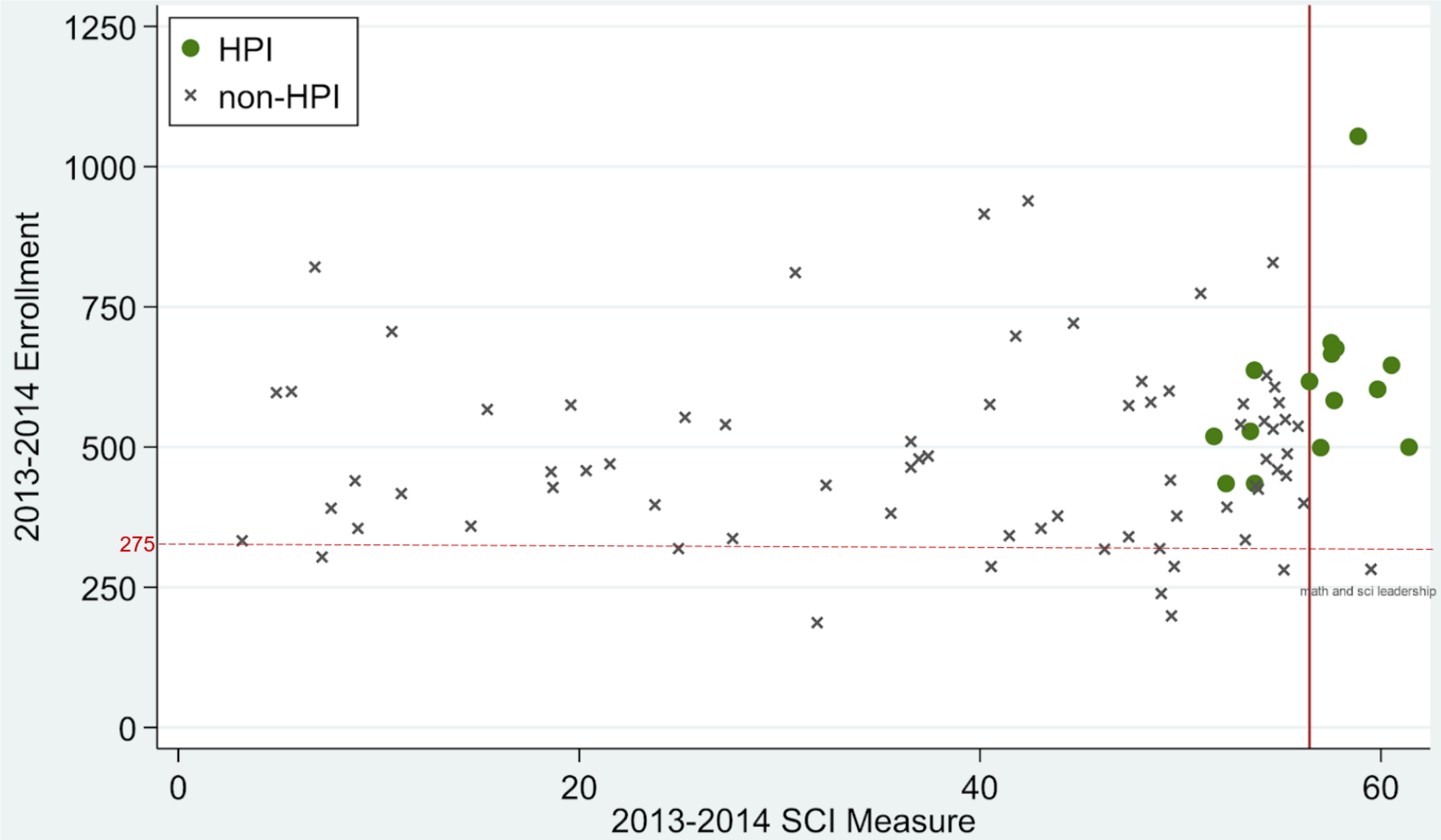
	HPI Schools		Non-HPI Schools		
	At Least 1 Match Found	No Matches Found	Used as a Match	Not Used as a Match	All Schools
Elementary	12	0	12	48	72
Middle	8	1	8	2	19
K-08	3	0	3	7	13
High	5	0	5	2	12
All Schools	28	1	28	59	116

Table 16. RQ 3E –CITS Results using Pre-Trend Matching: HPI Shock and Trend Effects on Student Achievement

		ELA Assessments		Math Assessments	
Matching Approach	School Level	Shock Effect	Trend Effect	Shock Effect	Trend Effect
1:1 NN Matching wo/ Replacement	Elementary	-0.018 *(ns)*	+0.023 *(ns)*	+0.062 *(ns)*	-0.009 *(ns)*
	Middle	-0.132 *(ns)*	+0.109 *	+0.010 *(ns)*	+0.039 *(ns)*
	K-08	+0.024 *(ns)*	+0.064 *(ns)*	+0.152 *	+0.013 *(ns)*
	High	+0.035 *(ns)*	-0.016 *(ns)*	-0.234 *(ns)*	+0.016 *(ns)*
	All	+0.014 *(ns)*	+0.036 *(ns)*	+0.067 *(ns)*	-0.001 *(ns)*
1:M Caliper Matching w/ Replacement	Elementary	-0.033 *(ns)*	+0.016 *(ns)*	+0.056 *(ns)*	+0.007 *(ns)*
	Middle	-0.098 *(ns)*	+0.098 *(ns)*	+0.040 *(ns)*	-0.035 *(ns)*
	K-08	+0.061 *(ns)*	+0.067 *(ns)*	+0.169 *(ns)*	+0.036 *(ns)*
	High	-0.097 *(ns)*	-0.013 *(ns)*	-0.178 *(ns)*	+0.013 *(ns)*
	All	-0.005 *(ns)*	+0.031 *(ns)*	+0.048 *(ns)*	+0.005 *(ns)*

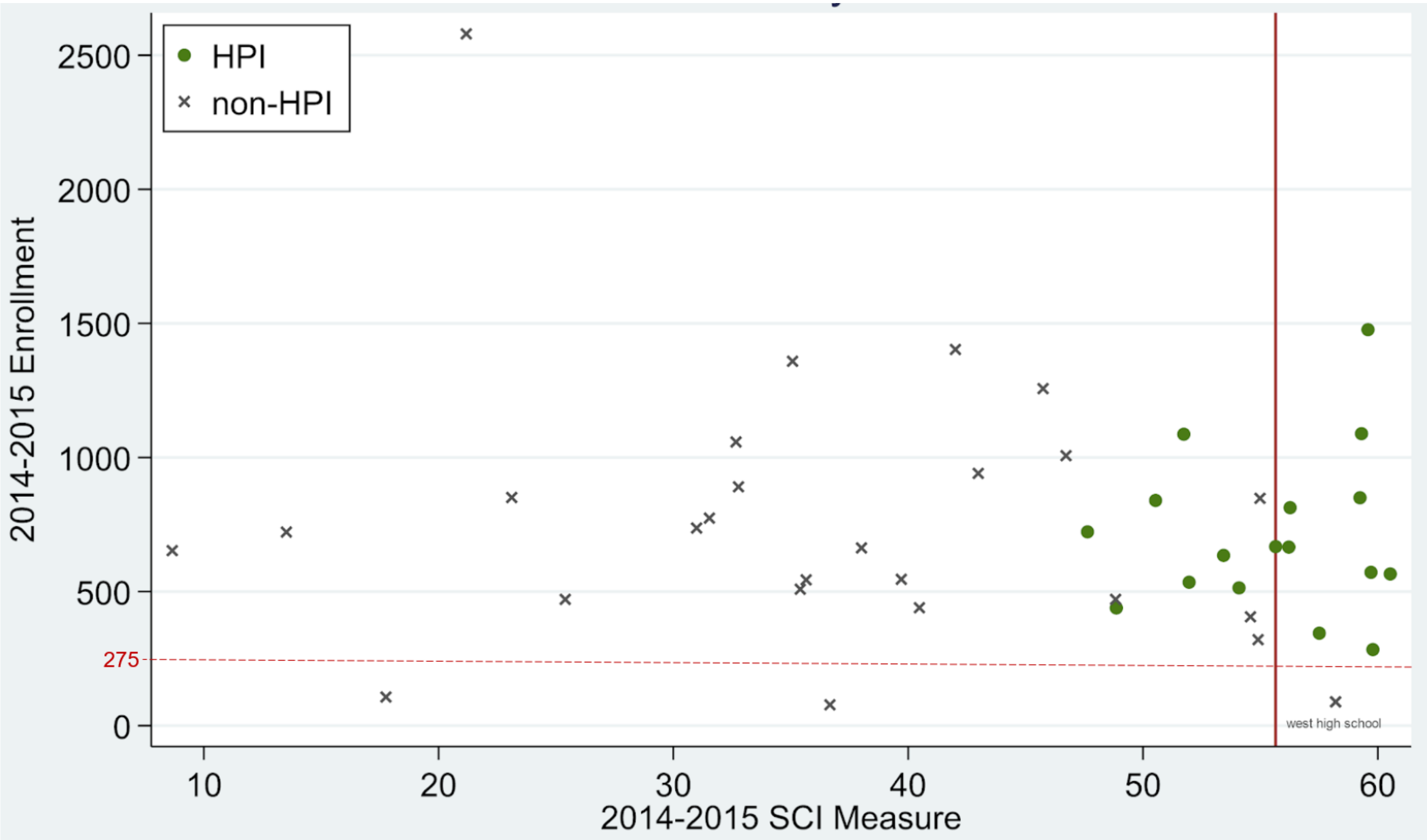
Figures

Figure 1. Elementary Schools Selected for HPI as a Function of Enrollment and SCI Measures



Note. The red vertical line marks the SCI value of the 10th highest HPI school. Mathematics and Science Leadership Academy had SCI scores that qualified for the program; it was likely excluded due to low enrollment.

Figure 2. Secondary Schools Selected for HPI as a Function of Enrollment and SCI Measures



Note. The vertical red line marks the SCI value of the 10th highest HPI school. West High School qualified for the program, but was likely excluded either due to low enrollment or because it was already in phase-out and designated for closure.

Figure 3. Average In-School Retention Rate and Average % of FRPL-Eligible Students in 3 Years Prior to HPI (does not include charter schools)

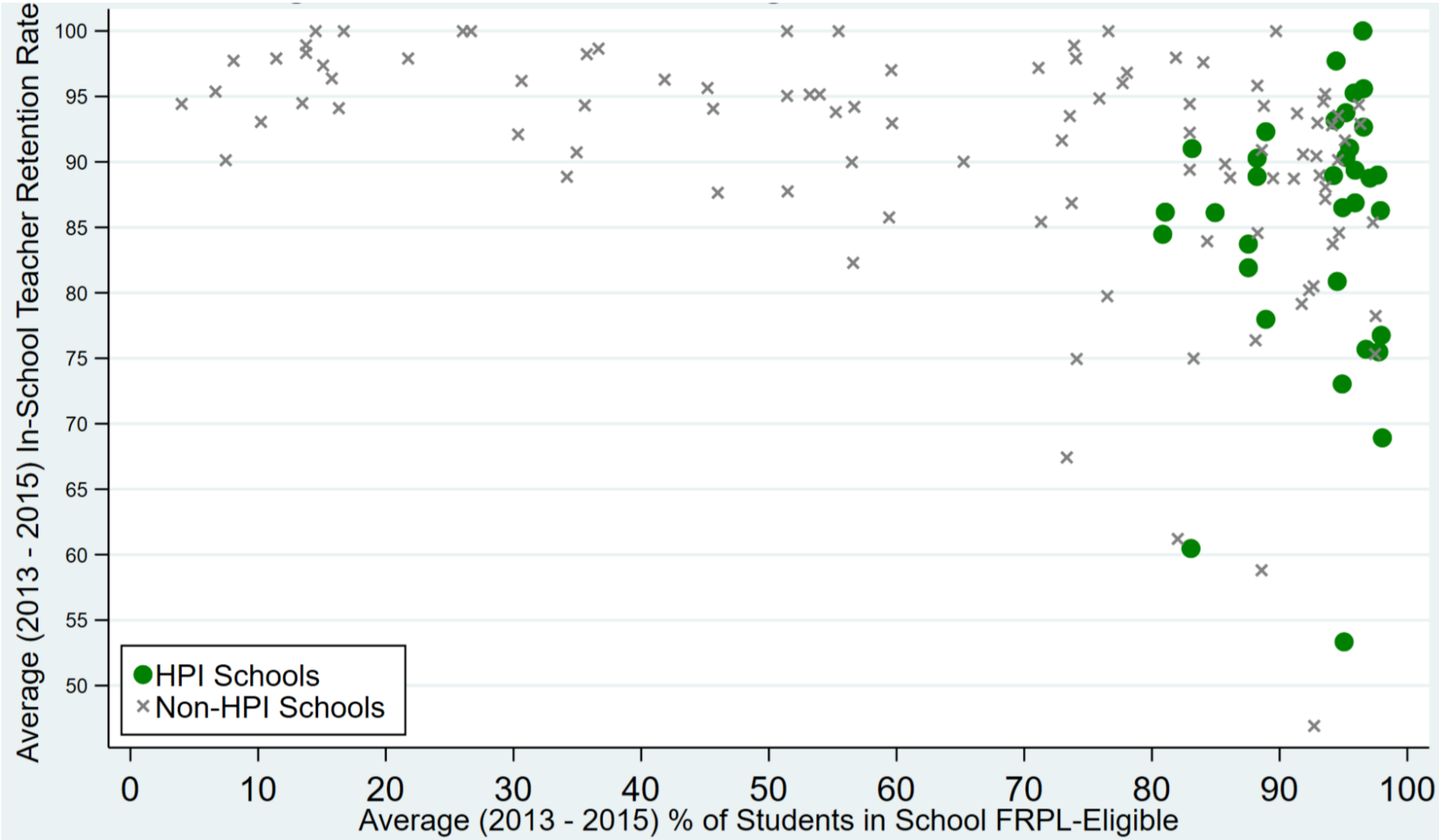


Figure 4. Number of DPS Schools including Main Analytic, Alternative/Charter Schools 2002 - 2020

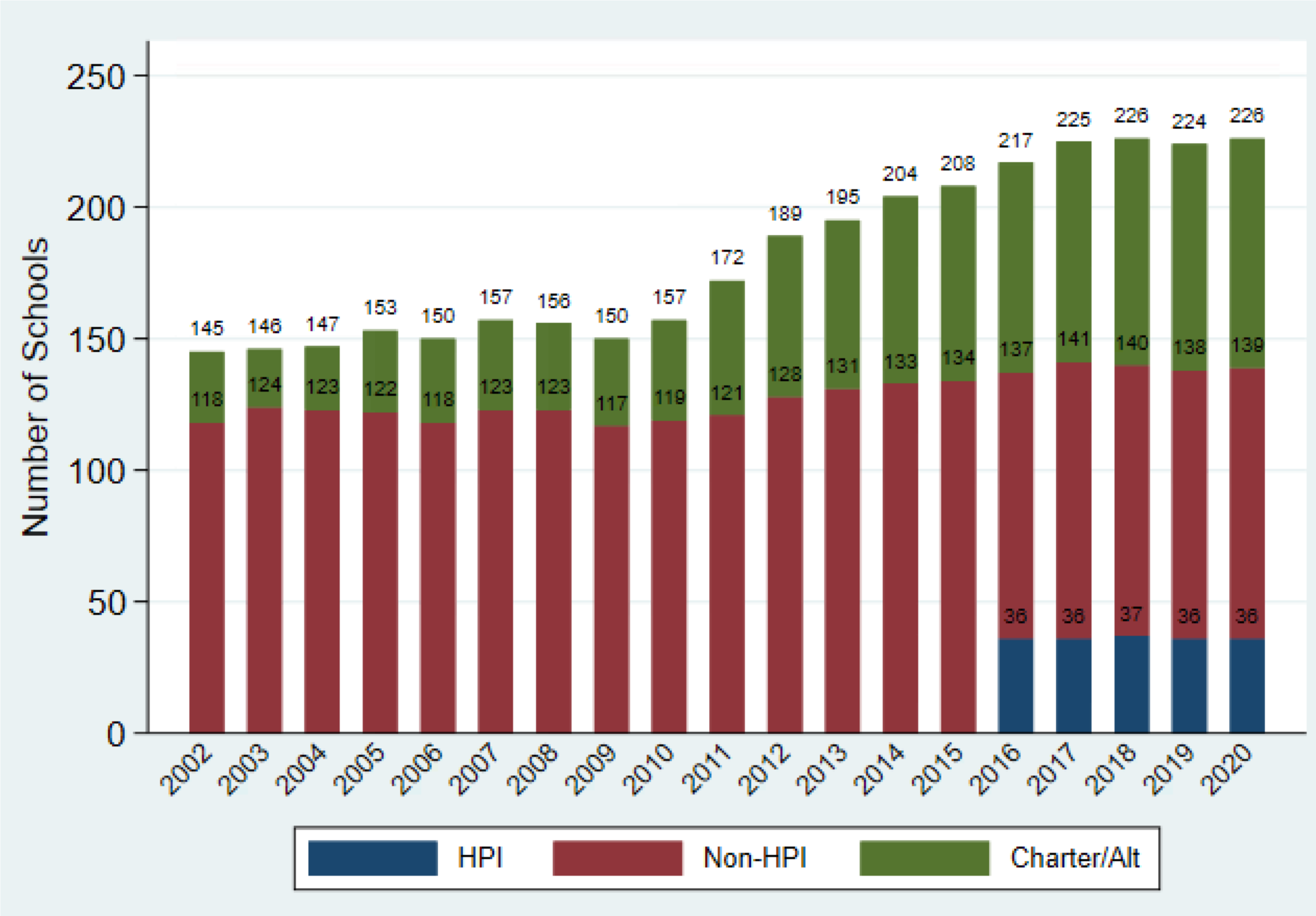


Figure 5. RQ 1A –Teacher Retention Over Time, 2002–2019, Separately for In-School and In-District Retention Rates

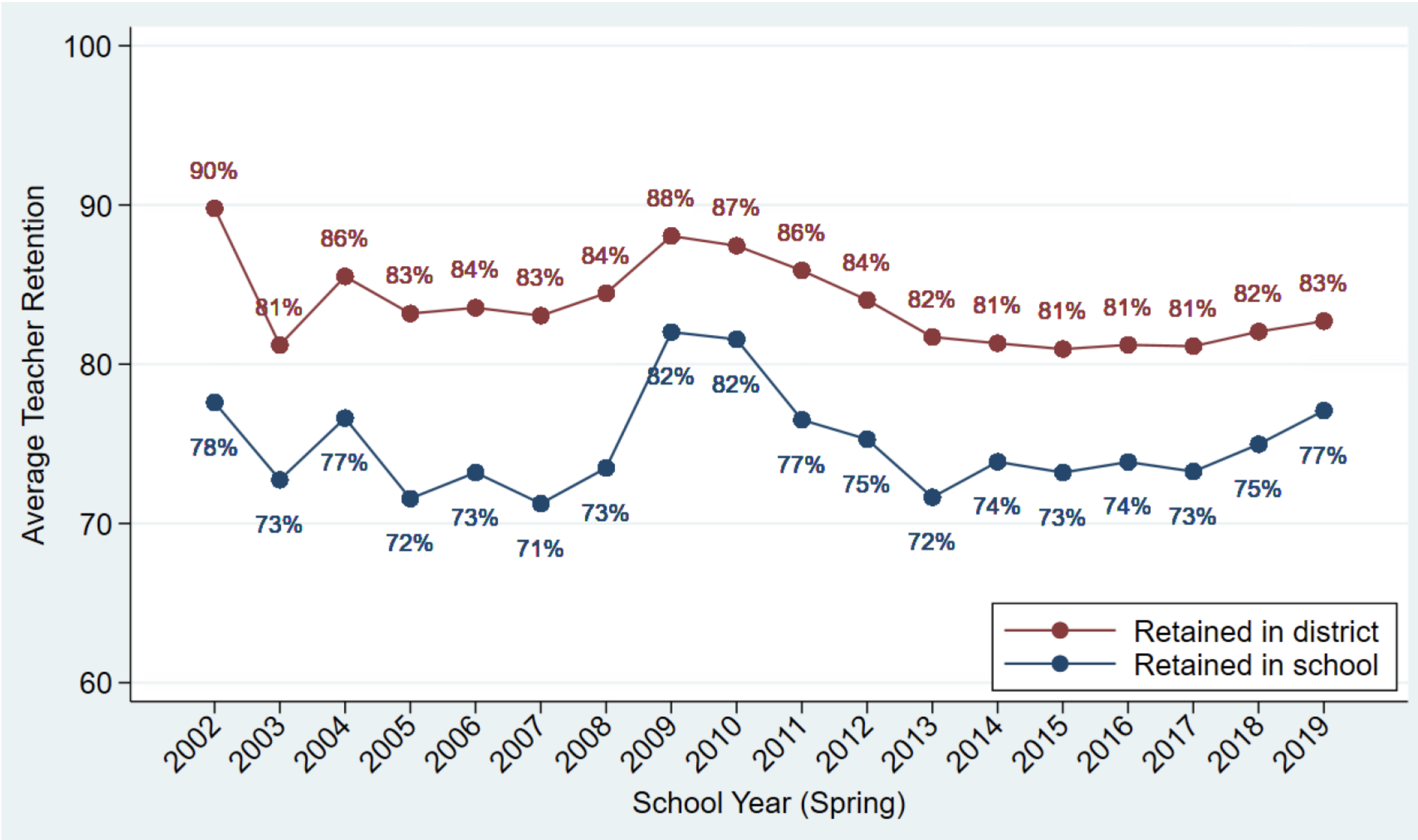


Figure 6. RQ 1B – Variability in In-School Retention Rates across DPS schools 2006 to 2019

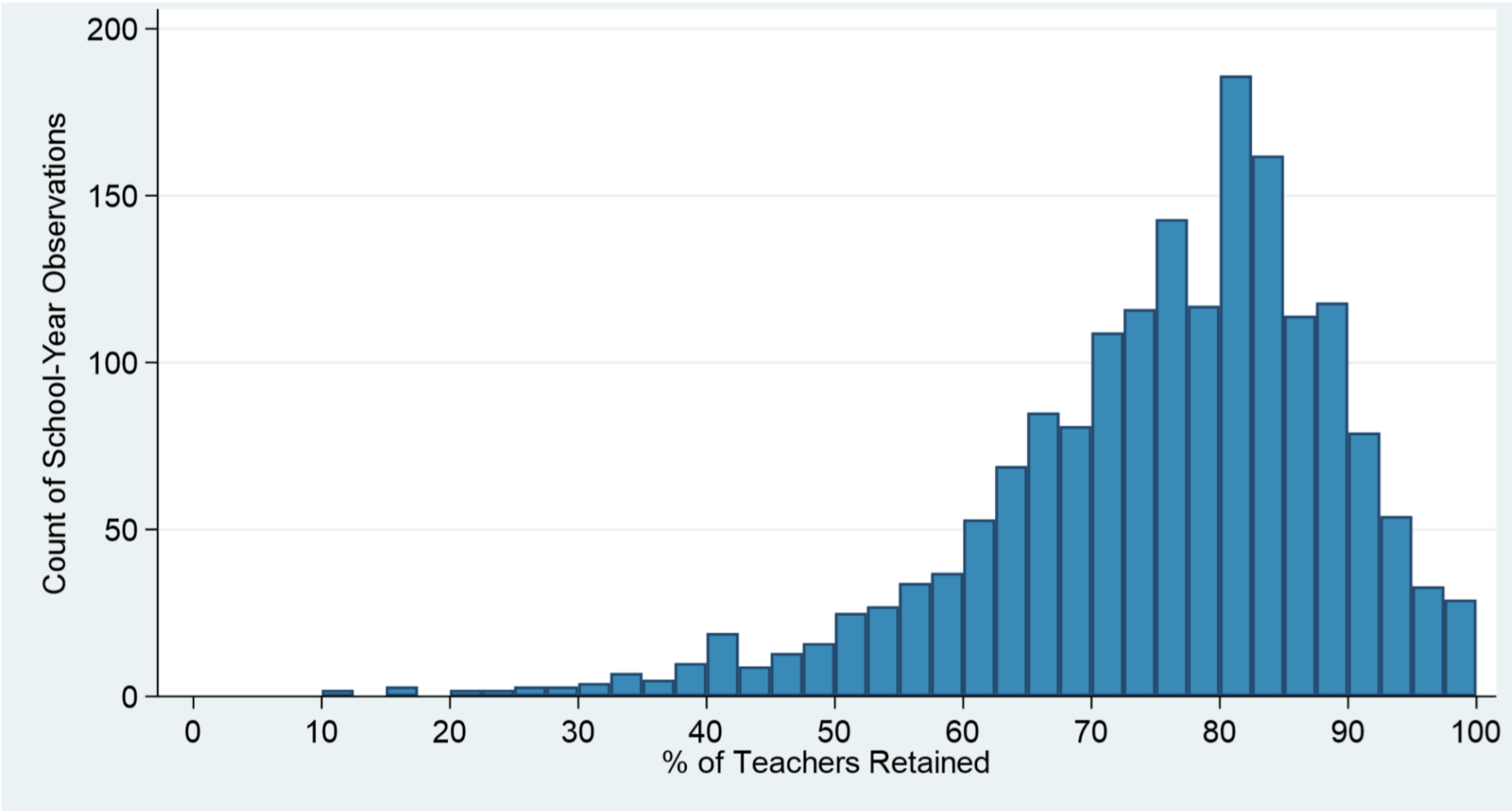
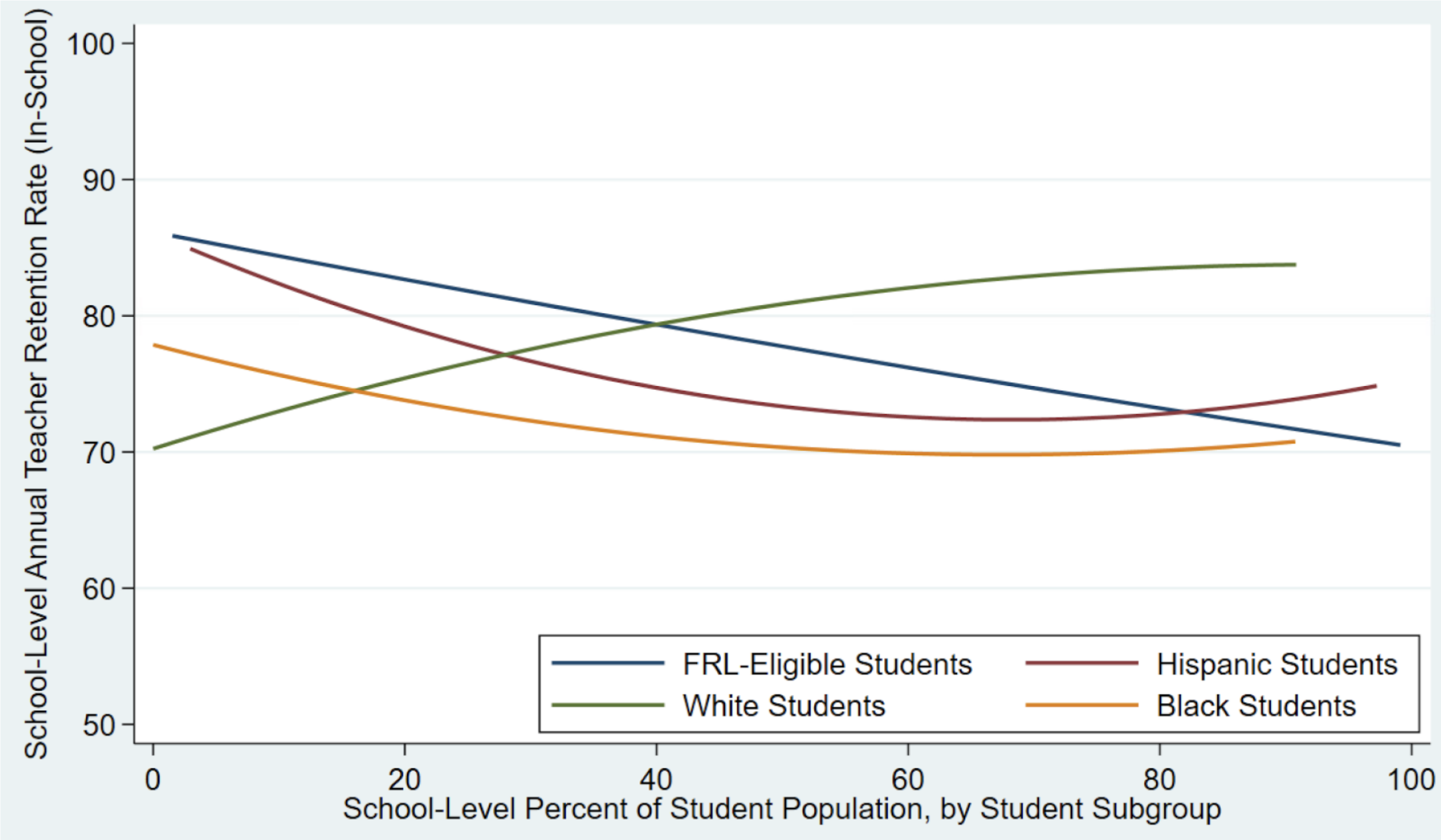
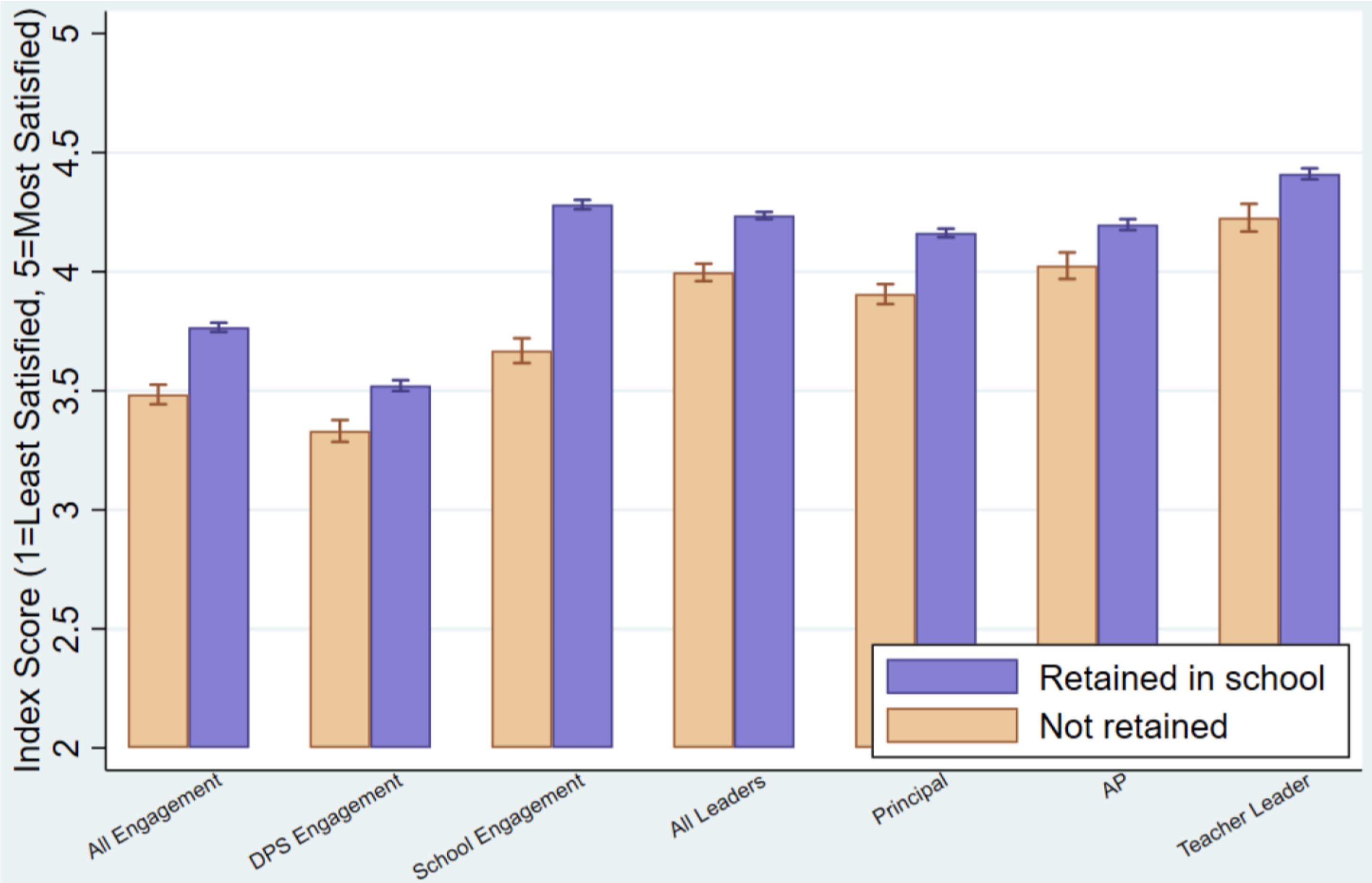


Figure 7. RQ 1C—Relationship Between DPS Schools' Annual Retention Rate and the School's Student Demographics



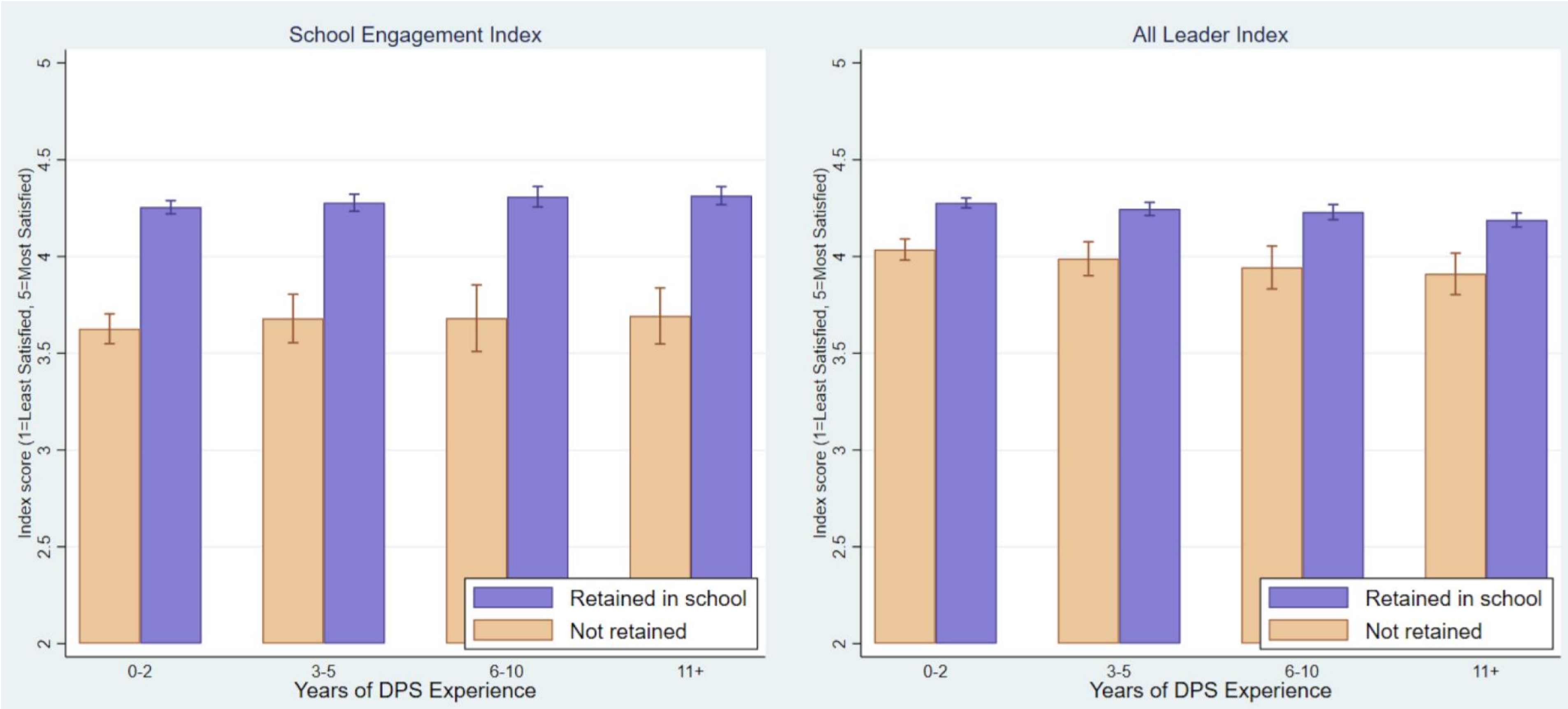
Note: This figure captures the relationship between school-level teacher retention rates and schools' levels of school-level student demographics for all schools in all years in the panel (2006 - 2019). Although the values for individual schools that were used to make these lines are not shown, we can see how student demographics and teacher retention are often associated at the school-level.

Figure 8. RQ 1C—Relationship between Teachers' Perceptions of Their School (Leadership, Engagements) and Teacher In-School Retention



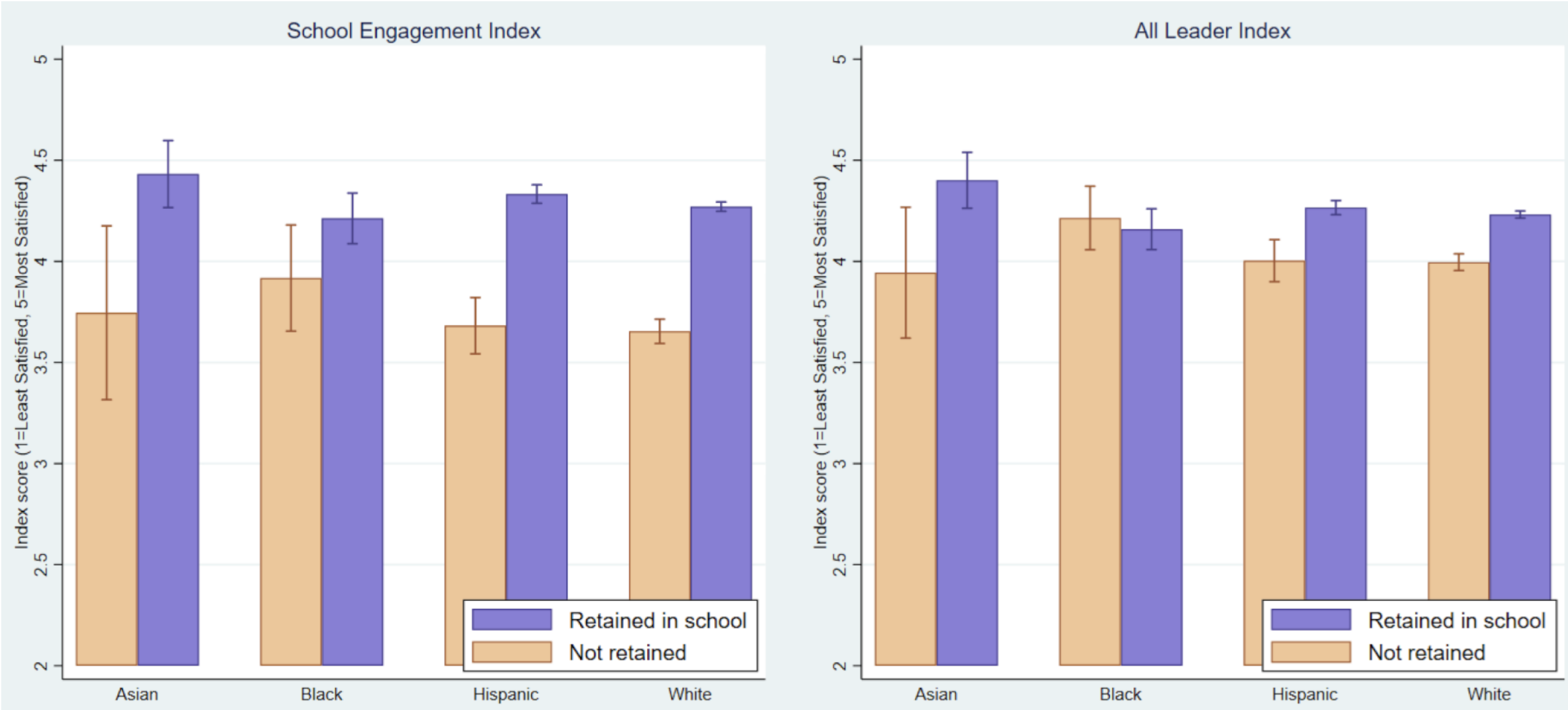
Note: In each survey year, we produce 7 different measures of teacher perceptions and satisfaction: A measure of each teacher's engagement overall, with DPS, and with their school their school, as well as a measure of how they perceive school leaders (all leaders, the principal, the assistant principal (AP), and teacher leaders). As an example, we produce a teacher's measure of their perception of their school leaders by taking the mean of their CollaboRATE responses to items related to school leaders, which are all coded on a scale from 1 ("Strongly disagree") to 5 ("Strongly agree"). These questions address teacher engagement, empowerment, perception of school/district values and commitments, and attitudes about team/school leader function. The school engagement index and DPS engagement index are composites of five questions identified in the DPS 2020 Thrive Guide. The questions are: (1) I enjoy my work at school/DPS. (2) I feel valued as an employee of my school/DPS. (3) I would recommend my school/DPS to others as a good place to work. (4) My job has a positive impact on my school/DPS. (5) I am proud to tell people I work for my school/DPS. Of these, only questions 1 and 2 were included in the 2018 survey.

Figure 9. RQ 1C — Does the Relationship between CollaboRATE and Retention Differ by Years of DPS Experience?



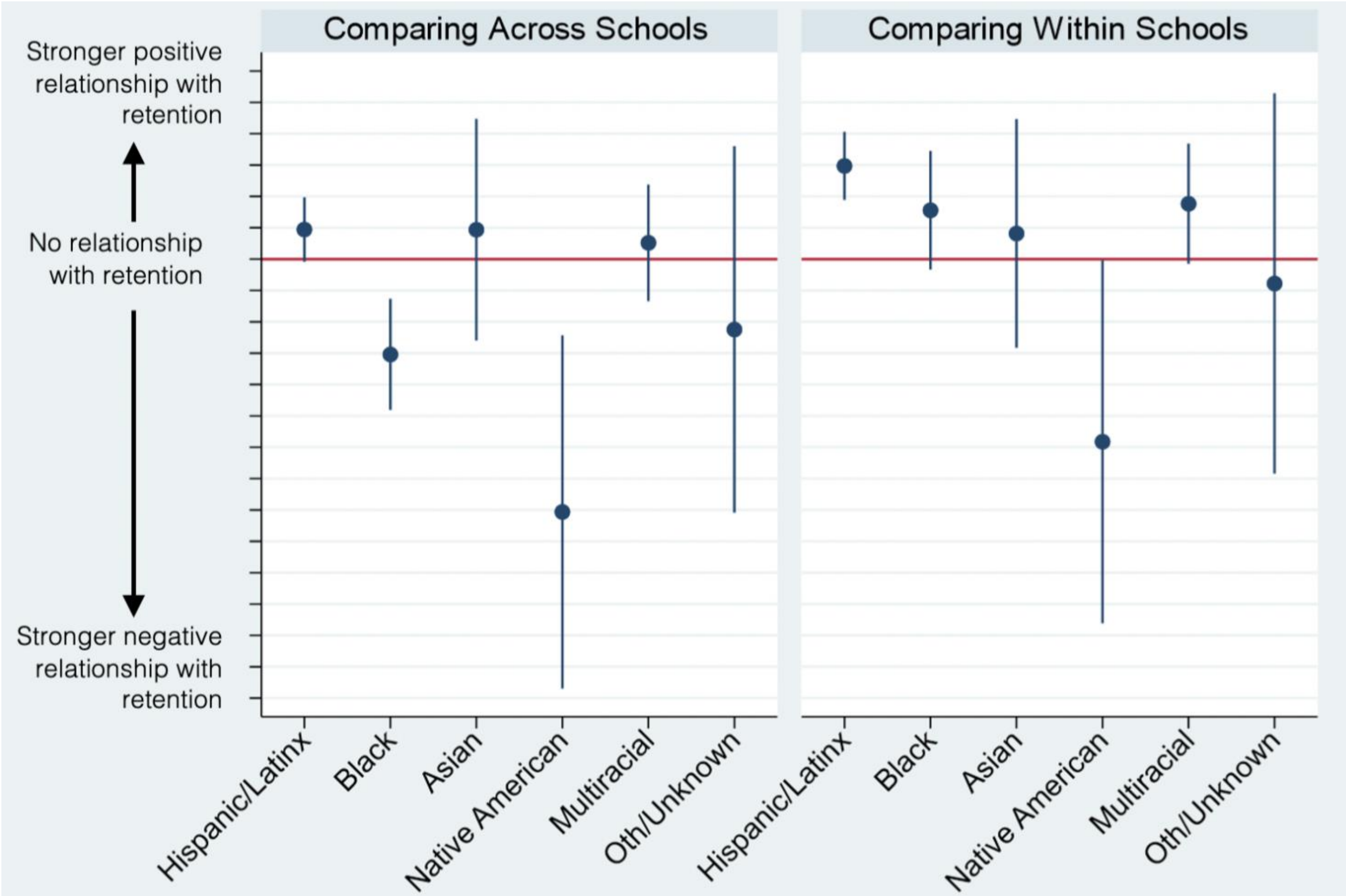
Note: this figure uses two of the key indices from Figure 8 and compares average index scores for staying and leaving teachers in by categories of prior in-district experience.

Figure 10. RQ 1C—Does the Relationship between CollaboRATE and Retention Differ by Teacher Race?



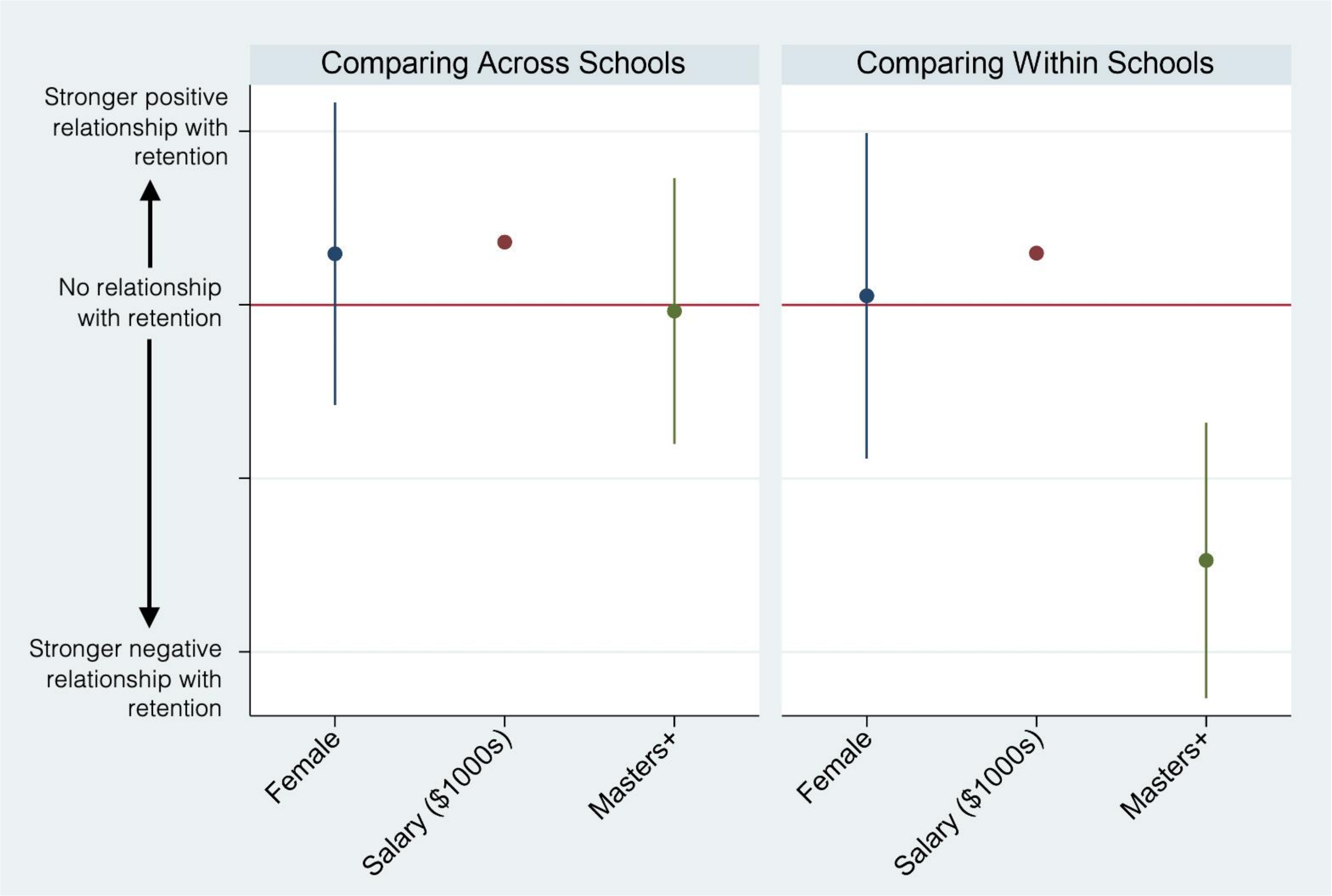
Note: This figure uses two of the key indices from Figure 8 and compares average index scores for staying and leaving teachers in four racial groups of DPS teachers. Not all racial groups could be considered due to small counts of teachers.

Figure 11. RQ 1D – Does Teacher Race/Ethnicity Predict In-School Teacher Retention?



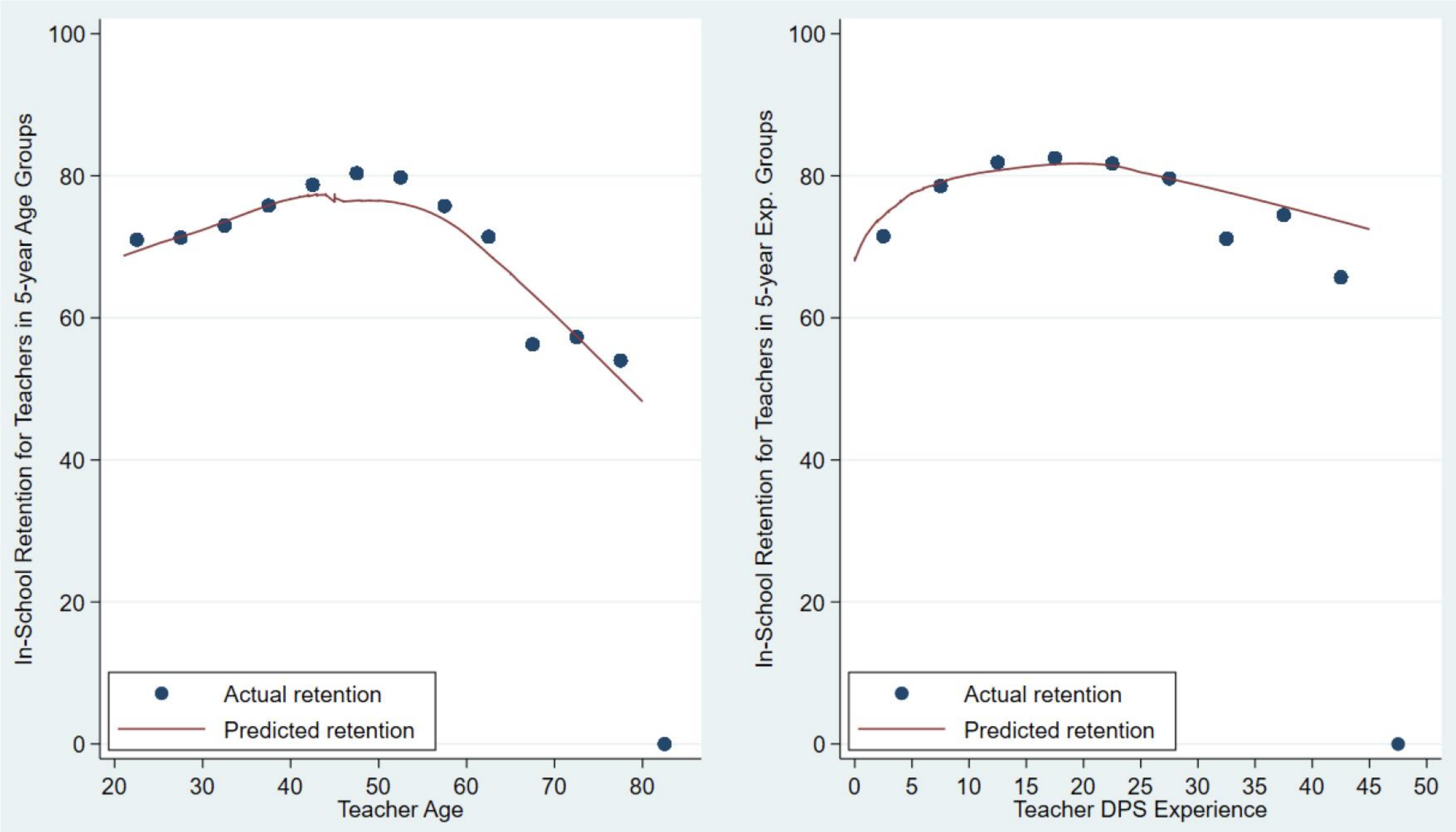
Note. Results in the two panels in this figure are from single logistic regression models that include information on teacher race/ethnicity. The left panel (Comparing Across Schools) does not include school fixed effects, while the right panel (Comparing Within Schools) does include them. The vertical lines around each dot (representing a regression coefficient) are 95% confidence intervals. The retention of different racial groups are shown in comparison with the retention of White teachers. White teachers are not selected as the reference group because their outcomes should be perceived as normative, but rather to focus on comparisons between the outcomes of historically-underserved teacher subgroups relative to this systematically more privileged group. In addition, by making the largest teacher subgroup the reference category, it maximizes the precision of estimates as approximately 73% of DPS teachers are White.

Figure 12. RQ 1D – Do Other Teacher Characteristics Predict In-School Teacher Retention?



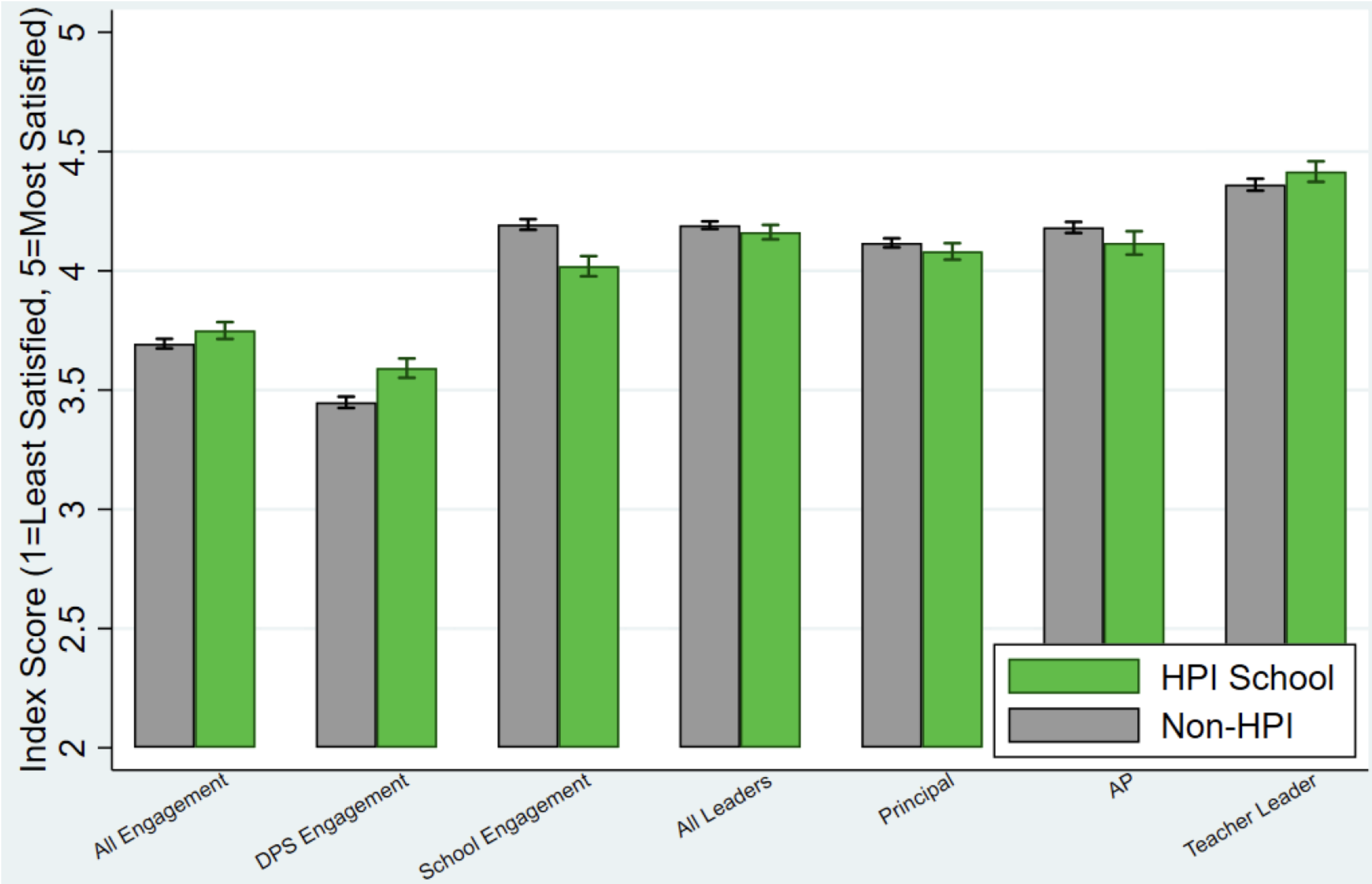
Note. Results in this figure are all from separate logistic regression models The left panel (Comparing Across Schools) does not include school fixed effects, while the right panel (Comparing Within Schools) does include them. The vertical lines around each dot (representing a regression coefficient) are 95% confidence intervals. The estimate for "salary" does have a confidence interval, but it is so small as to be not visible on the graph.

Figure 13: RQ 1D – Percent of Teachers Retained in School at Different Teacher Age Bins (Left Panel) and Teacher Experience (Right)



Note. These graphs show the average in-school retention for teachers in each 5-year age or experience "bin." This allows us to see, for example, that the average retention of teachers with 0-4 years of experience is about 70%. The bins allow for a simpler visual, while the predicted retention is based off of the age and retention of each individual teacher.

Figure 14. RQ 2A — CollaboRATE Results in HPI and Non-HPI Schools



Note. Data are drawn from 5,972 teacher surveys from non-HPI schools and 1,829 teacher surveys from HPI schools administered in 2018 and 2019. Not every teacher has a value for every index, due to skipped questions and sections of CollaboRATE

Figure 15. RQ 2C – Pre- and Post-HPI Trends in In-School Teacher Retention Rates in HPI Schools, by School Level

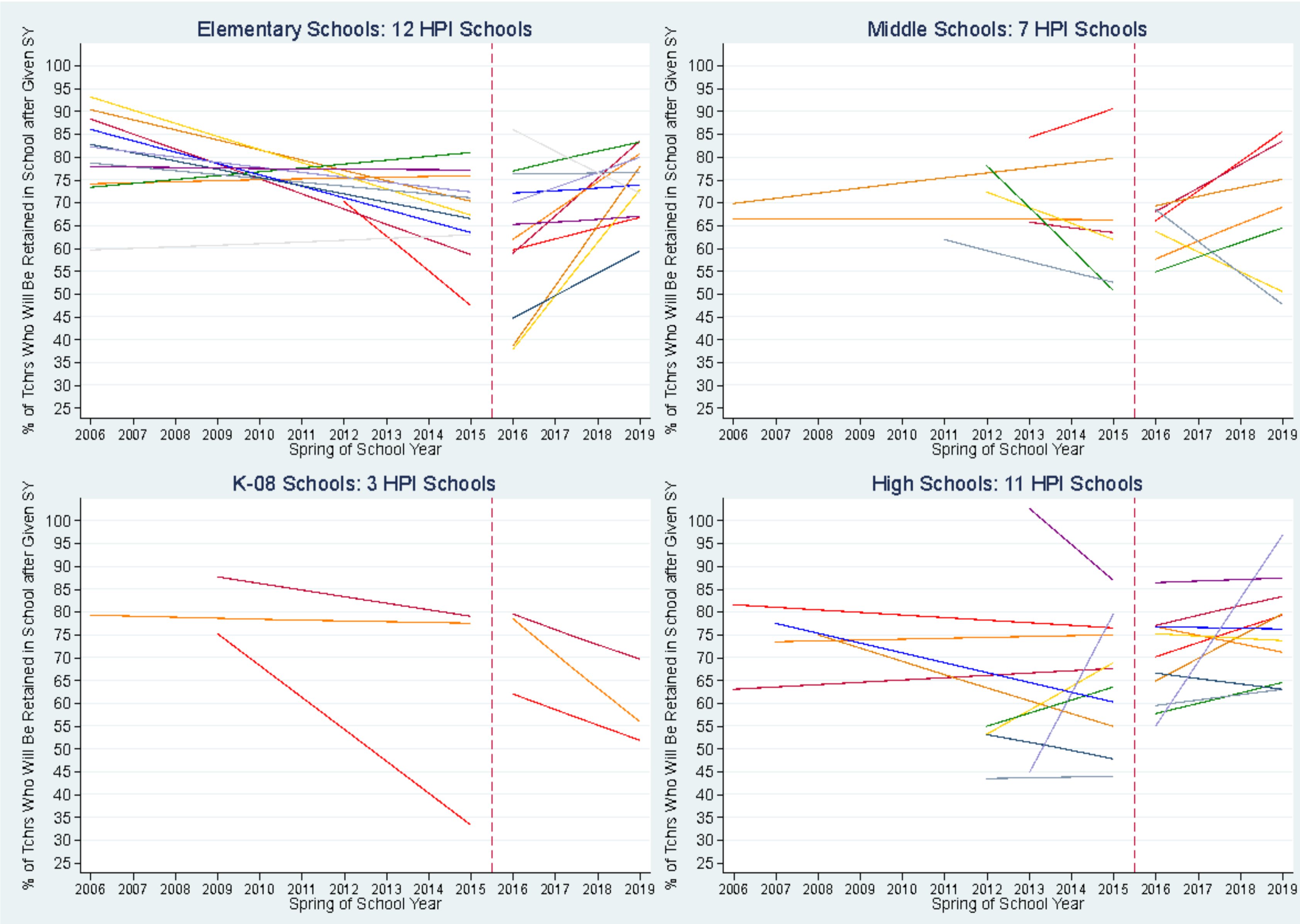
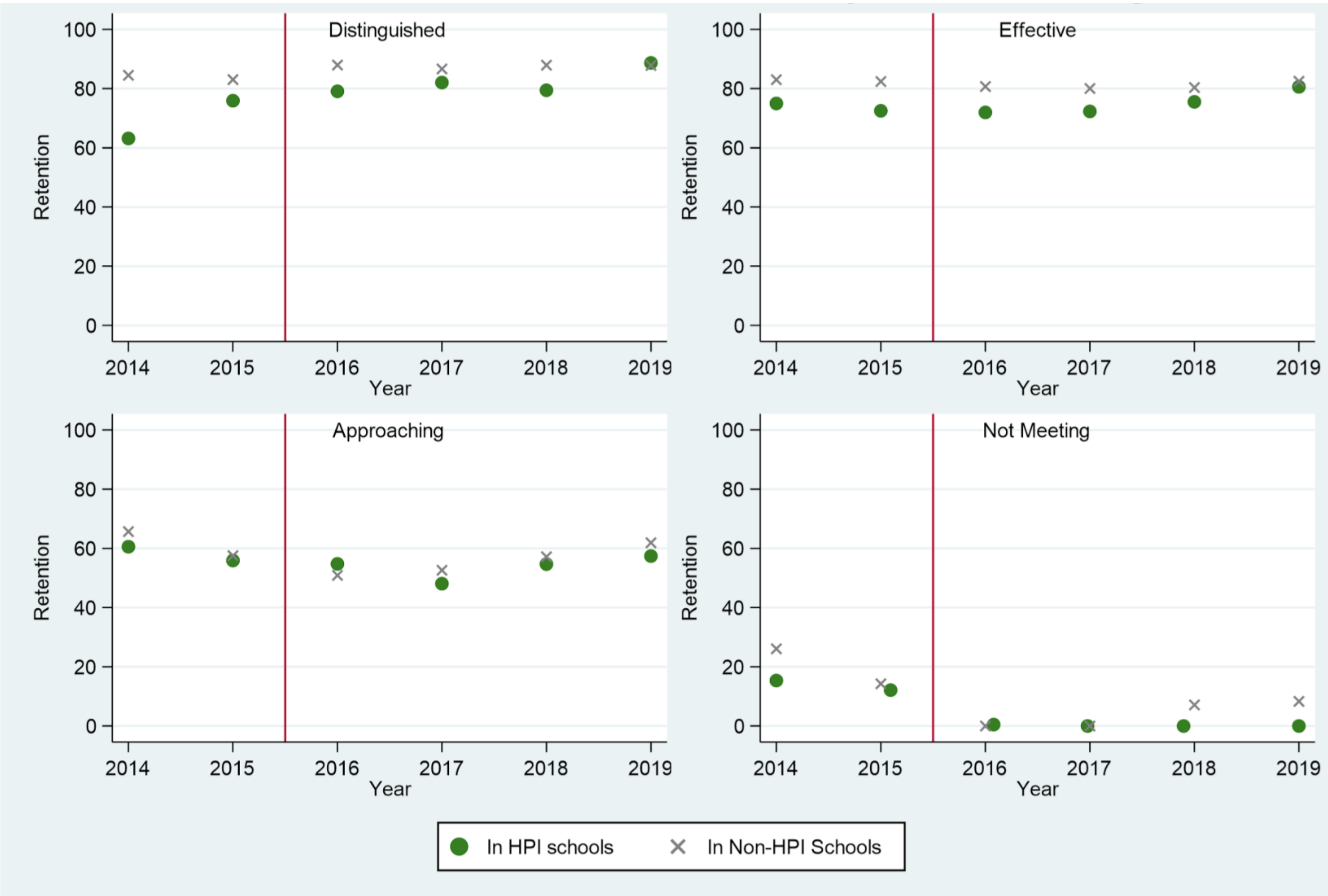
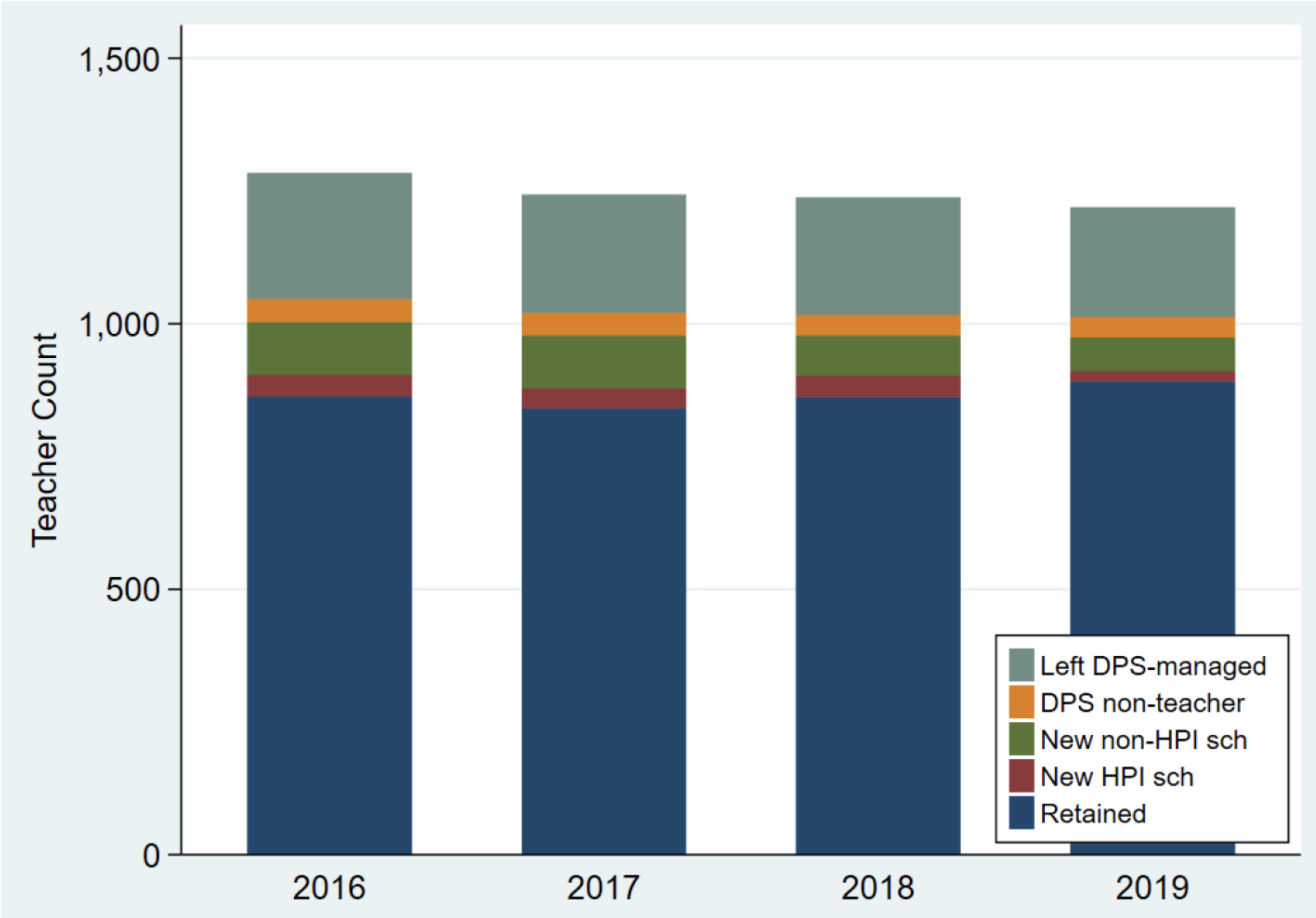


Figure 16. RQ 2C—Teacher Retention by LEAP Rating over time in HPI and non-HPI schools



Note: This graph shows in-school retention in HPI schools and non-HPI schools from 2014 - 2019. This graph is for descriptive purposes only and can be used to see trends, but not to make any conclusion about what caused those trends. Because LEAP is relatively recent, we cannot establish credible pre-trends in by-LEAP retention like we can with overall retention, so we cannot use the CITS method to make a causal inference. For more information about differences by LEAP rating, see Appendix E

Figure 17. RQ 2D – Retention and Mobility of Teachers in HPI School, over the 4 HPI Years



Note: *This graph tracks the movement (or lack thereof) of teachers in HPI schools throughout the program. Some teachers were retained, others went to new schools within DPS (either new HPI schools or new non-HPI schools), others continued to work for DPS in a different position, and the final group left district-managed schools altogether for schools outside of DPS or for DPS charter schools.*

Figure 18. RQ 3A – Example of Pre-Trend Matching for an HPI Elementary School

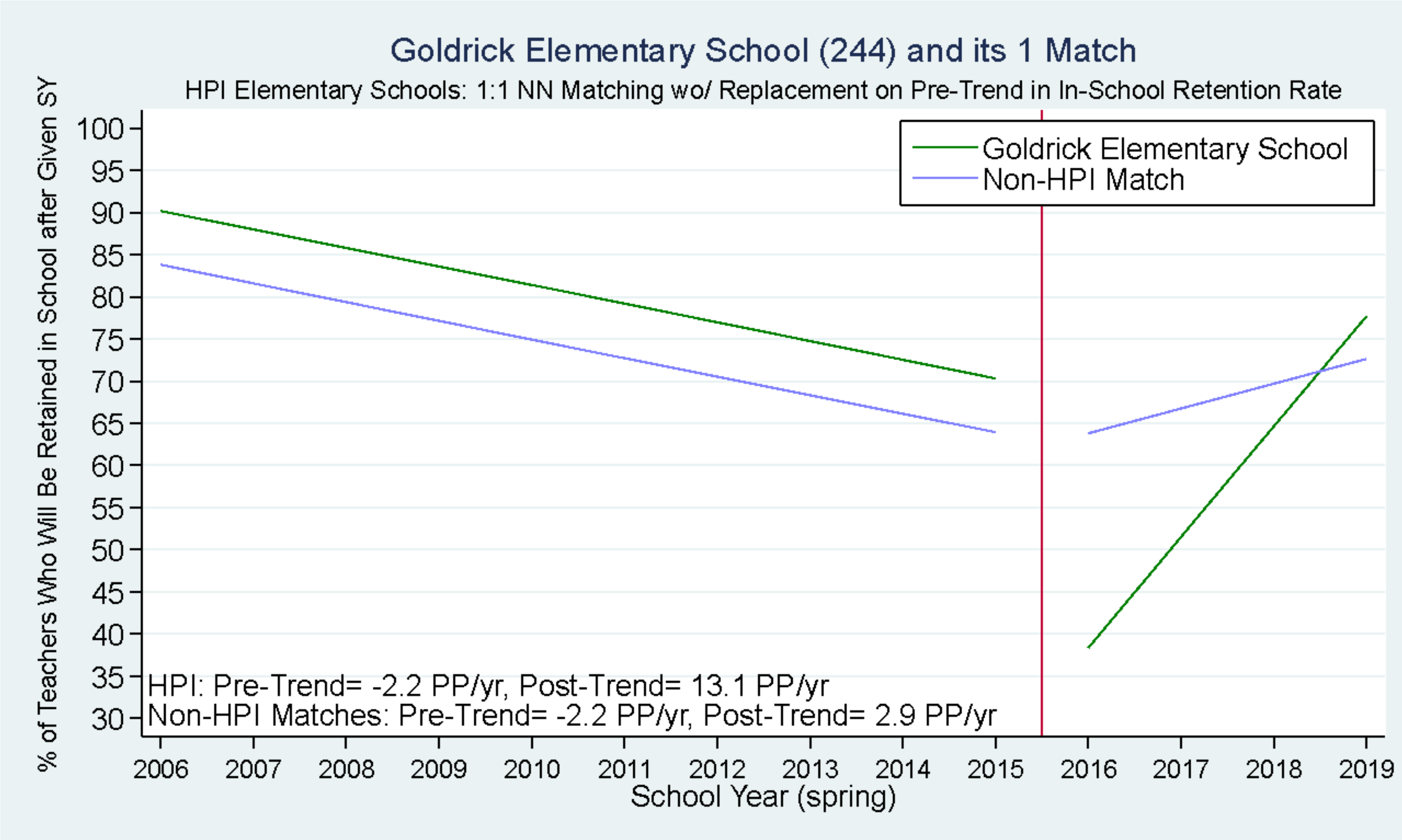


Figure 19. RQ3A – Elementary Schools, Ranked on Trend in In-School Retention Rate in Pre-HPI Years (2006 – 2015)

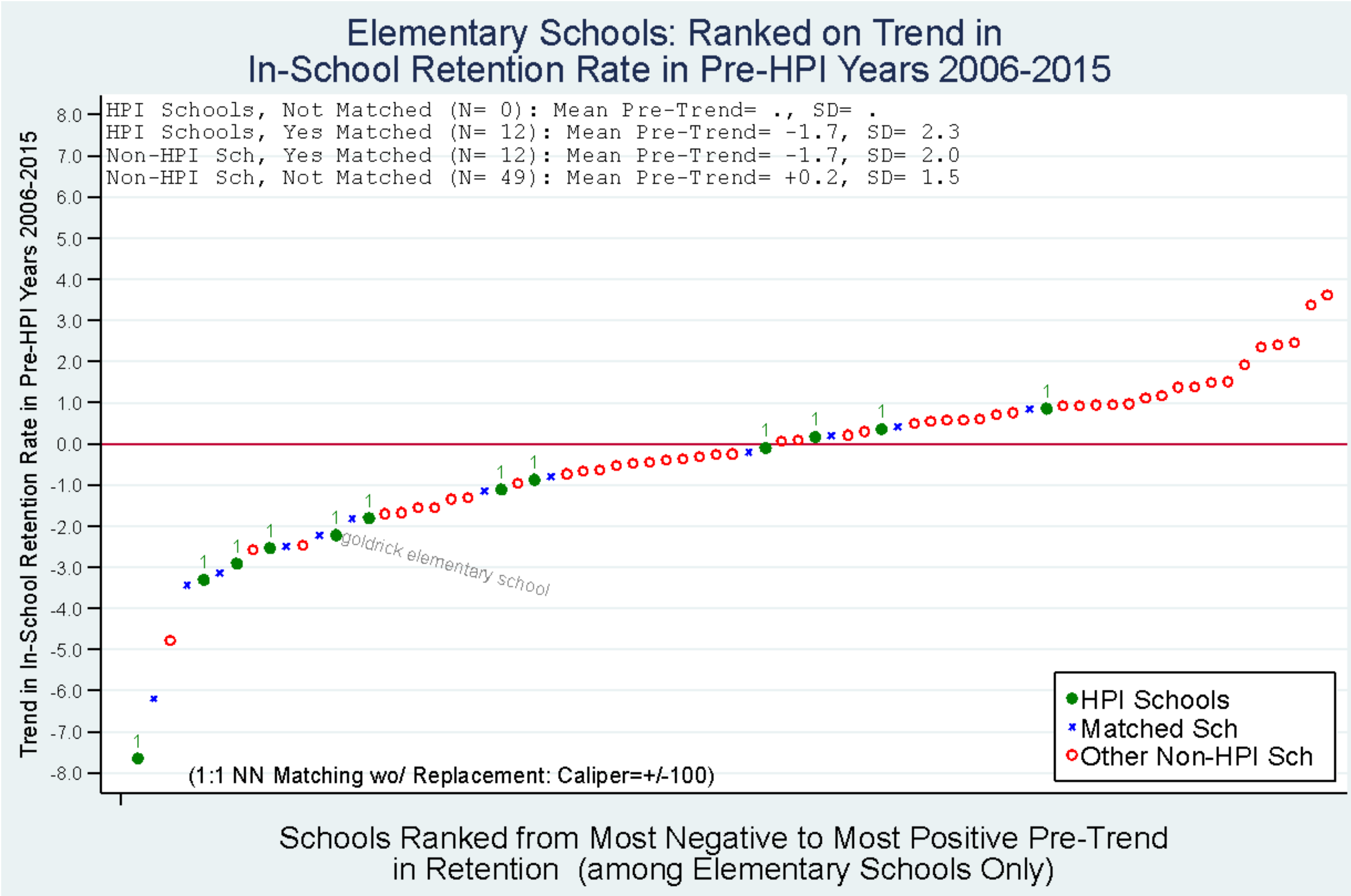


Figure 20. RQ 3A – CITS Estimate of HPI Impacts on In-School Teacher Retention for Elementary Schools

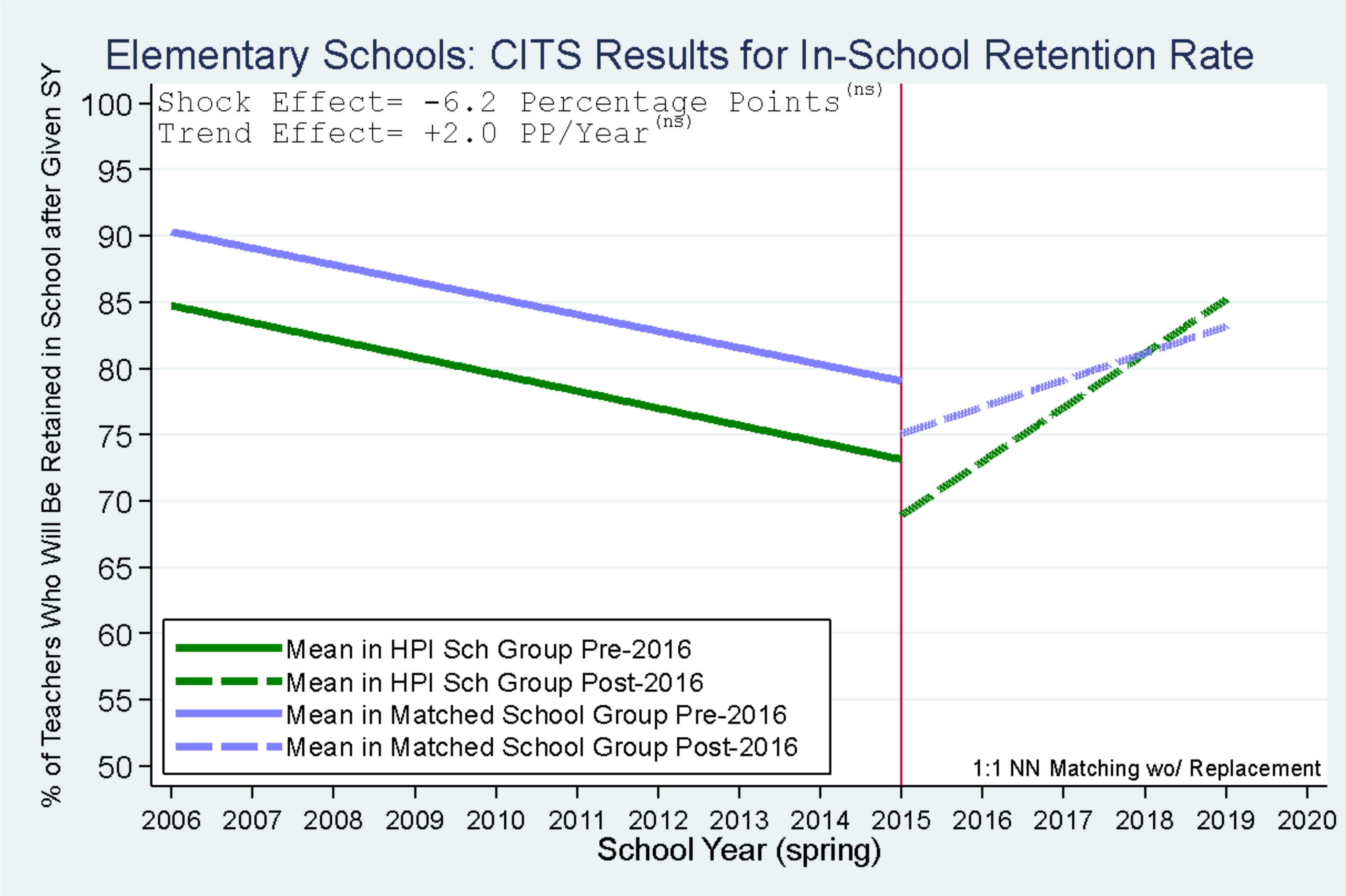


Figure 21. RQ 3B – Elementary Schools Ranked by Trend in In-District Retention Rate in Pre-HPI Years (2006 – 2015)

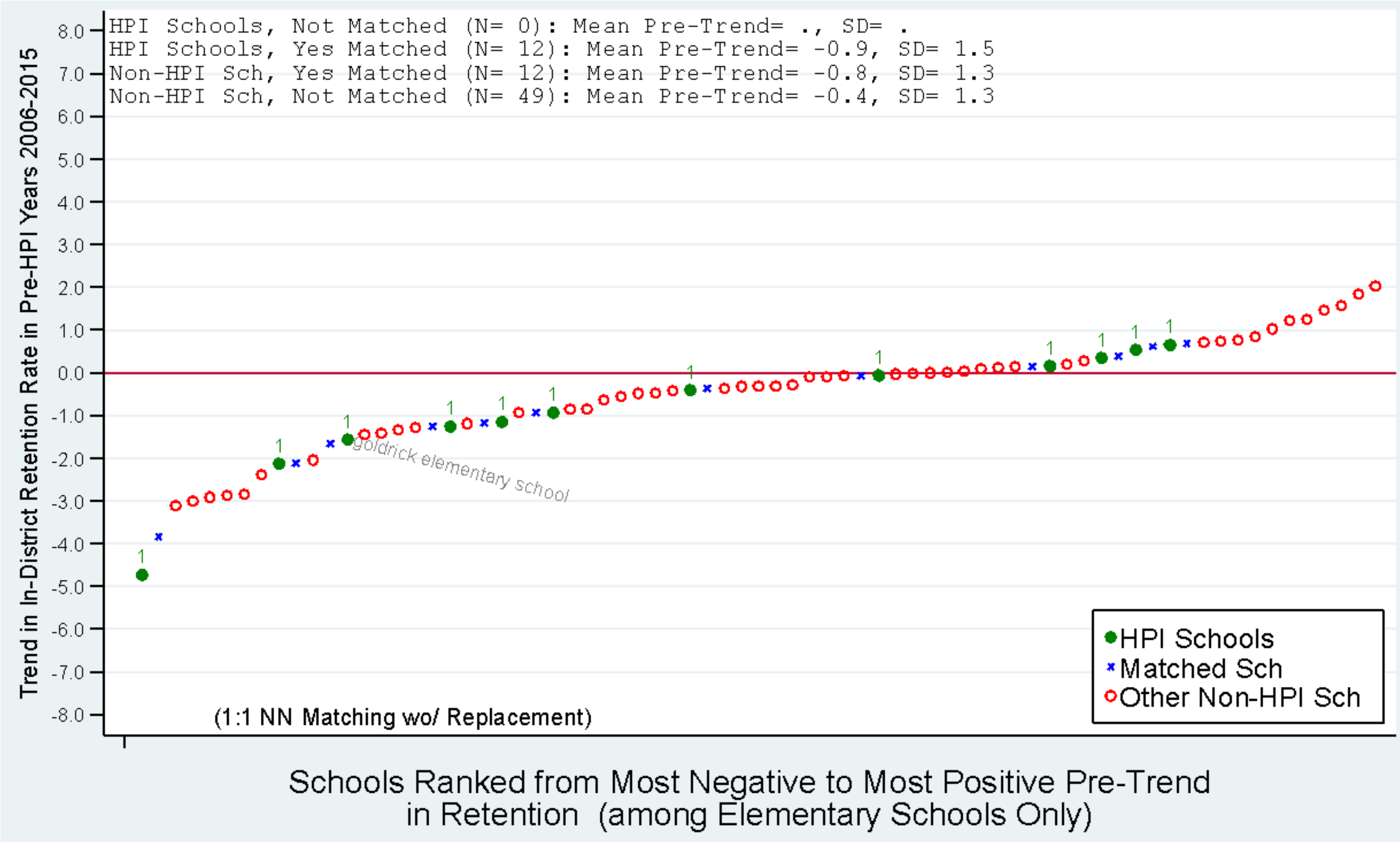


Figure 22 RQ 3B – High Schools Ranked by Trend in In-District Retention Rate in Pre-HPI Years (2006 – 2015)

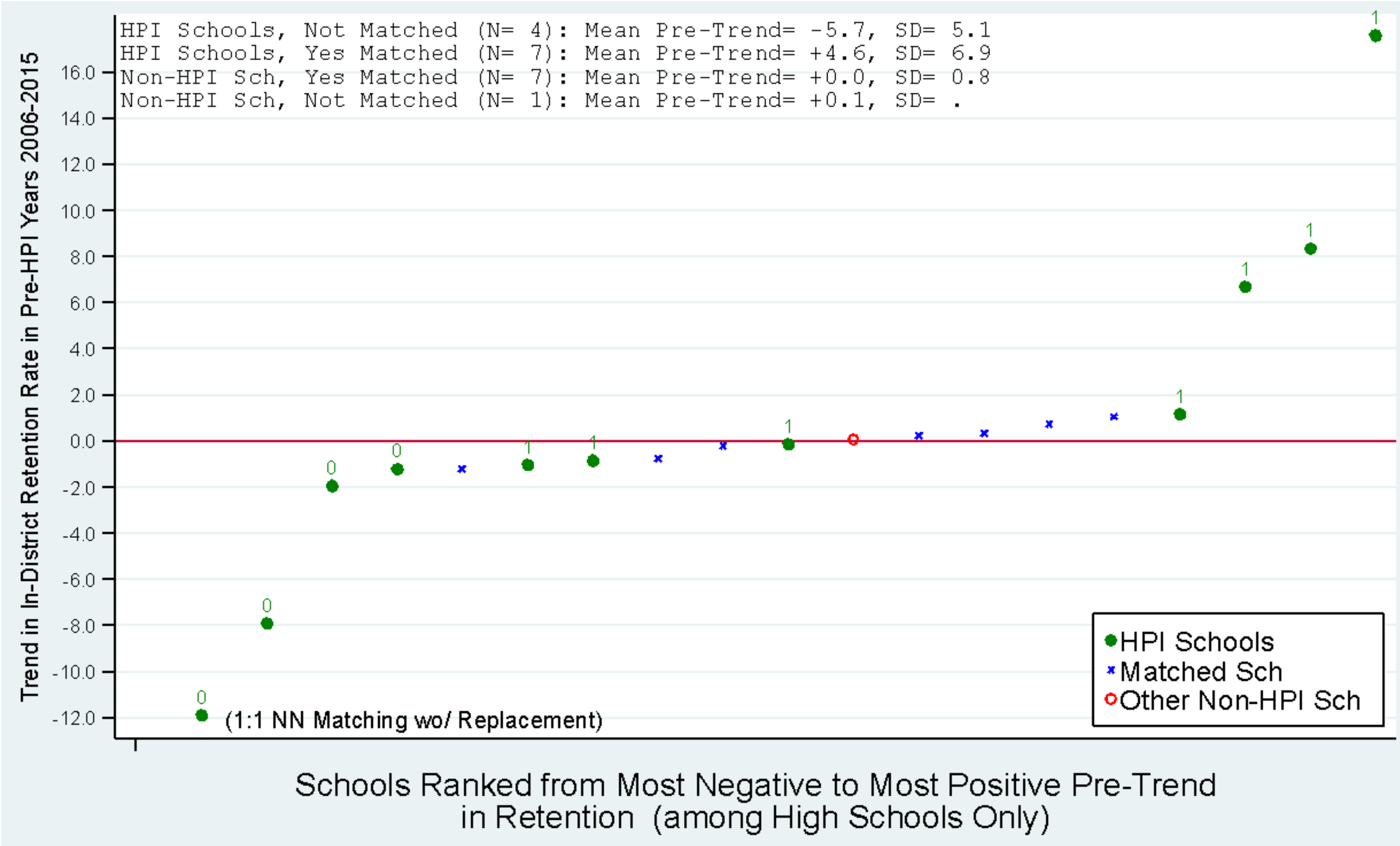
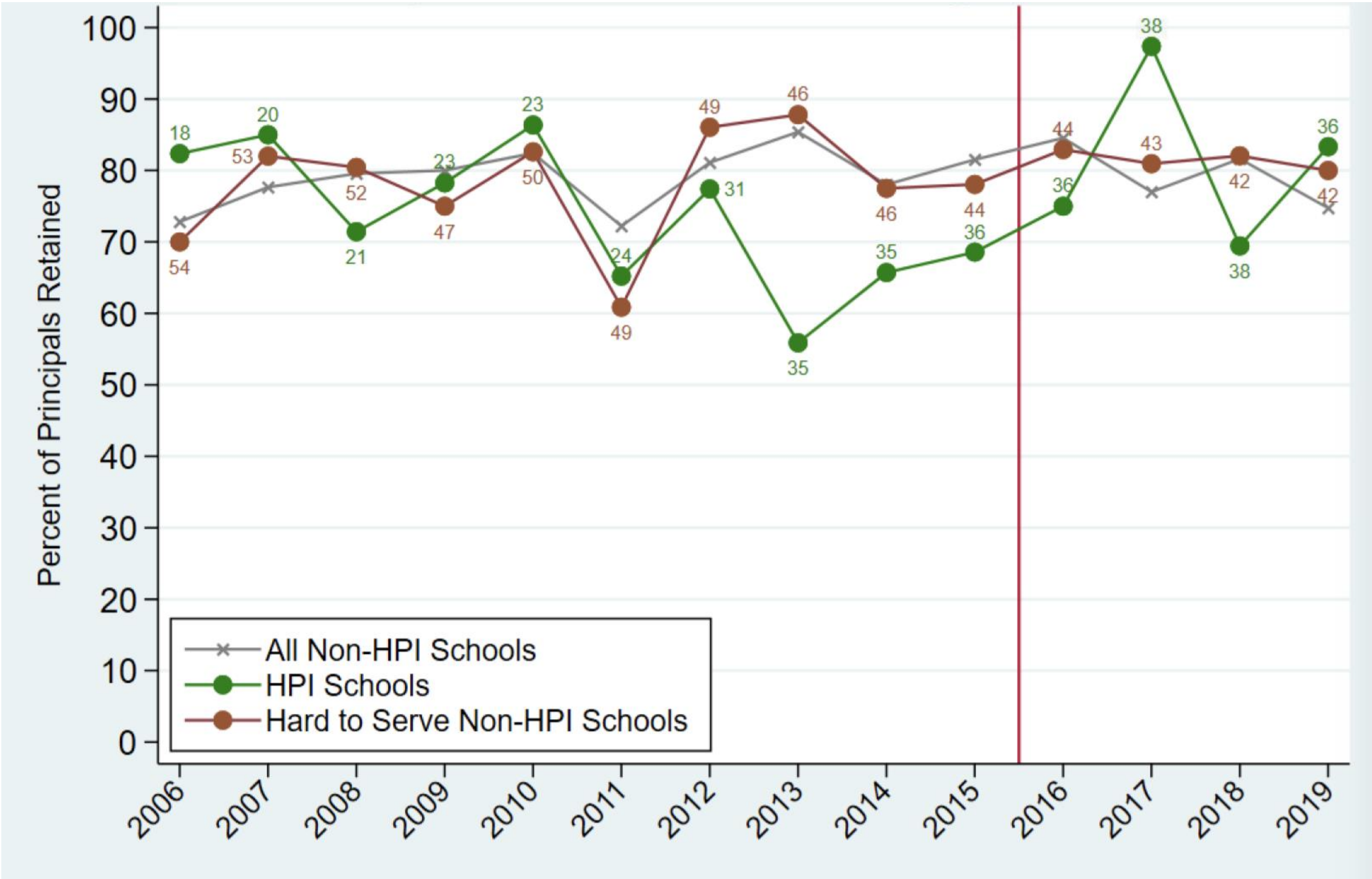


Figure 23. RQ 3D. Principal Retention In-School by HPI Status



Note. Prior to 2016, schools are counted as HPI if they will eventually receive HPI status. Marker labels represent the total number of schools in that category and year. HPI school counts are lower in earlier years due to eventual HPI schools having not yet opened. Similarly, non-HPI Hard to Serve school counts prior to 2016 do not include schools that were Hard to Serve in those years but that will eventually receive HPI status. HPI school counts fluctuate between 2016 and 2019 due to two new schools receiving HPI status in 2017 (Kepner Beacon and Bear Valley) and two HPI schools closing in 2019 (Kepner Middle and Henry World).

APPENDICES

Appendix A: Work Scope 76
Appendix B: TWC Mission and Report Author Biographies 81
Appendix C: Matching Tables and Figures for Middle, K-08, and High Schools 83
Appendix Table C1: In-School Teacher Retention: 1:1 NN Matching without replacement 83
Appendix Figure C1: Middle Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015 83
Appendix Figure C2: K-08 Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015 83
Appendix Figure C3: High Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015 84
Appendix Table C2: In-District Teacher Retention: 1:1 NN Matching without replacement: 85
Appendix Figure C4: Middle Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015 85
Appendix Figure C5: K-08: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015 86
Appendix Figure C6: High Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015 86
Appendix Table C3: In-School Teacher Retention: M:1 Caliper Matching with replacement: 87
Appendix Figure C7: Elementary Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015 87
Appendix Figure C8: Middle Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015 87
Appendix Figure C9: K-08 Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015 88
Appendix Figure C10: High Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015 88
Appendix Table C4: In-District Teacher Retention: M:1 Caliper Matching with replacement: 90
Appendix Figure C11: Elementary Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015 90
Appendix Figure C12: Elementary Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015 91
Appendix Figure C13: K-08 Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015 91
Appendix Figure C14: High Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015 92
Appendix D: CITS Analysis using Propensity Score Matching (PSM) 93
Appendix E: Regression Discontinuity Analysis 99
Appendix F: CollaboRATE Survey Items 106

Appendix A: Work Scope

Highest Priority Incentive Scope of Work
CU Boulder Faculty Allison Atteberry and Mimi Engel
Updated June 2019

This work scope is proposed by Professors Allison Atteberry and Mimi Engel, faculty at the University of Colorado Boulder (CU). The goal of the proposed analyses is to help DPS and Denver Classroom Teachers Association (DCTA) explore and understand the effects of the Highest Priority Incentive (HPI) program, which began in 2015. We will build on the work that DPS analysts have begun with the aim of understanding whether and to what extent the HPI program has affected student, teacher, and school outcomes. The primary objective in Phase I is to estimate the effect of the first years of the HPI program. The "Tentative Agreement Between School District #1 Denver Public Schools and Denver Classroom Teachers Association Counter Proposal: 02/14/2019 5:30AM" document states the following requirement: "DPS and DCTA will commission a joint research project conducted during the term of this agreement to examine the root causes of educator retention and turnover throughout the district's High Priority Schools in order to identify possible solutions" (page 5). This work scope is intended to meet that requirement.

Research Questions

Research Objective 1: Describe longitudinal patterns in teacher retention in, and exit from, DPS.

- What percentage of teachers leave their schools and/or the District each year?
- How has this changed over time?
- How does this vary across schools?
- Which school-level factors predict teacher retention/exit, including malleable (e.g., leadership, climate, DPS initiatives) and non-malleable factors (e.g., school location, student demographics)?
- How does this vary across teachers?
- Which teacher-level factors predict teacher retention/exit from a school or DPS (e.g., years of experience, demographic characteristics, subject, grade level)?
- Why do teachers report leaving (exit surveys)? Has this changed over time?

Research Objective 2: Estimate Causal Effects of HPI

- How do HPI-eligible teachers/schools differ from non-HPI schools/teachers? Are there comparable non-HPI schools in DPS?
- How many teachers receive HPI incentives, for how long, and at which schools?
- Does the HPI program help school leaders attract teachers to high priority schools?
- Do teachers who are eligible to receive these incentives tend to remain in their positions longer than otherwise similar teachers in otherwise similar schools?
- When teachers exit HPI schools, what do they do next (e.g., new non-HPI school, new HPI school, exit DPS)?
- Do teachers in HPI schools in general, and, teachers receiving HPI bonuses in particular report higher job satisfaction than similar counterparts who do not?

1

- Is HPI participation (school/ teacher-level) associated with changes in student achievement?

About the HPI Program

In 2015, DPS launched two incentives for over 1,500 teachers and specialized service providers (SSPs) in 30 high-poverty schools, broadly referred to as the Highest Priority Incentives program (HPI). The goal of this program is to attract and retain great teachers particularly in the District's highest-poverty schools. The incentives went into effect at the start of the 2015-16 school year and are paid in addition to all current ProComp incentives. There are two separate HPI incentives:

- ***Highest-Priority Monthly Incentive*** – ("Monthly Incentive")
  An incentive paid monthly for teachers and SSPs who work at one of the highest-priority schools (up to \$3000, \$2250, or \$1500 per year for teachers rated Distinguished, Effective, or Approaching, respectively).
- ***Highest-Priority Retention Incentive*** – ("Annual Incentive")
  A once-per-year, one-time incentive payment provided for teachers and SSPs who return to a highest-priority school (up to \$1000, \$750, or \$500 per year for teachers rated Distinguished, Effective, or Approaching, respectively).

Scope of Work

Phase I - Estimate Effects of HPI on Student and Teacher Outcomes

Overview.
The goal of the Phase I work scope is to leverage existing quantitative data to estimate the overall effect of the HPI program on outcomes in DPS. Using existing data, we can conduct a retrospective evaluation of the HPI program's effects from 2015 through 2018. Phase I results can be used to guide potential Phase II analyses that could involve the collection of *new* data. Initial findings on the effects of the HPI program estimated from these quantitative analyses will be used to determining whether we will collect additional data.

Data Sources.
To answer the proposed research questions, we will utilize existing data provided by DPS and/or DCTA. These include student achievement outcomes, teacher satisfaction and retention outcomes (both in-school and in-district), and school outcomes (e.g., stability of the teaching staff, school climate). We will include key control variables in the analysis to account for other potential systematic differences in the teachers and schools exposed to the HPI program, as well as the students who attend those schools. In addition, we will explore existing survey data from either DCTA or DPS related to teachers' opinions about this program, the school, and their decisions about where to work. The surveys will be evaluated for the appropriateness of the topics covered, question wording, and response rates. If they add value to answering the current research questions, we will incorporate them into the analysis.

Methodological Approach.
The objective of the Phase I analyses is to test the null hypothesis that the HPI program had no effect. The alternative hypothesis, that HPI program had an effect, will be explored at the

District, school, teacher, and student-levels. It is possible that we will find that the HPI program may have had an effect at one level (e.g., teachers) but not another (e.g., students). The focus of the research design will be to isolate any potential effects of the HPI program from the non-random, systematic sorting of students, teachers, and schools that experienced the HPI program. Data availability will determine which methodological approach(es) are most appropriate, however we will explore leveraging variation in exposure to HPI using fixed effects, comparative interrupted time series analyses, propensity score matching, and/or a regression discontinuity design. While a detailed cost-benefit analysis is outside the scope of the current study, we will contextualize any potential benefits of the program within its overall costs.

Proposed Timeline.
The CU-Boulder research team intends to have initial analyses proposed in Phase I completed within six months from when the necessary data is received from DPS.

Phase II - Explore Mechanisms for Phase I Findings [Tentative]

At this stage, we do not yet know what we will find with regard to the effects of the HPI program. If we *do* find effects, Phase II analyses could focus on understanding whether particular subgroups (of schools, teachers, and students) experienced differential effects. We could also potentially design and administer surveys and/or focus group conversations to explore the underlying mechanisms that might help us understand patterns of results. If we do not find effects of the HPI program, we could design surveys and/or focus groups to understand why the program did not work as intended (i.e., differentiating between implementation barriers, or assumptions that did not bear out with regard to the theory of change undergirding the HPI program).

After the completion of Phase I, results will be presented to both DCTA and DPS, and we will consider whether there is interest in pursuing Phase II. The design of any new data collection materials would be done in consultation with DPS and DCTA stakeholders.

Solicitation of Feedback and Review

The current work scope is intended to provide useful information to both DCTA and DPS stakeholders about the HPI program. To that end, we will interact with both groups to receive input on the scope of work and to review progress on executing research goals and objectives. This will be particularly salient once Phase I is complete and decisions need to be made about Phase II. Although additional requests may surface during the course of this work, the partner institutions would require that each additional request is negotiated to mitigate scope creep.

The CU Team – Atteberry and Engel, along with CU analyst support staff – are responsible for creating preliminary analysis plans (i.e., Work Scope), conducting data analysis, and drafting results for initial, internal review by DPS and DCTA. Results will be shared via memo prior to review meetings. The CU team will stay in regular communication with contacts at DPS and DCTA to help ensure that the report contents and structure align with the organizational needs and priorities of the District. A preliminary draft of each work product will be shared with key internal

DPS stakeholders and DCTA leadership. The CU Team will incorporate feedback and share revised drafts for DCTA and DPS review. The CU researchers will develop and propose dissemination plans. Following approval, the CU Team will disseminate results for academic audiences. DPS and DCTA leaders will determine how best to communicate results to their constituents and stakeholders.

Phase I Final Report

The report will include the following:
(1) A description of what is known about how HPI schools were selected and how those schools are observably different from other schools.
(2) A brief overview of the theory of action behind HPI.
(3) A description of all data sources used and definitions of key analytic variables.
(4) A discussion of the assumptions required for causal inference and methodology used in the analyses.
(5) A presentation of findings, insights, and the analytical basis for recommendations concerning the research topics outlined in this Work Scope
(6) An Executive Summary, summarizing key insights and recommendations. The Executive Summary will be organized based on discussions with DPS and DCTA to ensure that any recommendations or analytical insights will be optimally presented to support decision-making about the future of the HPI program.

Personnel

Allison Atteberry, Assistant Professor, CU-Boulder.

Allison Atteberry is an assistant professor of Research & Evaluation Methodology at the CU Boulder School of Education (SoE). Dr. Atteberry's work focuses on the identification, selection, development, and retention of teachers who have measurable impacts on student achievement. Dr. Atteberry has previous experience conducting teacher policy research in DPS (studying ProComp, DPS' pay-for-performance compensation system). The relationship that developed from that work is the foundation for the current collaboration. Dr. Atteberry is a 2017 National Academy of Education (NAEd)/Spencer Postdoctoral Fellow. Her research has been funded by the Smith Richardson Foundation, the Laura and John Arnold Foundation, and the Russell Sage Foundation, among others.

Dr. Atteberry has managed large research collaborations. She is PI on a 7-year RCT of full- vs. half-day preschool in Westminster, CO, a district near Denver, ($880,000 budget). For this study, she oversees the administration of over 1000 one-on-one child assessments per year, trains classroom observers, and designs and administers parent and teacher surveys. She is co-PI on a replication of this RCT in Pomona, CA. Dr. Atteberry has conducted teacher policy research in a number of districts including New York City and Prince William Unified School District.

Mimi Engel, Associate Professor, CU Boulder.

Mimi Engel is an associate professor of Research & Evaluation Methodology at the CU Boulder SoE. The central aim of Dr. Engel's research is to provide new information about policies, programs, contextual, and administrative factors that have the potential to improve students' school-related outcomes, particularly among students from traditionally underserved populations. Her interests span several areas in educational policy including teacher labor markets and early childhood. Dr. Engel uses both quantitative and qualitative methods to conduct policy analysis research and program evaluations.

Dr. Engel has conducted extensive research on the teacher workforce in the Chicago Public Schools and was a research associate at the Consortium on Chicago School Research; a model for researcher-practitioner partnerships. She has also used nationally representative data sets and proprietary data in her work on the teacher workforce. Dr. Engel's research has been funded by the WT Grant Foundation, the Spencer Foundation, and the Heising-Simons Foundation, among others.

Signatures of Agreement

DCTA Representative

Name: Robert Gould

Title: DCTA Vice President [Acting President]

Date: 6/13/2019

Signature:

DPS Representative

Name: Sarah Almy

Title: Executive Director, Teacher & Leader Learning

Date: 6/14/2019

Signature

5

Appendix B: TWC Mission and Report Author Biographies

TWC Mission

The Teacher Workforce Collaborative (TWC) is a research-practice partnership (RPP) between Denver Public Schools (DPS) and the University of Colorado Boulder (CU). Building on relationships that began in 2014, TWC was initiated in 2017. Our work aims to narrow Denver's large and persistent achievement gaps. The mechanism for doing so— strengthening the District's teacher workforce—is our partnership's focus.

TWC is focused on strengthening the DPS teacher workforce because increasing access to highly-effective teachers is among the most promising means for improving outcomes for the District's most disadvantaged students. Research documents that teachers are key determinants of student outcomes and that teacher effects vary widely. An extensive evidence base shows that effective teachers are not distributed equitably within and across schools or districts. Thus, DPS' efforts to retain top teachers are a high-leverage focal point that can have direct, sizable impacts on the lives of youth. Moreover, the District's ability to ensure that students from traditionally-underserved populations have access to top teachers is critical to reducing achievement and opportunity gaps.

Principal Investigators

Allison Atteberry, Assistant Professor, CU Boulder.

Allison Atteberry is an assistant professor of Research & Evaluation Methodology at the CU Boulder School of Education (SoE). Dr. Atteberry's work focuses on the identification, selection, development, and retention of teachers who have measurable impacts on student achievement. Dr. Atteberry has previous experience conducting teacher policy research in DPS (studying ProComp, DPS' pay-for-performance compensation system). The relationship that developed from that work is the foundation for the current collaboration. Dr. Atteberry is a 2017 National Academy of Education (NAEd)/Spencer Postdoctoral Fellow. Her research has been funded by the Smith Richardson Foundation, the Laura and John Arnold Foundation, and the Russell Sage Foundation, among others.

Dr. Atteberry has managed large research collaborations. She is PI on a 7-year Randomized Control Trial of full- vs. half-day preschool in Westminster, CO, a district near Denver, ($880,000 budget). For this study, she oversees the administration of over 1000 one-on-one child assessments per year, trains classroom observers, and designs and administers parent and teacher surveys. She is co-PI on a replication of this RCT in Pomona, CA. Dr. Atteberry has conducted teacher policy research in a number of districts including New York City and Prince William Unified School District.

Mimi Engel, Associate Professor, CU Boulder.

Mimi Engel is an associate professor of Research & Evaluation Methodology at the CU Boulder SoE. The central aim of Dr. Engel's research is to provide new information about policies, programs, contextual, and administrative factors that have the potential to improve students' school-related outcomes, particularly among students from traditionally underserved populations. Her interests span several areas in educational policy including the teacher workforce and early childhood. Dr. Engel uses both quantitative and qualitative methods to conduct policy research and program evaluations.

Dr. Engel has conducted extensive research on the teacher workforce in the Chicago Public Schools and was a research associate at the Consortium on Chicago School Research; a model for researcher-practitioner partnerships. She has also used nationally representative data sets and proprietary data in her work on the teacher workforce. Engel's research has been funded by the Institute of Education Sciences, the William T Grant Foundation, the Spencer Foundation, and the Heising-Simons Foundation, among others.

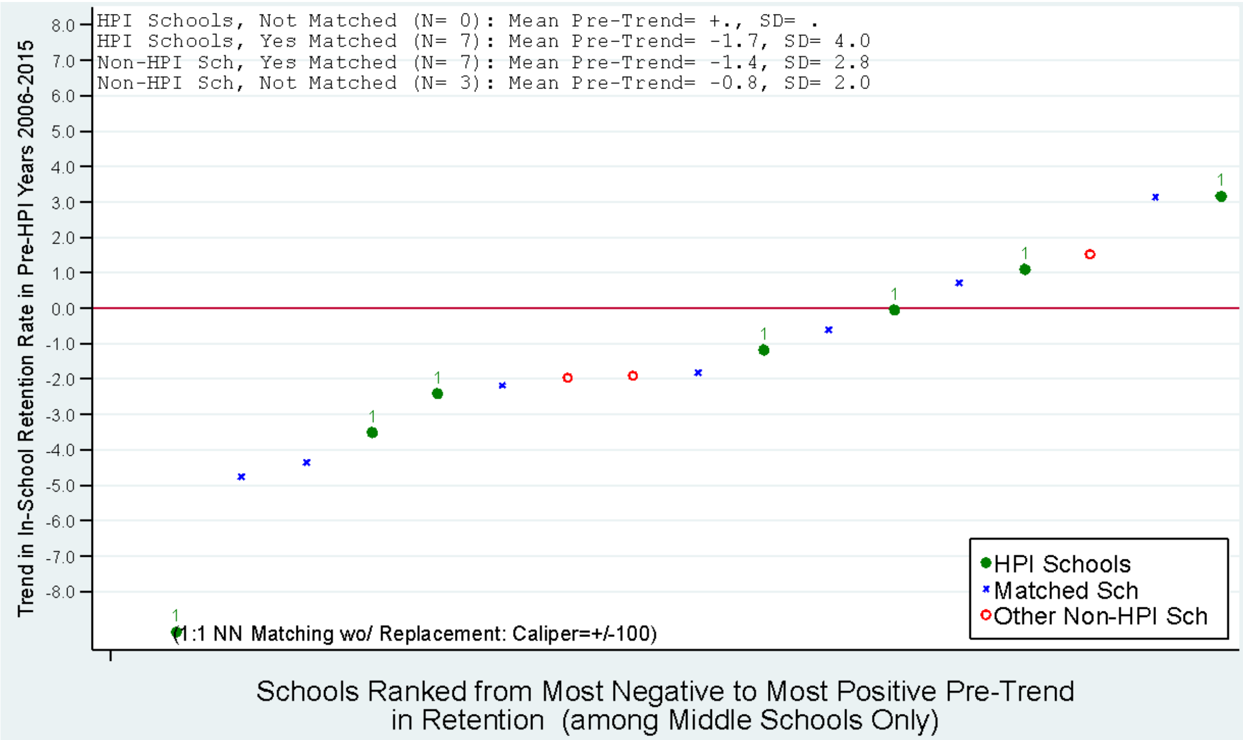
Appendix C: Matching Tables and Figures for Middle, K-08, and High Schools

Appendix Table C1: In-School Teacher Retention: 1:1 NN Matching without replacement

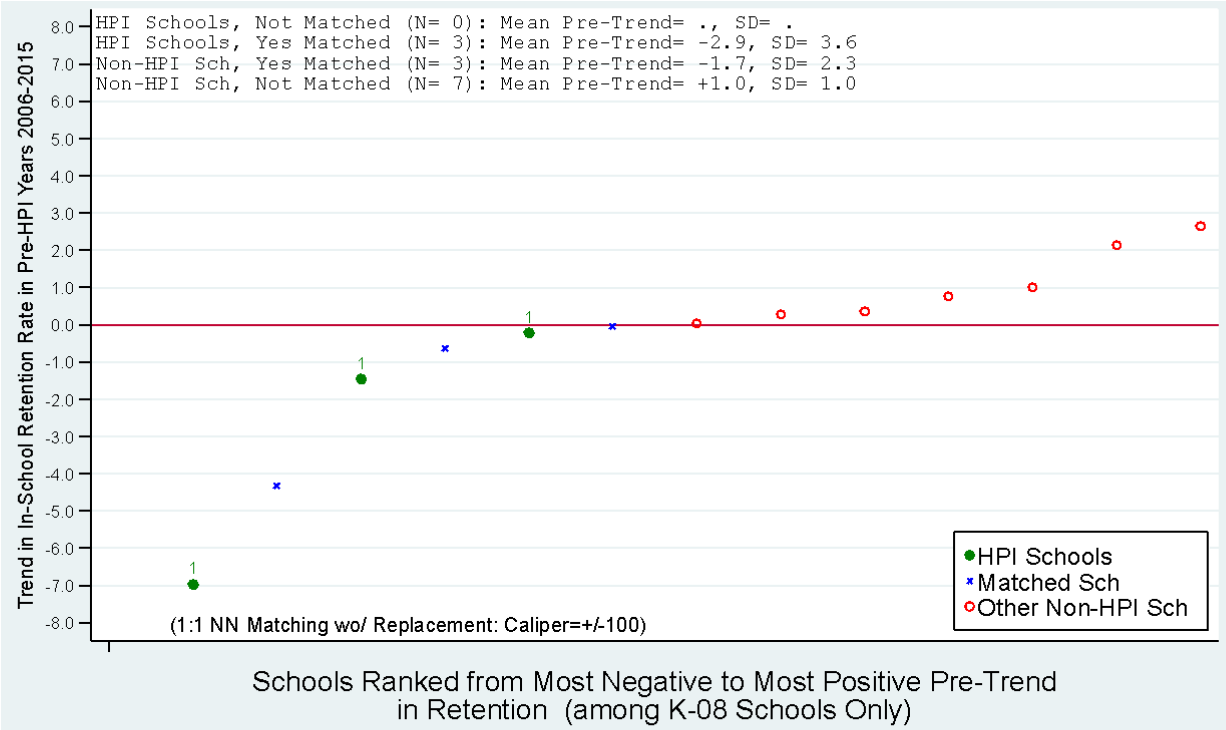
In-School Teacher Retention: Number of Matches Using 1:1 NN without Replacement

	HPI Schools		Non-HPI Schools		
	At Least 1 Match Found	No Matches Found	Used as a Match	Not Used as a Match	All Schools
Elementary	12	0	12	49	73
Middle	7	0	7	3	17
K-08	3	0	3	7	13
High	8	3	8	0	19
All Schools	30	3	30	59	122

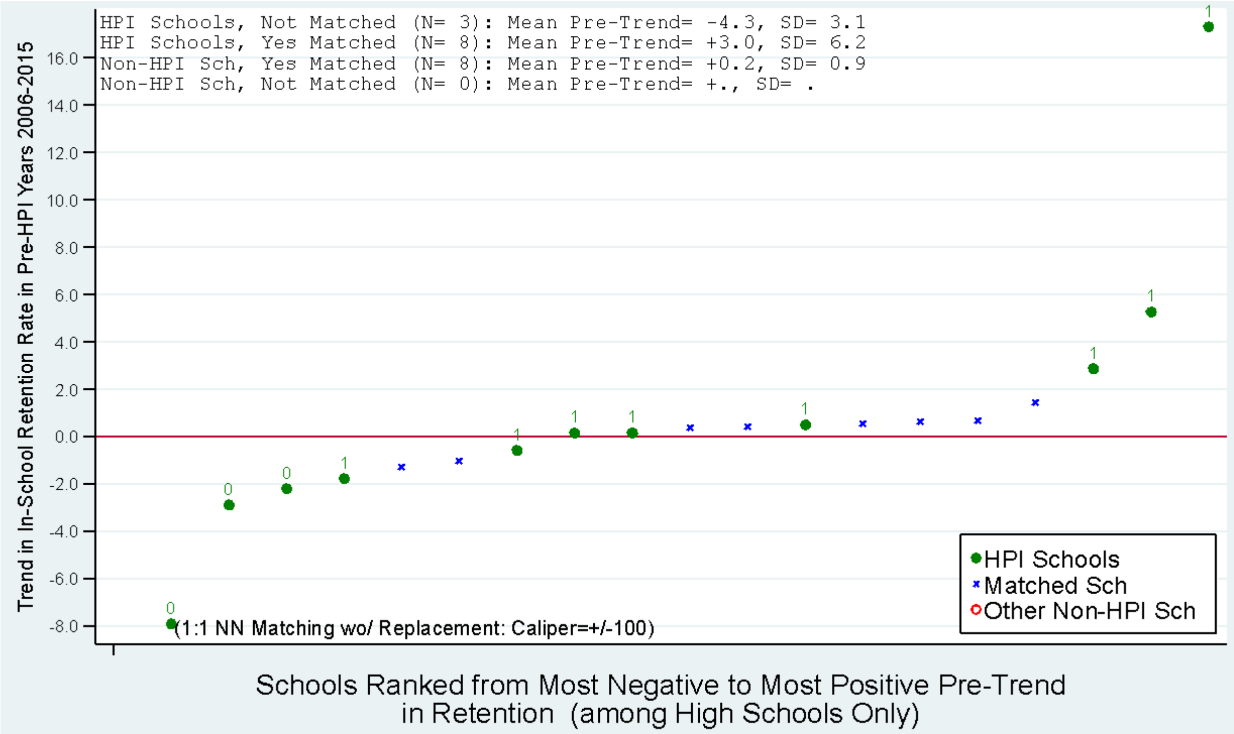
Appendix Figure C1: Middle Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015



Appendix Figure C2: K-08 Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015



Appendix Figure C3: High Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015

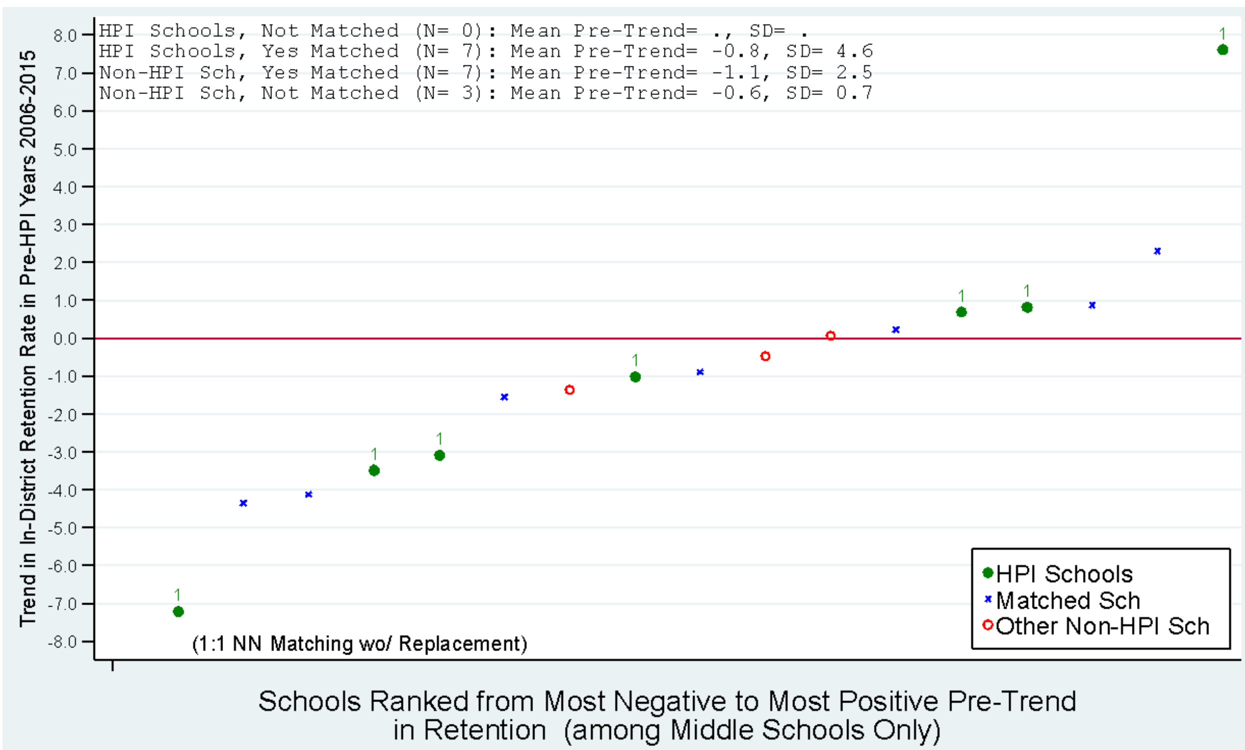


Appendix Table C2: In-District Teacher Retention: 1:1 NN Matching without replacement:

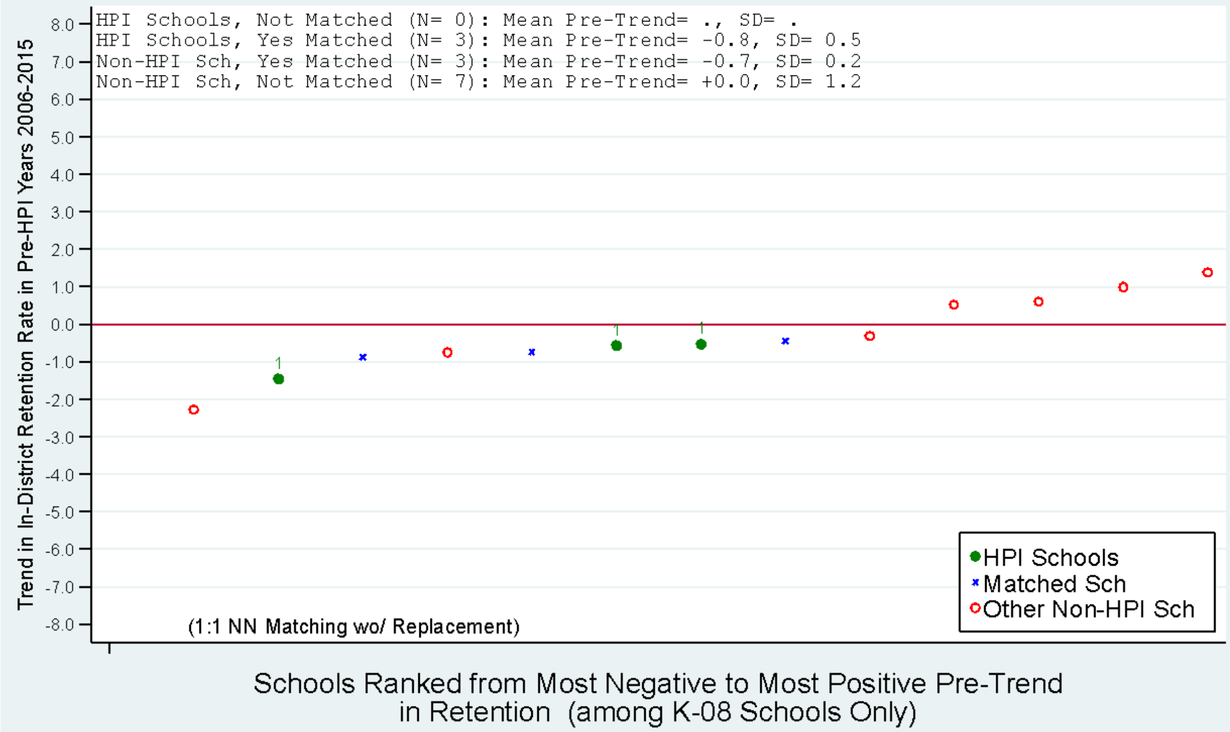
In-District Teacher Retention: Number of Matches Using 1:1 NN without Replacement

	HPI Schools		Non-HPI Schools		
	At Least 1 Match Found	No Matches Found	Used as a Match	Not Used as a Match	All Schools
Elementary	12	0	12	49	73
Middle	7	0	7	3	17
K-08	3	0	3	7	13
High	7	4	7	1	19
All Schools	29	4	29	60	122

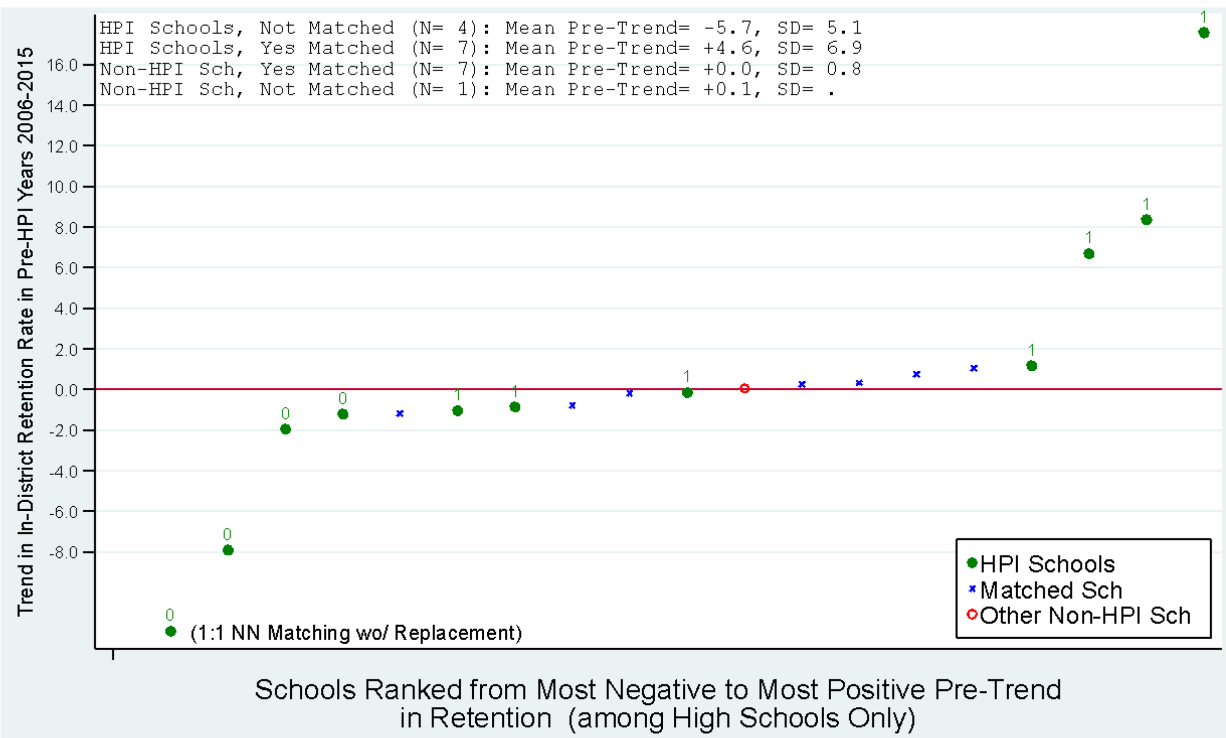
Appendix Figure C4: Middle Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015



Appendix Figure C5: K-08: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015



Appendix Figure C6: High Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015

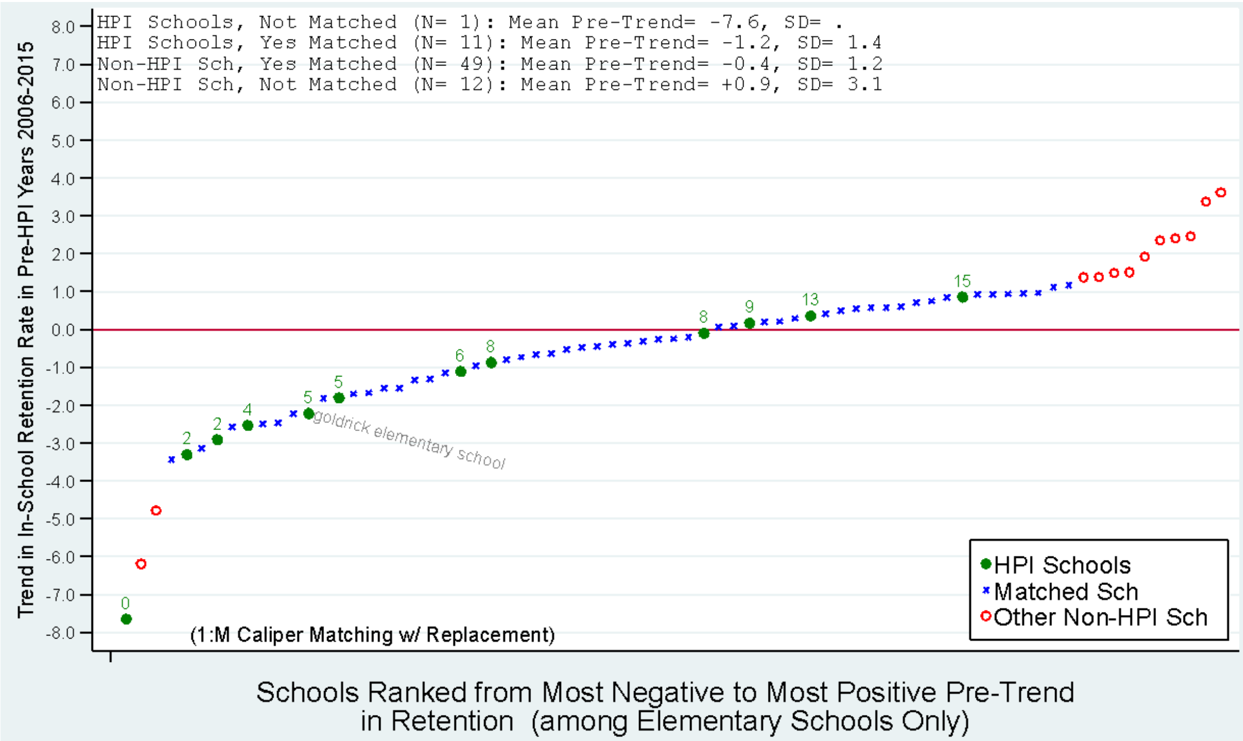


Appendix Table C3: In-School Teacher Retention: M:1 Caliper Matching with replacement:

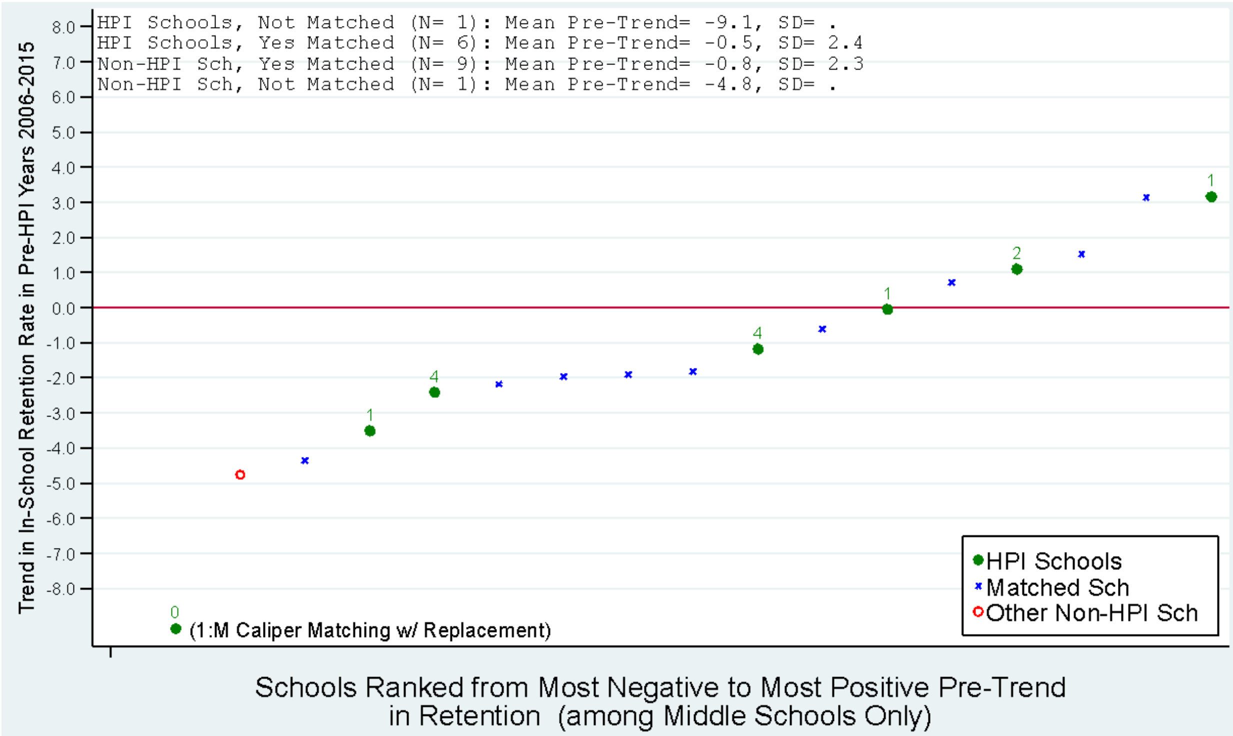
In-School Teacher Retention: Number of Matches Using Caliper Method with Replacemen

	HPI Schools		Non-HPI Schools		
	At Least 1 Match Found	No Matches Found	Used as a Match	Not Used as a Match	All Schools
Elementary	11	1	49	12	73
Middle	6	1	9	1	17
K-08	2	1	2	8	13
High	8	3	8	0	19
All Schools	27	6	68	21	122

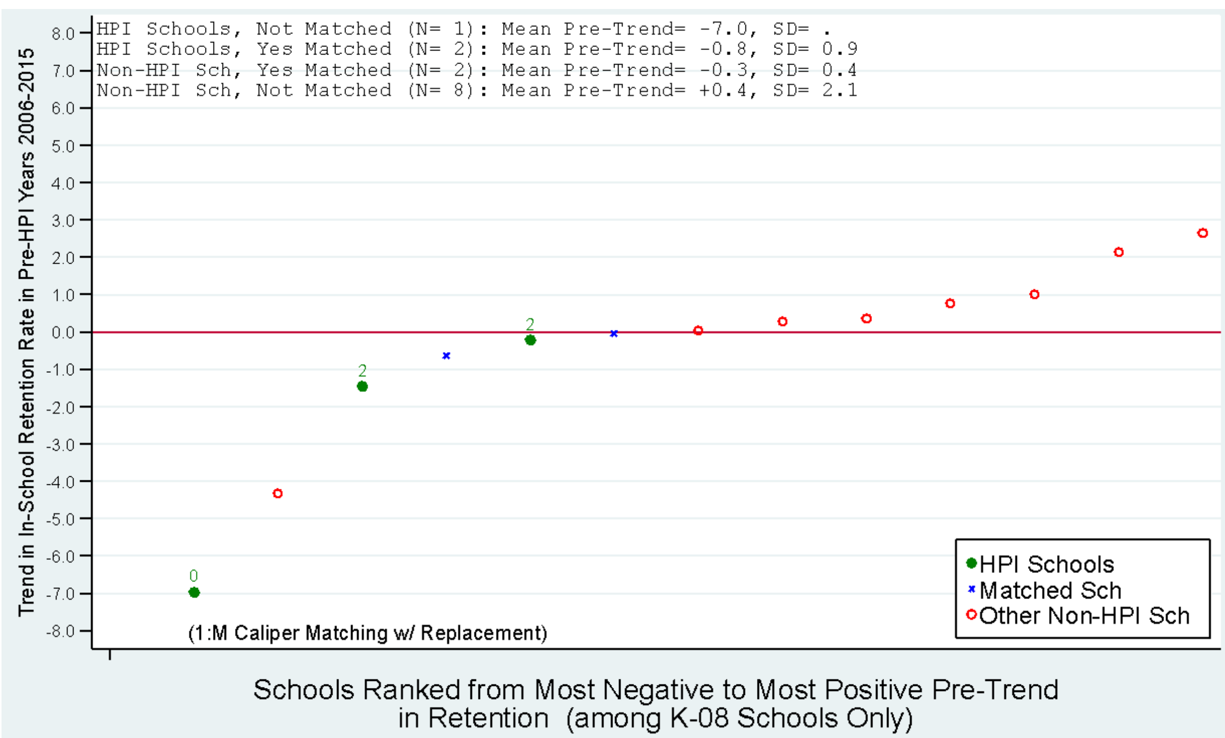
Appendix Figure C7: Elementary Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015



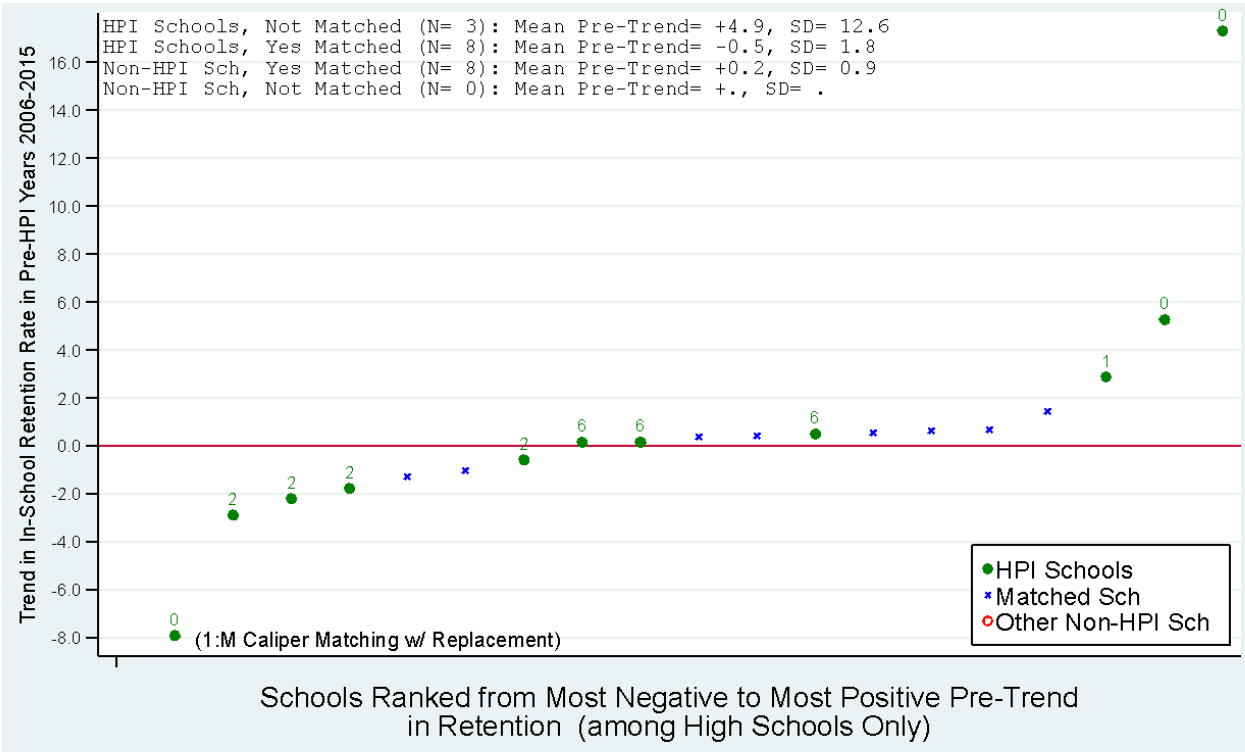
Appendix Figure C8: Middle Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015



Appendix Figure C9: K-08 Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015



Appendix Figure C10: High Schools: Ranked on Trend in In-School Retention Rate in Pre-HPI Years 2006-2015

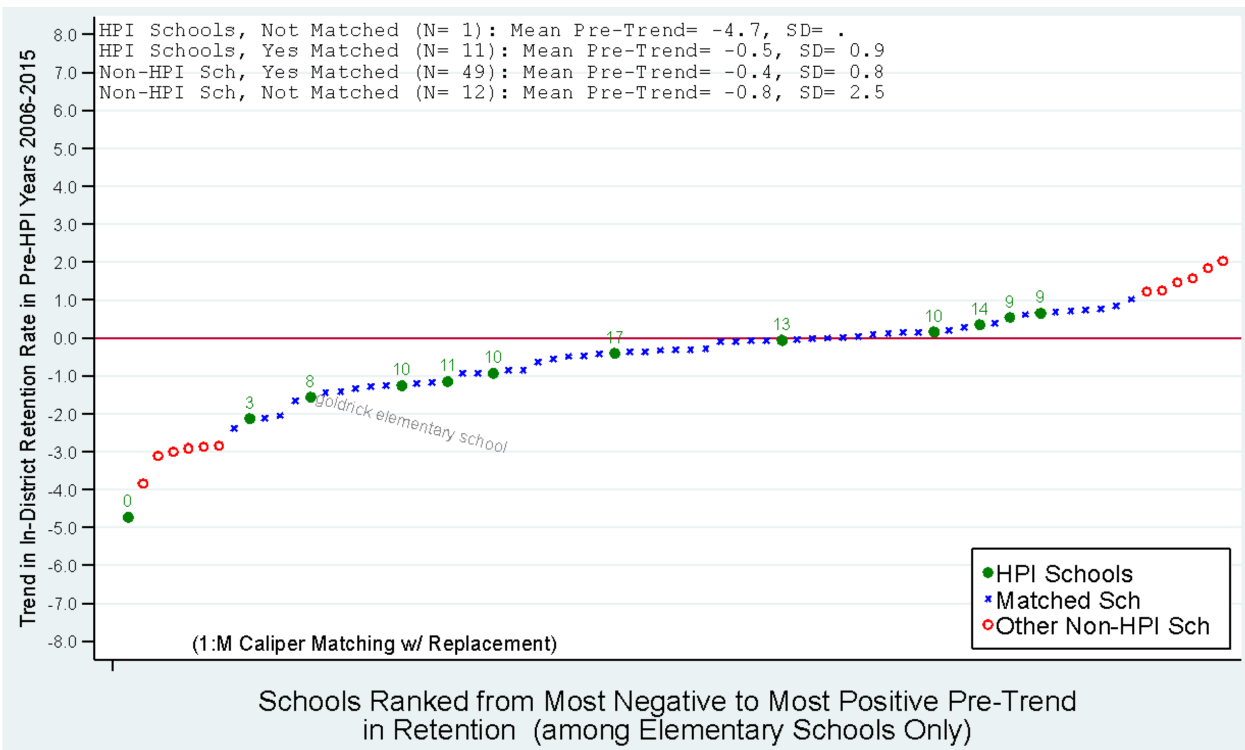

HPI Schools, Not Matched (N= 3): Mean Pre-Trend= +4.9, SD= 12.6
HPI Schools, Yes Matched (N= 8): Mean Pre-Trend= -0.5, SD= 1.8
Non-HPI Sch, Yes Matched (N= 8): Mean Pre-Trend= +0.2, SD= 0.9
Non-HPI Sch, Not Matched (N= 0): Mean Pre-Trend= +., SD= .
Trend in In-School Retention Rate in Pre-HPI Years 2006-2015
(1:M Caliper Matching w/ Replacement)
HPI Schools
Matched Sch
Other Non-HPI Sch
Schools Ranked from Most Negative to Most Positive Pre-Trend
in Retention (among High Schools Only)

Appendix Table C4: In-District Teacher Retention: M:1 Caliper Matching with replacement:

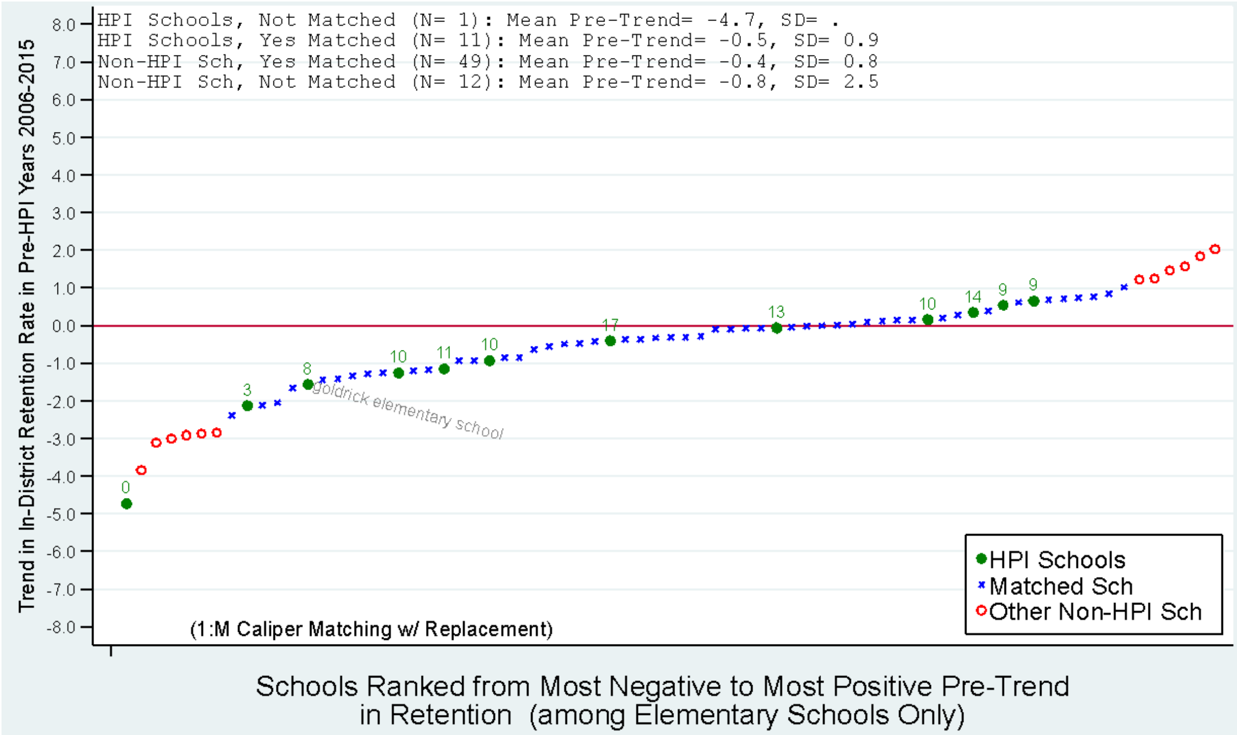
In-District Teacher Retention: Number of Matches Using Caliper Method with Replacement

	HPI Schools		Non-HPI Schools		
	At Least 1 Match Found	No Matches Found	Used as a Match	Not Used as a Match	All Schools
Elementary	11	1	49	12	73
Middle	6	1	9	1	17
K-08	2	1	2	8	13
High	8	3	8	0	19
All Schools	27	6	68	21	122

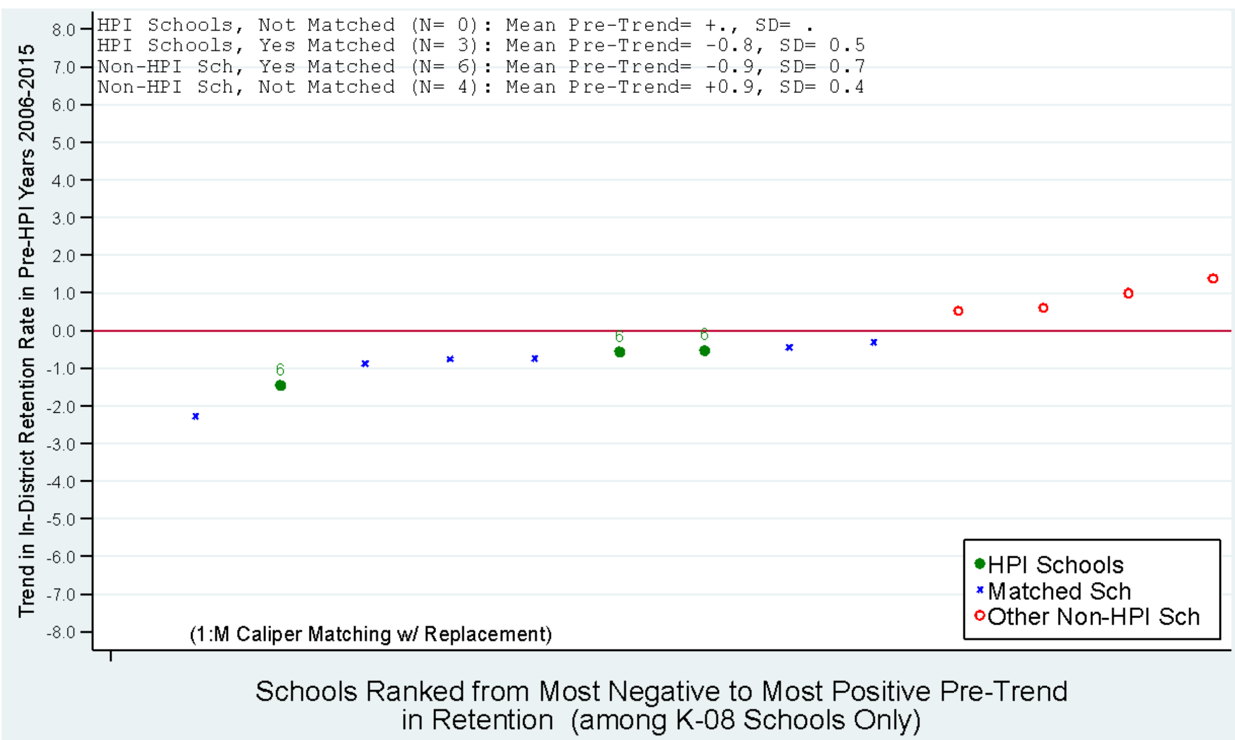
Appendix Figure C11: Elementary Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015



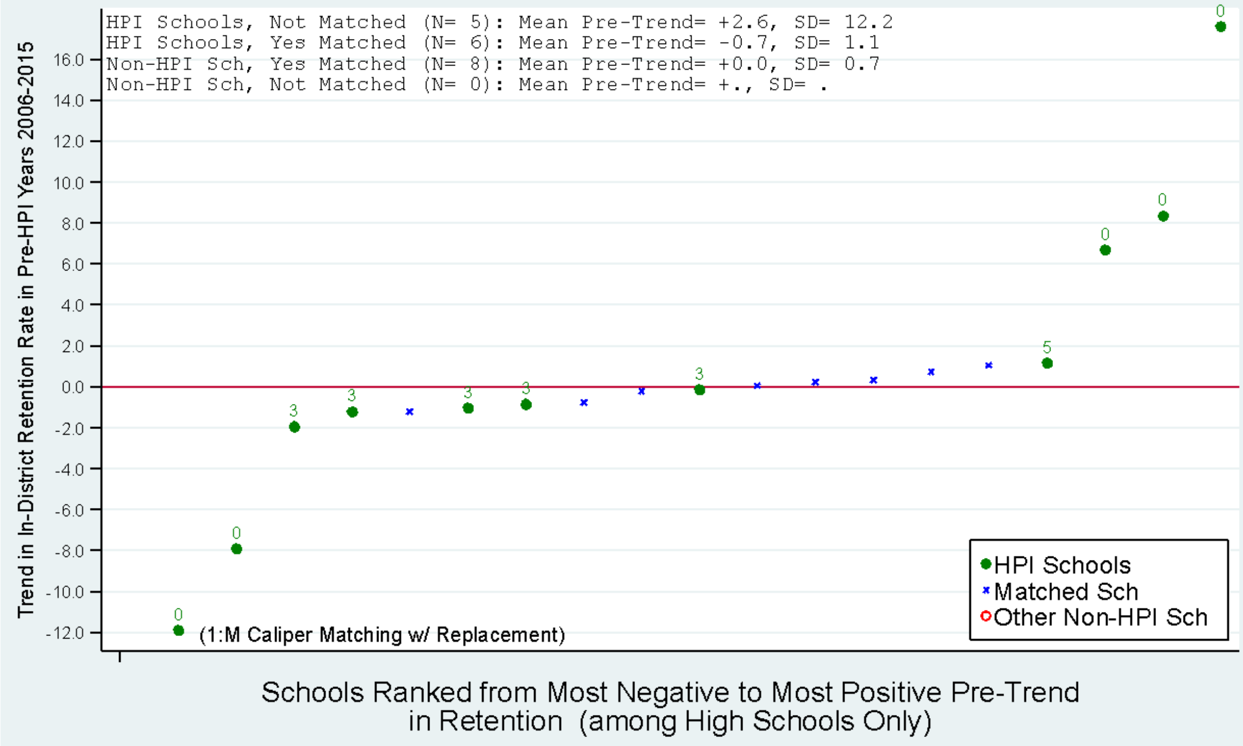
Appendix Figure C12: Elementary Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015



Appendix Figure C13: K-08 Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015



Appendix Figure C14: High Schools: Ranked on Trend in In-District Retention Rate in Pre-HPI Years 2006-2015



Appendix D: CITS Analysis using Propensity Score Matching (PSM)

HPI schools differ systematically from their non-HPI counterparts, as HPI schools were selected because they were considered to be among the highest needs schools in DPS. Because of the inherent differences between HPI and non-HPI schools, constructing a comparable group (the comparison group) consisting of non-HPI schools that are similar to HPI schools in as many ways as possible *aside* from their HPI status proved to be a challenge. Because of this, our preferred methodological approach, featured in the main body of the report, matches on pre-HPI trends in teacher retention. We did, however, also attempt to use propensity score matching (PSM). For PSM, we estimated the probability of being designated HPI (the propensity score), controlling for a range of characteristics, including pre-intervention trends in school demographics, enrollment, FTE, and teacher retention.

Once propensity scores are estimated, a comparison group is constructed by matching HPI and non-HPI schools with similar propensity scores. There are several methods by which this can be accomplished, each resulting in a somewhat different selection of schools comprising the matched group. For our analysis, we used two different matching algorithms: k-nearest neighbor matching and radius matching, each of which we specified multiple times using different matching criteria. The varied combinations of selection model and matching algorithm we tested resulted in up to 64 different permutations through which matched comparison groups could be identified. This process was completed independently for elementary, middle, K-8, and high schools.

In order to sufficiently reduce bias when using PSM to construct the comparison group for CITS, the matched comparison group must satisfy two important requirements. First, the distribution of propensity scores across HPI and non-HPI comparison schools must be as similar as possible. Second, balance must be achieved in terms of the means and distributions of each group's measured independent variables. When these requirements are not met, we cannot assume that comparison schools are sufficiently similar to their HPI counterparts to yield unbiased estimates of program impacts. In attempting to use PSM for our CITS analyses, we were unable to construct a sufficiently similar non-HPI comparison group due to several factors, including relatively small sample sizes and the fact that HPI and non-HPI schools tend to differ substantially from each other. Appendix Table D1 illustrates the former challenge, displaying total counts of traditional, district-run schools at each grade level as well as the number of schools available for matching after eliminating those with insufficient years of pre- or post-intervention data. As the table shows, the sample size of schools that remain are too small to allow for viable statistical analyses.

Appendix Figure D1 reports the balance of pre-intervention covariate trends across all HPI and non-HPI elementary schools. For each covariate, a standardized bias of less than 25% would indicate balance (Ho et al., 2007). As the figure indicates, the majority of covariates do not achieve this level of balance. A further illustration of how HPI and non-HPI schools differ substantially is shown in Appendix Figure D2. This figure shows the very limited overlap between HPI and non-HPI schools' propensity scores, including a set of key covariates. This small degree of overlap indicates that most non-HPI schools had a very low likelihood of having been selected as HPI, providing support for the argument that HPI schools were, in fact, substantively different from their non-HPI counterparts (a likely goal of the HPI program). However, given that successful PSM requires substantial overlap to allow for comparing HPI schools to those that would have the *same likelihood of being HPI* (but are not), this figure illustrates the central reason that we do not present results from PSM CITS in the main body of the report; propensity matching is not a viable method for the CITS analytic approach in this case.

Our analysis specifies up to 64 different permutations of selection model and matching algorithm and evaluates each in its ability to address the challenges noted above: balance in covariates, similarity of propensity scores distributions, and sufficient number of schools in each group. Because the criteria typically recommended in literature (which we refer to here as "strict") resulted in very few viable model permutations, we used adjusted criteria and denoted whether each permutation met the "strict" or "adjusted" cutoffs. The majority of permutations did not meet either set of criteria, and those that did included few covariates. This is antithetical to the goals of PSM: It is preferable to use as many relevant covariates in the PSM models as possible to help assure that, once covariate balance is achieved, the matched comparison groups would be as similar as possible to the HPI group on as many observable characteristics as possible, aside from HPI status. Again, however, our PSM efforts did not succeed in finding matched comparisons in models with key covariates. In fact, for middle and high schools, only pre-intervention retention rates could be included in the PSM before the model failed to produce viable matches. For K-8 schools even that simple model was problematic.

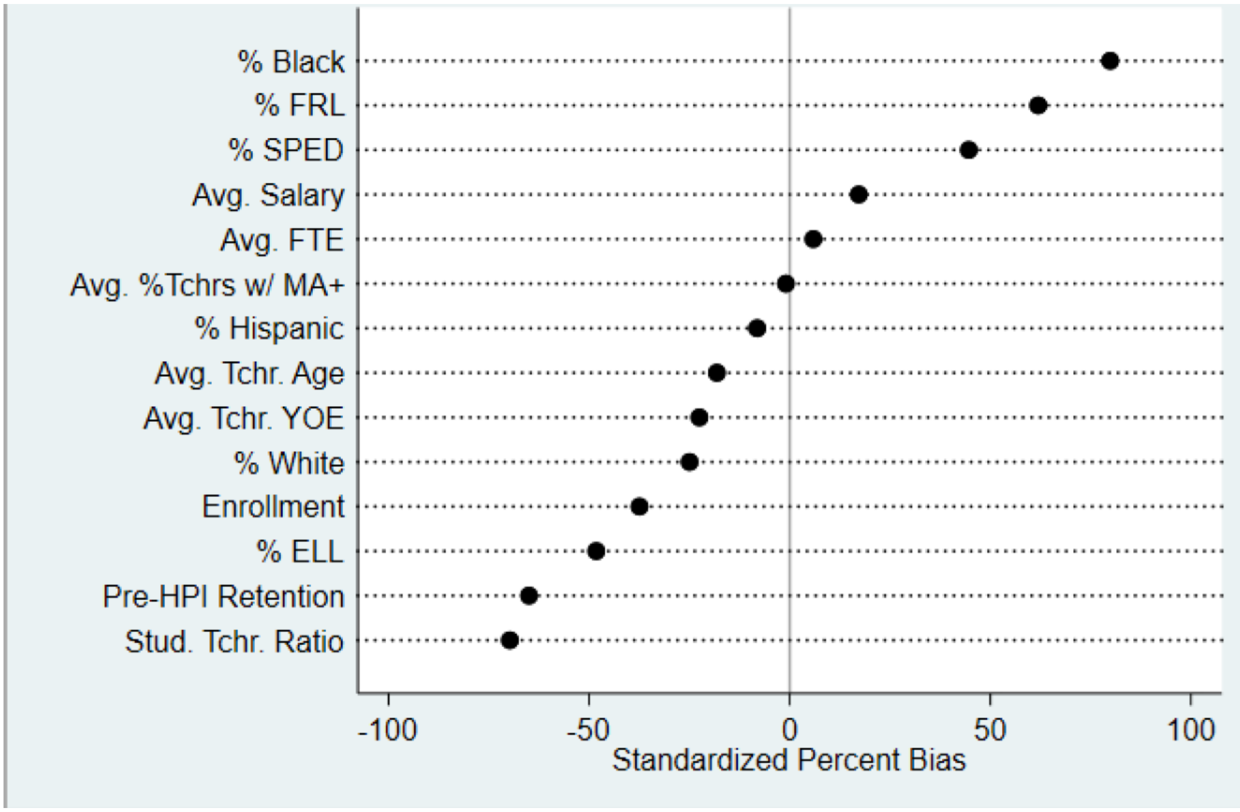
Using the matched HPI and non-HPI comparison groups from each of the qualifying PSM permutations, we specified a CITS for each to estimate the coefficients of interest for evaluating the effectiveness of the HPI program. Appendix Table D2 describes selection models and matching algorithms that met the balance criteria and reports the corresponding shock and trend effects, the coefficients of interest. As the table shows, no shock effects were statistically significant and trend effects were only significant at the $p<.05$ level in three elementary models. Both imprecision and small estimated effects with very few estimates indicating statistical significance indicate null results, similar to results presented in the main text for our preferred models.

Appendix Table D1: Schools Available for Matching, by Level and HPI Status

	HPI Schools		Non-HPI Schools	
	Total	Avail. For Matching	Total	Avail. For Matching
Elementary	13	12	82	62
Middle	11	9	24	11
K-8	3	3	12	10
High	11	11	18	8
All Schools	38	35	136	91

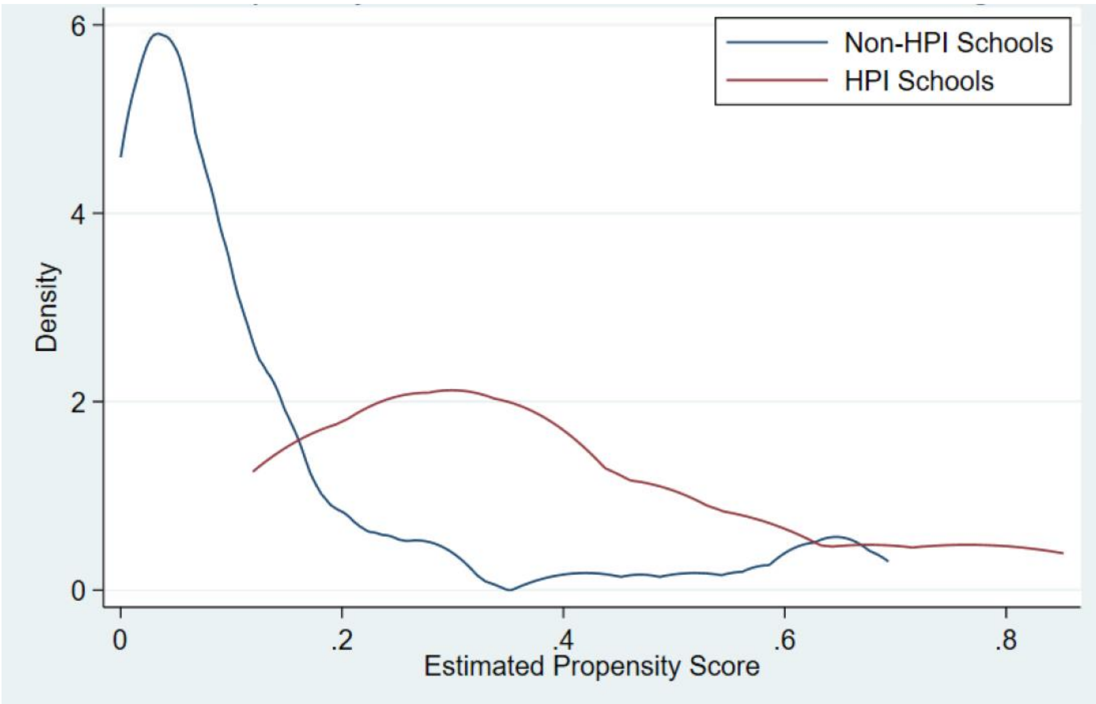
Note. Table includes traditional DPS-managed schools. Schools may be unavailable for matching due to insufficient data before or after the implementation of HPI. In order to have sufficient data, a school must have been open for at least two years prior to and at least two years after the 2015-16 school year.

Appendix Figure D1: Balance in Covariate Trends Across HPI and Non-HPI Schools Before Matching



Note. Covariates are typically considered balanced when percent standardized bias is 25 or less. All covariates are represented as pre-HPI trends from 2006 through 2015.

Appendix Figure D2: Propensity Score Distributions Before Matching



Note. Propensity scores estimated from selection model 4, elementary schools only.

Appendix Table D2: CITS Effects Estimates and PSM Balance Statistics for Selected Permutations of Selection Model and Matching Algorithm, by School Level

CITS Effects Estimates and PSM Balance Statistics for Selected Permutations of Selection Model and Matching Algorithm, by School Level

Matching Algorigthm	Selection Model	Shock Effect	Trend Effect	Parallel Trends	Balance Criteria	N HPI	N Matched	Mean % Bias	Median % Bias	Rubin's B	Rubin's R	Covariates Included in Model
Elementary Schools												
Radius Matching, Caliper=.03	2	-0.107	0.040 *	No	Adjusted	9	42	17.7	15.0	48.4	0.324	Retention + SCI
Radius Matching, Caliper=.03	4	-0.054	0.026	No	Adjusted	10	17	4.2	2.4	31.6	0.622	Retention + enrollment
Radius Matching, Caliper=.05	2	-0.118	0.044 *	No	Adjusted	9	49	10.8	7.0	35.1	0.264	Retention + SCI
Radius Matching, Caliper=.05	7	-0.059	0.030	Yes	Adjusted	12	23	18.9	17.8	52.3	0.665	Retention + SCI + race
Radius Matching, Caliper=.07	2	-0.116	0.045 *	No	Adjusted	9	54	14.4	10.4	39.1	0.288	Retention + SCI
Radius Matching, Caliper=.07	7	-0.060	0.030	Yes	Adjusted	12	28	18.2	15.3	52.7	0.554	Retention + SCI + race
K Nearest Neighbor, K=2	3	-0.095	0.041	No	Adjusted	11	18	23.5	21.2	56.7	0.429	Retention + race
K Nearest Neighbor, K=3	2	-0.069	0.016	Yes	Adjusted	10	19	15.2	16.9	51.5	0.278	Retention + SCI
K Nearest Neighbor, K=3	7	-0.088	0.037	Yes	Adjusted	12	16	13.8	11.3	52.0	0.629	Retention + SCI + race
K Nearest Neighbor, K=4	7	-0.085	0.039	Yes	Adjusted	12	17	18.5	15.1	53.7	0.436	Retention + SCI + race
Middle Schools												
Radius Matching, Caliper=.03	1	0.028	0.016	No	Strict	7	9	4.3	4.3	7.8	0.799	Retention only
Radius Matching, Caliper=.05	1	0.028	0.016	No	Strict	8	10	1.5	1.5	2.8	0.850	Retention only
Radius Matching, Caliper=.07	1	0.008	0.017	Yes	Strict	8	10	1.6	1.6	3.2	1.190	Retention only
K Nearest Neighbor, K=2	1	0.028	0.016	No	Strict	8	9	3.7	3.7	7.1	0.930	Retention only
K Nearest Neighbor, K=3	1	-0.027	0.014	No	Strict	8	10	1.5	1.5	3.1	1.194	Retention only
1:1 Nearest Neighbor	1	0.059	0.026	Yes	Adjusted	7	10	15.8	12.0	46.4	1.615	Retention + SCI
High Schools												
K Nearest Neighbor, K=2	1	0.138	-0.017	No	Strict	4	5	0.7	0.7	4.8	0.324	Retention only
K Nearest Neighbor, K=3	1	0.147	-0.006	No	Strict	4	6	2.9	2.9	22.3	0.412	Retention only
1:1 Nearest Neighbor	1	0.099	-0.019	No	Strict	4	3	0.0	0.0	0.0	0.372	Retention only

Note: Rubin's B = the absolute standardized difference of the means of the linear index of the propensity score in the treated and (matched) non-treated group. Rubin's R = the ratio of HPI to (matched) non-HPI variances of the propensity score index. Rubin (2001) recommends that B be less than 25 and that R be between 0.5 and 2 for the samples to be considered sufficiently balanced. We use these values along with mean % bias of 25 or below as "strict" balance criteria. "Adjusted" balance criteria allows B to be 60 or less but retains a mean % bias cutoff of 25. SCI variables denote those used to construct the School Characteristic Index (excluding volatility). "Parallel Trends" refers to whether or not the CITS model resulting from the PSM specifications satisfies the assumption that pre-intervention trends in the outcome of interest are parallel.

Appendix E: Regression Discontinuity Analysis

We explored whether a regression discontinuity (RD) design was a feasible means for estimating HPI effects. RD is a quasi-experimental method for causal inference that is viable when a continuous variable with a specific and known cutoff is used to determine assignment to "treatment" (i.e., HPI) or "control" (i.e., non-HPI) conditions. For HPI, our analyses indicate that the cutoff for being an HPI school was the tenth highest SCI score for both elementary and secondary schools. RD allows for the estimation of causal effects under the correct conditions because of this known cutoff. Schools that were just below or just above the SCI cutoff are, by design, going to be very similar in terms of SCI scores. Thus, the particular area, or bandwidth, of SCI scores just above and just below the cutoff for HPI designation results in a set of schools with very similar SCI scores, some of which were designated HPI schools and some of which were not. For example, the SCI cutoff for HPI designation for elementary schools was 56.4. For RD, schools just below and just above that cutoff are considered to be "as good as random" with regard to whether they were designated HPI or not, as there is little to no difference between a school with an SCI score of 56.4 (designated HPI) and a school with an SCI score of 56.3 (not HPI). Comparing outcomes for schools just above and just below the SCI cutoff could allow for the estimation of causal effects of the HPI program.

Only schools that were designated HPI based solely on SCI scores can be included in the RD analysis. Thus, this method is only applicable to the approximately 20 schools that were designated HPI using SCI scores. The HPI schools designated by DPS administration using additional information cannot be included in the RD, as the strength of this design is that it allows for the selection of a comparison group that is statistically equivalent to the treatment group through the use of a cutoff score. In this case, the comparison group would be schools just below the SCI cutoff score used to designate HPI schools.

In the case of HPI, we are unable to estimate the causal effects of the program using RD because of the relatively small number of schools (approximately 20) that were assigned to the program based on SCI scores. This small number, or small sample size, results in a lack of statistical power to detect program effects. Given the small sample size, the RD design would only work to detect very large program effects: around 6 percentage points difference between HPI and non-HPI schools in teacher retention, which would be considered a very large effect in terms of educational interventions and particularly in the case of incentive pay programs. Because HPI designation was at the school-level, teacher-level analysis does little to increase statistical power. Thus, RD analyses would be very unlikely to detect an effect of HPI unless that effect was very large.

Despite these limitations, we conducted a number of variations of RD-based analyses of the effects of HPI on teacher retention. However, as anticipated due to the small sample size and low statistical power, the results we estimate of the HPI program effects on In-School teacher retention are all null.

Data and Key Variables

While the data for this analysis come from the same source as the data for the primary CITS analysis, there are some specific variables and some changes to the analytic sample that are worth noting.

The SCI Index Score: The Potential Rating Variable

SCI is calculated separately for elementary, middle, and high schools (schools that span multiple levels, like K-8 schools, have multiple SCI values). Given that secondary schools were considered together for HPI selection purposes, a school's secondary SCI score is calculated as the average of its middle and high school SCI, where applicable.

To maximize statistical power, all eligible schools are included in the analysis sample. This required some rescaling of the SCI variable. Because 2013-2014 SCI was used for elementary school HPI designation, with one cutoff, and 2014-2015 SCI was used for secondary schools, with a different cutoff, we made the following adjustments. The correct years were used and the level-specific cutoffs were subtracted from SCI before the SCI scores were combined into a single SCI variable. For schools that spanned both elementary and secondary levels such as K-8 schools, the elementary and secondary scores were averaged. The final SCI variable is greater than or equal to zero for SCI-based HPI schools and negative for discretionary HPI schools or non-HPI schools. K-8 schools could qualify for HPI as either elementary or secondary schools.

Analysis Sample

HPI selection considers co-located schools as single schools, as the SCI rule was applied to co-located schools together. Because RD relies on the use of a cutoff to determine HPI selection, only schools selected using SCI scores alone are included. Math and Science Leadership Academy and West High School, both schools with SCI scores above the cutoff but which were not HPI schools due to low enrollment, are excluded, as are discretionary HPI schools. Further, only schools with pre-HPI SCI scores are included in the analysis. The two HPI K-8 schools are above the cutoffs for both elementary and secondary SCI scores, and thus "double-qualify". Because of this, there are only 18 SCI-based HPI schools. With these restrictions, the analytic sample includes 105 schools.

In-School Teacher Retention Rates

The outcome variable of interest remains In-School teacher retention. We can consider in-school retention for any year after the HPI program was implemented, which provides four years of post-program retention data for each school in the analysis sample. Because a school's teacher retention rate is likely correlated across years, standard errors are clustered at the school level.

For the school-level analysis, the outcome variable is the percentage of teachers retained into the following year. For the teacher-level analysis, the outcome variable is binary with a value of 1 if that teacher stays and 0 if that teacher exits the school.

Methodological Approach

The RD design leverages the difference between schools just above and just below the SCI cutoff. The equation below is used for the school-level analysis and models the school-level average teacher retention for school *s* in year *y* after the implementation of the program.

$$Model\ 1{:}\ Retention_{sy} = \beta_0 + \beta_1(SCI_s) + \beta_2(HPI_{sy}) + \beta_3(SCI_s \times HPI_{sy}) + S_{s(y)}\beta + \varepsilon_{sy}$$

In the above equation, $SCI_s$is the SCI variable described above, with the cutoff at zero, and $HPI_{sy}$is a binary variable indicating whether or not a school is an HPI school. By interacting SCI_s and HPI_{sy}, we allow the relationship between SCI and teacher retention to differ between HPI and non-HPI schools. Thus $\beta_1$represents how teacher retention changes with changes in SCI for non-HPI schools, and $\beta_3$represents the difference in the retention/SCI relationship between HPI and non-HPI schools. $\beta_0$represents the predicted value of teacher retention whenSCI_s and $HPI_{sy}$are both zero: the outcome value at the cutoff for non-HPI schools. $\beta_2$is the coefficient of interest, as it tells us the difference between the predicted value at the cutoff for non-HPI and HPI schools. This is the “difference” at the cutoff that is attributable to HPI.

$S_{s(y)}$represents a vector of time-variant and time-invariant school-level covariates to help 1) account for the influence of other factors on teacher retention, and 2) increase the precision of our estimates. These covariates include a school’s pre-HPI teacher retention for the 12/13 - 14/15 school years, whether a school was ever a top-performing school across the 12/13 - 14/15, indicators for school level, enrollment, and percentages of Hispanic, White, Black, and FRPL-eligible students.

The model below, which is similar, is at the teacher level.

$$Model\ 2{:}\ Retained_{psy} = \beta_0 + \beta_1(SCI_s) + \beta_2(HPI_{sy}) + \beta_3(SCI_s \times HPI_{sy}) + S_{s(y)}\beta + X_{p(y)}\beta + \varepsilon_{psy}$$

In this case the outcome, $Retained_{psy}$, is a binary indicator of whether a teacher was retained in their school (1) or not (0). The other variables are interpreted in the same way as they were for Model 1. The main difference is the ability to add $X_{p(y)}$, a vector of teacher-level covariates. These covariates include teacher race, teacher sex, teacher age (and a quadratic age term), TLC mentor and mentee tags, an indicator for whether a teacher has a master’s degree or higher, and a teacher’s in-district years of experience (including a quadratic experience term).

The teacher model shown is a linear probability model. We estimated both a linear probability and a logistic model and found that the results were substantively similar. We therefore show results of the linear probability model, as these are more easily interpreted.

RD Method: Sensitivity to Choice of Bandwidth

The causal estimate only uses the difference at the cutoff ($\beta_2$above) but uses schools with SCI values far from the cutoff to establish trends in SCI ($\beta_1$and β_3 above) which can affect the predicted values of the cutoff. If the SCI/retention relationship is very different at values that are far from the SCI cutoff, this could bias estimates. Because of this, each analysis is performed at a variety of bandwidths (BW) that determine which schools are close enough to be considered in the analysis. The widest bandwidth (60) includes all schools, while the smallest bandwidth (5) only considers schools with SCI values within 5 points of the cutoff. All bandwidths from 5 to 60 include all HPI schools but include different numbers of non-HPI schools as the maximum (transformed) SCI value is 5, while the minimum is -53.

RD Results: School-Level Analysis of In-School Teacher Retention

Appendix Figure E1 shows school-level post-HPI teacher retention for HPI and non-HPI schools according to their pre-HPI SCI value, with linear and quadratic lines of best fit. Each point on the graph represents a school-year observation, such that each school value can have a maximum of four observations on the graph. Linear and quadratic lines are shown in order to determine the best functional form for estimating the relationship between SCI and teacher retention. The figure indicates that the relationship is linear. The area of causal interest in the figure is at the cutoff line separating HPI and non-HPI schools. A discontinuity at the line, between very similar non-HPI and HPI schools on either side of the cutoff, would indicate an HPI effect. The figure shows a possible, very small negative effect, indicating that HPI schools have slightly lower teacher retention than similar non-HPI schools. However, this difference is indistinguishable from zero, or no effect.

Results of the linear regression at 5 different bandwidths are shown in Appendix Table E1. The main coefficients of interests are those on HPI, indicating no statistically significant effects. For example, at a bandwidth of 60 (including all schools), HPI schools just above the cutoff have retention at 2.21 percentage points lower than non-HPI schools just below the cutoff—but the standard errors (shown below the coefficients in parentheses) are large enough that this result is not statistically distinguishable from zero.

The same analysis was performed with the addition of school-level covariates, with results shown in the Appendix Table E2. Once again, the coefficient on HPI remains statistically insignificant at all bandwidths.

RD Results: Teacher-Level Analysis of In-School Teacher Retention (All Teachers)

The teacher-level analysis is very similar to the school-level analysis. The results of the linear probability model are shown in Appendix Table E3 for ease of interpretability (a logistic model was also estimated and produced substantively similar results). In this case, the coefficient on HPI describes the change in a teacher's probability of staying in their school. For example, using the largest bandwidth, this analysis suggests that a teacher at an HPI school is 0.48 percentage points more likely to stay in their school than a similar teacher in a similar non-HPI school. This result is not statistically distinguishable from zero, nor are results estimated at any bandwidth.

RD Results: Teacher-Level Analysis of In-School Teacher Retention (By Teacher LEAP Rating)

The dollar amount of the HPI bonus is determined by a teacher's most recent LEAP score (see Table 1 in main report). Thus, it is possible that HPI effects on teacher retention could vary by LEAP score categories. One way to disaggregate results is to use LEAP ratings from year y-1, as that determines the bonus in year y, and the logic of HPI suggests that receiving the bonus could cause teachers to want to remain in their school. An advantage of this approach is that teachers know their LEAP rating from year y-1 when deciding whether to stay in their school for year y+1. The disadvantage is that decisions to stay may be more likely to be based on what teachers expect to be paid in next year as opposed to what they received in the current year.

An alternative is to disaggregate results by LEAP rating from the same year as the retention variable. That LEAP score determines HPI incentive pay for the following year. While teachers may not know their LEAP score for the year in which they are determining whether to remain in their school, they will likely anticipate their scores based on ongoing evaluation. For this reason, we disaggregate retention results by the LEAP score of the same year (meaning, for example, analyzing how a teacher's decision to stay from the 2015/2016 school year to the 2016/2017 school year might change depending on their 2015/2016 LEAP score).

The statistical power of the analysis, due to the small sample size, was already a problem when considering all teachers. That problem only increases when considering teachers separately by LEAP category, as there are even less teachers in each analysis. Due to the small sample size and lack of statistical power, we find no pattern of significant results. There is no indication from this analysis that the HPI program had an effect on teachers in any LEAP category.

Appendix Table E1. RD results for the school-level linear regression with no covariates

	BW = 60	BW = 40	BW = 20	BW = 10	BW = 5
SCI	-0.0020 **	-0.0007	-0.0008	-0.0001	0.0210
	(0.0006)	(0.0013)	(0.0019)	(0.0038)	(0.0196)
HPI	-0.0221	-0.0358	-0.0325	-0.0375	-0.0718
	(0.0344)	(0.0362)	(0.0373)	(0.0394)	(0.0475)
SCI x HPI	0.0065	0.0052	0.0053	0.0046	-0.0166
	(0.0104)	(0.0105)	(0.0106)	(0.0112)	(0.0222)
N	411	359	271	206	154
R2	0.081	0.017	0.014	0.014	0.027

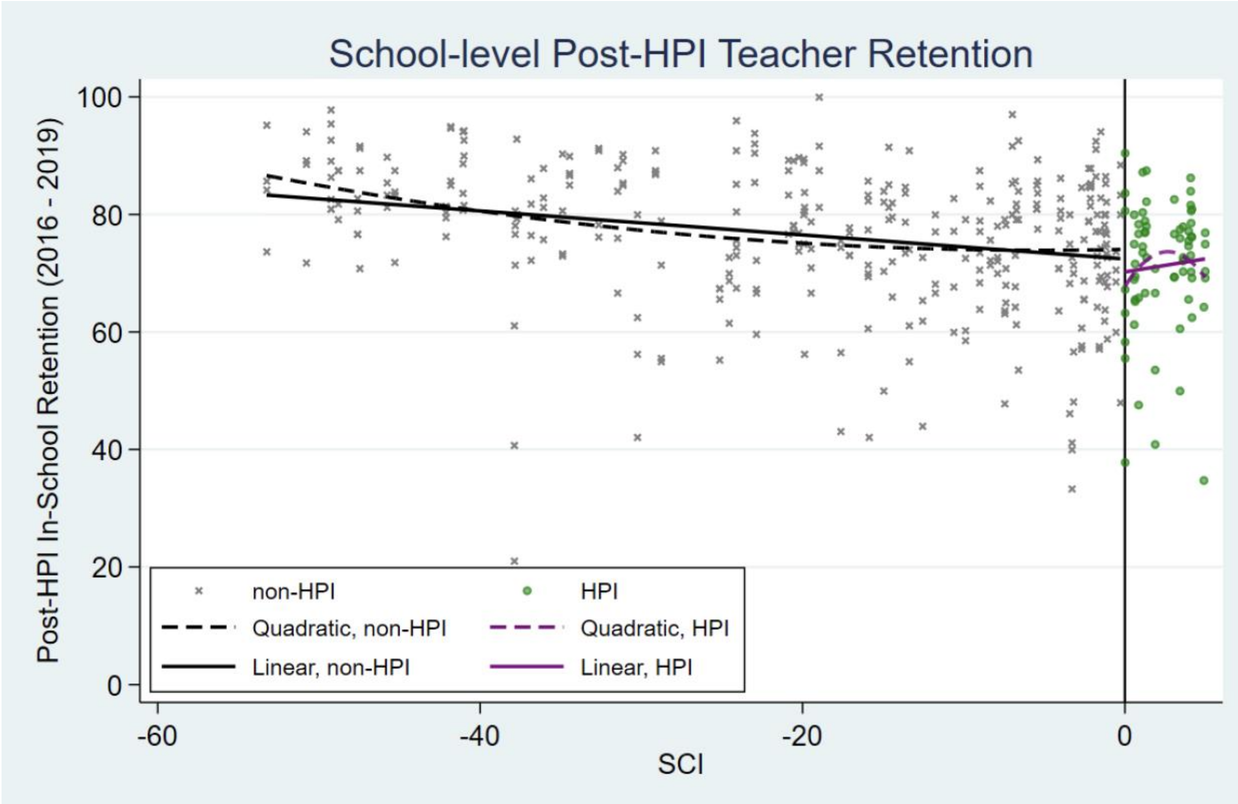
Appendix Table E2. RD results for the school-level linear regression with school-level covariates

	BW = 60	BW = 40	BW = 20	BW = 10	BW = 5
SCI	0.0051 *	0.0057 *	0.0049	0.0061	0.0379
	(0.0021)	(0.0022)	(0.0028)	(0.0058)	(0.0217)
HPI	-0.0066	-0.0148	-0.0120	-0.0103	-0.0689
	(0.0289)	(0.0287)	(0.0336)	(0.0310)	(0.0408)
SCI x HPI	-0.0020	-0.0026	-0.0019	-0.0050	-0.0375
	(0.0075)	(0.0074)	(0.0080)	(0.0087)	(0.0213)
N	374	322	246	185	145
R2	0.350	0.291	0.237	0.263	0.291

Appendix Table E3. RD results for the teacher-level linear regression with school- and teacher-level covariates

	BW = 60	BW = 40	BW = 20	BW = 10	BW = 5
SCI	0.0046 *	0.0052 *	0.0027	0.0069	0.0299
	(0.0019)	(0.0021)	(0.0025)	(0.0049)	(0.0182)
HPI	0.0048	-0.0045	-0.0047	-0.0160	-0.0636
	(0.0269)	(0.0273)	(0.0293)	(0.0294)	(0.0406)
SCI x HPI	-0.0027	-0.0022	0.0027	-0.0024	-0.0262
	(0.0068)	(0.0066)	(0.0066)	(0.0072)	(0.0181)
N	14192	12373	9304	6826	5465
R2	0.056	0.052	0.046	0.049	0.053

Appendix Figure E1. School-level post-HPI teacher retention for HPI and non-HPI schools according to their pre-HPI SCI value



Note: Linear and quadratic lines of best fit are shown for both HPI and non-HPI schools. For RD results to be valid, the functional form of the lines must represent the data well. We can see in this graph that the true relationship is more linear than quadratic.

Appendix F: CollaboRATE Survey Items

	CollaboRATE "Engagement" Survey Survey Question	Engagement Survey (X = Question Included in Survey-Year) 2017/18	2018/19
	Indicate how much you agree or disagree that [School or Department] is effective at demonstrating each of the Shared Core Values.		
	Students First: We put our kids' needs at the forefront of everything we do.	X	X
	Integrity: We tell the truth, and we keep our promises.	X	X
	Equity: We celebrate our diversity and will provide the necessary resources and supports to eliminate barriers to success and foster a more equitable future for all our kids.	X	X
	Collaboration: Together as a team, we think, we work, and we create in order to reach our goals.	X	X
	Accountability: We take responsibility for our individual and collective commitments; we grow from success; we learn from failure.	X	X
	Fun: We celebrate the joy in our work and foster in our students a joy and passion for learning to last their whole lives.	X	X
	Indicate how much you agree or disagree that DPS as a whole is effective at demonstrating each of the Shared Core Values.		
	Students First: We put our kids' needs at the forefront of everything we do.	X	X
	Integrity: We tell the truth, and we keep our promises.	X	X
	Equity: We celebrate our diversity and will provide the necessary resources and supports to eliminate barriers to success and foster a more equitable future for all our kids.	X	X
	Collaboration: Together as a team, we think, we work, and we create in order to reach our goals.	X	X
	Accountability: We take responsibility for our individual and collective commitments; we grow from success; we learn from failure.	X	X
	Fun: We celebrate the joy in our work and foster in our students a joy and passion for learning to last their whole lives.	X	X
	Please answer the following questions thinking about [School or Department].		
Engagement Index (School/ Dept.)	I enjoy my work at [School or Department]	X	X
Engagement Index (School/ Dept.)	I feel valued as an employee of [School or Department]	X	X
Engagement Index (School/ Dept.)	I would recommend [School or Department] to others as a good place to work.		X
Engagement Index (School/ Dept.)	My job has a positive impact on [School or Department].		X
Engagement Index (School/ Dept.)	I am proud to tell people I work for [School or Department].		X
	My workload is sustainable.	X	X
	I have a clear understanding of what is expected of me at work.	X	X
	I have the tools necessary to do my job effectively.	X	X
	I am involved in decisions that affect my work in [School or Department].	X	X
	My feedback is used to drive improvements in [School or Department].	X	X
	The people I work with are willing to help each other, even if it means doing something outside their usual activities.	X	X
	On our team, we feel responsible for each other's success.	X	X
	Employees in [School or Department] willingly provide candid and direct feedback to each other.	X	X
	[School or Department] is an inclusive place to work.	X	X

	Please answer the following questions thinking about DPS as a whole. Indicate how much you disagree or agree with each of the statements below.		
Engagement Index (District)	I enjoy my work at DPS.	X	X
Engagement Index (District)	I feel valued as an employee of DPS.	X	X
Engagement Index (District)	I would recommend DPS to others as a good place to work.	X	X
Engagement Index (District)	My job has a positive impact on DPS.	X	X
Engagement Index (District)	I am proud to tell people I work for DPS.	X	X
	I have the opportunity for growth and development at DPS.	X	X
	DPS cares about the personal well-being of its team members. (Well-being includes your physical, mental, and social health.)	X	X
	I find my job to be challenging and interesting.	X	X
	I believe in the DPS Shared Core Values.	X	X
	The district leadership of DPS has communicated a vision of the future that is motivating to me.	X	X
	I can see a clear link between my work and the top DPS priorities.	X	X
	The top priorities for DPS are likely to drive student achievement.	X	X
	At DPS, diversity and inclusiveness are appreciated and encouraged.	X	X
	No instruction included.		
	What are the top three reasons you continue to work for DPS?	X	X
	From this point on, how long do you see yourself working at DPS?	X	X
	What are the top two reasons that might make you consider leaving?	X	X
	What are the top actions DPS should take to make this a great place to work?	X	X
	Is there anything DPS can do to help you feel more valued?	X	X
	System questions.		
	School or Department	X	X
	DPSID	X	X

CollaboRATE Leader Surveys *(Respondents: Teachers)*	**Teacher Leader Survey (X = Included)**		**Principal Survey, AP Survey, LEAD Survey (X = Included)**	
Survey Question	**2017/18**	**2018/19**	**2017/18**	**2018/19**
Honors cultures and diverse backgrounds	X	X	X	X
Values and respects the students in my school	X	X	X	X
Empowers me to do my job	X	X	X	X
Ensures that students achievements are celebrated in our school	X		X	
Creates an environment that fosters passion for learning in students	X	X	X	X
(2018) Takes action to identify and close equity gaps at our school / (2019) Identifies and takes action to close equity gaps at our school	X	X	X	X
Prioritizes diversity and inclusiveness	X		X	
Values me	X	X	X	X
Is accessible to me	X	X	X	X
(2018)Resolves conflict effectively / (2019)Finds a way past conflict to a positive outcome	X	X	X	X
Is a person whom I trust	X	X	X	X
Trusts me	X		X	
Shares ownership and responsibility for achieving our schools goals	X	X	X	X
Models the DPS values	X	X	X	X
(2018) Learns and grows from feedback / (2019) Is reflective open to and grows from feedback	X	X	X	X
Is willing to make difficult decisions	X	X	X	X

	Remains positive in interactions and situations that may be frustrati	X		X	
	Motivates me	X		X	
	Does a good job keeping me informed about matters affecting me in my job	X		X	
	Holds our team accountable to high performance in an equitable manner	X			
	Holds all staff accountable to high performance in an equitable manner			X	
	Holds me accountable to high performance		X		X
	Ensures that teachers understand and consistently use student data to			X	
	Ensures that I understand and consistently use student data to drive effective instruction.	X	X		X
	Uses meaningful data and metrics to make decisions	X	X	X	X
	Ensures that I receive feedback and coaching that improve my job performance	X	X	X	X
	Ensures that I am supported in implementing best practice for English Language Acquisition.	X	X	X	X
	Promotes the use of available technology in the classroom	X		X	
	Supports me in implementing academic standards	X			
	Leads efforts to support teachers in the use of academic standards			X	
	Ensures that instructional time is prioritized			X	
	(2018) Appropriately prioritizes collaborative planning time / (2019) Implements systems to prioritize collaborative planning time	X	X	X	X
	Appropriately prioritizes individual planning time for teachers	X		X	
	Helps me solve problems	X		X	
	Ensures that clear systems and structures are in place to support a positive student culture.			X	X
	Is visible in the building			X	X
	Challenges existing processes in support of positive change	X		X	
	Sets clear goals for our team aligned to school goals	X			
	Sets clear goals for our school			X	
	Helps me feel heard in decision making at my school	X			
	Contributes to the improvements in my school through their role as a teac	X			
	Creates a welcoming environment and actively engages all families from diverse backgrounds and cultures.				X
	Communicates effectively about matters affecting me in my job		X		X
	Ensures that high quality classroom instructional strategies are implemented.		X		X
	Ensures high quality professional learning		X		X
	Regularly communicates progress towards our school's goals				X
	Regularly communicates progress towards our teams goals		X		
	Takes action on the most important strategic priorities for our school		X		
	Establishes and takes action on the most important strategic priorities for our school.				X
	Finds creative solutions to solve problems		X		
	Establishes strong partnerships with community organizations to mee			X	
	Encourages participation of families who are not native English speak			X	
	Actively engages parents			X	
	Overall is an effective leader	X	X	X	X
	What are your leader's strengths	X	X	X	X
	What can your leader improve to be more effective	X	X	X	X