Working Paper
September 8, 2023

Effect of Performance-based Compensation Systems on Principal Labor Markets

Hajime Mitani | *Rowan University*

This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation under NSF award NSF-DRL #1749275. Opinions reflect those of the author and do not necessarily reflect those AERA or NSF.

Prior research shows that high-need schools are more likely to be led by principals who are less qualified and less effective in many dimensions (e.g., Clotfelter et al., 2007; Grissom et al., 2019; Loeb et al., 2010; Papa et al., 2002). This principal sorting pattern raises concern among policymakers, parents, and local communities because it negatively affects student learning. For example, less effective principals usually visit classrooms for non-instructional purposes, provide little or no feedback for their teachers, do not encourage collaboration among teachers to enhance student learning, and execute irresponsible leadership (Oplatka, 2016; Wallace Foundation, 2013). These poor leadership practices deprive students of learning opportunities and lead to wider gaps in student learning. Given the importance of school leadership for school improvement (e.g., Branch et al., 2012; Grissom et al., 2021; Grissom et al., 2015; Hallinger & Heck, 1998; Murphy & Torre, 2014; Waters et al., 2003), it is pivotal for policymakers to rectify the inequitable distribution of principal quality.

The inequitable distribution is attributed mainly to the dearth of highly qualified applicants, differential turnover rates, and differential hiring and replacement across schools (e.g., Branch et al., 2012; Gates et al., 2006; Grissom et al., 2019; Roza, 2003). For example, in Tennessee, the mean principal turnover rate among low-poverty schools is only 17 percent, whereas it is 22 percent among high-poverty schools (Grissom et al., 2019).

Principal turnover is not necessarily an adverse event if departing principals are ineffective and replaced with more effective ones. Nevertheless, principals at high-poverty schools are, on average, replaced with beginning principals and less effective principals (Grissom et al., 2019). This type of replacement may cause disruptions in essential support areas in the school system through which principals influence school and student performance, such as parental-community ties, professional capacity of staff, a student-centered learning climate, and ambitious instruction (Bryk et al., 2010; Sebastian & Allensworth, 2012). Such disruptions could negatively impact teacher and institutional morale and, eventually, student performance (Meyer et al., 2009). Recent empirical studies found that principal turnover is associated with declines in student achievement for the next one to two years after the principal's departure (Bartanen et al., 2019; Béteille et al., 2012; Miller, 2013).

One solution policymakers have invested in to solve the problem is performance-based compensation systems (PBCS). PBCS is based on the principal-agent theory, which posits that district administrators (i.e., principal) motivate school principals (i.e., agent) through performance-based financial incentives such that the school principals improve their leadership practices to earn financial rewards and, in return, the administrators can meet desired performance goals (Goldhaber, 2007; Goldhaber et al., 2008). PBCS, such as those funded by the U.S. Department of Education's Teacher and School Leader (TSL) Incentive program and Teacher Incentive Fund (TIF) program, also proposes that it retains principals, particularly effective principals, for a more extended period and attracts such principals into PBCS schools (Goldhaber, 2007).

Despite the large amount of public and private money already invested in PBCS nationwide, little empirical evidence of its impacts exists (Goldhaber, 2007; Holley, 2009). There are two evaluation studies and one unpublished study that examined the impacts of PBCS. One of the evaluation studies investigated the effectiveness of PBCS implemented in the Pittsburgh Public Schools (Hamilton et al., 2012). This study found that the PBCS was associated with accelerated student achievement growth and positively changed principal labor markets. The other evaluation study evaluated TIF programs implemented in 10 evaluation districts with 131 schools nationwide and found no evidence that it improved the observational ratings of principals (Chiang et al., 2017). The study did not examine the impact on principal labor markets. The unpublished study (Author, 2023) investigated the impact of PBCS on principal job performance in Tennessee and found dynamic, positive effects on some of the job performance measures. Other studies on TIF programs and other PBCS programs have focused on their effects on teachers and student achievement and the evaluation of the design and implementation processes (e.g., Goodman & Turner, 2013; Heyburn et al., 2010; Hill & Jones, 2020; Shifrer et al., 2017).[1]

This study contributes to the current literature by providing new empirical evidence of the impacts of PBCS on principal labor markets, using longitudinal administrative data obtained from the Tennessee Department of Education (TDOE) via the Tennessee Education Research Alliance (TERA) from 2012 to 2020, merged with unique school- and district-level data on PBCS systematically collected from the TDOE and the districts. School districts in Tennessee implemented PBCS at different times. First, about 200 schools in 15 districts implemented PBCS between 2012 and 2015 through the federal Teacher Incentive Fund (TIF) program, coupled with the Race to the Top fund's competition under the Obama administration. Second, a handful of rural school districts received TIF grants and implemented PBCS in 2015. Not all schools in these districts participated in TIF-funded PBCS. Third, the TDOE enforced state law, T.C.A. § 49-3-306, in 2015, requiring all districts to "adopt and implement differentiated pay plans to aid in staffing hard-to-staff subject areas and schools and attracting and retaining highly qualified teachers" (Tennessee State Board of Education, 2017, p.1). Furthermore, State Board of Education Rule 0520-01-02-.02 "requires [local education agencies] to develop, adopt, and implement a differentiated pay plan under guidelines established by the State Board of Education (State Board) and subject to approval by the Department of Education (Department) to aid hard-to-staff subject areas and schools and in hiring and retaining effective teachers" (Tennessee State Board of Education, 2017, p.1). Although the state law specifically targets teacher compensation systems, some school districts incorporated PBCS for school administrators in developing new compensation systems.

The study focuses on principal labor markets because principal turnover and retention issues need to be mitigated first to make meaningful change happen in school. More specifically, this study investigates the following research questions. First, what is the impact of PBCS on principal turnover? Second, what is the impact among effective principals? Third, what is the impact among high-need schools? Fourth, does PBCS attract effective principals into PBCS schools? Finally, what is the impact of the maximum bonus amount on principal labor markets?

The results of this investigation may have implications for district administrators, state policymakers, and private foundations. Many districts and states nationwide implemented PBCS that includes a performance bonus component for school principals. District administrators may use the result as a reference to decide whether to start or continue PBCS in their districts to rectify the inequitable distribution of principal quality. The administrators may also use the result to determine the appropriate amount of performance bonus.

State policymakers and private foundations could use the result to determine whether to continue enforcing PBCS in their states and whether to continue investing in it. As the federal government, state governments, and private organizations invest a large amount of money in PBCS, the results from this study could be used to assess whether such investment is worthwhile.

This paper proceeds as follows. The next section describes an analytical framework used to analyze the impact of PBCS on principal labor markets, followed by a review of the limited literature on the effect of PBCS on school principals. Then, I describe the data and methods used in this study and report the findings. I conclude with discussions, limitations, and the direction for future research.

Analytical Framework

This study uses a cost-benefit analysis framework to understand school principals' turnover behaviors. Prior studies on school and district leader turnover in the education sector implicitly or explicitly used this framework (e.g., Grissom & Anderson, 2012; Grissom & Bartanen, 2019; Grissom & Mitani, 2016; Mitani, 2019). In the framework, principal turnover and retention result from a two-sided decision-making process. A school principal considers the benefits and costs of staying put at the current school relative to the benefits and the costs of the alternative career options (i.e., across-school transfer, between-district transfer, promotion/demotion, and exit). The costs and benefits that the principal considers include but are not limited to working conditions, salaries, accountability pressure and school performance, and student demographics (e.g., Grissom et al., 2019; Loeb et al., 2010; Mitani, 2018, 2019; Pendola, 2022). If the net benefit of staying at the current school exceeds the net benefit of the alternatives, the principal stays put, conditional on that the district agrees to keep the principal employed; otherwise, the principal leaves if they are given an offer from a new district.

Similarly, the district administrators compare the net benefit of retaining the principal at the current school with the net benefit of replacing them with a new, possibly more effective principal. If the former exceeds the latter, they retain the principal, conditional on the principal agreeing to remain in the school; otherwise, they replace the principal if there is no issue in the labor contract, and they can find a new principal. This two-sided decision process results in four possible outcomes among principals: stay put at the current school voluntarily or involuntarily and leave the current school voluntarily or involuntarily.

PBCS possibly affects principals in the following way. It is likely to affect their cost-benefit calculations. If a principal at a PBCS school expects that they can earn a performance bonus with a reasonable leadership effort, they will stay put at the current school. On the other hand, if a principal at a non-PBCS school finds that their odds of receiving a bonus at a PBCS school are high, they are likely to move into the PBCS school. Thus, PBCS increases the net benefit among some groups of principals, particularly if the maximum bonus amount is set high. Whether at a PBCS school or not, this group of principals are theoretically more effective leaders (e.g., Glazerman et al., 2011). In other words, effective leaders are likely to be more responsive to PBCS than less effective leaders, although a recent study did not find evidence of such behavioral patterns among teachers regarding their value-added scores (Hill & Jones, 2020).

How PBCS affects a principal's cost-benefit calculation may differ by school characteristics. Generally, the net benefit of working at high-need schools is considered lower than at low-need schools, resulting in high turnover rates among principals serving high-need schools (e.g., Grissom et al., 2019; Loeb et al., 2010). PBCS increases the net benefit among these principals, and it may do so more than principals serving low-need schools, due mainly to the different levels of the net benefits before PBCS is implemented. This suggests that principals serving high-need PBCS schools should be more likely to stay put than those serving low-need PBCS schools. It also implies that principals serving high-need non-PBCS schools are more likely to move into high-need PBCS schools than those serving low-need non-PBCS schools. This behavioral pattern may be more pronounced when the maximum bonus amount is high.

This discussion poses another possibility that principal effectiveness and school characteristics moderate the impact of PBCS together. For example, effective principals serving high-need PBCS schools may be more likely to stay put. On the other hand, effective principals serving high-need non-PBCS schools may be more likely to move into high-need PBCS schools. This sorting pattern would be the desired principal labor market outcome under PBCS.

However, the opposite labor market outcomes could also likely occur. For instance, less effective principals serving non-PBCS schools may move into PBCS schools because they are extrinsically motivated and less accurately calculate the net benefit. Effective principals serving high-need PBCS schools may transfer to low-need PBCS schools because earning financial rewards is easier. This is highly possible, as previous studies found that principals tend to transfer to low-need schools (e.g., Grissom et al., 2019; Loeb et al., 2010).

Two aspects of PBCS potentially make the interpretation of PBCS impact less straightforward. First, PBCS schemes may be flawed with false assumptions and conflict with a principal’s work philosophy (Kozlowski & Lauen, 2019). Educators and school leaders, and more generally, public service officers tend to be more intrinsically motivated and committed to their service work (e.g., Buelens & Van den Broeck, 2007; Levacic, 2009; Mintrop et al., 2018; Ritz et al., 2016). With such high public service motivation, they may view PBCS less preferably and may be less motivated by financial rewards alone. Some of the previous reports and studies found that educators and school leaders are not motivated by financial incentives (e.g., Hamilton et al., 2012; Kozlowski & Lauen, 2019; Mintrop et al., 2018). This conflict increases the net cost and may cause principals at PBCS schools to leave their schools, regardless of their effectiveness levels or school characteristics. On the other hand, principals at non-PBCS schools may decide to stay put at their schools.

Second, financial incentives in PBCS are generally not framed negatively, particularly the ones in Tennessee. This means that principals do not face a financial loss in their base salaries, even if they miss performance targets set by the districts. This framing contrasts the federal and state accountability incentives, which are generally negatively framed, particularly during the No Child Left Behind (NCLB) period. Principals and their schools during the period faced negative sanctions when their schools failed to meet adequate yearly progress in consecutive years. Mitani (2018, 2019) found evidence of the adverse effects of NCLB on principal turnover. Due to the positive framing, a principal at a PBCS school may stay put at the current school even if they expect a low likelihood of receiving a financial reward. Similarly, a principal at a non-PBCS school may remain in the current school even if they evaluate that they are less likely to meet performance targets.

The avenue through which PBCS possibly affects school principals' turnover behaviors and the two aspects of PBCS may make its impact less observable or less detectable by statistical tests. This study's research design cannot disentangle these multiple effects. Instead, the study only estimates the overall impact of PBCS.

Previous Work on the Impact of PBCS on School Principals

The literature on this research topic is scant, and only two evaluation studies have explored the impact of PBCS on school principals and been published (Goldhaber, 2007; Holley, 2009). One evaluation study examined the effectiveness of PBCS implemented in Pittsburgh Public Schools, the Pittsburgh Principal Incentive Program (PPIP) (Hamilton et al., 2012). PPIP was the central part of the Pittsburgh Urban Leadership System for Excellence (PULSE), and the district received funding from the federal Teacher Incentive Fund (TIF) program in 2007 to implement the program to improve the overall quality of school leadership in the district. PULSE consisted of six core components: the Pittsburgh Emerging Leaders Academy, the new administrator's induction program, Leader Academy for principals, assistant superintendent training and mentoring, performance-based evaluation, and performance-based compensation—

the last two components comprised PPIP (Hamilton et al., 2012). The two critical features of PPIP were that principals had an opportunity each year to earn (1) a base salary increase of up to $2,000 based primarily on their evaluation scores on leadership practices and (2) a bonus of up to $10,000 based primarily on student achievement growth. PPIP also provided bonuses for principals who took leadership jobs in high-need schools. Not only did the program offer financial incentives, but it also provided various support such as feedback and coaching from assistant superintendents and participation in professional growth projects.

The evaluation study found that principals' skills and practices and their schools' performance improved over time. However, their performance on leadership practices tied to the evaluation rubric and the bonus measures was constant. However, since PPIP offered professional support simultaneously, these findings cannot be solely attributable to financial incentives alone. The study also reported that most principals viewed the performance-based compensation component as problematic and did not think the financial incentives motivated them to change their behaviors. Regarding its impact on principal labor markets, the study found suggestive evidence that, although principal retention rates were constant, effective principals tended to move into higher-need schools or the central office. In contrast, less effective ones were demoted to assistant principals or left the district. This behavioral pattern is aligned with the analytical framework discussed in the previous section.

The other evaluation study examined the impact of TIF programs implemented in 10 districts between 2012 and 2015 across the nation, with a primary focus on student and teacher performance (Chiang et al., 2017). The study used random assignment to estimate the impact. It found that, although most principals earned a bonus, PBCS did not improve observational ratings of principals, which is consistent with the Pittsburgh study. The study did not examine the impact on principal labor markets.

One unpublished study (Author, 2023) examined the impact of PBCS on principal job performance in Tennessee, using longitudinal administrative data and unique PBCS data from 2012 to 2019, similar to the data used in the current study (see the data section for details). It analyzed the impact on principal evaluation scores and intermediate outcomes that should lead to school improvements, such as relational trust, teacher evaluation scores, and teacher retention rates and recruitment outcomes. The study found that, although PBCS did not improve principal job performance instantaneously, it improved the performance dynamically. For example, PBCS effects on the overall evaluation scores and the subjective rating scores increased in the second and/or third year after the PBCS implementation. Similar dynamic effects were found in teacher evaluation scores. Furthermore, PBCS effects were found to be more pronounced among high-poverty schools, suggesting that it improved the inequitable distribution of principal job performance.

Outside this limited literature, there are empirical studies that examined the relationship between school principals' mobility and their salaries. For example, Pendola (2022) analyzed the relationship between principal salary and turnover, using longitudinal administrative data from Texas, and found that principals have a high propensity to care about salary comparisons over and above basic salary. Principals' high sensitivity to comparative salaries was also observed in Missouri, where Baker et al. (2010) showed that school principals tended to stay put at their schools when their salaries were higher than their peers in the same labor markets. However, they were more likely to leave when their salaries were lower. Similarly, in California, Tran (2017) reported suggestive evidence that the salary satisfaction of a small sample of high school principals in the state depended on the salaries of their comparative peers and that lower salary satisfaction was associated with a higher intention to leave. Other related studies also found that principal mobility was associated with an increase in principals' salaries and that a raise in principal salary led to higher student performance (e.g., Akiba & Reichardt, 2004; Cullen & Mazzeo, 2007; Lavy, 2008).

In sum, these empirical studies collectively suggest that school principals respond to financial incentives and consider them in mobility decisions. This study will add new empirical evidence to this limited literature and advance our understanding of principals' labor market behaviors under financial incentives.

Methods

PBCS was implemented in multiple locations at different times in Tennessee. A standard approach to estimating the impact of PBCS on principal turnover in this setting is to use a two-way fixed effects (TWFE) difference-in-differences (DID) estimator. A basic model takes the following form:

$$Y_{isdt} = \rho_s + \varphi_t + \delta D_{st} + \beta P_{it} + \theta S_{st} + \pi X_{dt} + \varepsilon_{isdt}\ . \quad (1)$$

The probability that a principal *i* in school *s* in district *d* in year *t* leaves the school at the end of the school year is a function of school fixed effects ρ_s, time fixed effects φ_t, a binary PBCS treatment variable D_{st}, principal characteristics P_{it}, time-variant school characteristics S_{st}, district characteristics X_{dt}, and a random error term ε_{isdt}. Principal, school, and district characteristics included in the model are listed in Table 1, along with other school characteristics that were used in some analyses. Time-invariant characteristics are dropped from the analysis, and standard errors are clustered at the school level.

If all DID assumptions hold and the PBCS effect is homogeneous across schools and over time, a parameter estimate, $\hat{\delta}$, is unbiased and is equal to a weighted average of the average treatment (PBCS) effects (ATEs) across all possible two-by-two difference-in-differences estimators (Borusyak et al., 2022; de Chaisemartin & D'Haultfoeuille, 2020, 2022; Goodman-Bacon, 2021). Each weight is calculated by dividing a residual that is generated from regressing the treatment variable (PBCS) on school and time fixed effects by the average of all the weights (de Chaisemartin & D'Haultfoeuille, 2020). Some weights can become negative in the current study setting because some schools that serve as a control group in one DID comparison become a treatment group in another DID comparison (de Chaisemartin & D'Haultfoeuille, 2020). All the weights add up to one.

In the real world, however, the assumption of the homogeneous PBCS effect is less likely to hold because of substantial differences in principal labor markets from region to region and from time to time. For example, it is less reasonable to assume that the PBCS effect is the same between Metro Nashville Public Schools in Nashville, an urban city of close to 700,000 residents, and Van Buren County Schools in Van Buren County, a rural county of just over 6,000 residents. If the PBCS effect is heterogeneous across schools and time, the presence of the negative weights may result in a situation where $\hat{\delta}$ is negative, but all ATEs are positive, suggesting that $\hat{\delta}$ is biased (de Chaisemartin & D'Haultfoeuille, 2020, 2022). To investigate the severity of this problem, I estimated two types of measures that indicate robustness to the heterogeneity of the PBCS effect, following de Chaisemartin and D'Haultfoeuille (2020). The first measure was estimated by dividing the absolute value of the coefficient on the PBCS treatment variable, estimated by the standard TWFE DID approach above, by the standard deviation of the weights attached to the regression. The other measure was estimated by dividing the absolute value of the same coefficient on the PBCS treatment variable by a function of the weights. If these two measures are close to zero and many weights are negative, the problem is severe, and an alternative estimation approach is necessary. de Chaisemartin and D'Haultfoeuille (2020) provide detailed theoretical explanations about this analysis and its application.

I found evidence that treatment heterogeneity was a severe problem in the current context. These two measures were found close to zero in all models that used a binary PBCS treatment variable and an ordinal maximum bonus amount treatment variable (for Research Question 5). Nevertheless, the PBCS effect can still be robust to heterogeneity if the weights are not associated with the intensity of the treatment effects in all school-by-year cells, as shown in Assumption 7 in de Chaisemartin and D'Haultfoeuille (2020). Following de Chaisemartin and D'Haultfoeuille (2020), I used the inflation-adjusted district-level per-pupil expenditures as a proxy for the intensity of the effect and investigated the relationship. I found no evidence of the relationship. All results are reported in Online Appendix A Table 1.

Furthermore, the standard TWFE DID approach requires that the treatment be absorbing (e.g., de Chaisemartin & D'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun & Abraham, 2021). That is, the treatment continues until the end of the data analysis period and does not end in the middle. This requirement cannot be met in the current study context because some school districts in Tennessee discontinued PBCS in the middle of the period for various reasons, including a lack of funding. Still, some other districts discontinued PBCS and then started new PBCS later.

To address the severity of the heterogeneity problem and the above unmet requirement, I employed an alternative DID estimator, DID_M, recently developed by de Chaisemartin and

D'Haultfoeuille (2020), and used the same set of covariates specified in Equation 1. It takes the following form (covariates not shown for simplicity):

$$DID_M = \sum_{t=2}^{T}\left(\frac{N_{1,0,t}}{N_S} DID_{+,t} + \frac{N_{0,1,t}}{N_S} DID_{-,t}\right) \qquad (2).$$

N_S is the number of principals in school-by-year cells whose schools' PBCS status changed from *t-1* to *t*; $N_{1,0,t}$ is the number of principals in school-by-year cells whose schools had PBCS in *t* but did not in *t-1*; and $N_{0,1,t}$ is the number of principals in school-by-year cells whose schools did not have PBCS in *t* but did in *t-1*. $DID_{+,t}$ in the first term compares the evolution of the mean principal turnover outcome from *t-1* to *t* between principals serving schools that implemented PBCS in *t* and those serving schools that remained untreated. $DID_{-,t}$ in the second term compares the mean outcome evolution from *t-1* to *t* between principals serving schools that remained treated in both *t-1* and *t* and those serving schools that had PBCS in *t-1* but discontinued PBCS in *t*. The second term indicates that the DID_M estimator can incorporate "treatment leavers," which the standard TWFE DID estimator cannot. The estimator is a weighted average of these two DID estimators. It is unbiased, consistent, and robust to heterogeneous effects under the parallel trends assumption. Standard errors were clustered at the school level and were estimated through Bootstrap. Author (2023) used the same method in estimating the impact of PBCS on principal job performance.

I tested the parallel trends assumption using a placebo estimator, DID_M^{pl}, that de Chaisemartin and D'Haultfoeuille (2020) developed, along with the DID_M estimator. The placebo estimator takes a form very similar to Equation 2. It is a weighted average of $DID_{+,t}^{pl}$ and $DID_{-,t}^{pl}$ estimators. The first estimator compares the evolution of the mean principal turnover outcome from *t-2* to *t-1* between principals serving schools that did not have PBCS schools in *t-2* or *t-1* but implemented it in *t* and those serving schools that did not have PBCS during the three years. The second estimator compares the outcome evolution from *t-2* to *t-1* between principals serving schools that had PBCS during the three years and those serving schools that had PBCS in *t-2* and *t-1* but discontinued it in *t*. If the result is significant, the assumption is violated. I found no evidence of the violation in most of the models. Online Appendix A Table 2 reports the result of the main analysis (Research Question 1). When it was significant, I included school-specific linear time trends.

To answer the second and third research questions, I created an interaction term between the PBCS treatment variable and a binary principal effectiveness indicator (Research Question 2) and a binary high-poverty school indicator, and a binary high-students-of-color school indicator (Research Question 3).[3, 4] For the fourth research question, I transformed the principal-level data into the school-level data and used the mean principal effectiveness measures of new hires as outcome variables.[5] Because few schools had new principals in consecutive years, the DID_M estimator did not work well. For this reason, I used the standard TWFE DID approach (Equation 1).[6] The coefficient estimates were biased for the above reasons and cannot be interpreted as causal. In addition, to examine whether high-need PBCS schools could attract effective principals, I restricted the sample to high-need schools with new principals and created an interaction term between the PBCS treatment variable and the same binary principal effectiveness indicator variable. Finally, to answer the fifth research question, I replaced the binary PBCS treatment variable with an ordinal maximum bonus amount treatment variable.

Data

The primary data in this study came from state longitudinal administrative data made available by the Tennessee Department of Education (TDOE) via the Tennessee Education

Research Alliance housed at Vanderbilt University. The TDOE enrolled just over one million students in 1,894 public schools in 147 school districts, of which 1,846 were classified as regular public schools in 2020.[7] Thirty-one percent were located in urban areas, and 35% were in rural areas. Among students, 22% were Black, 12% were Hispanics, and 35% were recognized as economically disadvantaged students.[8]

The administrative data included principals' information about personal and professional characteristics (i.e., age, gender, race/ethnicity, highest degree earned, and salary).[9] Since the data did not include years of principal experience, I used administrative data from 2002 to create the experience variable. This method generated missing values for school principals who were already in leadership positions before 2002. As explained below, since the TDOE implemented the Tennessee Educator Acceleration Model (TEAM) in 2012, from which I utilized principal evaluation scores, the data analysis period resulted in from 2012 to 2020, with the last year being used to analyze turnover outcomes in 2019.

The administrative data also included school principals' job history and school assignment information, from which I created a school tenure variable and two types of turnover variables. The first type would take a value of unity if a principal did not return to the same school as a principal the following year and a value of zero otherwise. The second type would take a value of unity if the principal did not return to the same district as a principal the following year and a value of zero otherwise. I considered the second type of turnover because districts may view it as a loss to the districts if the principal took a non-leadership position (i.e., demotion) or exited the system, compared with within-district transfers (i.e., principals moving from one school to another within the same district), which would not cause any loss to them.[10] I used these two types of turnover variables as the outcome variables for Research Questions 1, 2, 3, and 5.

Principal Effectiveness Measures

I used input-based and output-based principal effectiveness measures. For the input-based measures, I used years of principal experience and principals' highest educational attainment. I created an indicator variable for holding an education specialist or doctoral degree. The output-based measures were based on data from the TEAM evaluation system, which was implemented in 2012. In the evaluation system, a school principal receives an overall evaluation score on a 1-to-5 scale, called Level of Effectiveness (LOE), every year. An LOE score is calculated based on scores in the subjective rating of principal leadership practices by a principal evaluator, which accounts for 50% of the LOE score, scores in the student growth measure (35%), and scores in the student achievement measure (15%). The subjective rating is based on a rubric pegged to the Tennessee Instructional Leadership Standards (TILS).[11] I used LOE scores and scores in the subjective rating separately as output-based principal effectiveness measures.

To answer the second research question, I created two indicator variables for effective principals. The first indicator variable is based on the LOE scores and takes a value of unity if the LOE score is at or above four and zero otherwise. Similarly, the second one is based on the scores in the subjective rating and takes a value of unity if the score is at or above four and zero otherwise. To answer the fourth research question, I took the school-level mean of principal effectiveness measures, both input- and output-based, among new hires.

School and District Characteristics

I merged these administrative and evaluation data with the school-level Common Core of Data (CCD) data files downloaded from the National Center for Education Statistics (NCES), the

U.S. Department of Education. The data files contained information on student demographic composition (i.e., race/ethnicity and eligibility for the federal free lunch program), enrollment size, school level, and school locale type (i.e., urban, suburban, town, and rural). I also used the TDOE's school profile data files to supplement the CCD's data on the free lunch program.[12] From the student demographic data, I created two indicator variables for high-need schools to answer the third research question. The first indicator variable is based on the percentage of students eligible for the free lunch program. It takes a value of one if the percentage was at or above 50% in 2012 and zero otherwise. The second indicator is based on the percentage of Black and Hispanic students and is coded similarly.

In addition, I merged the data with the TDOE's school accountability data files, which included the percentage of students who were classified as "on track" and "mastered" (formerly at proficient and above) in math and English Language Arts.[13] I combined these subject-level data by taking the grand mean for each school and standardized it within years and school levels.

For district characteristics, I merged the data with the district-level CCD data files that included data on the number of schools, the number of students, and student demographics. Data on the percentage of students eligible for the federal free lunch program were aggregated at the district level using the school-level data. I obtained data on district-level per-pupil expenditures from the TDOE's district profile data files.

Table 1 reports descriptive statistics for the principals and their schools and districts in the analytic sample during the data analysis period (i.e., 2012 to 2020). The mean principal age was 48, and close to 60% of the principals were female. Just over one-quarter were principals of color, and about a half had either an education specialist or doctoral degree. The mean years of principal experience was 5-6 years, and the mean salary was around $89,000. Their LOE scores ranged from one to five, with a mean of 3.81. Their scores in the subjective ratings had the same range and a similar mean, but the standard deviation was half that of the LOE scores.

About 50% of the schools these principals served were elementary schools, 16% middle schools, and 17% high schools. Sixteen percent of the schools served grades from elementary to middle school (i.e., K-8). Nearly 30% of the schools were in urban areas, whereas 38% were in rural areas. Among students enrolled in the schools, 45% were eligible for the federal free lunch program, and 28% were students of color. Just below 50% of the students at these schools were classified as "on track" or "mastered" in the state accountability system.

Districts in the state operated about 53 schools on average, and the largest school district had 290 schools. The mean district student enrollment was 42,068, and the mean per-pupil expenditure was $10,233.

PBCS Data

I systematically collected district- and school-level PBCS data by reviewing TIF-related reports and documents, differentiated pay plans posted on the TDOE websites, district budget reports, and local media websites. Furthermore, I collected additional information by directly contacting all district administrators (e.g., directors of schools, directors of human resources, and directors of financial services) and some school administrators (e.g., school principals). The school-level longitudinal PBCS data included information about the PBCS implementation status, the maximum bonus amount, and incentivized performance measures. From the data, I created a binary school-level PBCS treatment variable and an ordinal maximum bonus amount treatment variable (i.e., $0, $1 to $2,000, and $2001 and above).[14]

Table 2 reports PBCS data by year and program characteristics from 2012 to 2020.[15] A total of 314 schools had PBCS in place in 2012, accounting for 18% of the public schools in the state. The number of PBCS schools steadily increased over time. It culminated at 462 schools (26%) in 2015 when the second TIF-funded PBCS programs were implemented in a handful of school districts, and the TDOE enforced the state law regarding differentiated pay plans. Since then, the number gradually declined each year and decreased to 152 PBCS schools in 2020, which was only eight percent of the schools. The third and fourth columns show that some schools implemented PBCS, and other schools discontinued it each year.

The median maximum bonus amount ranged from $400 to $2,850 across schools and the years. The largest maximum bonus amount during the data period was $15,000 in 2012. A variety of principal job performance measures were used in PBCS, including TEAM evaluation scores (i.e., LOE scores), school-level Tennessee Value-Added Assessment System (TVAAS) scores, graduation rates (high schools only), ACT scores (high schools only), and other annual measurable objectives such as closing achievement gaps. In the data analysis period, just over three-quarters of the schools across the years used TEAM evaluation scores or TVAAS scores, or both as performance measures.

Results

I start by presenting the principals' turnover events during the data analysis period. Table 3 displays the percentages of principals in the analytic sample who left the positions (1st column) and the percentages of principals who moved to new schools within the same districts (2nd column) and across the districts (3rd column), and the percentages of principals who exited the system or changed the positions (4th column).[16] Panel A reports the percentages over time for all principals in the analytic sample. Panel B exhibits the percentages by the PBCS status for all years combined. The mean turnover rate was over 13% during the period, ranging from 10% to 16%. Exits or position changes accounted for many of the turnover events. The mean percentage of principals who exited or changed their positions was nine percent, and the percentage was at a maximum of 14% in 2018. Among transfers, within-district transfers were much more prevalent than across-district transfers (three percent and less than one percent, respectively), which may reflect the geography of the districts.

Turnover rates were statistically different between principals serving PBCS and non-PBCS schools. The mean turnover rate among principals serving PBCS schools was 17%, whereas only 12% among non-PBCS schools. This difference primarily came from the difference in within-district transfer and exit/position change rates. As the result from the main analysis below shows, this difference disappeared once various factors were accounted for, suggesting that the relationship between the PBCS status and the turnover rates was confounded.

Now I turn to the main analysis. Table 4 reports the result of Research Question 1. Models in Panel A used the binary PBCS treatment variable. Models in Panel B used the ordinal maximum bonus amount treatment variable (Research Question 5). The coefficient estimates mean the probabilities that the principal does not return to the same school (odd-numbered models) or the same district (even-numbered models) as the principal next year.

I found no evidence of the impact of PBCS on principal turnover. All coefficient estimates were statistically indistinguishable from zero, whether positive or negative. For example, PBCS increased the probability that a principal does not return to the same school as a principal (Model 1) by .01, with a 95% confidence interval [-.04, .07]. There is no systematic pattern in terms of the signs of the coefficients. Since the results were similar, whether the model used LOE scores or scores in the subjective ratings, I only reported the results for models with LOE scores in the subsequent analyses.

Next, I investigated the PBCS impact on effective principals (Research Question 2) and principals serving high-need schools (Research Question 3). Table 5 displays the results. The coefficient estimates mean the differences in the probabilities of turnover between effective principals serving PBCS and non-PBCS schools (Research Question 2) and between principals serving high-need and non-high-need schools (Research Question 3). Similar to the main result, I did not find any evidence that PBCS significantly affected principal turnover. All estimated coefficients were statistically indistinguishable from zero, although the signs of the coefficients were primarily positive. The results did not change even if I used the three different versions of the maximum bonus amount treatment variable (Research Question 5). Online Appendix B Table 2 reports the results.

To examine the impact of PBCS on effective principals serving high-need schools, which is one of the primary interests of PBCS, I restricted the analytic sample to principals serving high-need schools and estimated the impacts. Table 6 exhibits the results. There was no evidence of such PBCS impact. All estimates were imprecise and statistically indistinguishable from zero. However, the sizes of the coefficients became larger in most models, and their signs were all positive among effective principals serving high-poverty schools and mostly negative among effective principals serving high-color schools. I also used the three versions of the bonus amount treatment variable but found no evidence. Online Appendix B Table 3 shows the results.

All analyses so far have focused on the impact of PBCS on principal turnover. However, another critical aspect of PBCS is whether PBCS attracted effective principals into high-need schools such that it rectified the inequitable distribution of principal quality and performance. As a final analysis, I investigated the impact of PBCS on the mean principal quality and performance measures of new hires (Research Question 4). Table 7 shows the results for the following four types of output- and input-based principal quality measures: LOE scores, scores in the subjective ratings, highest degree earned, and years of principal experience. As I explained in the method section, I used the TWFE DID estimator, so the results may not be causal and are likely to be biased.

As the table shows, I found no evidence that PBCS affected the mean principal quality and performance measures among new hires. All estimated coefficients were statistically indistinguishable from zero. However, one unique pattern is that the coefficients were positive in many measures among high-need schools. For example, PBCS increased the mean LOE scores of new hires by 1.95 standard deviations among high-poverty schools (Model 5) and by 1.81 standard deviations among high-color schools (Model 9). Since the sample sizes were small, these coefficients were less precisely estimated. Nevertheless, these patterns are worth noting, given the purpose of PBCS.

Sensitivity Analyses

The main analysis found no evidence of the impact of PBCS on principal turnover and the mean principal quality and performance measures of new hires. However, these findings do not speak to the possible intertemporal effects of PBCS or may be sensitive to some factors. In this section, I examined the dynamic PBCS effects and the sensitivity of the findings.

Intertemporal Effects

The main analysis only estimated instantaneous effects. It is possible that the PBCS impact was observable a couple of years after the PBCS implementation. Author (2023) found that PBCS had dynamic, positive impacts on some principal performance measures. Following de Chaisemartin and D'Haultfoeuille (2023), I investigated the intertemporal PBCS effect on principal turnover. Figure 1 plots the estimated dynamic effects of PBCS (binary treatment) for the first type of turnover (i.e., not returning to the same school as a principal in the following year). I generally found a pattern that PBCS decreased the probability of principal turnover substantially in the second and third years, although the impact was less precisely estimated in many models.

The top three plots (i.e., (a), (b), and (c)) clearly show the dynamic PBCS effects.[17] For example, for the impact on all principals, the probability of principal turnover decreased by .02 in the first year, and the impact was statistically indistinguishable from zero. However, in the second year, the impact was more pronounced. The probability declined by .12 and was statistically significant. In the third year, it decreased by .09, although the impact was less precisely estimated ($p = .15$). Among effective principals, the probability increased by .05 in the first year ($p > .20$) but declined by .01 in the second year ($p > .20$) and by .08 in the third year ($p = .13$). Among principals serving high-poverty schools, the probability declined by .05 in the first year ($p = .15$), by .10 in the second year ($p = .11$), and by .13 in the third year ($p = .14$). The bottom three plots also show that the impact was intertemporal and more pronounced in the later years. However, their estimates were imprecise with the current sample. Panel A in Online Appendix B Table 4 reports the estimated coefficients. I also investigated the intertemporal effects of the second type of turnover and found similar results. Online Appendix B Figure 1 plots the estimated coefficients, and Panel B in Online Appendix B Table 4 reports the coefficients.

Principal Labor Markets in Urban and Rural Areas

Principal labor markets in Tennessee may look quite different from one region to another. For example, Metro-Nashville Public Schools (MNPS), an urban school district, operates more than 150 schools in an area of just over 500 square miles, surrounded by Cheatham County

Schools operating 14 schools, Williamson County Schools operating 51 schools, Rutherford County Schools operating 50 schools, Robertson County Schools operating 24 schools, Sumner County Schools operating 49 schools, and Wilson County Schools operating 25 schools. On the other hand, Hardeman County Schools, a rural school district, operates nine schools in an area of close to 670 square miles, surrounded by McNairy County Schools with nine schools, Chester County Schools with six schools, Madison County Schools with 24 schools, Haywood County Schools with seven schools, Fayette County Schools with seven schools, and Benton County Schools with six schools.[18] This simple comparison demonstrates possible differences in job opportunities between principals in urban and rural school districts. To test whether the main findings are sensitive to the geography of school districts, I estimated the impact of PBCS on all principals and effective principals serving urban and rural schools by creating an interaction term between the urban/rural indicator variable and the PBCS treatment variable.

I found very little evidence of the impact, whether rural or urban schools and whether effective or not. Because of the small sample sizes, all estimates were imprecise. Even among insignificant coefficients, I found no systematic pattern between principals serving urban and rural schools. One exception was that PBCS significantly increased the probability that a principal does not return to the same district as a principal by 1.21. However, the parallel trends assumption was violated in this model, and I used school-specific linear-time trends. So the coefficient cannot be interpreted as purely causal. Online Appendix B Tables 5 and 6 display the results for the binary PBCS treatment and the maximum bonus amount treatment, respectively.

Differences between PBCS funded by TIF and non-TIF PBCS

As described in the introduction, a handful of school districts implemented PBCS in 2012 and 2015 through the federal TIF grant programs, whereas other districts implemented it through their own funding sources. TIF-funded PBCS could be different from non-TIF PBCS in a systematic way that affects principals' turnover patterns, such as the funding level, design features, and accountability. Since most PBCS was not funded by TIF, I estimated the impact of PBCS, excluding schools that ever had TIF-funded PBCS in place during the data analysis period.

The main result mostly stayed the same. Almost all the estimates were imprecisely estimated and statistically indistinguishable from zero. However, one notable change is that many of the estimates, although insignificant, turned negative. That is, principals serving PBCS schools were found to be more likely to stay in the same schools. Online Appendix B Tables 7, 8, and 9 report the result.

Discussion and Conclusions

Prior research documented the inequitable distribution of principal quality and performance (e.g., Branch et al., 2012; Gates et al., 2006; Roza, 2003; Grissom et al., 2019). This concerns policymakers, as school principals play a critical role in improving school and student outcomes (e.g., Branch et al., 2012; Grissom et al., 2021; Grissom et al., 2015; Grissom et al., 2013; Hallinger & Heck, 1998; Waters et al., 2003). To address the problem, policymakers implemented PBCS in Tennessee. This study examined the impact of PBCS on principal turnover, the impact among effective principals, the impact among principals serving high-need schools, and the impact on the mean effective measures of newly hired principals, using state longitudinal administrative data and unique school- and district-level PBCS data from 2012 to 2020.

Summary of the Findings

This study found no evidence of the impact of PBCS on principal turnover, whether the binary PBCS or the maximum bonus amount treatment variable was used (Research Questions 1 and 5). Almost all the coefficients were imprecisely estimated and statistically indistinguishable from zero. The study also found null results among effective principals (Research Question 2), among principals serving high-need schools (Research Question 3), and among effective principals serving high-need schools (Research Questions 2 and 3). Moreover, it found that the mean principal quality and performance measures of newly hired principals at PBCS schools were statistically no different from the mean measures at non-PBCS schools. Nevertheless, it is worth noting that the estimated coefficients on the mean measures were mostly positive among high-need schools, suggesting that they could attract high-quality school principals. These findings were not sensitive to whether the schools were located in urban or rural areas or whether PBCS was funded by TIF.

One exception was the intertemporal effects of PBCS. The impact of PBCS became more pronounced in the second and third years than in the first year. In particular, PBCS significantly decreased the probability of the first type of turnover (i.e., not returning to the same school as a principal) by .12 in the second year and by .08 for the second type (i.e., not returning to the same district as a principal). The intertemporal effects were also present among effective principals and principals serving high-poverty schools, although the estimates were less precise ($.10 < p < .15$).

Discussion

This study's null findings, other than the intertemporal effects, are consistent with the two previous evaluation studies of the federal TIF grant programs (Chiang et al., 2017) and the Pittsburg Principal Incentive Program (PPIP) (Hamilton et al., 2012). Using random assignment, Chiang et al. (2017) analyzed principals' retention rates between the treatment (TIF) and control groups during the program period. They found that the retention rates between the treatment and control groups were statistically indistinguishable, whether the rates were calculated on a one-year, two-year, or three-year basis. Hamilton et al. (2012) similarly reported that mobility patterns were pretty constant and that there was no statistically significant association between the bonus amount and mobility.

The null results should not directly lead policymakers to discontinue PBCS. These null results may be attributable to multiple factors. First, the design of PBCS needs to be carefully examined. Table 2 reports that incentivized performance measures included principal evaluation scores and school-level value-added scores in most schools. Other measures not reported in the table included graduation rates, ACT scores, annual measurable goals, district-level performance measures, and the state's school recognition awards. Although the districts engaged diverse groups of stakeholders, including principals, to select performance measures, this process does not necessarily guarantee that principals were satisfied with the selection. Most, if not all, of the performance measures are based on measurable outcomes. However, to improve overall school performance, school leaders need to intervene in multiple essential support areas, such as professional capacity, school learning climate, parent, school, and community ties, instructional guidance, and relational trust (Bryk et al., 2010). Outcomes in these areas are not easily measurable and were not included in PBCS. Including these outcomes, whether measurable or not, may affect the principals' cost-benefit calculation and decisions to stay or leave, particularly among effective principals who demonstrate excellent leadership in intervening in the support areas.

The maximum bonus amount is another PBCS design feature that needs careful attention. In Tennessee, I found a wide range in the amount. However, the median maximum bonus amount among PBCS schools each year was always less than $3,000 and was only $400 in 2019. Given that the mean inflation-adjusted principal salary was about $89,000, it is unclear if the amount was large enough to give principals adequate incentives.

Second, PBCS is designed based on assumptions, and if these assumptions do not hold, it may not generate expected outcomes even if PBCS is reasonably designed. In particular, Kozlowski and Lauen (2019) strongly argue that PBCS is based on three flawed assumptions. Since their criticism focuses on teacher performance pay, I modified their claim to fit the context of PBCS for school leaders. The first flawed assumption they claim is that principals are motivated by money. While the current study does not have data to test this assumption, Hamilton et al. (2012) reported qualitative evidence that principals are not necessarily motivated by money. Furthermore, the answer to the fifth research question suggests that principals do not respond to financial incentives, although the maximum bonus amount was generally set low in the state.

The second flawed assumption is that principals do not work hard enough. The literature on principal time use and school leadership generally does not support this assumption. Recent studies and a systematic review describe extensive working hours and resulting mental health issues that principals face (e.g., DeMatthews et al., 2023; Hochbein et al., 2021; Steiner et al., 2022). The third flawed assumption is that principals know education production functions but do not make the necessary effort to improve school performance. This assumption does not appear to hold, given that principals work long hours extensively and are held accountable for student performance under the federal and state school accountability systems.

Given the improvement areas in the PBCS designs and the potential violation of the assumptions behind the PBCS logic, the null results should not be interpreted as program failure. Instead, this suggests that policymakers need to elaborate on the PBCS design features and assumptions. Careful redesign of the current PBCS programs may alter the null results.

In determining whether to continue PBCS, policymakers should also consider that PBCS has two primary goals. One is to improve principal job performance and probably more significantly among principals serving high-need schools. The other is to retain and attract effective principals, particularly at high-need schools. By achieving these two goals, PBCS can rectify the inequitable distribution of principal quality and performance. It is, therefore, more appropriate to evaluate these two aspects of PBCS together.

This study examined the second goal and found null results, except for the intertemporal effects in the second and third years of the PBCS implementation. Author (2023) examined the first goal and generally found no evidence of its instantaneous effect. However, the study found that PBCS improved principal job performance intertemporally. These two studies' findings suggest that PBCS is still a viable policy option to address the inequitable distribution, conditional on that policymakers carefully designing PBCS, reflecting more feasible assumptions regarding school principals' behaviors.

Limitations

The current study faces several limitations. First, principal turnover was estimated from the principals' perspectives. However, as the cost-benefit analysis framework describes, principal turnover is a two-sided decision process (i.e., principals and districts). To the extent that both sides were affected by PBCS, attributing principal turnover solely to principals' decisions may have biased the estimation results. Second, the DID_M estimator relies on principals whose treatment status changed from non-PBCS to PBCS and from PBCS to non-PBCS. Although not reported in the tables, the number of such principals was small relative to the full sample size. The larger number of "treatment switchers" will increase the statistical power to detect the difference in the probability of turnover between principals serving PBCS and non-PBCS schools. Third, as I mentioned in the previous section, the median maximum bonus amount was less than $3,000 and was small relative to the mean inflation-adjusted principal salary in many PBCS schools. As a result, the answer to Research Question 5 would not be conclusive. If the maximum bonus amount had varied more substantially with a wide range in the amount, the answer might have been different.

Conclusions

The inequitable distribution of principal quality and job performance negatively contributes to wider gaps in learning opportunities and outcomes between White students and students of color and between privileged and marginalized students. PBCS was implemented in Tennessee as a possible means to address the problem. This study provides no evidence of the instantaneous impact of PBCS on principal turnover. However, it found that PBCS decreased the probability of principal turnover in the later years of the PBCS implementation. Given the positive results from Author (2023) regarding its impact on principal job performance, PBCS still appears to be a policy option to consider. Nevertheless, it is pivotal that policymakers elaborate on PBCS designs and assumptions for future improvement.

Notes:

1: Although not in education, there are studies about the effect of PBCS on managers in the public sector, who are comparable to school principals in terms of role. These studies generally found mixed evidence of the impacts on employees' behaviors, turnover rates, and organizational performance (e.g., Perry et al., 2006; Perry et al., 2009; Weibel et al., 2010).

2: Principals may still feel a financial loss if they miss performance criteria. However, since their base salaries are not affected by failure to meet the criteria, I do not consider that PBCS is framed negatively.

3: I replaced the continuous variable with its indicator variable in each interaction model (i.e., principal effectiveness, percentage of students eligible for the federal free lunch program, and the percentage of students of color). I dropped the binary PBCS treatment variable. The coefficient on the interaction term means the difference in the outcome between principals serving PBCS and non-PBCS schools among effective principals, principals serving high-poverty schools, and principals serving high-students-of-color schools.

4: The data section describes how these indicator variables and the ordinal maximum bonus amount treatment variable were created.

5: The main analytical data were at the principal-by-year level, and there were no duplicates in terms of principals and years. However, there were duplicates in terms of schools and years.

6: Since the TWFE DID assumes that the treatment is absorbing, I deleted schools after they discontinued PBCS.

7: I refer to the academic year by the spring year hereafter.

8: Author's calculation based on the 2019-20 Common Core of Data, the National Center for Education Statistics, the U.S. Department of Education, except for the percentage of students classified as economically disadvantaged, whose data source was the TDOE.

9: Principal salaries and district-level per-pupil expenditures were adjusted for a 2020 constant dollar.

10: I assigned missing values to the turnover variables for principals whose schools were closed or turned into the Achievement School District at the end of the school year.

11: For the statistical properties of scores in the subjective rating, see Grissom et al. (2018). Online Appendix A Table 3 reports the rubric used from 2012 to 2019. Please note that not all school districts used the state-developed rubric for the subjective rating. A small number of school districts, such as the Hamilton County Department of Education and Bradley County Schools, used Project COACH. Unfortunately, their rubric was not publicly available.

12: The TDOE reported to the NCES the number of students eligible for the federal free lunch program and the number for the reduced lunch program for each school in the state separately until 2014. In 2015 and 2016, the TDOE only reported the combined number. From 2017 to 2019, the TDOE did not report the data on the lunch program at all. On the other hand, the TDOE makes publicly available data on the number of economically disadvantaged students in the school profile data file from 2010. However, the TDOE's definition of *economically disadvantaged students* changed in 2016 (Rainwater, 2017). Until 2015, the definition included all students receiving federal free and reduced lunches. In 2016, the definition included only "students who are directly certified to receive free lunch without the need to complete the household application" (Tennessee Department of Education, 2018, p. 68). The definition is similar to the number of students eligible for the federal free lunch program. For these reasons, I used the CCD data on the number of students eligible for the federal free lunch program from

2012 to 2014 and the TDOE's data on the number of economically disadvantaged students from 2015.

13: Accountability data were unavailable in 2016 for elementary and middle schools because the state assessment was suspended due to a technical glitch. I imputed this year's accountability data based on a regression analysis, using the previous years' accountability data, school and district characteristics, and year fixed effects. Since the standard errors in the DID_M estimator were estimated through Bootstrap, the uncertainty in this imputation method was simultaneously resolved.

14: The median maximum bonus amount was $2,000 across the schools and years. As a sensitivity check, I used three other ordinal bonus amount treatment variables with different cutoffs. The first one used three quartiles to create four categories (i.e., 1st quarter = $600, 2nd quarter = $2,000, 3rd quarter = $5,250) plus one category for no bonus. The second one used $4,500, which was close to the mean of the principal salary (=$4,419) to create three categories (i.e., $0, $1 to $4,500, and $4,501 and above). The last used $2,500 and $5,000 to create four arbitrary categories. Although the sizes of the estimated coefficients became larger when the second and third treatment variables were used, all estimates were statistically distinguishable from zero, consistent with the main result. The results are reported in Online Appendix B Table 1. All models used LOE scores.

15: The data in the table were calculated using the PBCS data only. They do not reflect the analytic sample.

16: The percentages in each category were lower than the data reported in Grissom et al. (2019) and Grissom and Bartanen (2019) due mainly to the different analytic samples used in their studies. I obtained a similar result when I used an unrestricted sample (i.e., full sample) similar to theirs.

17: The patterns in the bottom three plots were less clear because of the broader scales used due to the wider confidence intervals.

18: Data were taken from Census Reporter and the TDOE's 2022 school profile data file.

References

Author. (2023). Manuscript submitted for publication.

Akiba, M. & Reichardt, R. (2004). What predicts the mobility of elementary school leaders? An analysis of longitudinal data in Colorado. *Education Policy Analysis Archives, 12*(18), 1-20. https://doi.org/10.14507/epaa.v12n18.2004

Baker, B., Punswick, E., & Belt, C. (2010). School leadership stability, principal moves, and departures: Evidence from Missouri. *Educational Administration Quarterly, 46*(4), 523–557. https://doi.org/10.1177/0013161X10383832

Bartanen, B., Grissom, J. A., & Rogers, L. K. (2019). The impacts of principal turnover. *Educational Evaluation and Policy Analysis, 41*(3), 350-374. https://doi.org/10.3102/0162373719855044

Béteille, T., Kalogrides, D., & Loeb, S. (2012). Stepping Stones: Principal career paths and school outcomes. *Social Science Research, 41*(4), 904-919. https://doi.org/10.1016/j.ssresearch.2012.03.003

Borusyak, K., Jaravel, X., & Spiess, J. (2022). *Revisiting event study designs: Robust and efficient estimation*. Retrieved from https://arxiv.org/pdf/2108.12419.pdf

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). *Estimating the effect of leaders on public sector productivity: The case of school principals* (NBER Working Paper No. 17803). National Bureau of Economic Research. http://www.nber.org/papers/w17803

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

Buelens, M., & Van den Broeck, H. (2007). An analysis of differences in work motivation between public and private sector organizations. *Public Administration Review, 67*(1): 65–74. https://doi.org/10.1111/j.1540-6210.2006.00697.x

Chiang, H., Speroni, C., Herrmann, M., Hallgren, K., Burkander, P., & Wellington, A. (2017). *Evaluation of the Teacher Incentive Fund: Final report on implementation and impacts of pay-for-performance across four years* (NCEE 2017-4004). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Clotfelter, C., Ladd, H. F., Vigdor, J., & Wheeler, J. (2007). High-poverty schools and the distribution of teachers and principals. *North Carolina Law Review*, *85*, 1345-1379.

Cullen, J. B. & Mazzeo, M. J. (2007). *Implicit performance awards: An empirical analysis of the labor market for public school administrators* (working paper). Retrieved from http://www.kellogg.northwestern.edu/faculty/mazzeo/htm/txppals_1207.pdf.

de Chaisemartin, C., & D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review, 110*(9), 2964-2996. https://doi.org/10.1257/aer.20181169

de Chaisemartin, C., & D'Haultfoeuille, X. (2023). *Difference-in-differences estimators of intertemporal treatment effects*. http://dx.doi.org/10.2139/ssrn.3731856

de Chaisemartin, C., & D'Haultfoeuille, X. (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*. https://doi.org/10.1093/ectj/utac017

DeMatthews, D.E., Reyes, P., Carrola, P., Edwards, W., & James, L. (2023). Novice principal burnout: exploring secondary trauma, working conditions, and coping strategies in an urban district. *Leadership and Policy in Schools*, 22(1), 181-199. https://doi.org/10.1080/15700763.2021.1917624

Gates, S. M., Ringel, J. S., Santibanez, L., Guarino, C., Ghosh-Dastidar, B., & Brown, A. (2006). Mobility and turnover among school principals. *Economics of Education Review, 25*(3), 289-302. https://doi.org/10.1016/j.econedurev.2005.01.008

Glazerman, S., Chiang, H., Wellington, A. J., Constantine, J. M., & Player, D. (2011). *Impacts of performance pay under the teacher incentive fund: Study design report*. Mathematica Policy Research.

Godhaber, D. (2007). *Principal compensation: More research needed on a promising reform*. Center for American Progress.

Goldhaber, D., DeArmond, M., & Player, D., Choi, H. (2008). Why do so few public school districts use merit pay? *Journal of Education Finance, 33*(3), 262-289.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics, 225*(2), 254-277. https://doi.org/10.1016/j.jeconom.2021.03.014

Goodman, S. F., & Turner, L. J. (2013). The design of teacher incentive pay and educational outcomes: Evidence from the New York City bonus program. *Journal of Labor Economics, 31*(2), 409-420. 409-420. https://doi.org/10.1086/668676

Grissom, J. A., & Andersen, S. (2012). Why superintendents turn over. *American Educational Research Journal, 49*(6): 1146–1180. https://doi.org/10.3102/0002831212462622

Grissom, J. A., & Bartanen, B. (2019). Principal effectiveness and principal turnover. *Education Finance and Policy, 14*(3), 355-382. https://doi.org/10.1162/edfp_a_00256

Grissom, J. A., Bartanen, B., & Mitani, H. (2019). Principal sorting and the distribution of principal quality. *AERA Open, 5*(2), 1-21. https://doi.org/10.1177/2332858419850094

Grissom, J. A., Blissett R. S. L., & Mitani, H. (2018). Evaluating school principals: Supervisor rating of principal practice and principal job performance. *Educational Evaluation and Policy Analysis, 40*(3), 446-472. https://doi.org/10.3102/0162373718783883

Grissom, J. A., Egalite, A. J., & Lindsay, C. A. (2021). *How principals affect students and schools: A systematic synthesis of two decades of research*. Wallace Foundation.

Grissom, J. A., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis, 37*(1), 3-28. https://doi.org/10.3102/0162373714523831

Grissom, J. A., Loeb, S., & Master, B. (2013). Effective instructional time use for school leaders: Longitudinal evidence from observations of principals. *Educational Researcher*, *42*(8), 433-444. https://doi.org/10.3102/0013189X13510020

Grissom, J. A., & Mitani, H. (2016). Salary, performance, and superintendent turnover. *Educational Administration Quarterly, 52*(3): 351–391. https://doi.org/10.1177/0013161X15627677

Hallinger, P. & Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980-1995. *School Effectiveness and School Improvement, 9*(2), 157-191. https://doi.org/10.1080/0924345980090203

Hamilton, L. S., Engberg, J., Steiner, E. D., Nelson, C. A., & Yuan, K. (2012). *Improving school leadership through support, evaluation, and incentives*. RAND.

Heyburn, S., Lewis, J., & Ritter, G. (2010). *Compensation reform and design preferences of Teacher Incentive Fund grantees*. National Center on Performance Incentives, Vanderbilt University.

Hill, A. J., & Jones, D. B. (2020). The impacts of performance pay on teacher effectiveness and retention: Does teacher gender matter? *The Journal of Human Resources, 55*(1), 349-385. https://doi.org/10.3368/jhr.55.2.0216.7719R3

Hochbein, C., Mahone, A., & Vanderbeck, S. (2021). A systematic review of principal time use research. *Journal of Educational Administration, 59*(2), 215-232. https://doi-org.ezproxy.rowan.edu/10.1108/JEA-09-2019-0163

Holley, M. J. (2009). *A study of the school principal labor market in Arkansas: Implications for incentive-based compensation policies to improve principal quality*. (Doctoral dissertation). UMI Dissertation Publishing.

Kozlowski, K. P., & Lauen, D. L. (2019). Understanding teacher pay for performance: Flawed assumptions and disappointing results. *Teachers College Record, 121*(2), 1-38. https://doi.org/10.1177/016146811912100206

Lavy, V. (2008). Does raising the principal's wage improve the school's outcomes? Quasi-experimental evidence from an unusual policy experiment in Israel. *The Scandinavian Journal of Economics, 110*(4), 639-662. http://doi.org/10.1111/j.1467-9442.2008.00555.x

Levacic, R. (2009). Teacher incentives and performance: An application of principal-agent theory. *Oxford Development Studies, 37*(1), 33-46. https://doi.org/10.1080/13600810802660844

Loeb, S., Kalogrides, D., & Horng, E. L. (2010). Principal preferences and the even distribution of principals across schools. *Educational Evaluation and Policy Analysis, 32*(2), 205-229. https://doi.org/10.3102/0162373710369833

Meyer, M. J., & Macmillan, R. B., Northfield, S. (2009). Principal succession and its impact on teacher morale. *International Journal of Leadership in Education, 12*(2), 171-185. https://doi.org/10.1080/13603120802449660

Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review, 36(*1), 60-72. https://doi.org/10.1016/j.econedurev.2013.05.004

Mintrop, R., Ordenes, M., Coghlan, E., Pryor, L., & Madero, C. (2018). Teacher evaluation, pay for performance, and learning around instruction: Between dissonant incentives and resonant procedures. *Educational Administration Quarterly, 54*(1), 3-46. https://doi.org/10.1177/0013161X176965

Mitani, H. (2018). Principals' working conditions, job stress, and turnover behaviors under NCLB accountability pressure. *Educational Administration Quarterly, 54*(5), 822-862. https://doi.org/10.1177/0013161X18785874

Mitani, H. (2019). Principal turnover under NCLB accountability pressure. *Teachers College Record, 121*(2), 1-44. https://doi.org/10.1177/016146811912100203

Murphy, J., & Torres, D. (2014). *Creating productive cultures in schools*. Corwin.

Oplatka, I. (2016). "Irresponsible leadership" and unethical practices in schools: A conceptual framework of the "dark side" of educational leadership. In A. H. Normore, & J. S. Brooks (Eds.), *The dark side of leadership: Identifying and overcoming unethical practice in organizations* (*Advances in Educational Administration, Vol. 26*) (pp. 1-18). Emerald Group Publishing Limited. https://doi.org/10.1108/S1479-366020160000026001

Papa, F. C., Jr., Langford, H., & Wycoff, J. (2002). *The attributes and career paths of principals: Implications for improving policy*. State University of New York.

Pendola, A. (2022). Your pay or someone else's? Exploring salary dispersion, position, and principal turnover. *Education Finance and Policy, 17*(3), 511-540. https://doi.org/10.1162/edfp_a_00338

Perry, J. L., Engbers, T. A., & Jun, S. Y. (2009). Back to the future? Performance-related pay, empirical research, and the perils of persistence. *Public Administration Review, 69*(1), 39-51. https://doi.org/10.1111/j.1540-6210.2008.01939_2.x

Perry, J. L., Mesch, D., & Paarlberg, L. (2006). Motivating employees in a new governance era: The performance paradigm revisited. *Public Administration Review*, 66(4), 505-514. https://doi.org/10.1111/j.1540-6210.2006.00611.x

Rainwater, K. A. (2017, February 6). New definition means fewer Tennessee students are considered economically disadvantaged. *Chattanooga Times Free Press*. https://www.timesfreepress.com/news/local/story/2017/feb/06/state-redefines-student-poverty/411354/?bcsubid=f0b3b4ea-6322-4edc-aebb-fa74ed87c23e&pbdialog=reg-wall-login-created-tfp

Ritz A., Brewer, G. A., & Neumann, O. (2016). Public service motivation: A systematic literature review and outlook. *Public Administration Review, 76*(3), 414-426. https://doi.org/10.1111/puar.12505

Roza, M. (2003). *A matter of definition: Is there truly a shortage of school principals?* Center on Reinventing Public Education, University of Washington.

Sebastian, J., & Allensworth, E. (2012). The influence of principal leadership on instruction and student learning: A study of mediated pathways to learning. *Educational Administration Quarterly, 48*(4), 626-663. https://doi.org/10.1177/0013161X11436273

Shifrer, D., Turley, R. L., & Heard, H. (2017). Do teacher financial awards improve teacher retention and student achievement in an urban disadvantaged school district? *American Educational Research Journal, 54*(6), 1117-1153. http://doi.org/10.3102/0002831217716540

Steiner, E. D., Doan, S., Woo, A., Gittens, A. D., Lawrence, R. A., Berdie, L., Wolfe, R. Greer, L., & Schwartz, H. L. (2022). *Restoring teacher and principal well-being is an essential step for rebuilding schools: Findings from the State of the American Teacher and State of the American Principal Surveys*. RAND.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics, 225*(2), 175-199. https://doi.org/10.1016/j.jeconom.2020.09.006

Tennessee Department of Education. (2018). *Every Student Succeeds Act: Building on success in Tennessee – ESSA State Plan*. Author.

Tennessee State Board of Education. (2017). *Strategic compensation policy: 5.600*. Retrieved from https://www.tn.gov/content/dam/tn/stateboardofeducation/documents/2019-sbe-meetings/november-15%2C-2019-sbe-meeting/11-15-19%20II%20I%20Strategic%20Compensation%20Policy%205.600%20Attachment%20Clean%20Copy.pdf

Tran, H. (2017). The impact of pay satisfaction and school achievement on high school principals' turnover intentions. *Educational Management Administration & Leadership, 45*(4), 621–638. https://doi-org.ezproxy.rowan.edu/10.1177/1741143216636115

Wallace Foundation. (2013). *The school principal as leader: Guiding schools to better teaching and learning*. Author.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced Leadership: What 30 years of research tells us about the effect of leadership on student achievement* (Working paper). Mid-continent Research for Education and Learning.

Weibel, A., Rost, K., & Osterloh, M. (2010). Pay for performance in the public sector – Benefits and (hidden) costs. *Journal of Public Administration Research and Theory, 20*(2), 387-412.

Table 1

Descriptive statistics

	N	Mean	SD	Min	Max
Principal characteristics					
Age	10,470	48.19	8.37	22	82
Female	10,619	0.58	0.49	0	1
Race/ethnicity					
Non-white	10,543	0.26	0.44	0	1
Degree level					
EdS or doctorate	10,652	0.48	0.50	0	1
Years of principal experience	10,564	5.61	3.72	1	33
Tenure	10,564	4.33	3.12	1	18
Salary	10,653	88,838	17,337	0	235,511
Principal evaluation scores					
Level of effectiveness	8,931	3.81	1.06	1	5
Average subjective ratings	8,928	3.86	0.54	1	5
School characteristics					
School level					
Elementary school	10,557	0.46	0.50	0	1
Middle school	10,557	0.16	0.37	0	1
High school	10,557	0.17	0.38	0	1
Elementary - middle school	10,557	0.16	0.37	0	1
Elementary - high school	10,557	0.01	0.12	0	1
Middle - high school	10,557	0.03	0.17	0	1
Locality					
Urban	10,653	0.29	0.45	0	1
Suburban	10,653	0.16	0.37	0	1
Town	10,653	0.17	0.37	0	1

Rural	10653	0.38	0.49	0	1
Percent free lunch	10,592	45.23	22.78	0	100
Percent color	10,639	28.05	30.28	0	100
School enrollment size	10,652	605	370	0	2,789
School accountability					
Percent on-track and mastered reading	8,906	43.61	17.79	0	100
Percent on-track and mastered math	8,843	45.88	18.95	0	99
Percent on-track and mastered combined	8,931	44.58	17.90	0	99
District characteristics					
Number of schools	10,653	52.90	67.43	1	290
Percent free lunch	10,653	41.97	16.47	0	98
Percent color	10,653	27.78	26.14	0	97
Enrollment size	10,653	42,068	75,750	2	557,015
Per-pupil expenditure	10,652	10,233	1,288	7,725	15,791

Note. Data from 2012 to 2020 were used, except for principal evaluation scores and school accountability measures. These data were unavailable in 2020 due to COVID-19. The percentage of color includes black and Hispanic students. The school accountability data in 2016 were imputed using multiple regression analyses. Principal salary and per-pupil expenditure were adjusted for a 2020 constant dollar. District characteristics were not weighted by the number of schools.

Table 2

Characteristics of PBCS

	# PBCS schools	Percentage	N. of schools that implemented PBCS for the first time	N. of schools that discontinued PBCS at the end of the year	Maximum bonus amount			Performance measures			
					Minimum	Median	Maximum	N. of schools that used TEAM administrator evaluation scores	N. of schools that used school-level TVAAS scores	N. of schools that used either TEAM or TVAAS scores or both	Percentage of schools either TEAM or TVAAS scores or both
2012	314	18	314	0	1,000	2,000	15,000	169	133	221	70
2013	320	18	6	4	1,000	2,850	10,000	160	140	228	71
2014	333	18	18	10	1,000	2,850	10,000	164	151	240	72
2015	462	26	145	162	250	2,000	10,250	245	184	359	78
2016	337	18	42	61	250	620	10,250	140	93	209	62
2017	279	15	11	28	300	2,000	6,000	128	131	235	84
2018	241	14	0	16	300	2,000	9,000	109	118	216	90
2019	233	13	9	103	250	400	9,000	98	120	205	88
2020	152	8	19	NA	150	600	9,000	83	36	110	72
Total	2671	16	NA	NA	150	2,000	15,000	1,296	1,106	2,023	76

Note. The total number of PBCS schools in each year is not necessarily equal to the sum of the number of PBCS schools in the previous year and the number of new PBCS schools in the current year minus the number of schools that discontinued PBCS at the end of the previous year because there were schools whose treatment status switched on and off multiple times. The third wave of PBCS implementation took place at the district level in most cases. As a result, the medians reflect the district sizes (i.e., the number of schools in the district).

Table 3

Principal turnover by types and PBCS status

Panel A: All principals	All turnover	Within-district transfer	Across-district transfer	Exit or position change
2012	15.33	5.76	0.56	9.01
2013	10.38	3.09	0.17	7.12
2014	10.32	3.19	0.66	6.47
2015	13.32	3.99	0.58	8.75
2016	12.19	2.71	0.23	9.25
2017	15.25	2.69	0.45	12.11
2018	16.43	2.21	0.33	13.89
2019	14.25	3.69	0.50	10.06
All years	13.33	3.41	0.43	9.49
Panel B: By PBCS status	**All turnover**	**Within-district transfer**	**Across-district transfer**	**Exit or position change**
Principals at PBCS schools	17.11***	4.63***	0.28	12.20***
Principals at Non-PBCS schools	12.39	3.11	0.46	8.82
All principals	13.33	3.41	0.43	9.49

Note. The analytic sample was used to calculate percentages. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 4

Impact of PBCS on principal turnover

Panel A: PBCS - binary	LOE scores used		Scores in subjective ratings used	
	Model 1	Model 2	Model 3	**Model 4**
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
Effect of PBCS	0.01	-0.01	0.00	0.01
	(0.03)	(0.03)	(0.03)	(0.04)
N	4626	4262	5972	5972
Panel B: PBCS - bonus	LOE scores used		Scores in subjective ratings used	
	Model 5	Model 6	Model 7	Model 8
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
Effect of PBCS	-0.02	0.01	0.02	0.06
	(0.09)	(0.08)	(0.09)	(0.08)
N	4639	4639	5965	5965

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 5

Impact of PBCS on effective principals and principals serving high-need schools

Panel A: Effective principals	PBCS - binary		PBCS - bonus	
	Model 1	Model 2	Model 3	Model 4
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(effective)	0.03	0.02	0.02	0.02
	(0.03)	(0.04)	(0.05)	(0.04)
N	6243	6243	5941	5941
Panel B: Principals at high-poverty schools	PBCS - binary		PBCS - bonus	
	Model 5	Model 6	Model 7	Model 8
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(high-poverty)	0.00	0.00	-0.13	-0.03
	(0.04)	(0.04)	(0.09)	(0.07)
N	4401	4401	4204	4204
Panel C: Principals at high-color schools	PBCS - binary		PBCS - bonus	
	Model 9	**Model 10**	Model 11	Model 12
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(high-color)	0.10	0.09	0.10	0.07
	(0.10)	(0.13)	(1.44)	(3.45)
N	3544	3544	4195	4195

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * p < .10, ** p < .05, *** p < .01.

Table 6

Impact of PBCS among effective principals serving high-need schools

Panel A: Effective principals serving high-poverty schools	PBCS - binary		PBCS - bonus	
	Model 1	Model 2	Model 3	Model 4
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(effective)	0.03	0.02	0.04	0.05
	(0.03)	(0.03)	(0.10)	(0.11)
N	3188	3188	3015	3015
Panel B: Effective principals serving high-color schools	PBCS - binary		PBCS - bonus	
	Model 5	Model 6	Model 7	Model 8
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(effective)	-0.06	0.00	-0.21	-0.14
	(2.68)	(2.43)	(1.94)	(1.97)
N	873	873	806	806

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * $p < .10$, ** $p < .05$, *** $p < .01$.

Table 7

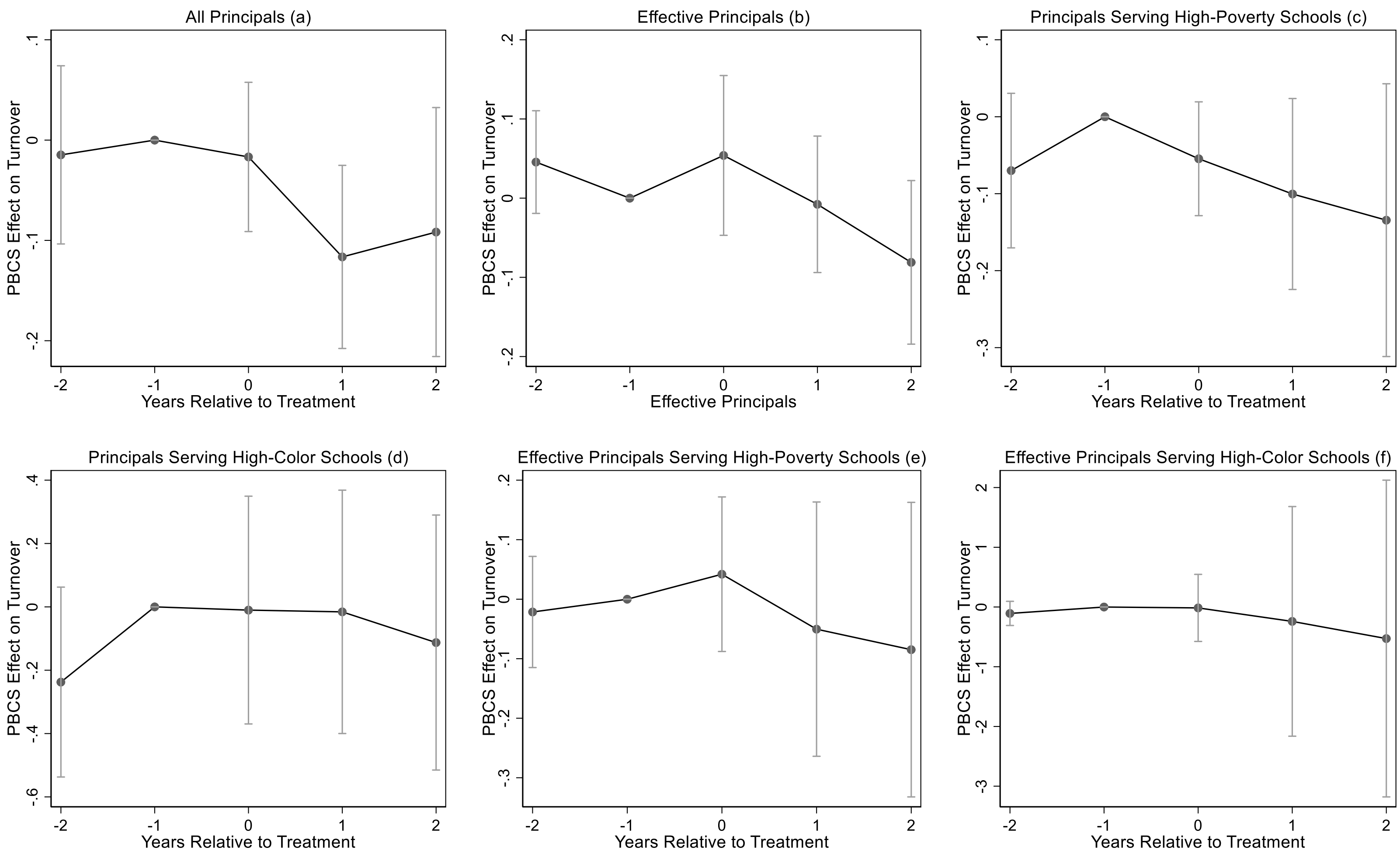
Impact of PBCS on mean principal quality and performance measures of newly hired principals

Panel A: All schools	Model 1	Model 2	Model 3	Model 4
	LOE scores	Scores in subjective ratings	Education specialist or doctoral degree	Years of principal experience
Effect of PBCS	2.16	-0.50	-0.04	-0.18
	(8.31)	(7.87)	(0.09)	(0.49)
N	325	356	1915	1936
Panel B: High-poverty schools	**Model 5**	**Model 6**	**Model 7**	**Model 8**
	LOE scores	Scores in subjective ratings	Education specialist or doctoral degree	Years of principal experience
Effect of PBCS	1.95	0.11	0.10	-0.16
	(2.65)	(2.12)	(0.13)	(0.57)
N	290	310	1692	1709
Panel C: High-color schools	**Model 9**	**Model 10**	**Model 11**	**Model 12**
	LOE scores	Scores in subjective ratings	Education specialist or doctoral degree	Years of principal experience
Effect of PBCS	1.81	-0.03	0.13	0.77
	(7.62)	(3.18)	(0.19)	(0.73)
N	292	313	1717	1734

Note. For LOE scores and scores in subjective ratings, new principals in the system were not included, as their previous years' scores were unavailable.

Figure 1

Intertemporal effects of PBCS on principal turnover



Online Appendix A

Online Appendix A Table 1

Tests on the severity of the heterogeneity issue

Panel A: PBCS - binary	LOE scores used		Scores in subjective ratings used	
	Model 1	Model 2	Model 3	Model 4
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
N of average treatment effects on the treated	144	144	144	144
Positive weights	113	113	114	114
Negative weights	31	31	30	30
Proportion negative weights	0.22	0.22	0.21	0.21
Sum of negative weights	-34.41	-34.41	-29.19	-29.19
First measure	0.01	0.00	0.01	0.00
Second measure	0.00	0.00	0.00	0.00
Correlation of district-level PPE	-0.11	-0.11	-0.11	-0.11
t-statistic	-1.34	-1.34	-1.27	-1.27
Panel B: PBCS - bonus	LOE scores used		Scores in subjective ratings used	
	Model 5	Model 6	Model 7	Model 8
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
N of average treatment effects on the treated	103	103	103	103
Positive weights	90	90	92	92
Negative weights	13	13	11	11
Proportion negative weights	0.13	0.13	0.11	0.11
Sum of negative weights	-23.12	-23.12	-27.61	-27.61
First measure	0.02	0.02	0.02	0.02
Second measure	0.00	0.00	0.00	0.00
Correlation of district-level PPE	0.07	0.07	0.08	0.08
t-statistic	0.52	0.52	0.58	0.58

Note. All models used the principal level of effectiveness scores. PPE stands for per-pupil expenditures in a 2020

Online Appendix A Table 2

Tests on the parallel trends assumption

Panel A: PBCS - binary	LOE scores used		Scores in subjective ratings used	
	Model 1	Model 2	Model 3	Model 4
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
Placebo effect	0.04	0.04	0.05	0.07**
	(0.04)	(0.03)	(0.04)	(0.03)
N	2961	2961	4161	4161
Panel B: PBCS - bonus	LOE scores used		Scores in subjective ratings used	
	Model 5	Model 6	Model 7	Model 8
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
Placebo effect	0.00	-0.02	0.03	0.01
	(0.03)	(0.04)	(0.02)	(0.03)
N	2777	2777	3949	3949

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$.

Online Appendix A Table 3

State-developed rubrics for the subjective ratings of leadership practices

Year		Dimension
2012	A	Continuous school improvement
	B	Culture for teaching and learning
	C	Instructional leadership and assessment
	D	Professional growth
	E	Management of the school
	F	Ethics
	G	Diversity
	H	Quality of teacher evaluation
2013	A	Quality of teacher evaluation
	B	Instructional leadership
	C	Continuous improvement
	D	Culture for teaching and learning
	E	Talent and operations management
	F	Diversity
	G	Ethics
2014	A	Quality of teacher evaluation
	B	Instructional leadership
	C	Continuous improvement
	D	Culture for teaching and learning
	E	Talent and operations management
	F	Diversity
	G	Ethics
2014	A	Instructional leadership for continuous improvement
Pilot	B	Culture for teaching and learning
	C	Professional learning and growth
	D	Resource management
2015	A	Instructional leadership for continuous improvement
	B	Culture for teaching and learning
	C	Professional learning and growth
	D	Resource management
2016	A	Instructional leadership for continuous improvement
	B	Culture for teaching and learning
	C	Professional learning and growth
	D	Resource management
2017	A	Instructional leadership for continuous improvement
	B	Culture for teaching and learning
	C	Professional learning and growth

	D	Resource management
2018	A	Instructional leadership for continuous improvement
	B	Culture for teaching and learning
	C	Professional learning and growth
	D	Resource management
2019	A	Instructional leadership for continuous improvement
	B	Culture for teaching and learning
	C	Professional learning and growth
	D	Resource management

Online Appendix B

Online Appendix B Table 1

Impact of PBCS on principal turnover – three maximum bonus amount treatment variables

	Bonus (quartiles)		Bonus (mean)		Bonus (arbitrary)	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS	-0.05	0.01	0.15	0.18	0.35	0.40
	(0.19)	(0.15)	(0.12)	(0.13)	(0.77)	(0.73)
N	4545	4545	4433	4433	4500	4500

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * $p < .10$, ** $p < .05$, *** $p < .01$.

Online Appendix B Table 2

Impact of PBCS among effective principals and principals serving high-need schools – three maximum bonus amount treatment variables

Panel A: Effective principals	Bonus (quartiles)		Bonus (mean)		Bonus (arbitrary)	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(effective)	0.30	0.32	0.04	0.05	0.32	0.33
	(2.82)	(2.75)	(0.20)	(0.14)	(1.20)	(0.89)
N	5940	5940	5921	5921	5924	5924
Panel B: Principals serving high-poverty schools	Bonus (quartiles)		Bonus (mean)		Bonus (arbitrary)	
	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(high-poverty)	-0.28	-0.09	0.08	0.15	0.75	0.79
	(99.87)	(90.41)	(0.15)	(0.14)	(1.93)	(1.93)
N	4196	4196	4125	4125	4240	4240
Panel C: Principals serving high-color schools	Bonus (quartiles)		Bonus (mean)		Bonus (arbitrary)	
	Model 13	Model 14	Model 15	Model 16	Model 17	Model 18
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(high-color)	-0.12	-0.26	0.07	0.18	0.09	0.20
	(2.62)	(2.80)	(3.81)	(2.28)	(1.42)	(7.71)
N	4185	4185	4173	4173	4181	4181

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * $p < .10$, ** $p < .05$, *** $p < .01$.

Online Appendix B Table 3

Impact of PBCS among effective principals serving high-need schools – three maximum bonus amount treatment variables

Panel A: Effective principals serving high-poverty schools	Bonus (quartiles)		Bonus (mean)		Bonus (arbitrary)	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
Effect of PBCS	0.46	0.65	0.07	0.08	1.19	1.19
	(0.77)	(0.73)	(2.31)	(2.34)	(1.45)	(1.48)
N	2999	2999	3006	3006	3005	3005
Panel B: Effective principals serving high-color schools	Bonus (quartiles)		Bonus (mean)		Bonus (arbitrary)	
PBCS - bonus	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
Effect of PBCS	0.13	0.15	-0.16	0.00	0.04	0.05
	(0.71)	(0.67)	(2.08)	(2.03)	(1.04)	(1.02)
N	788	788	808	808	807	807

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * $p < .10$, ** $p < .05$, *** $p < .01$.

Online Appendix B Table 4

Intertemporal effects of PBCS on principal turnover

Panel A: Turnover (1st type)	All principals	Effective principals	Principals serving high-poverty schools	Principals serving high-color schools	Effective principals serving high-poverty schools	Effective principals serving high-color schools
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
PBCS 1st year	-0.02	0.05	-0.05	-0.01	0.04	-0.02
	(0.04)	(0.05)	(0.04)	(0.18)	(0.07)	(0.29)
PBCS 2nd year	-0.12**	-0.01	-0.10	-0.02	-0.05	-0.24
	(0.05)	(0.05)	(0.06)	(0.20)	(0.11)	(0.98)
PBCS 3rd year	-0.09	-0.08	-0.13	-0.11	-0.08	-0.53
	(0.06)	(0.05)	(0.09)	(0.21)	(0.13)	(1.35)
N	NA	NA	NA	NA	NA	NA
Panel B: Turnover (2nd type)	All principals	Effective principals	Principals serving high-poverty schools	Principals serving high-color schools	Effective principals serving high-poverty schools	Effective principals serving high-color schools
	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
PBCS 1st year	-0.03	0.04	-0.04	-0.10	0.03	0.07
	(0.04)	(0.05)	(0.04)	(0.17)	(0.05)	(0.30)
PBCS 2nd year	-0.08*	-0.02	-0.05	0.08	-0.03	-0.13
	(0.04)	(0.05)	(0.06)	(0.18)	(0.07)	(0.89)
PBCS 3rd year	-0.07	-0.07	-0.08	-0.01	-0.03	-0.22
	(0.05)	(0.05)	(0.07)	(0.19)	(0.11)	(1.28)
N	NA	NA	NA	NA	NA	NA

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Since the sample sizes varied from year to year after PBCS implementation, NAs were used. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * $p < .10$, ** $p < .05$, *** $p < .01$.

Online Appendix B Table 5

Impact of PBCS by school locality

Panel A: Principal serving urban schools	All principals		Effective principals	
	Model 1	**Model 2**	Model 3	Model 4
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(urban)	0.11	1.21***	-0.03	0.09
	(0.12)	(0.45)	(0.33)	(0.24)
N	505	505	195	195
Panel B: Principals serving rural schools	All principals		Effective principals	
	Model 5	Model 6	Model 7	Model 8
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(rural)	0.01	0.00	0.10	0.05
	(0.02)	(0.03)	(0.07)	(0.07)
N	1896	1896	473	473

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * $p < .10$, ** $p < .05$, *** $p < .01$.

Online Appendix B Table 6

Impact of PBCS by school locality – maximum bonus amount

Panel A: Principal serving urban schools	All principals		Effective principals	
	Model 1	Model 2	Model 3	Model 4
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(urban)	0.04	0.19	0.05	0.24
	(0.47)	(0.57)	(2.40)	(2.57)
N	642	642	195	195
Panel B: Principals serving rural schools	All principals		Effective principals	
	Model 5	Model 6	Model 7	Model 8
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(rural)	-0.07	-0.07	0.64	0.90
	(0.08)	(0.09)	(129.61)	(77.81)
N	1859	1859	461	461

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * $p < .10$, ** $p < .05$, *** $p < .01$.

Online Appendix B Table 7

Impact of non-TIF-funded PBCS among all principals and effective principals

Panel A: All principals	PBCS - binary		PBCS - bonus	
	Model 1	**Model 2**	**Model 3**	**Model 4**
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS	-0.03	-0.02	-0.08	0.04
	(0.05)	(0.04)	(0.83)	(0.86)
N	3382	3382	3277	3277
Panel B: Effective principals	PBCS - binary		PBCS - bonus	
	Model 5	Model 6	Model 7	Model 8
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(effective)	0.01	0.03	16.24**	16.27**
	(0.03)	(0.03)	(6.61)	(6.61)
N	5308	5308	4353	4353

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * $p < .10$, ** $p < .05$, *** $p < .01$.

Online Appendix B Table 8

Impact of non-TIF-funded PBCS among principals serving high-need schools

Panel A: Principals at high-poverty schools	PBCS - binary		PBCS - bonus	
	Model 1	**Model 2**	**Model 3**	**Model 4**
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(high-poverty)	-0.08	-0.05	-0.23	-0.04
	(0.07)	(0.07)	(3.99)	(3.98)
N	3008	3008	2908	2908
Panel B: Principals at high-color schools	PBCS - binary		PBCS - bonus	
	Model 5	**Model 6**	Model 7	**Model 8**
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(high-color)	0.00	-0.05	-0.02	-0.06
	(0.16)	(0.17)	(0.13)	(0.13)
N	1485	1485	2206	2206

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * $p < .10$, ** $p < .05$, *** $p < .01$.

Online Appendix B Table 9

Impact of non-TIF-funded PBCS among effective principals serving high-need schools

Panel A: Effectve principals at high-poverty schools	PBCS - binary		PBCS - bonus	
	Model 1	Model 2	Model 3	Model 4
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(high-poverty)	0.00	0.01	6.59	6.62
	(0.04)	(0.04)	(35.05)	(35.82)
N	2677	2677	2123	2123
Panel B: Effective principals at high-color schools	PBCS - binary		PBCS - bonus	
	Model 5	Model 6	Model 7	Model 8
	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)	Turnover (did not return to the same school as a principal in the following year)	Turnover (did not return to the same district as a principal in the following year)
PBCS * I(high-color)	0.39	0.44	7.67	7.79
	(3.67)	(3.67)	(9.95)	(9.95)
N	415	415	224	224

Note. All models included year fixed effects, school fixed effects, and time-variant characteristics of principals, schools, and school districts. Principal and school characteristics only included time-variant characteristics. Standard errors were estimated through Bootstrap, clustered at the school level, and reported in parentheses. Models bolded included school-specific linear trends due to the violation of the parallel trends assumption. * $p < .10$, ** $p < .05$, *** $p < .01$.

Online Appendix B Figure 1

Intertemporal effects of PBCS – maximum bonus amount