ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-190

Does Principal Professional Development Improve Schooling Outcomes? Evidence from Pennsylvania's Inspired Leadership Induction Program

Matthew P. Steinberg
George Mason University

Haisheng Yang
Abt Associates

Principals shape the academic setting of schools. Yet, there is limited evidence on whether principal professional development improves schooling outcomes. Beginning in 2008-09, Pennsylvania's Inspired Leadership (PIL) induction program required that newly hired principals complete targeted in-service professional development tied to newly established state leadership standards within five years of employment. Using panel data on all Pennsylvania students, teachers, and principals, we leverage variation in the timing of PIL induction across principal-school cells and employ difference-in-differences and event study strategies to estimate the impact of PIL induction on teacher and student outcomes. We find that PIL induction increased student math achievement through improvements in teacher effectiveness, and that the effects of PIL induction on teacher effectiveness were concentrated among the most economically disadvantaged and urban schools in Pennsylvania. Principal professional development had the greatest impact on teacher effectiveness when principals completed PIL induction during their first two years in the principalship. We also find evidence that teacher turnover declined in the years following the completion of PIL induction. We discuss the implications of our findings for principal induction efforts.

VERSION: November 2021

Suggested citation: Steinberg, Matthew P., and Haisheng Yang. (2021). Does Principal Professional Development Improve Schooling Outcomes? Evidence from Pennsylvania's Inspired Leadership Induction Program. (EdWorkingPaper: 20-190). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/z46p-cq19

Does Principal Professional Development Improve Schooling Outcomes? Evidence from Pennsylvania's Inspired Leadership Induction Program

Matthew P. Steinberg
George Mason University

Haisheng Yang
Abt Associates

October 19, 2021

The authors thank the Pennsylvania Department of Education for providing data access and, in particular, David Volkman, Carla Wilson, and Thresa Smith for their guidance and insights about the Pennsylvania Inspired Leadership (PIL) induction program. Matthew Steinberg (corresponding author) is an associate professor at George Mason University (msteinb6@gmu.edu). Haisheng Yang is a Senior Analyst at Abt Associates (haisheng_yang@abtassoc.com).

Abstract

Principals shape the academic setting of schools. Yet, there is limited evidence on whether principal professional development improves schooling outcomes. Beginning in 2008-09, Pennsylvania's Inspired Leadership (PIL) induction program required that newly hired principals complete targeted in-service professional development tied to newly established state leadership standards within five years of employment. Using panel data on all Pennsylvania students, teachers, and principals, we leverage variation in the timing of PIL induction across principal-school cells and employ difference-in-differences and event study strategies to estimate the impact of PIL induction on teacher and student outcomes. We find that PIL induction increased student math achievement through improvements in teacher effectiveness, and that the effects of PIL induction on teacher effectiveness were concentrated among the most economically disadvantaged and urban schools in Pennsylvania. Principal professional development had the greatest impact on teacher effectiveness when principals completed PIL induction during their first two years in the principalship. We also find evidence that teacher turnover declined in the years following the completion of PIL induction. We discuss the implications of our findings for principal induction efforts.

Keywords: Principal induction; professional development; educator mobility; student achievement; teacher effectiveness

Introduction

Effective school leadership is vital to school improvement efforts. Effective principals attract and retain more effective teachers (Ladd, 2011; Loeb et al., 2012), promote teacher learning and instructional development (Robinson et al., 2008; Steinberg & Sartain, 2015), and improve staff motivation, commitment, and working conditions (Leithwood et al., 2008). Further, student achievement improves when principals have greater decision-making autonomy in their schools (Bloom et al., 2015; Clark, 2009; Steinberg, 2014; Steinberg & Cox, 2017). Indeed, while a widely-cited report on principal effectiveness concluded that principal leadership is second only to classroom instruction (among all school-level factors) in its influence on student learning (Leithwood et al., 2004), a more recent synthesis of the empirical evidence on principal effectiveness from the past two decades suggests that prior evidence has likely understated the importance of school principals (Grissom et al., 2021). Yet, despite the widely acknowledged importance of principal human capital (Ladd, 2011; Loeb et al., 2012; Leithwood et al., 2004; Grissom et al., 2021), little empirical evidence exists linking principal professional development to teacher and student outcomes.

In 2007, the Pennsylvania Public School Code was amended to provide principals with targeted professional development designed to place an effective school leader in all Pennsylvania schools. The policy reform, known as *Act 45*, directed the Pennsylvania Department of Education (PDE) to establish a principal induction program which focused on developing the capacity of school leaders to improve student achievement. Beginning in January 2008, all school principals employed for the first time were required to complete the Pennsylvania Inspired Leadership (PIL) induction program within the first five years of employment. According to PDE leadership, the PIL induction program, which was tied to newly

established state leadership standards, was designed to endow new principals with the skills necessary to be effective school leaders.[1]

In this paper, we rely on administrative data on all students, teachers, and principals in Pennsylvania public schools during the 2008-09 through 2015-16 school years to address the following questions: (1) Does principal professional development affect teacher retention? (2) Does principal professional development improve teacher effectiveness? (3) Does principal professional development improve student achievement? Because principal professional development via PIL induction was not randomly assigned across schools in Pennsylvania, we employ event study and difference-in-differences strategies to estimate the effect of principal professional development on teacher and student outcomes.

In the context of our quasi-experimental empirical strategy, we define treatment at the school-year level and leverage variation in the timing of principal completion of PIL induction across principal-school cells. Specifically, our source of identification relies on changes in within-principal-by-school outcomes for principals who remain in the same school and complete professional development activities in different school years. Since treatment is defined at the school-year level and because we rely on variation in the timing of PIL induction across principal-school cells, we estimate the effect of a principal's completion of PIL induction within the same school setting on teacher and student outcomes. And by defining treatment in this way, we aim to avoid concerns that non-random principal transitions might bias estimates of PIL induction. This treatment definition also enables us to model multiple treatment events at the school level in cases where the same school experienced multiple treatment events – i.e.,

[1] Author's communication with David Volkman, Pennsylvania Department of Education Executive Deputy Secretary and PIL Program Leader (June 16, 2016).

principal completion of PIL induction – across the study period. Further, we construct multiple comparison groups to approximate the counterfactual for what would have happened in treatment schools in the absence of principal professional development. The first comparison group includes all school-year observations in the 2008-09 through 2015-16 school years and thus includes all Pennsylvania principals, many of whom completed or were in the process of completing their continuing professional development requirements during the study period. The second comparison group (a subset of the first) restricts school-year comparisons to those containing just novice and early career (i.e., PIL eligible) principals, thereby addressing the possibility that the completion of principal professional development among more experienced principals might understate PIL induction effects.

We find that PIL induction improved student achievement in math through improvements in teacher human capital. Specifically, PIL induction increased student math achievement by 0.01-0.03 student-level standard deviations, and these increases are linked to approximately a 0.20 standard deviation improvement in teachers' effectiveness in math. Improvements in teacher math effectiveness were concentrated among Pennsylvania schools serving the most economically disadvantaged and minority students in urban districts. This pattern of findings suggests that the benefits of principal professional development were not only distributed unevenly across the state's schools but that the returns to principal professional development were greatest in schools serving the state's most economically disadvantaged students. We further find that the impact of principal professional development on improvements in teacher math effectiveness were greatest when principals completed PIL induction during their first two years in the principalship. In contrast, PIL induction had no discernible effect on teacher effectiveness and student achievement in English language arts (ELA). Though there is no

average effect of PIL induction on teacher turnover, teacher turnover declined by approximately 2 percentage points (or 18 percent) in the second and third years following the completion of PIL induction. Together, these results suggest that principal professional development may not only lead to short-term and sustained improvements in teacher effectiveness and student achievement, but may also stabilize a school's teaching force by reducing teacher turnover in the wake of improvements in principal human capital. An important caveat, however, is that despite the robustness of our results across multiple specification checks addressing, among other endogeneity concerns, the non-random selection into PIL induction (i.e., whether or not PIL eligible principals completed PIL induction), the potential endogeneity related to the timing into PIL induction (i.e., when PIL eligible principals completed PIL induction during their early career tenure) is a residual concern that our empirical approach is more limited in its ability to address.

This paper contributes needed empirical evidence on the efficacy of state-level principal induction efforts designed to provide in-service professional development to novice and early-career principals. By examining the consequences of a statewide policy reform in Pennsylvania which codified principal induction for all novice principals statewide, we provide rigorous quasi-experimental evidence on the efficacy of principal induction to improve teacher and student outcomes. Evidence from this paper should inform nascent policy efforts designed to improve principal human capital through early-career induction and professional development.

Related Literature and Policy Context

Though education scholars and school practitioners widely agree on the importance of effective principal leadership, rigorous empirical evidence examining the effect of in-service principal professional development on teacher and student outcomes is limited (Murphy &

Vriesenga, 2006). This limited evidence base reflects state and local policy efforts which typically focus on teacher, rather than principal, professional development despite the acknowledged importance of developing and fostering effective principals (Manna, 2015). Historically, principal professional development has been subsumed under teacher professional development in terms of both content and funding. In fact, most professional development provided to principals is similar to that which is provided to teachers, reflecting state and local education agencies' failure to meaningfully distinguish between principal and teacher professional development (Haller et al., 2016; Manna, 2015; Rowland, 2017).

The recent reauthorization of the Elementary and Secondary Education Act (ESEA) in 2015, known as the Every Student Succeeds Act (ESSA), incentivizes states and local school districts to invest in principal professional development. ESSA authorizes approximately $2.3 billion annually to states to improve teacher and principal human capital. State education agencies can reserve up to three percent of these ESSA funds to improve aspects of principal professional development through preservice programs, differential pay scales, and induction for early career school leaders (Haller et al., 2016; Herman et al., 2017). Yet, in order to maximize the efficacy of ESSA funds so that induction efforts can improve principal human capital, state and local education agencies must implement principal training and induction programs with an established evidence base.[2]

To situate our study within the broader policy and research landscape, we organize our discussion of principal professional development as follows. First, we describe how principal

[2] ESSA stipulates three tiers of evidence on educational interventions. Tier I (strong evidence) is evidence derived from a well-implemented randomized control trial. Tier II (moderate evidence) is evidence derived from a single well-designed and implemented quasi-experimental study. Tier III (promising evidence) is evidence derived from at least a single well-designed and implemented study that controls for selection bias. Our study would likely be characterized as Tier II evidence based on ESSA's designations for evidentiary rigor. The only Tier I evidence (to our knowledge) comes from Herrmann et al. (2019), which we discuss below.

quality matters to a range of student and schooling outcomes. Next, we describe the extant evidence on in-service principal professional development; particular attention is given to the National Institute of School Leadership (NISL) Executive Development Program, a principal professional development program adopted by many states and from which Pennsylvania drew select coursework for inclusion in its PIL induction program. We then describe the scope and nature of state-level policy efforts around principal induction and early career professional development, situating Pennsylvania's PIL induction effort in this national policy climate.

Principal Effectiveness

Principal effectiveness and school quality are inextricably linked. Effective principals improve student achievement, develop teacher talent, and manage the organization and mission of schools (Coelli & Green, 2012; Branch et al., 2012; Leithwood et al., 2008; Grissom et al., 2021). Though the estimated variance of principal effects on student test score performance varies depending on both the research setting and statistical modeling choices (Austin et al., 2019; Chiang et al., 2016; Dhuey & Smith, 2014, 2018; Grissom et al., 2015; Laing et al., 2016), a host of evidence shows that principals have a meaningful impact on their students' academic outcomes. For example, compared to the average principal, a principal who is one standard deviation above average improves average student achievement from the 50th to the 58th percentile in one academic year (Branch et al., 2012). Similarly, Coelli and Green (2012) estimate that a principal who is one standard deviation above mean principal quality can improve graduation rates by 2.6 percentage points and English standardized tests scores by 2.5 percentage points. More recent evidence from a meta-analysis examining results from 6 studies covering more than 22,000 principals in four states (and two urban school districts) finds that replacing a below-average elementary school principal (at the 25th percentile of effectiveness) with an

above-average principal (at the 75th percentile) would result in an additional 2.9 months of math learning and 2.7 months of reading learning each year (Grissom et al., 2021). Beyond student achievement, principals are instrumental in retaining teachers (Branch et al., 2012; Miller, 2013) and reducing student absenteeism (Bartanen, 2020). In New York City, researchers found that a one standard deviation increase in perceived administrator quality decreases a teacher's likelihood of exit by 44 percent (Boyd et al., 2011), while Bartanen (2020) shows in Tennessee that replacing a less-effective principal (at the 25th percentile of principal value-added) with a more-effective principal (at the 75th percentile) can reduce student absences by 1.4 instructional days. And though prior research finds that the magnitude of principal effects on student achievement were second only to teacher quality among all school-level inputs (Leithwood et al., 2004), Grissom et al. (2021) show that Leithwood et al. (2004) may have understated the importance of school principals. Indeed, Grissom et al. (2021) highlight that while the impact of an effective principal on student achievement is nearly the same as that of an effective teacher, the scope of principal effectiveness is more important than the effectiveness of any single teacher for a school as a whole since principal effects are averaged over an entire school rather than a single classroom.

Effective principals create a shared vision for schools, promote school-wide goals, and set high-performance expectations (Leithwood et al., 2008). They develop educator human capital and provide individualized support for staff development. Effective principals manage the instructional mission of the school through targeted instructional support to teachers (Leithwood et al., 2008). Indeed, instructional leadership has been identified as the most direct influence principals have on student achievement (Robinson et al., 2008). For example, intensive instructional coaching – such as pre- and post-observation conferences where principals provide

detailed feedback to teachers about their instructional performance – has been found to be positively associated with student achievement (Grissom et al., 2013; Steinberg & Sartain, 2015). Programmatic efforts such as the Wallace Foundation's School Administration Manager (SAM) project aim to provide school leaders with greater time to dedicate to instructional leadership in place of managerial tasks such as office work, building management and student discipline (Turnbull et al., 2009). Yet, in a recent meta-analysis, Liebowitz & Porter (2019) find that principal behaviors other than instructional leadership may be similarly important mechanisms for improving student outcomes. These principal behaviors include the development and maintenance of student and family relationships; organizational management, such as budgetary tasks, facility planning, and managing noninstructional staff; administrative tasks, such as compliance activities, standardized assessment implementation, and school schedule management as well as student service management, student supervision, and managing school attendance; and external relations, such as communicating with the district office, community members, partners, or other outside stakeholders and fundraising efforts (Liebowitz & Porter, 2019).

Principal Professional Development

Two avenues of professional development attempt to develop and improve principal human capital. First, principals may participate in *pre-service* training prior to starting the principalship. Second, principals may participate in *in-service* training once they assume the principalship; we differentiate between two dimensions of in-service training: *induction* and *ongoing professional development*. Induction is in-service professional development targeted to novice and early-career principals; ongoing professional development is in-service professional development targeted to more experienced principals (i.e., those not in the early stages of their

careers as principals) and is typically a requirement for principals to maintain their active principal licensure. Notably, not all principals receive systematic induction support (or even in-service training of any kind); principal support and training depends on state- and district-specific policy efforts (see Table 1 for summary of state-level principal induction efforts). We focus on in-service professional development as the current study examines principal induction for novice and early-career principals in Pennsylvania.[3]

While little systematic evidence exists on the efficacy of principal professional development via in-service induction for early career principals (Manna, 2015; Rowland, 2017), evidence on the efficacy of in-service professional development programs for school leaders is mixed. In one study, a professional development program which focused on developing principal capacity to conduct structured observations of teachers' classroom instruction and provide targeted feedback to teachers was randomly assigned to school dyads (within districts) among a sample of 100 elementary schools across eight school districts in five states (Herrmann et al., 2019). Though the professional development program provided nearly 200 hours of professional development over two years, half of it through individualized coaching, Herrmann et al. (2019) find that principals' practices changed little as a result of the professional development, with no discernible effect on either teacher retention or student achievement. In an investigation of the New Leaders Program which recruits, trains, and provides induction support for novice principals, Gates et al. (2014) find that spending three years or more with a New Leaders principal increases student achievement by 0.7 to 1.3 percentile points; notably, however,

[3] In their review of pre-service training since 2007, Ni et al. (2017) identified 52 published articles across several education journals and research/advocacy organizations. Of these articles, 38 are implementation studies that focus on understanding the programmatic elements of pre-service programs. In the 14 studies that focused on outcomes, only two studies examine the association between pre-service principal training programs and student achievement (Corcoran et al., 2012; Gates et al., 2014).

principal induction was a component of the New Leaders program, limiting insight into the unique effect of in-service principal induction on student achievement. Among a sample of rural schools in northern Michigan, Jacob et al. (2015) study the impact of McREL's Balanced Leadership Professional Development (BLPD) Program – a widely used professional development program for school leaders – on principal efficacy, leadership practices, the instructional climate of the school, staff turnover, and student achievement. The BLPD program, which provides principals with research-based training on 21 leadership responsibilities and was randomly assigned across 126 rural northern Michigan schools, was found to have no effect on either the instructional climate of schools (as reported by teachers) or on student achievement; yet, the program was found to reduce teacher and principal turnover. To the best of our knowledge, ours is the first study to focus exclusively on the effects of a statewide principal induction program on a range of schooling outcomes, although other work has studied principal preparation programs in which induction was a component (Gates et al., 2014). Indeed, state-level policy efforts which have been focused on principal induction and in-service training have received scant attention in the research literature on school leaders. This paper aims to contribute rigorous evidence on Pennsylvania's efforts to improve principal human capital via a statewide principal induction policy.

National Institute of School Leadership

The National Institute of School Leadership (NISL) Executive Development Program (EDP) is a widely used principal professional development program.[4] The primary goal of NISL

[4] NISL EDP includes 24 days of instruction, consisting of 12 two-day units organized into three courses: World Class Schooling; Focus on Teaching and Learning; and Sustaining Transformation through Capacity and Commitment. Historically, NISL included a fourth course – Driving for Results – which was designed to improve data-driven decision-making within schools. However, this course has since been removed from the EDP coursework.

EDP is to provide principals with the skills and knowledge to create and implement instructional coaching to promote a high-performing school environment. Three NISL courses – World Class Schooling; Focus on Teaching and Learning; and Sustaining Transformation through Capacity and Commitment – emphasize blended learning and require principals to participate in online professional learning communities, prepare for in-person sessions by completing readings and pre-work, and design an Action Learning Project (Corcoran, 2017).[5] Course delivery for NISL EDP can vary, based on whether district or state policymakers opt for NISL staff to facilitate principal trainings or for district/state staff to be trained to facilitate the program themselves. Further, NISL can be implemented at different policy levels. 10 states have adopted EDP statewide for the purpose of principal professional development.[6] Pennsylvania is the only state to use select NISL coursework for its principal induction program, which we detail in the next section.

Evidence on the relationship between NISL EDP and student achievement, though suggestive, is positive. In Wisconsin, Corcoran (2017) finds that students experienced greater achievement growth in schools with NISL EDP trained principals compared to students in schools without NISL EDP trained principals. In Massachusetts, Nunnery et al. (2011A) similarly find that student achievement is greater in schools with NISL EDP trained principals. Nunnery et al. (2011B) find that, in Pennsylvania, schools with EDP trained principals had associated gains in annual student proficiency rates of 0.48 percentage points in mathematics and

[5] The Action Learning Project asks principals to apply lessons learned in EDP to an issue specific to their school environments. This practice is supported by recent research that suggests that "application-oriented" activities, in which principals apply lessons from coursework to their own school environments, are highly effective for principal professional development (Korach & Cosner, 2017).

[6] The 10 states include Arizona, Kentucky, Louisiana, Massachusetts, Minnesota, Missouri, Mississippi, Oklahoma, Rhode Island, and Virginia. Districts in another 14 states – Alabama, California, Colorado, Florida, Georgia, Illinois, Maryland, Nevada, New Hampshire, New Mexico, North Carolina, Tennessee, Texas, and Wisconsin – have adopted EDP (or a subset of NISL courses) for the purpose of principal professional development (National Institute for School Leadership, 2017).

0.54 percentage points in English Language Arts, relative to comparison schools without EDP trained principals.

While these studies inform our understanding of how PIL induction in Pennsylvania may affect student achievement, key differences exist between PIL induction and studies of NISL EDP. First, PIL induction relies on two courses from the EDP curriculum – World Class Schooling and Driving for Results – while existing evidence examines the efficacy of NISL EDP (Corcoran, 2017; Nunnery et al., 2010; Nunnery et al. 2011A; Nunnery, 2011B). Second, existing evidence on NISL EDP relies on study designs that limit the generalizability of study findings, relying on a single-cohort design (Nunnery et al., 2010; Nunnery et al., 2011B), a single school district (Corcoran, 2017), or a select number of schools (Nunnery et al., 2011A). Third, existing studies of NISL EDP rely on school-level aggregates, rather than student-level data, to examine the association between principal professional development and student achievement. Ours is the first empirical investigation of PIL induction, and aims to improve upon this prior work in two important ways. First, while Nunnery et al. (2011B) relied on school-level proficiency rates as the outcome measure, we employ rich microdata on students, teachers, and principals to estimate the effect of PIL induction on teacher and student outcomes; the use of student-level microdata avoids concerns related to aggregation bias while improving the statistical precision of the estimated effect of principal professional development. Second, our study spans a longer time period and includes the population of Pennsylvania schools, whereas Nunnery et. al (2011B) rely on a select sample of Pennsylvania schools during a more abbreviated study period. In the sections that follow we describe the data and empirical strategy, which leverages individual-level data for the population of Pennsylvania principals and a more

rigorous quasi-experimental approach to uncovering the effect of principal professional development under Act 45 on student and teacher outcomes.

State-Level Principal Induction Policies

Even in the absence of a rigorous evidence base on in-service principal professional development, many states have enacted principal induction policies. As of 2016, 20 states had introduced principal induction requirements via state-level policy reforms; two additional states (Illinois and Kentucky) had policy stipulations for induction programs but were unfunded mandates (Goldrick, 2016). Further, three states (Alabama, Connecticut, and New Mexico) have some form of principal induction, but unlike the 20 states with formal induction programs, these programs are not a required component of the principalship (Goldrick, 2016). A key component of many state-level principal induction efforts is that new principals are assigned a principal mentor who provides feedback on a new principal's practice. Seventeen states include mentorship as part of their induction process; of the 15 state policies that include coursework, only three – Hawaii, Pennsylvania, and South Carolina – require specific coursework. The duration of the induction period also varies, although most states require that principals complete induction within 2 years (see Table 1).[7]

<Table 1 about here>

While principal professional development has, historically, received less policy attention than teacher professional development (Manna, 2015; Rowland, 2017), many states have begun to take advantage of ESSA's dedicated school leadership funding to create new principal professional development opportunities. For example, North Dakota employed its ESSA funding

[7] Alongside state-level initiatives to improve principal professional development, school districts and private entities also provide opportunities for professional learning (Herman et al., 2017).

from Title II to create a Leadership Academy to provide principal professional development for employed principals and a new mentorship program for novice principals (Espinoza & Cardichon, 2017).[8] The policy expansion of principal professional development and, specifically, induction programs, motivate the current study's efforts to understand one state's efforts – Pennsylvania – to improve schooling outcomes through targeted principal induction.

Principal Induction in Pennsylvania

The Pennsylvania Inspired Leadership (PIL) induction program was introduced through Pennsylvania's *Act 45* of 2007, which dramatically changed the professional development requirements for newly hired principals in Pennsylvania. Prior to *Act 45*, *Act 48* of 1999 granted principals – both novice and more experienced principals – a variety of professional development options to maintain their active certification status.[9] Under *Act 45*, principal professional development requirements were revised to include more formal coursework – via the National Institute for School Leadership (NISL) – to focus on newly established leadership standards and to establish the PIL induction program for newly hired principals. Beginning in January 2008, all school principals employed for the first time (on or after January 1, 2008) were required to complete the PIL induction program within their first five years of employment (see Table A1 for a comparison of *Act 45* and *Act 48* requirements). Therefore, a key feature of the PIL induction program is that principals could complete PIL induction in any years during their first five years of employment. In the empirical approach section, we document how our

[8] For more detail on state spending of ESSA funds to improve principal training, see: https://learningpolicyinstitute.org/sites/default/files/product-files/Investing_Effective_School_Leadership_BRIEF.pdf

[9] Under *Act 48*, principals could choose from the following professional development options, which they were required to complete within every five year period in order to maintain their active certification status: (i) earn six credits of collegiate study; (ii) earn six credits of PDE-approved continuing professional education courses; (iii) complete 180 hours of continuing professional education programs, activities or learning experiences through a PDE approved provider; or (iv) any combination of the above (see Table A1).

identification strategy leverages both variation in the timing of PIL induction at the principal-school level (i.e., variation across principal-school cells) and multiple comparison groups to estimate the effect of PIL on teacher and student outcomes.

The PIL induction program requires principals to complete two NISL courses designed to meet the three core leadership standards established by *Act 45* (Table A2 summarizes Pennsylvania's core and corollary leadership standards under *Act 45*) within the first five years of employment as a school principal. Pennsylvania adopted NISL coursework for its PIL induction program to support the state's newly implemented standards-based approach to school leadership. In the 2004-05 academic year, Pennsylvania's governor and Secretary of Education tasked a group of educators, policymakers, and researchers to establish leadership performance standards (Pennsylvania Department of Education, 2016). Together, this group determined a set of leadership standards deemed necessary for school leaders to improve student achievement and matched these standards to NISL courses constituting the PIL induction requirements.

The first NISL course, World-Class Schooling, is designed to provide principals with the strategic planning tools to implement a vision of high-quality teaching and student achievement. The course curriculum for World-Class Schooling is aligned with the first two core leadership standards: the school leader has the knowledge and skills to think and plan strategically, creating an organizational vision around personalized student success; and the school leader has an understanding of standards-based systems theory and design and the ability to transfer that knowledge to the leader's job as the architect of standards-based reform in the school. The second course, Driving for Results, is designed to provide principals with training to examine school data, including student achievement data, to identify school, teacher, and individual student needs. The course curriculum for Driving for Results is aligned with the third core

leadership standard: the school leader has the ability to access and use appropriate data to inform decision-making at all levels of the system. Each of the two courses count for 60 professional development hours, last for six days (36 total hours per course) and have 6 hours of pre-course assignments (Table A3 provides more detail on the NISL courses that comprise both PIL induction and NISL EDP).

Principals are required to complete both NISL courses (i.e., PIL induction) within their first five years of employment as a principal in Pennsylvania. Completion of the PIL induction program is tied to principals' administrative licenses; if newly hired principals fail to complete PIL, they are unable to renew their licenses and can no longer continue employment as principals. Moreover, the two NISL courses that constitute PIL induction emphasize the skills that Pennsylvania and PDE policymakers believed early career principals needed to succeed. Indeed, *Act* 45, with its newly implemented leadership standards coupled with PIL induction (via NISL coursework) tied to those standards, represented Pennsylvania's effort to inject greater rigor and accountability into in-service principal professional development.

Consistent with *Act 45*, principals hired after January 1, 2008 are defined as PIL-eligible principals. Among PIL-eligible principals, those who complete two NISL courses – World Class Schooling and Driving for Results – are categorized as having completed PIL induction. Thus, Pennsylvania principals fall into two distinct groups: (i) principals hired on or after January 1, 2008 who are required to complete the PIL Induction Program within their first five years of employment;[10] and (ii) principals employed prior to January 1, 2008 who must complete their continuing professional development requirements established by *Act 48* proportional to their

[10] After the initial five years of employment, these principals continue to fulfill their 180 hours of professional development requirement in PIL-approved courses.

employment post-January 1, 2008 (e.g., if a principal was employed for only two years prior to January 1, 2008, then he/she must complete 60 percent of the remaining professional development hours in the PIL program, unless he/she completed more than 40 percent of the required hours in the first 2 years).[11] Although *Act 45* was passed in 2008, PIL-eligible principals can complete PIL at any time within their first five years as principals. As such, the pre-treatment period is defined as school years prior to a PIL-eligible principal's completion of PIL, while the post-treatment period is defined as school years after PIL completion. Thus, the year of PIL completion may vary across principals. In our empirical approach section (below), we describe in greater detail how we model the timing of PIL completion in the context of our identification strategy.

Data & Sample

We construct a panel dataset for all students, teachers, and principals in all traditional and charter public schools in the state of Pennsylvania for the 2008-09 through 2015-16 school years. For each student, we observe a unique student identifier, allowing us to follow students across time; a unique school identifier; teacher identifiers; birth date, which allows us to construct student age; demographic information (race, gender); grade level; free/reduced-price lunch status (eligibility and receipt); poverty status (whether a student receives supplementary government

[11] PIL induction and continuing professional development are administered within one of eight administrative regions in Pennsylvania (a map of the administrative regions can be found at: https://www.education.pa.gov/Teachers%20-%20Administrators/PA%20Inspired%20Leaders/Pages/default.aspx). The course offerings can be broadly defined as: (i) PIL induction coursework; (ii) NISL, non-PIL induction coursework; and (iii) non-NISL professional development coursework. PIL induction coursework includes World-Class Schooling and Driving for Results. NISL, non-PIL induction coursework includes Focus on Teaching and Learning and Sustaining Transformation through Capacity and Commitment. Non-NISL professional development coursework includes coursework in school leadership (e.g., effective communication, setting goals and expectations for a school), data use within schools, understanding early childhood education, and emphasizing the need for student equity and career readiness. Notably, these professional development courses are not developed by NISL, and can be developed and delivered by any entity that receives approval (e.g., universities) from PDE to provide principal professional development.

services, such as TANF or SNAP); English language learner (ELL) status; special education status; and gifted status. We observe student achievement outcomes – both scaled scores and proficiency levels for math and English language arts (ELA) – for all students in grades 3-8. For all analyses of student achievement and teacher effectiveness, we rely on math and ELA test scores from the Pennsylvania System of School Assessment (PSSA) for students in grades 3-8 (which we standardize at the subject*grade*year level).

For teachers, we observe a unique teacher identifier; a school identifier; demographic information (race, gender); date of birth; educational attainment (i.e., highest degree completed); experience (total years of educational experience in Pennsylvania); and courses taught, allowing us to link individual teachers to individual students for the purposes of constructing teacher effectiveness measures. For principals, we observe a unique principal identifier; a unique school identifier; demographics (race, gender); date of birth; educational experience (total years of educational experience in Pennsylvania); and educational attainment. Importantly, we observe PIL induction coursework taken by PIL eligible principals, as well as ongoing professional development coursework taken by principals hired before January 2008. Specifically, coursework data includes a unique principal identifier, course numbers, course names, course start dates, course end dates, and the credit hours a course is worth. The following NISL courses are included among the professional development coursework data: World-Class Schooling; Focusing on Teaching and Learning; Driving for Results; and Sustaining Transformation through Capacity and Commitment (see Table A3).

Sample

We construct our school-year sample from all traditional and charter public schools in Pennsylvania during the 2008-09 through 2015-16 school years. We limit the sample to school-

year observations with one principal in a given school year (schools with multiple assistant principals are retained in the sample) and drop those school-year observations where the principal is employed at multiple schools in a given school year. Thus, our main analytic sample ("All Schools") consists of 20,636 school-year observations across 3,183 unique public schools and 4,877 unique principals.[12] We also construct a sub-sample of our main analytic sample ("PIL Eligible") which includes school-year observations where the principal is PIL Eligible (i.e., a school's principal became a principal for the first time in Pennsylvania after January 2008 and did not complete a PIL course as an assistant principal).[13] The PIL Eligible sample consists of 7,148 school-year observations across 1,736 unique public schools and 1,886 unique principals. For estimates of teacher effectiveness (i.e., the sample including teachers with available VAM), the All Schools sample includes 9,434 school-year observations (45.7 percent of the main All Schools sample) across 2,255 unique schools and 3,085 unique principals; the PIL Eligible sample includes 3,731 school-year observations (52.2 percent of the main PIL Eligible sample) across 1,205 unique schools and 1,312 unique principals.[14]

[12] We start with 22,693 school-year observations and drop 20 observations due to restricting school-year observations to those where the principal was a principal in no more than one school in a given year. We drop an additional 2,037 school-year observations by restricting school-year observations to one principal per year per school. This gives us 20,636 school-year observations.

[13] Principals who had prior PIL exposure as assistant principals are included in the All Schools comparison group; principals who had prior PIL exposure as assistant principals are excluded from the PIL Eligible sample. If principal professional development via PIL induction improves teacher and student outcomes, the inclusion of principals who completed PIL as an assistant principal in the All Schools sample will understate estimated PIL effects. By excluding principals who completed PIL as an assistant principal from the PIL Eligible sample, we aim to estimate the effect of PIL induction among novice principals of record without prior professional development experience. There are 654 principals who completed PIL as an assistant principal who are included in the All Schools comparison group and excluded from the PIL Eligible sample.

[14] For each of the two school-year samples (i.e., All Schools and PIL Eligible), we construct a teacher-by-year panel and student-by-year-by-subject panel to estimate the effect of PIL induction on teacher outcomes (effectiveness (VAM) and turnover) and student outcomes (math and ELA achievement), respectively. In the teacher-by-year panel, we include all full-time K-12 teachers (in models estimating teacher effectiveness (VAM), only teachers in grades 4-8 are included). In the student-by-year-by-subject panel, we include all students in grades 3 through 8 with available math and/or ELA test scores from the Pennsylvania System of School Assessment (PSSA).

Table 2 (Panel A) summarizes the characteristics of a school's principal by PIL eligibility and PIL treatment status (i.e., PIL or No PIL). PIL treatment status is defined as the completion of the PIL induction program (i.e., completion of two NISL courses – World Class Schooling and Driving for Results) at the school-year level within principal-school cells.[15] Among the main sample ("All Schools"), treatment schools contain principals who are, on average, younger, more likely to be female, more racially diverse, less experienced in Pennsylvania public education and less likely to hold an advanced degree than comparison schools. Among the PIL Eligible sample, principals are, as expected, much more similar across treatment and comparison schools; specifically, while principals in treatment schools are statistically more likely to be White and have more educational experience in Pennsylvania, these differences are not substantively meaningful in magnitude.[16] Table 2 (Panel B) summarizes the characteristics of a school's students by PIL eligibility and treatment status. Among the main sample ("All Schools"), treatment schools serve, on average, a larger share of students who are economically disadvantaged (based on eligibility for free/reduced-price lunch) and lower-achieving (as measured by proficiency on annual state math and ELA exams) than comparison schools. Among the PIL Eligible sample, treatment schools serve, on average, a larger share of lower-achieving students than comparison schools.

<Table 2 about here>

Empirical Approach

[15] Specifically, treatment status (i.e., PIL) is defined at the school-year level and indicates the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells).
[16] Although Act 45 required all principals hired after the 2008-09 school year to complete PIL induction within five years of employment, evidence indicates that compliance was imperfect among PIL eligible principals. Among the 1,886 PIL eligible principals, 600 completed PIL induction (32 percent). Of the 702 PIL eligible principals employed for five or more years, 347 (49 percent) completed PIL induction.

Act 45 dictated that newly hired principals (as of January 2008) complete PIL induction within their first five years of employment. Thus, principals who begin the principalship in the same academic year may complete PIL induction in different years during their first five years as principal. We leverage variation in the timing of PIL induction across principal-school cells and employ difference-in-differences and event study strategies to estimate the impact of PIL induction on teacher effectiveness, teacher turnover, and student achievement.

Difference-in-Differences

To estimate the effect of PIL induction on teacher effectiveness and teacher turnover (Appendix B details the construction of teacher-level value added estimates (VAM) of teacher effectiveness), we estimate the following difference-in-differences specification:

$$(1)\quad Y_{jspt} = \beta_1 PIL_{pst} + \boldsymbol{X}_{jt}\boldsymbol{\Gamma} + \boldsymbol{Z}_{st}\boldsymbol{\gamma} + \phi_s + \delta_{dt} + \mu_{jspt}$$

where Y is an outcome (either teacher turnover or teacher effectiveness) for teacher *j* in school *s* with principal *p* in school year *t*. For estimates of teacher turnover (where Y equals 1 if teacher *j* exits school *s* at the end of school year *t*, and 0 otherwise), we include all full-time classroom teachers. For subject-specific estimates of teacher effectiveness (where Y equals teacher *j*'s math (ELA) VAM score from school year *t*), we include teachers in grades 4-8 with available math and/or ELA VAM.[17] **X** is vector of time-varying teacher characteristics, including age, gender, race, years of experience (in Pennsylvania) and educational attainment. **Z** is a vector of time-varying school characteristics, including the percent of economically disadvantaged students (i.e., the share of a school's students who are free or reduced-price lunch eligible), the

[17] Grade 6-8 teachers with available math and/or ELA VAM are included for the entire study period (i.e., 2008-09 through 2015-16 school years); grade 4-5 teachers with available math and /or ELA VAM are included from the 2013-14 through 2015-16 school years due to data limitations prohibiting the linking of grade 4-5 teachers to students prior to the 2013-14 school year.

percent of racial/ethnic minority students, the percent of students receiving specialized services (i.e., ELL, IEP, gifted), and school size (enrollment). The variable ϕ_s is a school fixed effect which controls for all time-invariant (and unobserved) school-level heterogeneity; δ_{dt} is a district*year fixed effect which controls for year-specific idiosyncratic shocks common to all schools within a school district; and μ_{jspt} is a random error term. To account for serial correlation across teachers in the same schools, we cluster standard errors at the school level.

The treatment variable of interest is PIL_{pst}, which is an indicator that equals 1 in the school year *t* (and in all subsequent years) that principal *p* completed PIL induction at school *s*. Since treatment is defined at the school-year level and because we rely on variation in the timing of PIL induction across principal-school cells, β_1 estimates the effect of principal *p* completing PIL induction within the same school *s*. This approach to defining treatment mitigates concern that non-random principal transitions at the school level might bias estimates of PIL induction. Yet, the timing of PIL induction (at the principal-school level) may be non-random since PIL-eligible principals typically select the school years in which to complete PIL coursework within the state's mandated five-year window. Notably, the presence of non-random selection into PIL induction in this setting is akin to the non-random timing of school exit and mobility by principals in other settings (Bartanen et al., 2019; Miller, 2013). In the *Threats to Identification* section below, we discuss the potential endogeneity due to the non-random timing of principal completion of PIL induction.

Next, we estimate the effect of PIL induction on student achievement according to the following difference-in-differences specification:

(2) $$Achievement_{ispt} = \beta_1 PIL_{pst} + \boldsymbol{V}_{it}\boldsymbol{\Gamma} + \boldsymbol{Z}_{st}\boldsymbol{\gamma} + \phi_s + \delta_{dt} + \mu_{ispt}$$

where *Achievement* is the academic achievement, in either math or English language arts (ELA), of student *i* attending school *s* led by principal *p* in school year *t*. Students' scaled scores are standardized within year-subject-grade in our sample to account for test differences across years. For student math and ELA achievement outcomes, estimates are based on students in grades 3 through 8 with available math and/or ELA test scores from the Pennsylvania System of School Assessment (PSSA). **V** is a vector of time-varying student characteristics, including: age, race, gender, grade level, free/reduced price-lunch eligibility status, poverty status (whether a student receives supplementary governmental services, such as TANF or SNAP), special education status, English language learner (ELL) status, and gifted status. All other variables are defined as in equation (1), and we cluster the standard errors at the school level to account for serial correlation across students within the same school.

Event Study

We complement the difference-in-differences estimates with an event study approach. The event study approach not only relaxes the assumption that PIL treatment effects are time-invariant and thereby permits the average treatment effect (i.e., the difference-in-differences estimator) to be modeled flexibly post-treatment, but also enables an empirical assessment of the parallel trends assumption underlying the difference-in-differences approach. Thus, to assess one identifying assumption of the difference-in-differences approach (i.e., parallel pre-trends) and to estimate the potentially dynamic effects of PIL induction, we disaggregate the treatment indicator (PIL_{pst}) from equations (1) and (2) in the following event study framework. For teacher-level outcomes, we estimate the following specification:

$$(3)\ Y_{jspt} = \sum_{r=-4+,r!=-1}^{r=4+} \beta_r \left(PIL_{ps,t+r}\right) + \boldsymbol{X}_{jt}\boldsymbol{\Gamma} + \boldsymbol{Z}_{st}\boldsymbol{\gamma} + \phi_s + \delta_{dt} + \mu_{jspt}$$

where Y is an outcome (either teacher turnover or teacher effectiveness) for teacher *j* in school *s* with principal *p* in school year *t*. In equation (3), β_r estimates the year-specific effects of PIL induction in the *r* years before and after the completion of PIL induction. All other variables are defined as in equation (1) and standard errors are clustered at the school level.

For student-achievement outcomes, we estimate the following specification:

$$(4)\ Achievement_{ispt} = \sum_{r=-4+,r!=-1}^{r=4+} \beta_r\left(PIL_{ps,t+r}\right) + \boldsymbol{V}_{it}\boldsymbol{\Gamma} + \boldsymbol{Z}_{st}\boldsymbol{\gamma} + \phi_s + \delta_{dt} + \mu_{ispt}$$

where *Achievement* is the academic achievement, in either math or English language arts (ELA), of student *i* attending school *s* led by principal *p* in school year *t*. In equation (4), β_r estimates the year-specific effects of PIL induction in the *r* years before and after the completion of PIL induction. All other variables are defined as in equation (2) and standard errors are clustered at the school level. For estimates of equations (3) and (4), we assess the parallel trends assumption underlying the difference-in-differences approach by testing the joint significance of the pre-treatment year-specific effects (i.e., $\beta_{r=-2} = \beta_{r=-3} = \beta_{r=-4+}$), the reference category for which is the year prior to the completion of PIL induction (i.e., $\beta_{r=-1}$).

Multiple Comparison Groups

To examine the robustness of our results, we estimate equations (1) – (4) on two school-year samples. The first is the *All Schools* sample, which includes all school-year observations in the 2008-09 through 2015-16 school years. The *All Schools* sample includes all Pennsylvania principals, many of whom have completed or are in the process of completing their continuing professional development requirements during the study period. Thus, one concern is that prior completion of or contemporaneous participation in ongoing principal professional development among more experienced principals might understate PIL induction effects. To address this

concern, we construct a second analytic sample, the *PIL Eligible* sample, which is a subsample of the main sample and includes school-year observations where the school principal is PIL Eligible (i.e., a school's principal became a principal for the first time in Pennsylvania after January 2008 and did not complete a PIL course as an assistant principal). Thus, because the *PIL Eligible* sample restricts school-year comparisons to those containing just novice and early career principals without any prior principal professional development training, we are able to account for the possibility that the completion of principal professional development among more experienced principals contained in the *All Schools* sample might understate PIL induction effects.

Threats to Identification

In the context of our empirical approach, two primary threats to identification may introduce bias into estimated PIL induction effects: non-random selection into treatment and the staggered timing of treatment across principal-by-school cells. First, the timing of PIL induction (at the principal-school level) is likely non-random since PIL-eligible principals typically select the school years in which to complete PIL coursework within the state's mandated five-year window. For example, a principal who begins in a more challenging school environment (e.g., a school that serves higher-poverty and lower-achieving students) may delay the completion of PIL until after they have dedicated time to learning about and adapting their leadership style to their new school environment; in this case, we would be concerned that the timing of PIL induction would downwardly bias estimates of the PIL effect if there are positive effects from the principal's initial leadership efforts. On the other hand, if more motivated (and higher ability) principals systematically select into PIL induction earlier in their principal tenure, then the PIL effect would be upwardly biased. Indeed, differences in the timing of principal selection into

completing PIL coursework could stem from unobserved differences in principal skills, beliefs, or disposition toward professional development that could independently drive differences in teacher and student outcomes beyond the professional development provided via PIL coursework. And, while event study estimates permit analysis of the parallel trends assumption without imposing a particular functional form on pre-treatment outcome trends, the presence (or absence) of parallel trends in the outcome measures is neither a necessary nor sufficient condition to defend the assumption that there is no endogenous selection into treatment (Kahn-Lang & Lang, 2020). To address this concern, we replace the school fixed effect in the difference-in-difference and event study models with a principal-by-school fixed effect, allowing us to control for time-invariant differences across principal-by-school cells that may be correlated with both the timing of principal selection into treatment and the teacher and student outcomes within the same school setting.

Further, it's notable that the presence of unobserved, time-varying heterogeneity at the principal-by-school level that is correlated with the timing of PIL induction may still introduce bias into estimated PIL effects. More specifically, the timing of when PIL eligible principals select into PIL induction is not determined exogenously. To address this, we re-estimate our primary specification on a sample of school-by-year observations that contain just PIL eligible principals who completed PIL induction. While this robustness check addresses concerns related to endogenous selection into PIL induction on the extensive margin (whether or not PIL eligible principals completed PIL induction), the endogeneity of the timing into PIL induction (when PIL

eligible principals completed PIL induction) that is correlated with unobserved and time-varying characteristics at the principal-by-school level remains a residual concern.[18]

Second, estimated PIL effects rely on the staggered timing of principal selection into PIL induction across principal-by-school cells. A recent literature has emerged which documents how the staggered timing of treatment across multiple principal-by-school groups and school years (i.e., group and time dimensions) implicitly places greater weight on principal-by-school units that experience the treatment for longer time periods as well as cases where treated units may be assigned negative weight due to the treatment-comparison cells in which they occupy (Baker et al., 2021; Callaway & Sant'Anna, 2020; de Chaisemartin & D'Haultfoeuille, 2020; Goodman-Bacon, 2018). To examine the extent to which our primary difference-in-differences estimates are robust to the staggered timing of treatment, we implement the procedure introduced by de Chaisemartin & D'Haultfoeuille (2020) to estimate two-way fixed effects estimators with heterogeneous treatment effects via their *did_multiplegt* package in Stata.

And, in addition to school principals, other school and district leaders (e.g., assistant principals and district superintendents) are also required to participate in professional development (see Principal Induction in Pennsylvania section for more detail). Thus, there may be concern that the potential endogenous selection into PIL among school principals may be

[18] We estimate our primary empirical specification on a school-by-year sample containing just PIL eligible principals who completed PIL induction. While this approach removes endogenous selection into PIL induction, the endogeneity of the timing into PIL induction (i.e., when PIL eligible principals completed PIL induction during their early career tenure) is a residual concern that our empirical approach is more limited in its ability to address. Table A7 reports descriptive statistics on principal and school characteristics by the year in which a PIL eligible principal completed PIL induction and Table A8 reports the estimated effects of PIL induction among this PIL-completer sample. Table A8 shows that our primary results are robust to restricting the sample to just those K-12 traditional and charter public schools with PIL eligible principals (those hired during or after the 2008-09 school year) who completed PIL induction within the first five years of the principalship. Table A7 shows that PIL eligible principals who completed PIL earlier in their principalship were more likely to be non-White and to lead schools serving higher-poverty and lower-achieving students than PIL eligible principals who completed PIL later in their principalship, suggesting that the heterogeneous benefits of completing PIL induction earlier in a principal's tenure (see Table 11) were concentrated among more disadvantaged Pennsylvania schools.

correlated with the timing of principal professional development among other school/district leaders. In the following sections, we present evidence suggesting that this is unlikely to be a concern. Indeed, the inclusion of district-by-year fixed effects controls for idiosyncratic shocks such as the timing of district leader professional development, and the robustness of our primary estimates across multiple sensitivity checks – in particular, the robustness of our estimates to staggered treatment timing – should allay these concerns.

Results

Did the completion of PIL induction improve teacher effectiveness in math and ELA instruction? Difference-in-differences estimates indicate that PIL induction improved teacher effectiveness in math by, on average, 0.012-0.019 student-level standard deviations among the *All Schools* and *PIL Eligible* samples, respectively (Table 3, column 1), corresponding to an effect size of approximately 0.10-0.20 standard deviations of teacher math VAM. Not only do we find that these estimates are robust to the inclusion of principal-by-school fixed effects (column 2), school-specific linear time trends (column 5) and (in the *PIL Eligible* sample) staggered treatment timing (column 6), but we also find that estimates based on the *PIL Eligible* sample are approximately 50 percent larger in magnitude than estimates based on the *All Schools* sample. Differences in the magnitude of PIL effects reflect differences in the composition of the comparison groups between the *All Schools* and *PIL Eligible* samples. The *All Schools* comparison group includes all principals, many of whom have completed or are in the process of completing their continuing professional development requirements. In contrast, the *PIL Eligible* sample includes just novice and early-career principals who have yet to complete any principal professional development requirements. Thus, we focus our remaining discussion of PIL effects on the *PIL Eligible* sample which excludes from the comparison group school leaders who have

completed (or are in the process of completing) ongoing professional development requirements (we continue to report results based on the *All Schools* sample for purposes of comparison).

Though teacher math effectiveness improved, on average, following the completion of PIL induction, improvements in teacher effectiveness may not emerge until years after the completion of principal professional development. Alternatively, if the human capital benefits of principal professional development are short-lived and fade over time, the average effect of PIL induction may mask potentially important dynamic effects of principal professional development in the years after PIL induction. Event study results, which reveal no pre-PIL trend among the *All Schools* sample (p-value=0.88) and marginally significant pre-PIL effects among the *PIL Eligible* sample (p-value=0.054) (see Figure 1 and Table A4, column 1), follow a very similar pattern of positive impacts across the two samples with the estimated PIL effects always larger among the *PIL Eligible* sample than the *All Schools* sample. The event study results indicate that improvements in teacher math effectiveness, which first occur in the year of PIL completion, persist in the years following the year of PIL completion; there is no evidence of significant post-PIL trends among either the *All Schools* or *PIL Eligible* comparison groups (see Table A4). Thus, the event study results reveal an upward shift in teacher math effectiveness following principal completion of PIL induction, and evidence that principal human capital acquired under PIL induction does not fade over time. We find no evidence of any significant or substantive improvements in teacher ELA effectiveness (see Table 3 and Figure 2).

<Table 3 about here>

<Figures 1 & 2 about here>

The main results suggest that teacher math effectiveness improved because of PIL induction. Yet, changes in teacher composition could drive the observed PIL effect. If, for

example, PIL completion is associated with the recruitment or retention of more effective teachers, then the main results would conflate changes in teacher composition within a school with the PIL induction effect, and our main estimates of PIL induction would be upwardly biased. This is because the difference-in-differences model compares changes in average teacher effectiveness (at the school-year level) before and after principal completion of PIL induction; thus, any change in average teacher effectiveness following PIL that are due to compositional changes would bias the PIL effect.

To examine whether our main results of PIL induction are driven, at least in part, by changes in teacher composition within schools, we re-estimate equation (1) by including teacher (or teacher-by-school) fixed effects. In models that include teacher (or teacher-by-school) fixed effects, the identifying variation comes from within-teacher (or within-teacher-by-school) responses to the principal-by-school level treatment. And by restricting the identifying variation to changes that occur within-teacher (or within-teacher-by-school) cells in response to a principal's completion of PIL within the same school, we estimate whether PIL induction improved the effectiveness of a school's current teachers – those teachers present in the school during the pre- and post-PIL years of a school's principal.[19] Thus, comparing the point estimates on teacher VAM from difference-in-differences estimates with and without teacher fixed effects provides insight into whether the PIL effect reflects improvements in current teacher effectiveness or the recruitment and retention of more effective teachers.[20]

[19] Since the teacher-by-school fixed effects restrict variation to within-teacher-by-school cells in the school in which a PIL eligible principal completed PIL, teacher observations with no within-school variation in PIL status do not contribute to the estimated PIL effect. This approach allows for insight into whether PIL induction improved the effectiveness of a school's current teachers who taught in the same school in the period before and after the school's principal completed PIL induction.

[20] We compare the point estimates on teacher VAM between difference-in-differences (DD) estimates with and without teacher (or teacher-by-school) fixed effects. If, for example, the DD estimates without teacher (or teacher-

Evidence from Table 3 suggests that changes in teacher composition following the completion of PIL induction are unlikely to be a concern. Indeed, among the *PIL Eligible* sample, estimates based on models with teacher and teacher-by-school fixed effects indicate that PIL induction improves the math effectiveness of a school's current teachers by 0.020 student-level standard deviations (Table 3, Panel B, columns 3 and 4); these estimates are nearly identical to the primary difference-in-differences estimate of 0.019. We again find no discernible effect of PIL induction on teacher ELA effectiveness (Table 3, Panel B, columns 9 and 10), a result that is consistent with our main results from Table 3. The robustness of results across the difference-in-differences estimates with and without teacher fixed effects suggest that PIL improved the math effectiveness of a school's current teachers rather than the impact of PIL being generated by changes in teacher composition via the recruitment and retention of more effective teachers. This is further supported by the null effect on teacher turnover, which we discuss next.

Given that effective principals improve teacher retention (Ladd, 2011; Loeb et al., 2012), we next examined the impact of PIL induction on teacher turnover. Table 4 and Figure 3 summarize these results. First, we find no evidence of a pre-PIL trend among the two comparison groups, providing empirical support for the key identifying assumption of the difference-in-differences approach (see Figure 3 and Table A5, column 1). And, while we find

by-school) fixed effects are positive while the DD estimates with teacher (or teacher-by-school) fixed effects are not different from zero, then this would suggest that there were changes in the composition of teacher effectiveness following PIL completion, while also revealing that PIL had no discernible impact on the effectiveness of a school's current teachers. In contrast, if the DD estimates without teacher (or teacher-by-school) fixed effects are substantively the same in magnitude as the DD estimates with teacher (or teacher-by-school) fixed effects, then this would indicate that PIL improved the effectiveness of a school's current teachers rather than being driven by post-PIL changes in the average effectiveness of a school's teachers.

no effect, on average, of PIL induction on teacher turnover– an estimate which is robust to the inclusion of principal-by-school fixed effects (column 2), school-specific linear time trends (column 3) and staggered treatment timing (column 4), we do find that teacher turnover declined significantly in the years following PIL completion. Indeed, in the second and third years after PIL completion, teacher turnover declined by approximately 0.02 percentage points, representing an approximately 18 percent decline relative to the mean teacher turnover rate (of 11 percent) among Pennsylvania teachers during the study period. The fact that teacher turnover declines in the years subsequent to PIL completion within the same schools suggests that principal professional development may not only lead to short-term and sustained improvements in teacher effectiveness but may also stabilize a school's teaching force by reducing teacher turnover in the wake of improvements in principal human capital.[21]

<Table 4 about here>

<Figure 3 about here>

Teacher effectiveness and teacher turnover are two important channels through which improvements in principal human capital, via professional development, might operate to improve student achievement. Indeed, prior evidence suggests that teacher turnover reduces student achievement (Atteberry et al., 2017), while evidence herein indicates that PIL induction improved teacher math effectiveness. Therefore, to what extent did PIL induction generate

[21] To provide additional insight into the patterns of teacher turnover, we also examined whether the impact of PIL induction on teacher turnover varied as a function of teacher effectiveness (i.e., VAM). To do so, we estimate the impact of PIL induction on teacher turnover by quartile of teacher effectiveness (quartiles of teacher math and ELA VAM are constructed at the school year level) in the context of our primary difference-in-differences and event study specifications. We find no evidence that the impact of PIL induction on teacher turnover varied heterogeneously as a function of teacher effectiveness (results are available upon request). This provides additional evidence that PIL induction improved the (math) effectiveness of a school's current teachers rather than reflecting any compositional changes in the effectiveness of a school's teachers (via, for example, the recruitment and retention of more effective teachers) in the wake of a principal's completion of PIL induction.

improvements in student achievement? Table 5 and Figures 4-5 summarize these results. On average, student math achievement improved by an imprecisely estimated .01-.02 student-level standard deviations, a result that is robust across specifications that control for principal-by-school fixed effects (column 2), school-specific linear time trends (column 3) and staggered treatment timing (column 4). As with teacher ELA effectiveness, we find no effect on student ELA achievement.

<Table 5 about here>

<Figures 4 & 5 about here>

Notably, the student achievement estimates summarized in Table 5 are based on all grade 3-8 students, independent of whether a school-grade-year observation contains teachers for which a teacher-level VAM score can be calculated. To provide insight into the link between improvements in principal human capital (via PIL induction), changes in teacher effectiveness and, ultimately, changes in student achievement, we estimate the returns to PIL induction on student achievement among the same school-grade-year observations for which teacher VAM can be calculated (i.e., *VAM Teacher Sample*). These results are summarized in Table 6. We find that PIL induction improved student math achievement on the order of approximately 0.02-0.03 student-level standard deviations, and these estimates are again robust (though, slightly smaller in magnitude) to principal-by-school fixed effects (column 2), school-specific linear time trends (column 3) and staggered treatment timing (column 4). In terms of the dynamic nature of the achievement effects, event study estimates shown in Figure 4, Panel B (and Table A6) reveal that the impact of PIL induction on student math achievement first appears in the year in which a school's principal completed PIL induction and persists in the years after PIL completion. These results not only provide additional support for the positive impact of PIL induction on student

achievement, but also highlight a key mechanism through which principal professional development improves student achievement; namely, through improvements in teachers' instructional effectiveness.

<Table 6 about here>

Prior evidence finds that stable principal leadership improves student achievement and decreases teacher turnover (Bartenan et al., 2019; Miller, 2013). Thus, an additional concern relates to the endogeneity of principal persistence within a school. Indeed, the estimated impacts of PIL induction might reflect not only improvements in principal human capital due to PIL induction but also human capital accumulation reflecting the positive returns to principal experience within the same school context. We address this concern by re-estimating equations (1) and (2) among principals who have at least three, four, or five consecutive years of experience within the same school and restrict analysis to the PIL Eligible sample. Table 7 provides evidence that the main effects of PIL induction on teacher effectiveness (from Table 3) are robust across PIL eligible principals whose longevity within a school varies (Table 7, Columns 1 and 2). The effect of PIL on teacher math effectiveness ranges from 0.02-0.03 student-level standard deviations, consistent with our primary estimates reported in Table 3 (Panel B). We again find that PIL induction had no impact on teacher ELA effectiveness. We find no average effect of PIL on teacher turnover and these effects are not only invariant to principal longevity (column 3) but are also consistent with our primary estimates of teacher turnover (from Table 4). Further, we find that PIL induction improves student math achievement among all students on the order of 0.02-0.03 standard deviations (Table 7, column 4) and by .04-.06 standard deviations among the sample of school-grade-year observations for which teacher VAM can be calculated (i.e., *VAM Teacher Sample*), with no discernible effect on student ELA

achievement; both sets of results are consistent with our main achievement results (Tables 5 and 6). Finally, to more explicitly adjust for potential confounding due to principal experience, we re-estimate equations (1) and (2) controlling directly for principal years of experience; all results remain unchanged (and are available upon request).

<Table 7 about here>

Heterogeneous Effects of PIL Induction

Schools in Pennsylvania vary in the characteristics of the students that they serve and the geographic settings in which they are located (see Table 2). Thus, the effect of PIL induction might also vary across schools serving different student populations. Indeed, prior evidence suggests that principals adjust their leadership behaviors based on the characteristics of their school settings (Goldring et al., 2008; May et al., 2012). We examine five dimensions of school settings for which the consequences of principal professional development might differentially impact teacher and student outcomes: (i) poverty; (ii) racial/ethnic minority; (iii) achievement; (iv) geographic location (i.e., urbanicity); and (v) sector (charter or traditional public schools). Table 8 summarizes the heterogeneous effects of PIL induction on teacher outcomes (effectiveness and turnover) as a function of school poverty, racial/ethnic minority, and achievement; Tables 9 summarizes the heterogeneous effects of PIL induction on student achievement; and Table 10 summarizes the heterogeneous effects of PIL induction as a function of school geographic location and sector.

In Table 8, we show that the impact of PIL induction on teacher math effectiveness is concentrated among schools serving the most economically disadvantaged and minority students. Indeed, PIL induction improves teacher math effectiveness more in schools with the highest share of students eligible for free/reduced-price lunch and in schools with the highest share of

racial/ethnic minority students. In contrast, PIL induction had no discernible impact on teacher effectiveness in schools serving the most economically advantaged students and in schools serving the lowest share of minority students. And though there is little consistent evidence of heterogeneous effects of PIL induction on either teacher turnover (Table 8) or student achievement (Table 9), improvements in teacher math effectiveness were concentrated among urban Pennsylvania schools (Table 10, Panel A); this result is consistent with evidence (from Table 8) that the effect of PIL induction on teacher math effectiveness was concentrated among the state's most economically disadvantaged schools. The heterogeneity analysis also suggests that principals in charter schools may have benefitted less from PIL Induction than principals in TPS schools with respect to teacher (ELA) effectiveness (Table 10, Panel B). This might, in part, be a function of the fact (see Table 2) that PIL take-up among charter school principals is particularly low relative to the share of all principals in charter schools (i.e., among the PIL Eligible sample, 8 percent of school-year observations are charter schools, but only 3 percent of PIL completers are in charter schools). At the same time, charter principals typically have greater autonomy than TPS principals; yet, the results suggest that the benefits of professional development among all Pennsylvania principals were concentrated among those principals in TPS schools with less autonomy. Thus, PIL induction may have improved the leadership skills of TPS principals who completed PIL more than charter principals who also completed PIL.

<Tables 8-10 about here>

Finally, the effects of professional development might also depend on the timing of when early-career principals complete PIL induction. Specifically, principals who complete PIL induction earlier in their tenure may benefit differently than principals who complete PIL induction later in their (early-career) tenure. For example, principals who delay the completion

of PIL induction until later in their (early-career) tenure may be more resistant to incorporating the skills taught via PIL induction than those who are in their first or second year in the principalship. Table 11 presents evidence on the potentially differential returns to PIL induction by the year in which a principal completed PIL. We find that the impact of principal professional development on teacher effectiveness is concentrated among principals who complete PIL induction in their first two years as principal. Indeed, among school-year observations in which the principal is in his/her first two years as school leader, PIL induction improves teacher math effectiveness by 0.043 student-level standard deviations (or nearly 0.40 standard deviations of teacher effectiveness). In contrast, PIL induction has no significant impact on teacher effectiveness among school-year observations with principals in at least their 3rd year as school leader. We find no consistent evidence of differential effects of PIL induction by the year of PIL completion on teacher turnover or student achievement.

<Table 11 about here>

Conclusion

Principals are among the most important inputs to the operation and performance of schools. Yet, little work has examined whether efforts to improve principal human capital via in-service induction and professional development can positively affect a range of schooling outcomes. In this paper, we examine a statewide policy reform in Pennsylvania aimed to improve principal human capital through targeted professional development for novice principals tied to the state's leadership standards – the PIL induction program. Relying on difference-in-differences and event study strategies, we estimate the effect of PIL induction on teacher effectiveness, teacher turnover and student achievement. We find that PIL induction improved teacher math effectiveness and student math achievement, and that the effects of PIL induction

on teacher effectiveness were concentrated among the most economically disadvantaged schools in urban Pennsylvania school districts. Notably, we show that improvements to teacher math effectiveness are not driven by changes in teacher composition or by non-random variation in principal longevity within a school. In contrast, we find no impact of PIL induction on student ELA achievement or teacher ELA effectiveness. And, while we find no evidence that PIL induction decreased teacher turnover in the year of PIL completion, teacher turnover declined by approximately 2 percentage points in the years following PIL induction. Finally, principal professional development had the greatest impact on teacher effectiveness when principals completed PIL induction during their first two years as principal.

The heterogeneous effects of PIL induction – which are concentrated among schools serving the most economically disadvantaged students located in urban school districts in the state of Pennsylvania – speak to the need to tailor principal induction to school contexts. Not all schools are created equal, and prior research shows that principal behavior varies with school contexts (Goldring et al., 2008; May et al., 2012). Our investigation of PIL induction reveals heterogeneous returns to teacher effectiveness based on school characteristics and the timing of PIL completion in the early part of a principal's career. Though we are unable to explicitly identify the specific reasons why the greatest returns to principal professional development are concentrated in the state's most disadvantaged schools, these findings together reveal the importance of considering school context and the local human capital needs of principals in very different school settings when designing and implementing principal induction policies. And while we show that our results are robust across multiple specification checks, it's notable that the endogeneity of the timing into PIL induction remains a residual limitation of our empirical approach.

Results from this study therefore have important policy implications for improving the human capital of school leaders and, in particular, the design of principal induction programs. First, policymakers should recognize the critical importance of supporting the professional development of school leaders in schools serving the most economically students. This is in light of evidence that targeted professional development for principals who lead the most disadvantaged schools has the potential to generate the greatest improvements in teacher effectiveness and student achievement. Second, policymakers should consider mandating the completion of (and provide the necessary supports for) principal professional development during a principal's first two years in the principalship. This is in light of evidence that the returns to principal professional development are concentrated among principals who completed PIL induction during their first two years in the principalship. Indeed, of the twenty states with state-level principal induction policies, 14 states require that novice principals complete induction requirements during their first two years as school leader (see Table 1). This design feature of principal induction efforts in these states suggests the potential for immediate and significant improvements in principal human capital and schooling outcomes in schools with novice principals. Third, principal induction programs should incorporate coursework that focus on developing school leaders' capacities to implement and support high-quality instruction among their schools' teachers (as in NISL's World Class Schooling course) and which augments principals' efforts to use administrative and school-specific data to identify and support the instructional and learning needs of teachers and students (as in NISL's Driving for Results course). Indeed, recent evidence indicates that principals can drive the greatest improvements in school outcomes when they engage with their teachers around instructional practice and establish data-driven, school-wide instructional programs to facilitate and inform these instructionally-

focused interactions with teachers (Grissom et al., 2021). And though Pennsylvania's PIL induction effort lacked a mentorship component, additional benefits to principal professional development may be realized by incorporating ongoing mentorship of novice and early-career principals into principal induction programs. In fact, 17 (of 20) states with principal induction programs (see Table 1) currently assign new principals a mentor to provide feedback on their practice.

Though this study offers a rigorous empirical investigation of statewide principal induction in Pennsylvania, a deeper understanding of the implementation and effects of principal induction would likely benefit from additional analyses. For example, prior evidence finds that when principals participate in effective preparation programs, they gain an understanding of effective leadership practices and later use those leadership practices in their schools (Orr & Orphanos, 2011). Indeed, we do not observe important aspects of PIL implementation and the extent of principals' experiences in PIL coursework, including the quality of PIL instructors who deliver the course content, the extent of principal engagement in the PIL coursework, nor the quality of course instruction or specific course materials. For example, were principals at high-poverty schools more (or less) likely to apply certain leadership practices than principals in schools serving more economically advantaged students? These, among other questions, could support state and local policymakers in their efforts to refine and improve principal induction programs targeted at improving the academic settings of schools, the effectiveness of teachers and, ultimately, student academic achievement. Nonetheless, this paper is the first to study a statewide principal induction reform and contributes much needed evidence on how principal professional development shapes teacher and student outcomes.

References

Atteberry, A., Loeb, S., & Wyckoff, J. (2015). Do first impressions matter? Predicting early career teacher effectiveness. *AERA Open*, *1*(4), 1-23.

Atteberry, A., Loeb, S., & Wyckoff, J. (2017). Teacher churning: Reassignment rates and implications for student achievement. *Educational Evaluation and Policy Analysis*, *39*(1), 3-30.

Austin, W., Chen, B., Goldhaber, D., Hanushek, E., Holden, K., Koedel, C., Ladd, H., Luo, J., Parsons, E., Phelan, G., Rivkin, S., Sass, T., & Turaeva, M. (2019). Path to the principalship and value-added: A cross-state comparison of elementary and middle school principals. Technical Report, CALDER: Washington, DC.

Baker, A., Larcker, D.F., & Wang, C.C. Y. (March 29, 2021). How much whould we trust staggered difference-in-differences estimates? European Corporate Governance Institute – Finance Working Paper No. 736/2021, Rock Center for Corporate Governance at Stanford University Working Paper No. 246, Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3794018

Bartanen, B. (2020). Principal quality and student attendance. *Educational Researcher*, 49(2), 101-113.

Bartanen, B., Grissom, J. A., & Rogers, L. K. (2019). The Impacts of Principal Turnover. *Educational Evaluation and Policy Analysis*, *41*(3), 350-374.

Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does management matter in schools? *The Economic Journal*, 125(584), 647-674.

Boyd, D., Grossman, G., Ing, M., Lankford, H., Loeb, S. & Wyckoff, J. (2011). The influence of school administrators on teacher retention decisions. *American Educational Research Journal*, *48*(2) 303-333.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2012). Estimating the effect of leaders on public sector productivity: The case of school principals (No. w17803). *National Bureau of Economic Research.*

Callaway, B., & Sant'Anna, P.H.C. (2020). Difference-in-differences with multiple time periods. Forthcoming at *Journal of Econometrics*.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593-2632.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633-79.

Chetty, R., Friedman, J., & Rockoff, J. (2014c). Discussion of the American Statistical Association's Statement (2014) on using value-added models for educational assessment. *Statistics and Public Policy*, *1*(1), 111-113.

Chiang, H., Lipscomb, S., & Gill, B. (2016). Is school value added indicative of principal quality? *Education Finance and Policy*, 11(3), 283-309.

Clark, D. (2009). The performance and competitive effects of school autonomy. *Journal of Political Economy 117*(4), 745–783.

Coelli, M., & Green, D. A. (2012). Leadership effects: School principals and student outcomes. *Economics of Education Review, 31*(1), 92-109.

Cohodes, S. R., Setren, E., & Walters, C. (2021). Can successful schools replicate? Scaling up Boston's charter sector. *American Economic Journal: Economic Policy*, 13(1): 138-167.

Corcoran, R. P. (2017). Preparing principals to improve student achievement. *Child & Youth Care Forum*, *46*(5), 769-781.

Corcoran, S. P., Schwartz, A. E., & Weinstein, M. (2012). Training your own: The impact of New York City's aspiring principals program on student achievement. *Educational Evaluation and Policy Analysis*, *34*(2), 232-253.

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110 (9), 2964-96.

Dhuey, E., & Smith, J. (2014). How school principals influence student learning. *Empirical Economics*, 54(2), 851-882.

Dhuey, E., & Smith, J. (2018). How important are school principals in the production of student achievement? *Canadian Journal of Economics*, 47(2), 634-663.

Espinoza, D., & Cardichon, J. (2017). Investing in effective school leadership: How states are taking advantage of opportunities under ESSA. *Palo Alto, CA: Learning Policy Institute*.

Gates, S. M., Hamilton, L. S., Martorell, P., Burkhauser, S., Heaton, P., Pierson, A., Baird, M., Vuollo, M., Li, J., Lavery, D., Harvey, M., & Gu, K. (2014). Preparing Principals to Raise Student Achievement: Implementation and Effects of the New Leaders Program in Ten Districts. *RAND Corporation*.

Goldrick, L. (2016). Support from the start: A 50-state review of policies on new educator induction and mentoring. *New Teacher Center*.

Goldring, E., Huff, J., May, H., & Camburn, E. (2008). School context and individual characteristics: what influences principal practice?. *Journal of Educational Administration*, *46*(3), 332-352.

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. NBER Working Paper No. 25018, National Bureau of Economic Research, Cambridge, MA.

Grissom, J. A., & Bartanen, B. (2019). Principal effectiveness and principal turnover. *Education Finance and Policy*, *14*(3), 355-382.

Grissom, J.A., Egalite, A.J., & Lindsay, C.A. (2021). How Principals Affect Students and Schools: A Systematic Synthesis of Two Decades of Research. New York: The Wallace Foundation. Retrieved from: http://www.wallacefoundation.org/principalsynthesis.

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.

Grissom, J. A., & Loeb, S. (2017). Assessing principals' assessments: Subjective evaluations of teacher effectiveness in low-and high-stakes environments. *Education Finance and Policy*, *12*(3), 369-395.

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.

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American Journal of Education*, *119*(3), 445-470.

Guarino, C. M., Maxfield, M., Reckase, M. D., Thompson, P. N., & Wooldridge, J. M. (2015). An evaluation of empirical Bayes's estimation of value-added teacher performance measures. *Journal of Educational and Behavioral Statistics*, *40*(2), 190-222.

Haller, A., Hunt, E., Pacha, J., & Fazekas, A. (2016). Lessons for States: The" Every Student Succeeds Act" (ESSA) Increases Focus on and Investment in Supporting Principal Preparation and Development. *Center for the Study of Education Policy*.

Herman, R., Gates, S.M., Arifkhanova, A., Bega, A., Chavez-Herreias, E.R., Han, E., Harris, M., Tamargo, J. and Wrabel, S.L. (2017). School Leadership Interventions Under the Every Student Succeeds Act: Evidence Review. *RAND Corporation.*

Herrmann, M., Clark, M., James-Burdumy, S., Tuttle, C., Kautz, T., Knechtel, V., Dotter, D., Wulsin, C. S., and Deke, J. (2019). *The Effects of a Principal Professional Development Program Focused on Instructional Leadership: Appendices* (NCEE 2020-0002). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2015). Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis*, *37*(3), 314-332.

Kahn-Lang, A., & Lang, K. (2020). The promise and pitfalls of differences-in-differences: Reflections on 16 and Pregnant and other applications. *Journal of Business & Economic Statistics*, 38(3), 613-620.

Kane, T. J., & Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation (No. w14607). *National Bureau of Economic Research.*

Korach, S., & Cosner, S. (2017). Developing the leadership pipeline: Comprehensive leadership development. *Handbook of Research on the Education of School Leaders*, 262-282.

Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*, *54*(1), 1-36.

Ladd, H. F. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement?. *Educational Evaluation and Policy Analysis*, *33*(2), 235-261.

Laing, D., Rivkin, S.G., Schiman, J.C., & Ward, J. (2016). Decentralized governance and the quality of school leadership. NBER Working Paper No. 22061, National Bureau of Economic Research, Cambridge, MA.

Leithwood, K., Harris, A., & Hopkins, D. (2008). Seven strong claims about successful school leadership. *School Leadership & Management, 28*(1), 27–42.

Leithwood, K., Seashore Louis, K., Anderson, S., & Wahlstrom, K. (2004). Review of research: How Leadership Influences Student Learning. *The Wallace Foundation.*

Liebowitz, D., & Porter, L. (2019). The effect of principal behaviors on student, teacher, and school outcomes: A systematic review and meta-analysis of the empirical literature. *Review of Educational Research*, 89(5), 785-827.

Loeb, S., Kalogrides, D., & Béteille, T. (2012). Effective schools: Teacher hiring, assignment, development, and retention. *Education Finance and Policy, 7*(3), 269-304.

Manna, P. (2015). Developing Excellent School Principals to Advance Teaching and Learning: Considerations for State Policy. *The Wallace Foundation*.

May, H., Huff, J., & Goldring, E. (2012). A longitudinal study of principals' activities and student performance. *School Effectiveness and School Improvement*, *23*(4), 417-439.

Miller, A. (2013). Principal turnover and student achievement. *Economics of Education Review*, *36*, 60-72.

Morganstein, D., & Wasserstein, R. (2014). ASA statement on value-added models. *Statistics and Public Policy*, *1*(1), 108-110.

Murphy, J., & Vriesenga, M. (2006). Research on school leadership preparation in the United States: An analysis. *School Leadership and Management*, *26*(2), 183-195.

National Institute of School Leadership (2016). NISL Executive Development Program. Retrieved from http://www.education.pa.gov/Documents/Teachers-Administrators/PA%20Inspired%20Leaders-PIL/NISL%20Overview%20Description.pdf

National Institute of School Leadership (2017). Retrieved from http://www.nisl.org/

Ni, Y., Hollingworth, L., Rorrer, A., & Pounder, D. (2017). The evaluation of educational leadership preparation programs. *Handbook of Research on the Education of School Leaders*, 285-308.

Nunnery, J. A., Ross, S. M., & Yen, C. J. (2010). The Effect of the National Institute for School Leadership's Executive Development Program on School Performance Trends in Pennsylvania. *Online Submission*.

Nunnery, J. A., Ross, S. M., Chappell, S., Pribesh, S., & Hoag-Carhart, E. (2011A). The Impact of the NISL Executive Development Program on School Performance in Massachusetts: Cohort 2 Results. *Online Submission*.

Nunnery, J. A., Yen, C. J., & Ross, S. M. (2011B). Effects of the National Institute for School Leadership's Executive Development Program on School Performance in Pennsylvania: 2006-2010 Pilot Cohort Results. *Online Submission*.

Orr, M. T., & Orphanos, S. (2011). How graduate-level preparation influences the effectiveness of school leaders: A comparison of the outcomes of exemplary and conventional leadership preparation programs for principals. *Educational Administration Quarterly*, *47*(1), 18-70.

Pennsylvania Department of Education (2016). PA Inspired Leadership Legislation: Frequently Asked Questions. Retrieved from: http://www.education.pa.gov/Teachers%20-%20Administrators/PA%20Inspired%20Leaders/Pages/PIL-Legislation-FAQ.aspx#tab-1

Pennsylvania Department of Education (2016). Principal Induction Program. Retrieved from: http://www.education.pa.gov/Teachers%20-%20Administrators/PA%20Inspired%20Leaders/Pages/Principals'-Induction-Program.aspx#tab-1

Pennsylvania Department of Education (2018). Invitation to Qualify: Information and Instructions. Retrieved from: https://www.education.pa.gov/Documents/Teachers-Administrators/PA%20Inspired%20Leaders-PIL/PIL%20-%20Invitation%20to%20Qualify.pdf

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on school outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, *44*(5), 635–674.

Rowland, C. (2017). Principal Professional Development: New Opportunities for a Renewed State Focus. *Education Policy Center at American Institutes for Research*.

Steinberg, M. P. (2014). Does greater autonomy improve school performance? Evidence from a regression discontinuity analysis in Chicago. *Education Finance and Policy, 9*(1), 1-35.

Steinberg, M. P., & Cox, A. B. (2017). School autonomy and district support: how principals respond to a tiered autonomy initiative in Philadelphia public schools. *Leadership and Policy in Schools, 16*(1), 130-165.

Steinberg, M. P., & Kraft, M. A. (2017). The sensitivity of teacher performance ratings to the design of teacher evaluation systems. *Educational Researcher, 46*(7), 378-396.

Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching project. *Education Finance and Policy, 10*(4), 535-572.

Turnbull, B.J., Haslam, M.B., Arcaira, E.R., Riley, D.L., Sinclair, B., & Coleman, S. (October 2009). Evaluation of the school administration manager project. Technical report, Policy Studies Associates, Inc., Washington, D.C.

Tables & Figures

Table 1. State-Level Principal Induction Policies

State	Coursework	Mentorship	Tied to Licensure	Dosage/Duration of Induction
Arkansas	Yes	Yes	Yes	New principals must complete mentorship within first 3 years
California	Yes	Yes	Yes	New principals must complete 2 years of induction
Colorado	Yes	Yes	Yes	New principals must complete induction within first 3 years
Delaware	Yes	Yes	Yes	New principals receive 30 hours of mentorship in the first year; additional induction may be completed within first 3 years
Hawaii	Yes	Yes	n/a	New principals must participate in Hawaii's New Principal Academy within the first 2 years of principalship
Iowa	Yes	Yes	n/a	New principals must complete mentorship, which is required for the first year and may last upwards of 3 years
Kansas	Yes	Yes	Yes	New principals must complete one year of mentorship
Maryland	Yes	Yes	n/a	New principals must complete one year of induction
Massachusetts	Yes	Yes	Yes	New principals must complete one year of induction
Missouri	Yes	No	n/a	New principals must complete two years of induction
New Jersey	Yes	Yes	Yes	New principals must complete a one-year residency program
New York	No	Yes	n/a	New principals must complete one year of mentorship
Pennsylvania	Yes	No	Yes	New principals have 5 years to complete the PIL Induction program
South Carolina	Yes	No	n/a	New principals must complete one year of induction
Texas	Yes	Yes	n/a	New principals must complete one year of induction
Utah	No	Yes	n/a	New principals must complete one year of mentorship
Vermont	No	Yes	n/a	New principals must complete two years of mentorship
Virginia	No	Yes	n/a	New principals must complete one year of mentorship
West Virginia	Yes	Yes	n/a	New principals must complete one year of induction
Wisconsin	No	Yes	n/a	New principals must complete up to five years of mentorship

Notes. State-level policy summary of principal induction derived from Goldrick (2016). *Coursework* indicates whether a principal is required to complete formal coursework as part of principal induction. *Mentorship* indicates whether a principal is required to receive mentorship as part of principal induction. *Induction Tied to Licensure* indicates whether completion of induction is a requirement of principals to obtain and maintain their principal license. Cells with "n/a" indicate that information on principal licensure was not available. *Dosage/Duration of Induction* indicates the type and length of induction required of new principals. Goldrick (2016) can be downloaded from: https://newteachercenter.org/wp-content/uploads/2016CompleteReportStatePolicies.pdf; individual state summaries can be downloaded from: https://newteachercenter.org/policy/state-policy-reviews/.

Table 2. Principal and School Characteristics, by PIL Treatment Status

	All Schools			PIL Eligible		
	All	**No PIL**	**PIL**	**All**	**No PIL**	**PIL**
Panel A: Principal Characteristics						
Age	46.65 (8.6)	46.91 (8.65)	43.36*** (7.11)	43.64 (8.14)	43.71 (8.39)	43.36 (7.11)
Female	.45	.44	.52***	.46	.45	.52
White	.88	.88	.86**	.85	.85	.86***
Black	.10	.10	.11*	.12	.12	.11
Hispanic	.01	.01	.01*	.02	.02	.01***
Other Race	.01	.01	.02***	.01	.01	.02**
Experience	19.05 (9.16)	19.32 (9.23)	15.66*** (7.38)	15.31 (8.23)	15.22 (8.44)	15.66* (7.38)
Bachelor's Degree	.13	.13	.18***	.17	.17	.18
Advanced Degree	.86	.87	.82***	.82	.82	.82
Panel B: School Characteristics						
Enrollment	610.85 (368.37)	615.84 (372.27)	547.48*** (307.84)	590.18 (368.93)	601.58 (382.83)	547.48*** (307.84)
Female	.48 (.04)	.48 (.04)	.48 (.04)	.48 (.05)	.48 (.05)	.48 (.04)
Age	10.83 (3.32)	10.87 (3.32)	10.40*** (3.18)	10.90 (3.35)	11.03 (3.38)	10.40*** (3.18)
White	.71 (.31)	.71 (.31)	.69** (.32)	.67 (.34)	.67 (.34)	.69** (.32)
Minority	.23 (.30)	.23 (.30)	.24 (.30)	.27 (.32)	.27 (.33)	.24*** (.30)
FRPL	.44 (.27)	.43 (.27)	.49*** (.28)	.48 (.28)	.47 (.28)	.49* (.28)
IEP	.16 (.06)	.16 (.07)	.16 (.05)	.16 (.07)	.17 (.07)	.16 (.05)
ELL	.03 (.05)	.03 (.05)	.03** (.05)	.03 (.06)	.03 (.06)	.03 (.05)
Gifted	.04 (.05)	.04 (.05)	.03*** (.03)	.03 (.04)	.03 (.04)	.03** (.03)
Math Proficiency	.65 (.24)	.66 (.23)	.56*** (.26)	.60 (.25)	.61 (.25)	.56*** (.26)
ELA Proficiency	.67 (.19)	.68 (.18)	.62*** (.20)	.64 (.20)	.64 (.20)	.62*** (.20)
Charter	.04	.04	.03	.08	.09	.03***

City	.19	.19	.21*	.23	.23	.21
Suburban	.45	.45	.42**	.43	.43	.42
Rural	.25	.24	.29***	.25	.24	.29***
Town	.11	.12	.08***	.09	.1	.08
Principals	4,877	4,831	584	1,886	1,840	584
Schools	3,183	3,168	642	1,736	1,636	642
School*Years	20,636	18,925	1,711	7,148	5,582	1,711

Notes. In Panel A, proportions are reported, except for age and experience, which report means (standard deviation). In Panel B, school-level means (standard deviation) reported, except for charter status and urbanicity, which report proportions. The sample includes K-12 traditional and charter public schools present in any school year during the 2008-09 through 2015-16 school years. *PIL Eligible* includes schools with principals hired during or after the 2008-09 school year who did not complete World Class Schooling or Driving for Results as assistant principals. Treatment status (i.e., *PIL*) is defined at the school-year level and indicates the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). Differences between *PIL* and *No PIL*, within a sample, are statistically significant at *10%, **5%, and ***1%.

Table 3. Effect of PIL Induction on Teacher Effectiveness

	Math VAM						ELA VAM					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: All Schools												
PIL	.012*** (.003)	.010** (.004)	.013*** (.004)	.013*** (.003)	.016*** (.006)	.005 (.007)	.001 (.002)	-.002 (.003)	.003 (.003)	.003 (.003)	.003 (.004)	.000 (.001)
Outcome Mean (SD)	.00 (.09)						.00 (.06)					
Teacher*Years	59,906	59,906	59,906	59,906	59,906	59,906	71,219	71,219	71,219	71,219	71,219	71,219
Teachers	23,513	23,513	23,513	23,513	23,513	23,513	27,647	27,647	27,647	27,647	27,647	27,647
School*Years	9,434	9,434	9,434	9,434	9,434	9,434	9,412	9,412	9,412	9,412	9,412	9,412
Schools	2,255	2,255	2,255	2,255	2,255	2,255	2,257	2,257	2,257	2,257	2,257	2,257
R^2	.233	.252	.663	.662	.267		.254	.273	.618	.616	.287	
Panel B: PIL Eligible												
PIL	.019*** (.005)	.018*** (.007)	.020*** (.006)	.020*** (.005)	.031*** (.009)	.019** (.007)	-.000 (.004)	-.003 (.004)	.001 (.004)	.000 (.004)	.002 (.006)	-.005 (.009)
Outcome Mean (SD)	.00 (.09)						.00 (.06)					
Teacher*Years	23,348	23,348	23,348	23,348	23,348	23,348	27,204	27,204	27,204	27,204	27,204	27,204
Teachers	11,322	11,322	11,322	11,322	11,322	11,322	13,009	13,009	13,009	13,009	13,009	13,009
School*Years	3,731	3,731	3,731	3,731	3,731	3,731	3,729	3,729	3,729	3,729	3,729	3,729
Schools	1,205	1,205	1,205	1,205	1,205	1,205	1,204	1,204	1,204	1,204	1,204	1,204
R^2	.253	.262	.689	.686	.288		.278	.287	.660	.657	.308	
School FE	X		X	X	X	X	X		X	X	X	X
Principal*School FE		X						X				

Teacher FE	X				X			
Teacher*School FE		X				X		
School-Specific Linear Time Trends			X				X	
Staggered Treatment Timing				X				X

Notes. Each column represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include district-year fixed effects. Sample includes all K-12 traditional and charter public schools present in any school year during the 2009-09 through 2015-16 school years. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years; *PIL Eligible* includes schools with principals hired during or after the 2008-09 school year who did not complete World Class Schooling or Driving for Results as assistant principals. PIL is an indicator variable, defined at the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). Staggered treatment timing implements the estimator proposed by De Chaisemartin and d'Haultfoeuille (2020) to account for negative weighting which may arise from staggered treatment timing. Coefficients statistically significant at *10% **5% and ***1% levels.

Table 4. Effect of PIL Induction on Teacher Turnover

	(1)	(2)	(3)	(4)
Panel A: All Schools				
PIL	-.003 (.003)	-.002 (.004)	.003 (.004)	.000 (.005)
Outcome Mean		.11		
Teacher*Years	785,572	785,572	785,572	785,572
Teachers	154,530	154,530	154,530	154,530
School*Years	20,617	20,617	20,617	20,617
Schools	3,177	3,177	3,177	3,177
R^2	.064	.072	.071	
Panel B: PIL Eligible				
PIL	.000 (.005)	-.000 (.005)	.010* (.006)	.004 (.009)
Outcome Mean		.12		
Teacher*Years	261,013	261,013	261,013	261,013
Teachers	79,473	79,473	79,473	79,473
School*Years	7,139	7,139	7,139	7,139
Schools	1,733	1,733	1,733	1,733
R^2	.064	.072	.080	
School FE	X		X	X
Principal*School FE		X		
School-Specific Linear Time Trends			X	
Staggered Treatment Timing				X

Notes. Each column represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include district-year fixed effects. Sample includes all K-12 traditional and charter public schools present in any school year during the 2009-09 through 2015-16 school years. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years; *PIL Eligible* includes schools with principals hired during or after the 2008-09 school year who did not complete World Class Schooling or Driving for Results as assistant principals. PIL is an indicator variable, defined at

the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). Staggered treatment timing implements the estimator proposed by De Chaisemartin and d'Haultfoeuille (2020) to account for negative weighting which may arise from staggered treatment timing. Coefficients statistically significant at *10% **5% and ***1% levels.

Table 5. Effect of PIL Induction on Student Achievement (All Students)

	Math Achievement				ELA Achievement			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: All Schools								
PIL	.011 (.008)	.018** (.008)	.016** (.007)	.011 (.009)	.001 (.007)	-.007 (.007)	.003 (.006)	-.002 (.007)
Outcome Mean (SD)	.00 (1.00)				.00 (1.00)			
Student*Years	5,052,949	5,052,949	5,052,949	5,052,949	5,061,602	5,061,602	5,061,602	5,061,602
Students	1,671,839	1,671,839	1,671,839	1,671,839	1,660,139	1,660,139	1,660,139	1,660,139
School*Years	16,441	16,441	16,441	16,441	16,441	16,441	16,441	16,441
Schools	2,555	2,555	2,555	2,555	2,555	2,555	2,555	2,555
R^2	.396	.399	.399		.412	.414	.416	
Panel B: PIL Eligible								
PIL	.011 (.012)	.015 (.012)	.016 (.012)	.016 (.015)	-.005 (.010)	-.010 (.010)	-.002 (0.009)	-.003 (.011)
Outcome Mean (SD)	-.06 (1.00)				-.06 (1.01)			
Student*Years	1,722,010	1,722,010	1,722,010	1,722,010	1,720,700	1,720,700	1,720,700	1,720,700
Students	826,920	826,920	826,920	826,920	823,829	823,829	823,829	823,829
School*Years	5,690	5,690	5,690	5,690	5,690	5,690	5,690	5,690
Schools	1,392	1,392	1,392	1,392	1,392	1,392	1,392	1,392
R^2	.413	.414	.416		.429	.430	.431	
School FE	X		X	X	X		X	X
Principal*School FE		X				X		

School-Specific Linear Time Trends	X		X	
Staggered Treatment Timing		X		X

Notes. Each column represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include district-year fixed effects. Sample includes all K-12 traditional and charter public schools present in any school year during the 2009-09 through 2015-16 school years. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years; *PIL Eligible* includes schools with principals hired during or after the 2008-09 school year who did not complete World Class Schooling or Driving for Results as assistant principals. PIL is an indicator variable, defined at the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). Staggered treatment timing implements the estimator proposed by De Chaisemartin and d'Haultfoeuille (2020) to account for negative weighting which may arise from staggered treatment timing. *All Students* based on full sample of students. Coefficients statistically significant at *10% **5% and ***1% levels.

Table 6. Effect of PIL Induction on Student Achievement (VAM Teacher Sample)

	Math Achievement				ELA Achievement			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: All Schools								
PIL	.018* (.010)	.012 (.012)	.035*** (.012)	.016 (.013)	.007 (.009)	-.012 (.010)	.016* (.009)	-.003 (.011)
Outcome Mean (SD)			-.02 (1.00)				.00 (1.00)	
Student*Years	3,495,078	3,495,078	3,495,078	3,495,078	3,505,330	3,505,330	3,505,330	3,505,330
Students	1,585,785	1,585,785	1,585,785	1,585,785	1,575,287	1,575,287	1,575,287	1,575,287
School*Years	9,434	9,434	9,434	9,434	9,412	9,412	9,412	9,412
Schools	2,255	2,255	2,255	2,255	2,257	2,257	2,257	2,257
R^2	.422	.423	.424		.438	.439	.440	
Panel B: PIL Eligible								
PIL	.028* (.016)	.040** (.017)	.052*** (.021)	.046 (.023)	.001 (.013)	.005 (.015)	.012 (.014)	.007 (.018)
Outcome Mean (SD)			-.08 (1.00)				-.05 (1.01)	
Student*Years	1,307,278	1,307,278	1,307,278	1,307,278	1,303,749	1,303,749	1,303,749	1,303,749
Students	715,799	715,799	715,799	715,799	712,143	712,143	712,143	712,143
School*Years	3,767	3,767	3,767	3,767	3,752	3,752	3,752	3,752
Schools	1,213	1,213	1,213	1,213	1,210	1,210	1,210	1,210
R^2	.432	.433	.434		.449	.449	.450	
School FE	X		X	X	X		X	X
Principal*School FE		X				X		

School-Specific Linear Time Trends	X		X	
Staggered Treatment Timing		X		X

Notes. Each column represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include district-year fixed effects. Sample includes all K-12 traditional and charter public schools present in any school year during the 2009-09 through 2015-16 school years. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years; *PIL Eligible* includes schools with principals hired during or after the 2008-09 school year who did not complete World Class Schooling or Driving for Results as assistant principals. PIL is an indicator variable, defined at the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). Staggered treatment timing implements the estimator proposed by De Chaisemartin and d'Haultfoeuille (2020) to account for negative weighting which may arise from staggered treatment timing. *VAM Teacher Sample* based on sample that includes school-grade-year observations with teacher VAM (i.e., *VAM Teacher Sample*). Coefficients statistically significant at *10% **5% and ***1% levels.

Table 7. Effects of PIL Induction, by Principal Longevity

	VAM			Achievement (All Students)		Achievement (VAM Teacher Sample)	
	Math	**ELA**	**Teacher Turnover**	**Math**	**ELA**	**Math**	**ELA**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: 3+ Years							
PIL	.019*** (.007)	-.002 (.005)	.001 (.005)	.020 (.012)	-.006 (.010)	.046** (0.018)	.012 (.016)
Unit*Year	18,118	21,229	206,890	1,367,859	1,366,451	1,019,142	1,011,340
Unit	8,114	9,400	56,587	644,845	642,337	540,419	535,801
School*Year	2,866	2,860	5,613	4,478	4,478	2,892	2,849
School	799	801	1,123	903	903	801	801
R^2	.251	.275	.068	.404	.420	.422	.439
Panel B: 4+ Years							
PIL	.031*** (.011)	-.002 (.007)	.001 (.006)	.024* (.014)	-.005 (.012)	.066*** (.022)	.008 (.020)
Unit*Year	13,590	16,065	165,881	1,089,321	1,088,621	788,009	781,914
Unit	5,850	6,842	42,773	511,724	509,685	415,977	411,817
School*Year	2,179	2,162	4,466	3,553	3,553	2,200	2,165
School	568	567	802	641	641	568	568
R^2	.256	.282	.063	.397	.413	.417	.434
Panel C: 5+ Years							
PIL	.031** (.015)	-.000 (.010)	-.001 (.007)	.025 (.018)	-.010 (.015)	.059** (0.027)	.006 (.026)
Unit*Year	10,514	12,712	130,357	867,750	867,512	626,333	621,576
Unit	4,414	5,248	31,939	409,506	407,937	330,295	327,058

School*Year	1,668	1,652	3,474	2,772	2,772	1,688	1,661
School	414	413	564	451	451	414	414
R^2	.256	.277	.062	.398	.413	.416	.434

Notes. Each column within a panel represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include school and district-year fixed effects. *3+ Years* includes schools with principals with three or more years of experience at that school; *4+ Years* includes schools with principals with four or more years of experience at that school; *5+ Years* includes schools with principals with five or more years of experience at that school. The analytic sample is the *PIL Eligible* sample which includes schools with principals hired during or after the 2008-09 school year who did not complete World Class Schooling or Driving for Results as assistant principals. *PIL* is an indicator variable, defined at the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). Coefficients statistically significant at *10% **5% and ***1% levels.

Table 8. Heterogeneous Effects of PIL on Teacher Outcomes, by School Characteristics

	VAM								
	Math			ELA			Teacher Turnover		
	Low	Medium	High	Low	Medium	High	Low	Medium	High
Panel A: FRPL									
PIL	.010 (.007)	-.004 (.013)	.024** (.011)	-.012 (.008)	-.019*** (.007)	.003 (.007)	.006 (.009)	-.007 (.009)	-.001 (.009)
Tercile Mean (SD)	.16 (.08)	.39 (.06)	.77 (.18)	.16 (.08)	.39 (.06)	.77 (.18)	.16 (.08)	.39 (.06)	.76 (.18)
Teacher*Years	6,728	7,320	9,285	8,244	8,722	10,219	81,438	86,201	93,344
Teachers	3,188	3,850	5,304	3,872	4,514	5,841	27,222	31,326	35,946
School*Year	871	1,239	1,606	864	1,247	1,599	1,935	2,466	2,737
Schools	329	468	601	327	469	601	577	755	824
R^2	.261	.305	.252	.271	.326	.286	.051	.063	.099
Panel B: Minority									
PIL	-.001 (.014)	.005 (.009)	.028*** (.010)	-.023** (.010)	-.004 (.008)	.003 (.007)	.004 (.009)	-.001 (.008)	.001 (.008)
Tercile Mean (SD)	.16 (.08)	.39 (.06)	.77 (.18)	.16 (.08)	.39 (.06)	.77 (.18)	.16 (.08)	.39 (.06)	.76 (.18)
Teacher*Years	11,140	7,340	4,842	12,511	8,600	6,066	112,562	80,626	59,405
Teachers	7,130	4,988	3,090	7,964	5,753	3,775	52,675	39,091	23,973
School*Year	1,855	1,181	669	1,856	1,176	670	2,881	2,136	1,742
Schools	880	656	364	883	653	366	1,184	954	605
R^2	.290	.258	.259	.342	.265	.275	.058	.052	.095
Panel C: Achievement									
PIL	.013 (.013)	.021** (.011)	.019 (.016)	-.005 (.008)	.005 (.009)	-.041*** (.008)	-.000 (.011)	.015 (.014)	.003 (.014)

Tercile Mean (SD)	.38 (.15)	.71 (.06)	.87 (.04)	.38 (.15)	.71 (.06)	.87 (.04)	.37 (.16)	.71 (.06)	.88 (.04)
Teacher*Years	6,480	7,664	9,201	7,857	9,217	10,124	75,535	85,930	99,548
Teachers	3,222	3,852	4,933	3,839	4,584	5,433	24,461	29,142	35,294
School*Year	1,132	1,072	1,524	1,141	1,073	1,509	2,260	2,204	2,675
Schools	404	413	509	405	409	510	612	642	734
R^2	.273	.295	.293	.292	.326	.306	.106	.066	.065

Notes. Each column within a panel represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include school and district-year fixed effects. Estimates based on *PIL Eligible* sample. Teacher math (ELA) VAM constructed for grades 6-8 teachers in the 2008-09 through 2015-16 school years and grades 4-5 teachers in the 2013-14 through 2015-16 school years. *FRPL* is the proportion of students in a school*year who are eligible to receive free-or-reduced price lunch. *Minority* is the proportion of students in a school*year who are either Black or Hispanic. *Achievement* is the proportion of students in school*year who tested either proficient or advanced on the mathematics standardized exam. Terciles (*Low*, *Medium*, *High*) are based on school-by-year characteristics. *PIL* is an indicator variable, defined at the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). Coefficients statistically significant at *10% **5% and ***1% levels.

Table 9. Heterogeneous Effects of PIL on Student Achievement, by School Characteristics

	All Students						VAM Teacher Sample					
	Math			ELA			Math			ELA		
	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low	Medium	High
Panel A: FRPL												
PIL	.030* (.018)	-.003 (.016)	.019 (.021)	.012 (.014)	.011 (.012)	-.005 (.018)	.069** (.031)	-.017 (.025)	.053** (.024)	.008 (.026)	-.006 (.021)	.022 (.021)
Tercile Mean (SD)	.16 (.08)	.39 (.06)	.77 (.18)	.16 (.08)	.39 (.06)	.77 (.18)	.15 (.07)	.38 (.06)	.78 (.18)	.15 (.07)	.38 (.06)	.78 (.18)
Student*Years	515,789	557,633	648,588	516,701	559,031	644,968	362,628	434,133	510,517	363,153	431,620	500,344
Students	273,954	305,173	339,606	273,291	305,344	337,247	213,187	262,816	294,869	212,124	261,591	290,184
School*Years	1,453	1,914	2,323	1,453	1,914	2,323	812	1,323	1,632	807	1,303	1,590
Schools	440	587	691	440	587	691	306	492	611	304	491	603
R^2	.347	.316	.363	.335	.330	.373	.360	.330	.373	.367	.350	.393
Panel B: Minority												
PIL	.029 (.021)	.005 (.018)	.018 (.020)	-.008 (.013)	-.026 (.016)	-.011 (.016)	-.037 (.033)	.053* (.031)	.047* (.024)	-.062*** (.017)	.024 (.027)	.011 (.021)
Tercile Mean (SD)	.03 (.01)	.09 (.04)	.57 (.28)	.03 (.01)	.09 (.04)	.57 (.28)	.03 (.01)	.10 (.04)	.60 (.28)	.03 (.01)	.10 (.04)	.60 (.28)
Student*Years	636,449	544,066	541,495	633,190	545,292	542,218	656,170	382,908	268,200	646,912	380,577	267,628
Students	407,047	356,819	307,276	405,187	356,328	306,901	427,962	275,831	182,981	423,331	273,519	181,690
School*Years	2,147	1,845	1,698	2,147	1,845	1,698	2,111	1,086	570	2,071	1,066	563
Schools	942	840	583	942	840	583	959	620	315	951	615	313
R^2	.331	.352	.413	.338	.354	.417	.352	.366	.419	.371	.379	.434
Panel C: Achievement												

PIL	-.003 (.018)	-.014 (.018)	.012 (.020)	-.029* (.016)	-.014 (.017)	-.002 (.015)	.026 (.017)	-.005 (.024)	.066* (.035)	-.013 (.016)	-.009 (.027)	.055** (.023)
Tercile Mean (SD)	.38 (.15)	.72 (.06)	.88 (.04)	.38 (.15)	.72 (.06)	.88 (.04)	.40 (.16)	.74 (.06)	.88 (.03)	.40 (.17)	.74 (.06)	.88 (.03)
Student*Years	471,426	585,211	665,373	472,543	585,646	662,511	389,671	399,931	517,676	388,983	399,763	506,371
Students	241,411	314,197	342,333	241,326	313,559	340,060	226,296	236,372	289,104	225,408	235,269	284,196
School*Years	1,792	1,741	2,157	1,792	1,741	2,157	1,295	979	1,493	1,278	971	1,451
Schools	490	511	594	490	511	594	449	364	501	446	362	495
R^2	.419	.363	.309	.451	.370	.317	.428	.365	.331	.455	.381	.342

Notes. Each column within a panel represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include school and district-year fixed effects. Estimates based on *PIL Eligible* sample. *FRPL* is the proportion of students in a school*year who are eligible to receive free-or-reduced price lunch. *Minority* is the proportion of students in a school*year who are either Black or Hispanic. *Achievement* is the proportion of students in school*year who tested either proficient or advanced on the mathematics standardized exam. Terciles (*Low*, *Medium*, *High*) are based on school-by-year characteristics. *PIL* is an indicator variable, defined at the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). Coefficients statistically significant at *10% **5% and ***1% levels.

Table 10. Heterogeneous Effects of PIL, by Urbanicity and Sector

	Urbanicity				School Sector	
	City	**Suburb**	**Town**	**Rural**	**TPS**	**Charter**
Panel A: Math VAM						
PIL	.029** (.012)	.012** (.006)	.005 (.052)	-.000 (.013)	.010 (.006)	.015 (.032)
Teacher*Years	5,468	11,085	1,740	5,045	22,034	1,295
Teachers	3,086	5,162	882	2,420	10,454	877
School*Years	1,018	1,500	291	912	3,404	308
Schools	325	508	105	300	1,100	103
R^2	.256	.248	.274	.315	.260	.377
Panel B: ELA VAM						
PIL	.005 (.007)	-.008 (.006)	-.006 (.020)	-.020*** (.007)	-.005 (.005)	-.035** (.016)
Teacher*Years	5,986	13,189	2,030	5,990	25,826	1,364
Teachers	3,320	6,091	1,029	2,857	12,098	930
School*Years	1,010	1,501	293	916	3,403	312
Schools	326	506	105	300	1,099	104
R^2	.265	.261	.339	.355	.289	.360
Panel C: Teacher Turnover						
PIL	.011 (.010)	-.008 (.006)	-.004 (.024)	-.004 (.007)	.000 (.005)	.035 (.038)
Teacher*Years	57,980	123,826	22,213	56,994	245,582	15,431
Teachers	21,597	37,043	7,685	18,234	72,944	7,167
School*Years	1,608	3,081	669	1,781	6,597	542
Schools	435	759	196	461	1,594	139
R^2	.087	.055	.073	.061	.071	.109
Panel D: Math Achievement						
PIL	.016 (.023)	.016 (.013)	.022 (.035)	.056** (.023)	.017 (.012)	.080* (.044)
Student*Years	401,024	823,403	138,777	358,806	1,618,048	103,962
Students	199,626	411,754	74,187	184,143	781,618	56,194
School*Years	1,334	2,415	500	1,441	5,242	448
Schools	358	600	148	377	1,271	121
R^2	.374	.390	.308	.318	.412	.337
Panel E: ELA Achievement						

PIL	-.013 (.019)	-.015 (.011)	.005 (.035)	.015 (.014)	-.001 (.010)	-.027 (.045)
Student*Years	398,021	823,943	139,133	359,603	1,616,851	103,849
Students	197,638	410,583	74,329	184,275	778,558	56,102
School*Years	1,334	2,415	500	1,441	5,242	448
Schools	358	600	148	377	1,271	121
R^2	.398	.401	.324	.337	.429	.355
Panel F: Math Achievement (VAM Teacher Sample)						
PIL	.044* (.027)	.058*** (.015)	.040 (.086)	-.036 (.029)	.028 (.018)	.085** (.040)
Student*Years	335,057	618,373	98,795	255,053	1,218,490	88,788
Students	180,591	350,656	58,053	147,737	672,164	50,695
School*Years	1,025	1,518	300	924	3,429	338
Schools	331	512	105	302	1,105	108
R^2	.388	.405	.321	.331	.432	.340
Panel G: ELA Achievement (VAM Teacher Sample)						
PIL	.010 (.023)	.005 (.015)	.025 (.036)	-.056** (.022)	.007 (.017)	-.020 (.042)
Student*Years	327,452	616,205	97,060	254,400	1,208,956	86,161
Students	177,298	348,178	57,363	147,346	665,970	49,564
School*Years	1,001	1,499	290	910	3,380	320
Schools	326	508	105	301	1,099	105
R^2	.409	.419	.336	.357	.450	.361

Notes. Each column within a panel represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include school and district-year fixed effects. In Panels A-E, estimates based on *PIL Eligible* sample; in Panels F-G, estimates based on *PIL Eligible* sample that includes school-grade-year observations with teacher VAM (i.e., *VAM Teacher Sample*). *PIL* is an indicator variable, defined at the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). Coefficients statistically significant at *10% **5% and ***1% levels.

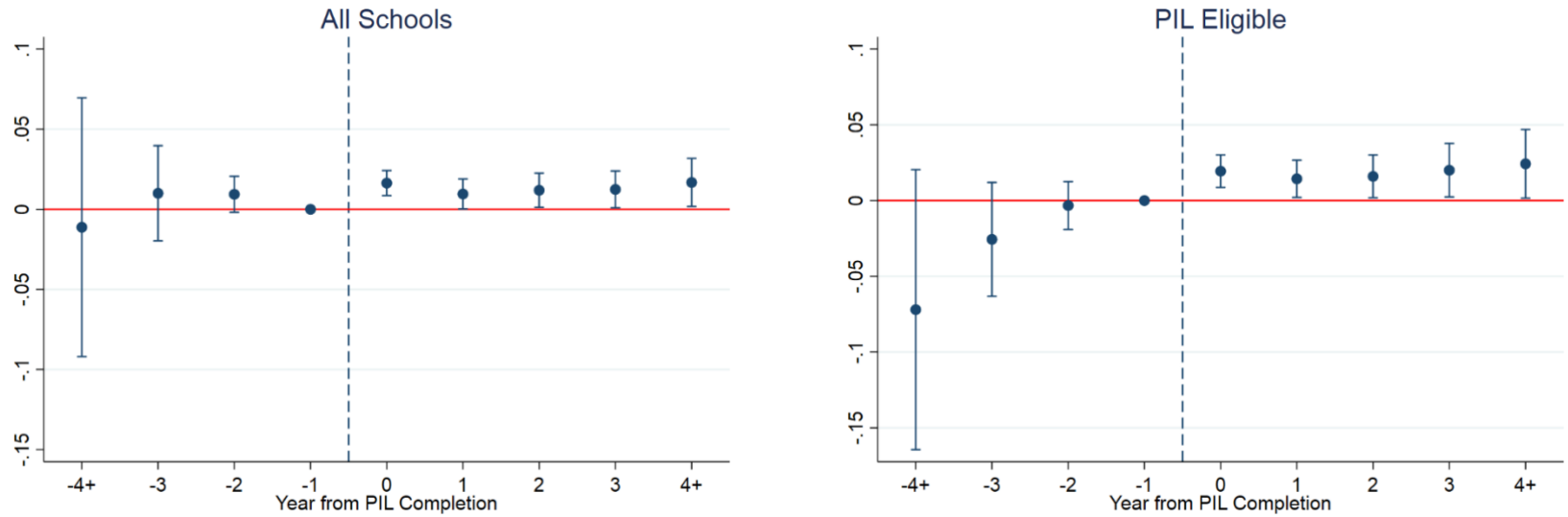
Table 11. Heterogeneous Effects of PIL, by Year Principal Completed PIL

	Year of PIL Completion		
	1-2 Years	**3-4 Years**	**5+ Years**
Panel A: Math VAM			
PIL	.043** (.019)	.016 (.011)	.013 (.023)
Teacher*Years	2,568	4,944	1,449
Teachers	1,335	2,415	713
School*Years	399	790	281
Schools	150	250	81
R^2	.289	.270	.298
Panel B: ELA VAM			
PIL	.023** (.011)	.005 (.007)	.009 (.013)
Teacher*Years	2,792	5,876	1,622
Teachers	1,469	2,806	790
School*Years	403	783	277
Schools	148	249	81
R^2	.340	.271	.335
Panel C: Teacher Turnover			
PIL	-.020 (.023)	.026* (.013)	.031 (.024)
Teacher*Years	22,347	52,288	22,086
Teachers	8,293	16,235	6,751
School*Years	683	1,537	645
Schools	202	360	133
R^2	.077	.073	.079
Panel D: Math Achievement			
PIL	.024 (.035)	.004 (.020)	.002 (.064)
Student*Years	163,762	365,962	118,166
Students	91,925	196,099	63,851
School*Years	591	1,269	517
Schools	172	303	103
R^2	.407	.393	.432
Panel E: ELA Achievement			

PIL	.015 (.023)	.008 (.017)	.029 (.049)
Student*Years	163,506	365,462	117,876
Students	91,838	195,294	63,545
School*Years	592	1,269	517
Schools	172	303	103
R^2	.424	.405	.447
Panel F: Math Achievement (VAM Teacher Sample)			
PIL	-.017 (.017)	.015 (.026)	.028 (.069)
Student*Years	123,919	263,413	75,615
Students	74,382	150,318	43,914
School*Years	411	794	283
Schools	150	251	81
R^2	.427	.411	.457
Panel G: ELA Achievement (VAM Teacher Sample)			
PIL	-.015 (.015)	.018 (.022)	.012 (.050)
Student*Years	123,847	263,781	74,980
Students	74,250	149,857	43,412
School*Years	412	796	279
Schools	150	252	81
R^2	.445	.424	.473

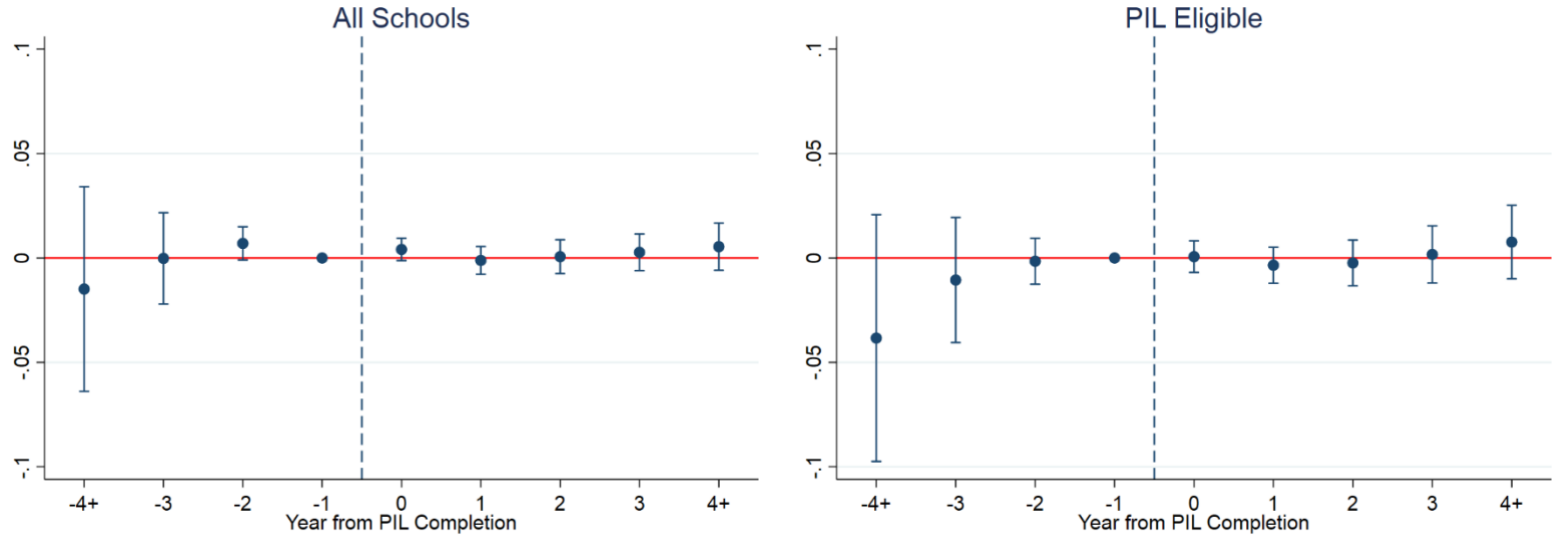
Notes. Each column within a panel represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include school and district-year fixed effects. In Panels A-E, estimates based on *PIL Eligible* sample; in Panels F-G, estimates based on *PIL Eligible* sample that includes school-grade-year observations with teacher VAM (i.e., *VAM Teacher Sample*). *PIL* is an indicator variable, defined at the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). *Year of PIL Completion* indicates the year during a principal's career he/she completed PIL induction. Coefficients statistically significant at *10% **5% and ***1% levels.

Figure 1. Event Study Estimates of PIL Induction on Teacher Effectiveness (Math VAM)



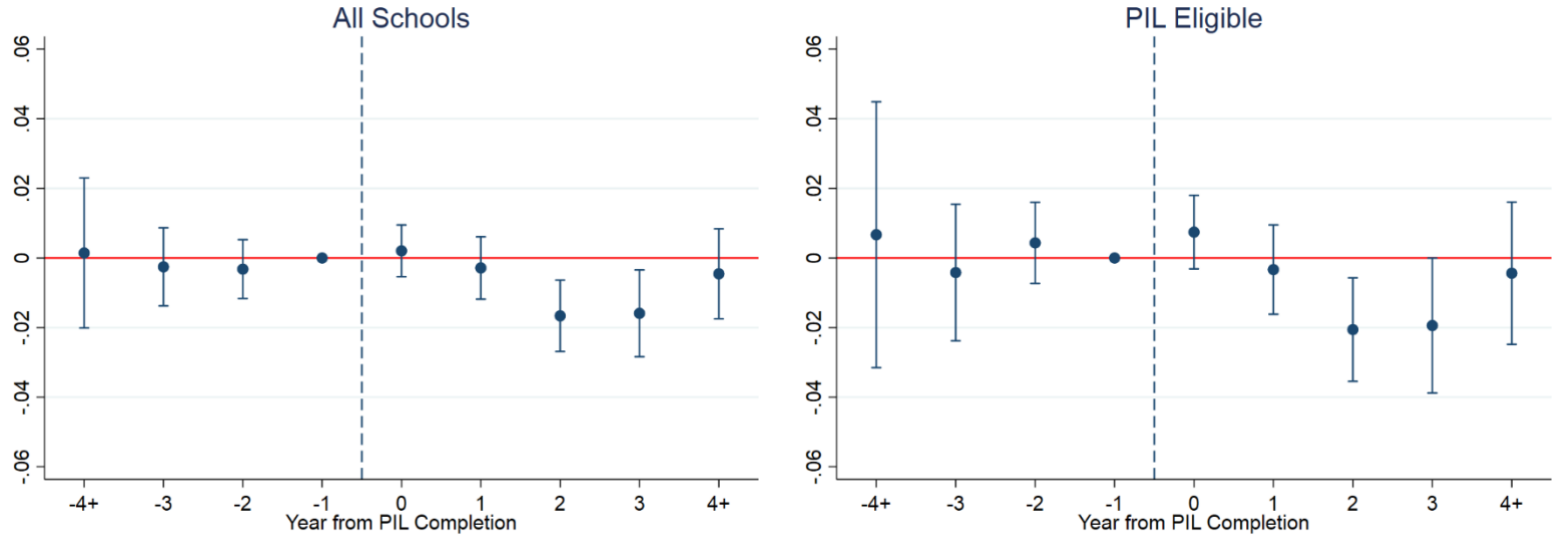
Notes. Figures report coefficient estimates and 95 percent confidence intervals for teacher effectiveness (math VAM). Coefficients reported relative to the year prior to PIL completion. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years. *PIL Eligible* is a subset of *All Schools* and includes school-year observations where the school principal is PIL Eligible (i.e., a school's principal became a principal for the first time in Pennsylvania after January 2008 and did not complete a PIL course as an assistant principal). See Table A4 for event study estimates.

Figure 2. Event Study Estimates of PIL Induction on Teacher Effectiveness (ELA VAM)



Notes. Figures report coefficient estimates and 95 percent confidence intervals for teacher effectiveness (ELA VAM). Coefficients reported relative to the year prior to PIL completion. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years. *PIL Eligible* is a subset of *All Schools* and includes school-year observations where the school principal is PIL Eligible (i.e., a school's principal became a principal for the first time in Pennsylvania after January 2008 and did not complete a PIL course as an assistant principal). See Table A4 for event study estimates.

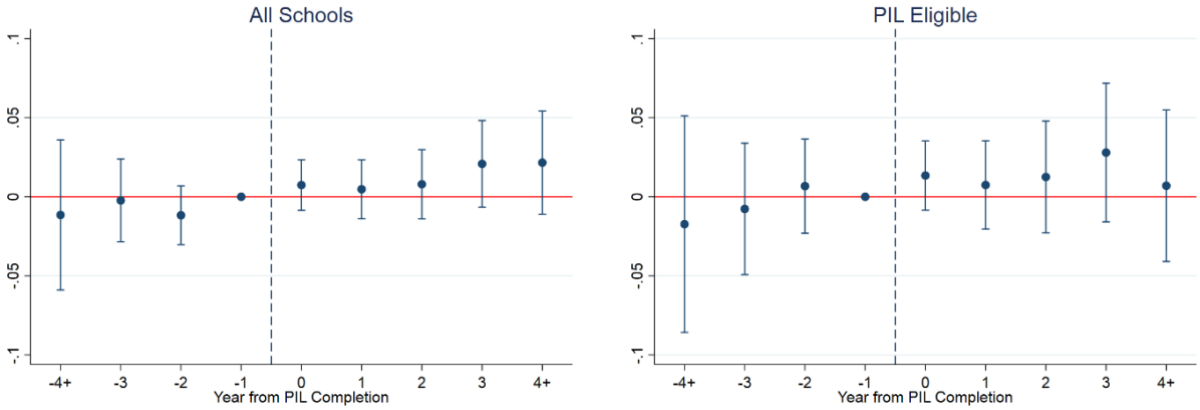
Figure 3. Event Study Estimates of PIL Induction on Teacher Turnover



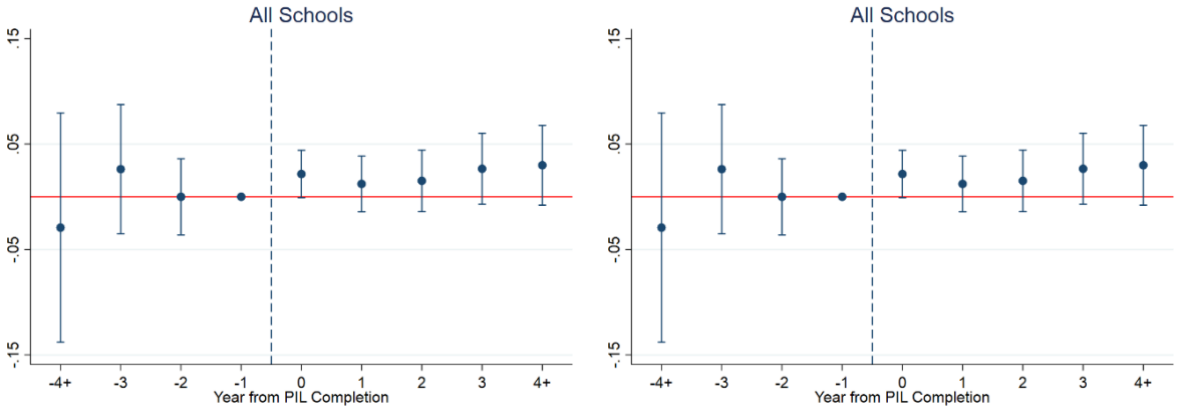
Notes. Figures report coefficient estimates and 95 percent confidence intervals for teacher turnover. Coefficients reported relative to the year prior to PIL completion. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years. *PIL Eligible* is a subset of *All Schools* and includes school-year observations where the school principal is PIL Eligible (i.e., a school's principal became a principal for the first time in Pennsylvania after January 2008 and did not complete a PIL course as an assistant principal). See Table A5 for event study estimates.

Figure 4. Event Study Estimates of PIL Induction on Student Math Achievement

<u>Panel A:</u> All Students



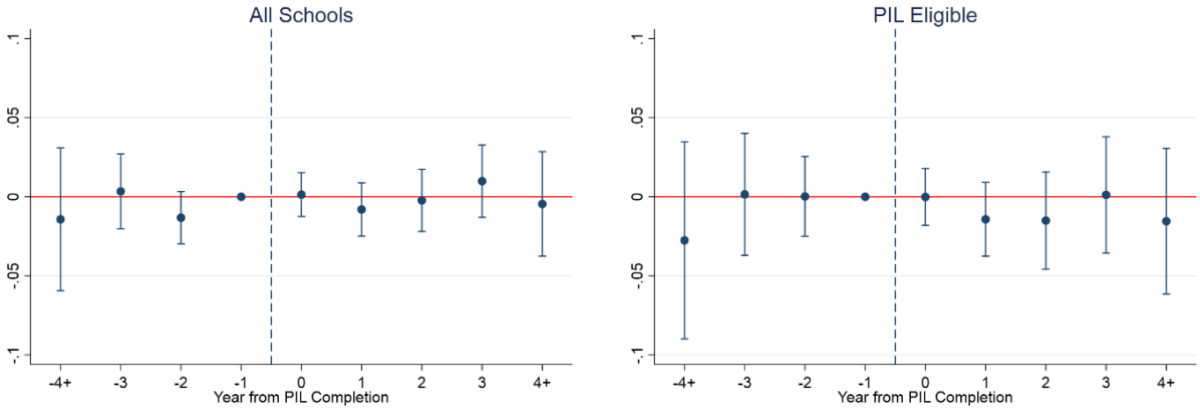
<u>Panel B:</u> VAM Teacher Sample



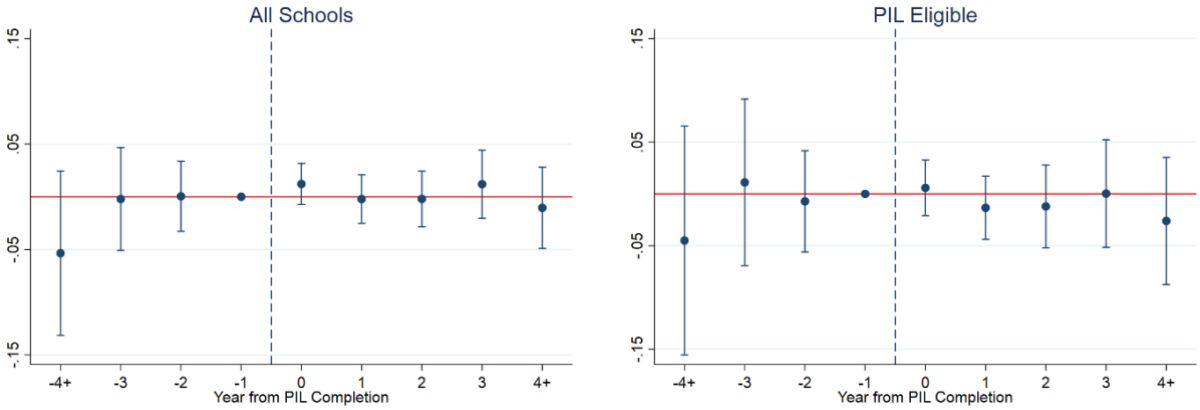
Notes. Figures report coefficient estimates and 95 percent confidence intervals for student math achievement. Coefficients reported relative to the year prior to PIL completion. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years. *PIL Eligible* is a subset of *All Schools* and includes school-year observations where the school principal is PIL Eligible (i.e., a school's principal became a principal for the first time in Pennsylvania after January 2008 and did not complete a PIL course as an assistant principal). In Panel A, estimates for *All Students* based on full sample of students. In Panel B, estimates for *VAM Teacher Sample* based on sample that includes school-grade-year observations with teacher VAM. See Table A6 for event study estimates.

Figure 5. Event Study Estimates of PIL Induction on Student ELA Achievement

Panel A: All Students



Panel B: VAM Teacher Sample



Notes. Figures report coefficient estimates and 95 percent confidence intervals for student ELA achievement. Coefficients reported relative to the year prior to PIL completion. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years. *PIL Eligible* is a subset of *All Schools* and includes school-year observations where the school principal is PIL Eligible (i.e., a school's principal became a principal for the first time in Pennsylvania after January 2008 and did not complete a PIL course as an assistant principal). In Panel A, estimates for *All Students* based on full sample of students. In Panel B, estimates for *VAM Teacher Sample* based on sample that includes school-grade-year observations with teacher VAM. See Table A6 for event study estimates.

Appendix A. Supplemental Tables & Figures

Table A1. Professional Development and Induction Requirements under Pennsylvania *Act 48* and *Act 45*

	Act 48	Act 45
Year Enacted	1999	2007
Effective Date	July 1, 2000	January 1, 2008
Professionals Affected	All educators in Pennsylvania that hold the following certificates: Instructional I and II, Educational Specialist I and II, Administrative, Supervisory, Letters of Eligibility, and all vocational certificates	Principals, Assistant/Vice Principals, Superintendents, Assistant Superintendents, Intermediate Unit Executive Director, Intermediate Unit Assistant Executive Director, Director of an Area Vocational-Technical School
Continuing Professional Development Requirements	Every five years, educators must either earn six hours of college credits, six credits of PDE approved professional development courses, 180 hours of professional development programs approved by PDE, or any combination of the three	Principals employed before January 1, 2008 must complete their professional development requirements proportional to their employment period (e.g., if a principal has worked was employed for one year before January 1, 2008, she must complete 80% of her professional development requirements in a PIL course)
Principal Induction Requirements	N/A	All principals employed for the first time on or after January 1, 2008 must complete the Pennsylvania Inspired Leadership Induction Program within five years of employment
Alignment to Pennsylvania Standards	N/A	Aligned to 3 core leadership standards and 6 corollary standards
Consequence of Not Meeting Requirements	Suspension of license, resulting in suspension of employment	Suspension of license, resulting in suspension of employment

Notes. Source: Pennsylvania Department of Education (www.education.pa.gov). Although Act 48 affects all Pennsylvania educators, Act 45 only affects those employed as principals, assistant/vice principals, superintendents, assistant superintendents, intermediate unit executive directors, intermediate unit assistant executive directors, or directors of an area vocational-technical schools (i.e., school or district leaders).

Table A2. Pennsylvania Leadership Standards

Core Leadership Standards	Corollary Leadership Standards
• The leader has the knowledge and skills to think and plan strategically, creating an organizational vision around personalized student success	• The leader knows how to create a culture of teaching and learning with an emphasis on learning
• The leader has an understanding of standards-based systems theory and design and the ability to transfer that knowledge to the leader's job as the architect of standards-based reform in the school	• The leader knows how to manage resources for effective results
• The leader has the ability to access and use appropriate data to inform decision-making at all levels of the system	• The leader knows how to collaborate, communicate, engage and empower others inside and outside of the organization to pursue excellence in learning • The leader knows how to operate in a fair and equitable manner with personal and professional integrity • The leader knows how to advocate for children and public education in the larger political, social, economic, legal and cultural context • The leader knows how to support professional growth of self and others through practice and inquiry

Notes. Source: Pennsylvania Department of Education (www.education.pa.gov). The Pennsylvania Inspired Leadership (PIL) induction program focuses on the 3 core leadership standards.

Table A3. National Institute of School Leadership (NISL) Coursework

Course Title	Course Description	Course Units
Panel A: PIL Courses		
World Class Schooling: Vision and Goals	This course focuses on providing principals the strategic planning tools to implement a vision of high-quality teaching and student achievement.	• The Educational Challenge: This unit emphasizes the need for all students to be college and career ready. • Principal as Strategic Thinker: This unit gives principals the tools to be strategic thinkers and effective decision-makers. • Elements of Standards-Aligned Instructional Systems: This unit emphasizes an understanding of standards, assessments, and how to align instruction to standards.
Driving for Results	Principals are trained to examine student achievement and many other types of data to identify school, teacher and individual student needs.	• Driving for Change • Leading for Results • Culminating Simulation
Panel B: Non-PIL Courses		
Focus on Teaching and Learning	This course focuses on the principal as an instructional leader; participants learn to "integrate curriculum, instruction, and assessment within the instructional core."	• Foundations of Effective Learning: This unit is designed to help participants understand "the relationship between ideas about learning, the alignment of standards, curriculum, instruction, and assessment." • Leadership in the Instructional core: Part 1: This unit provides participants the tools to implement and support effective English Language Arts and History instruction. • Leadership in the Instructional Core: Part 2: This unit provides participants the tools to implement and support effective Mathematics and Science instruction. • Coaching for High Quality Teaching: This unit gives participants practice in coaching and developing human capital within a school.
Sustaining Transformation through Capacity and Commitment	This course focuses on principals as organizational leaders of schools.	• Promoting the Learning Organization: This unit teaches principals to view schools as learning organizations and apply teacher accountability to improve instruction.

- Teams for Instructional Leadership: This unit teaches principals the importance of distributing leadership throughout the school and how to create and foster leadership teams.
- Ethical Leadership for Equity: This unit teaches principals how to make moral and ethical decisions despite facing external and operational pressures.
- Driving and Sustaining Transformation: This unit gives principals the skills to maintain changes over time.

Notes. Source: Pennsylvania Department of Education (www.education.pa.gov). The Pennsylvania Inspired Leadership (PIL) program consists of World-Class Schooling and Driving for Results from the 2008-09 academic year through the 2015-16 year. In 2016-17, PDE replaced Driving for Results with Focusing on Teaching and Learning as a PIL course.

Table A4. Event Study Estimates of PIL Induction on Teacher Effectiveness

	Math VAM					ELA VAM				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: All Schools										
4+ Years Before	-.011 (.041)	-.047 (.041)	-.021 (.062)	-.016 (.056)	-.042 (.044)	-.015 (.025)	-.029 (.025)	-.025 (.033)	-.022 (.028)	-.013 (.018)
3 Years Before	.010 (.015)	-.007 (.014)	-.004 (.019)	-.004 (.016)	.001 (.016)	-.000 (.011)	-.004 (.011)	.001 (.009)	.002 (.008)	-.009 (.013)
2 Years Before	.009 (.006)	.002 (.006)	.004 (.006)	.003 (.005)	.001 (.008)	.007* (.004)	.003 (.005)	.008 (.005)	.008* (.005)	.001 (.005)
Year of PIL	.016*** (.004)	.013*** (.005)	.014*** (.005)	.014*** (.004)	.018*** (.006)	.004 (.003)	.001 (.003)	.005 (.003)	.004 (.003)	.003 (.004)
1 Year After	.010** (.005)	.004 (.005)	.012** (.006)	.012** (.005)	.008 (.007)	-.001 (.003)	-.004 (.005)	.002 (.004)	.002 (.004)	-.004 (.006)
2 Years After	.012** (.005)	.005 (.006)	.015** (.007)	.015*** (.006)	.012 (.009)	.001 (.004)	-.005 (.006)	.005 (.005)	.004 (.004)	.006 (.007)
3 Years after	.012** (.006)	.003 (.007)	.011 (.008)	.011* (.006)	.011 (.011)	.003 (.004)	-.005 (.006)	.004 (.005)	.003 (.005)	.013 (.008)
4+ Years After	.017** (.008)	.008 (.009)	.015 (.010)	.015* (.008)	.014 (.015)	.005 (.006)	-.002 (.007)	.006 (.007)	.005 (.006)	.022* (.011)
P-Values from F-Statistic:										
Before PIL	.88	.29	.81	.80	.58	.47	.31	.31	.28	.53
Year of/After PIL	.61	.50	.92	.88	.51	.48	.69	.95	.92	.06
Outcome Mean (SD)			.00 (.09)					.00 (.06)		
Teacher*Years	59,906	59,906	59,906	59,906	59,906	71,219	71,219	71,219	71,219	71,219
Teachers	23,513	23,513	23,513	23,513	23,513	27,647	27,647	27,647	27,647	27,647
School*Years	9,434	9,434	9,434	9,434	9,434	9,412	9,412	9,412	9,412	9,412

Schools	2,255	2,255	2,255	2,255	2,255	2,257	2,257	2,257	2,257	2,257
R^2	.233	.252	.663	.662	.267	.255	.274	.618	.616	.287
Panel B: PIL Eligible										
4+ Years Before	-.072 (.047)	-.073 (.048)	-.098 (.060)	-.100** (.049)	-.200 (.125)	-.015 (.025)	-.029 (.025)	-.025 (.033)	-.022 (.028)	-.089* (.051)
3 Years Before	-.026 (.019)	-.033* (.018)	-.028 (.026)	-.027 (.021)	-.092 (.057)	-.000 (.011)	-.004 (.011)	.001 (.009)	.002 (.008)	-.039 (.033)
2 Years Before	-.003 (.008)	-.004 (.009)	-.005 (.010)	-.006 (.008)	-.039 (.031)	.007* (.004)	.003 (.005)	.008 (.005)	.008* (.005)	-.003 (.016)
Year of PIL	.019*** (.005)	.018*** (.007)	.019*** (.007)	.019*** (.006)	.050* (.026)	.004 (.003)	.001 (.003)	.005 (.003)	.004 (.003)	-.000 (.016)
1 Year After	.014** (.006)	.011 (.008)	.019** (.008)	.019*** (.007)	.057 (.047)	-.001 (.003)	-.004 (.005)	.002 (.004)	.002 (.004)	-.014 (.030)
2 Years After	.016** (.007)	.009 (.011)	.024** (.010)	.024*** (.008)	.066 (.067)	.001 (.004)	-.005 (.006)	.005 (.005)	.004 (.004)	-.014 (.044)
3 Years after	.020** (.009)	.011 (.013)	.025** (.012)	.025** (.010)	.072 (.087)	.003 (.004)	-.005 (.006)	.004 (.005)	.003 (.005)	-.009 (.059)
4+ Years After	.024** (.012)	.011 (.016)	.027* (.016)	.027** (.013)	.061 (.105)	.005 (.006)	-.002 (.007)	.006 (.007)	.005 (.006)	-.000 (.074)
P-Values from F-Statistic:										
Before PIL	.05	.03	.10	.04	.18	.32	.31	.31	.28	.07
Year of/After PIL	.66	.74	.95	.91	.83	.32	.69	.95	.92	.13
Outcome Mean (SD)			.00 (.09)					.00 (.06)		
Teacher*Years	23,348	23,348	23,348	23,348	23,348	27,204	27,204	27,204	27,204	27,204
Teachers	11,322	11,322	11,322	11,322	11,322	13,009	13,009	13,009	13,009	13,009
School*Years	3,731	3,731	3,731	3,731	3,731	3,729	3,729	3,729	3,729	3,729
Schools	1,205	1,205	1,205	1,205	1,205	1,204	1,204	1,204	1,204	1,204
R^2	.253	.262	.690	.687	.295	.278	.287	.661	.657	.315

School FE	X		X	X	X	X		X	X	X
Principal*School FE		X					X			
Teacher FE			X					X		
Teacher*School FE				X					X	
School-Specific Linear Time Trends					X					X

Notes. Each column represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include district-year fixed effects. Sample includes all K-12 traditional and charter public schools present in any school year during the 2009-09 through 2015-16 school years. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years; *PIL Eligible* includes schools with principals hired during or after the 2008-09 school year who did not complete World Class Schooling or Driving for Results as assistant principals. PIL is an indicator variable, defined at the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). P-Value from F-Statistic displays the p-values of F-tests of the joint significance of the pre-treatment effects (i.e., Before PIL) and of the post-treatment effects (i.e., Year of/After PIL). Coefficients statistically significant at *10% **5% and ***1% levels.

Table A5. Event Study Estimates of PIL Induction on Teacher Turnover

	(1)	(2)	(3)
Panel A: All Schools			
4+ Years Before	.001 (.011)	-.000 (.013)	-.004 (.013)
3 Years Before	-.003 (.006)	-.002 (.007)	-.005 (.006)
2 Years Before	-.003 (.004)	-.001 (.005)	-.002 (.005)
Year of PIL	.002 (.004)	.001 (.004)	.003 (.005)
1 Year After	-.003 (.005)	-.001 (.005)	.001 (.006)
2 Years After	-.017*** (.005)	-.016*** (.005)	-.015* (.008)
3 Years after	-.016** (.006)	-.014** (.007)	-.013 (.011)
4+ Years After	-.005 (.007)	-.003 (.007)	-.003 (.015)
P-Values from F-Statistic:			
Before PIL	.91	.96	.91
Year of/After PIL	.00	.00	.02
Outcome Mean		.11	
Teacher*Years	785,572	785,572	785,572
Teachers	154,530	154,530	154,530
School*Years	20,617	20,617	20,617
Schools	3,177	3,177	3,177
R^2	.064	.072	.071

Panel B: PIL Eligible			
4+ Years Before	.007 (.019)	.013 (.022)	.001 (.022)
3 Years Before	-.004 (.010)	-.008 (.011)	-.007 (.013)
2 Years Before	.004 (.006)	.002 (.006)	.001 (.007)
Year of PIL	.007 (.005)	.004 (.006)	.011* (.007)
1 Year After	-.003 (.007)	-.004 (.007)	.010 (.011)
2 Years After	-.021*** (.008)	-.021** (.008)	.000 (.016)
3 Years after	-.019** (.010)	-.019* (.010)	.010 (.023)
4+ Years After	-.004 (.010)	-.001 (.011)	.021 (.031)
P-Values from F-Statistic:			
Before PIL	.57	.34	.73
Year of/After PIL	.00	.00	.46
Outcome Mean		.12	
Teacher*Years	261,013	261,013	261,013
Teachers	79,473	79,473	79,473
School*Years	7,139	7,139	7,139
Schools	1,733	1,733	1,733
R^2	.080	.084	.091
School FE	X		X
Principal*School FE		X	
School-Specific Linear Time Trends			X

Notes. Each column represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include district-year fixed effects. Sample includes all K-12 traditional and charter public schools present in any school year during the 2009-09 through 2015-16 school years. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years; *PIL Eligible* includes schools with principals hired during or after the 2008-09 school year who did not complete World Class Schooling or Driving for Results as assistant principals. PIL is an indicator variable, defined at the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). P-Value from F-Statistic displays the p-values of F-tests of the joint significance of the pre-treatment effects (i.e., Before PIL) and of the post-treatment effects (i.e., Year of/After PIL). Coefficients statistically significant at *10% **5% and ***1% levels.

Table A6. Event Study Estimates of PIL Induction on Student Achievement

	All Schools						PIL Eligible					
	Math Achievement			ELA Achievement			Math Achievement			ELA Achievement		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: All Students												
4+ Years Before	-.011	.014	-.022	-.014	.003	-.017	-.017	.053	-.046	-.028	.026	-.050
	(.024)	(.037)	(.025)	(.023)	(.029)	(.028)	(.035)	(.061)	(.063)	(.032)	(.048)	(.057)
3 Years Before	-.002	.007	-.012	.003	.007	-.006	-.008	.013	-.004	.002	.007	-.008
	(.013)	(.012)	(.013)	(.012)	(.012)	(.013)	(.021)	(.019)	(.036)	(.020)	(.017)	(.029)
2 Years Before	-.012	-.001	-.015*	-.013	-.014*	-.020***	.007	.018	-.006	.000	.002	-.020
	(.009)	(.008)	(.008)	(.008)	(.007)	(.007)	(.015)	(.011)	(.018)	(.013)	(.010)	(.015)
Year of PIL	.007	.020***	.017**	.001	-.005	.005	.013	.022**	.027**	-.000	-.003	.007
	(.008)	(.007)	(.007)	(.007)	(.006)	(.006)	(.011)	(.010)	(.013)	(.009)	(.008)	(.010)
1 Year After	.005	.014	.017*	-.008	-.020***	-.005	.007	.018	.037*	-.014	-.021*	.006
	(.009)	(.009)	(.010)	(.009)	(.008)	(.009)	(.014)	(.014)	(.020)	(.012)	(.012)	(.018)
2 Years After	.008	.014	.023*	-.002	-.015	.008	.012	.018	.070**	-.015	-.022	.035
	(.011)	(.011)	(.014)	(.010)	(.010)	(.012)	(.018)	(.019)	(.030)	(.016)	(.016)	(.026)
3 Years after	.021	.034**	.046**	.010	-.001	.029*	.028	.039	.113***	.001	-.004	.065*
	(.014)	(.014)	(.019)	(.012)	(.012)	(.015)	(.022)	(.024)	(.040)	(.019)	(.019)	(.034)
4+ Years After	.022	.029*	.073***	-.005	-.018	.048**	.007	.014	.154***	-.015	-.021	.100**
	(.017)	(.017)	(.026)	(.017)	(.017)	(.023)	(.024)	(.028)	(.053)	(.023)	(.026)	(.045)
P-Values from F-Statistic:												
Before PIL	.56	.81	.88	.23	.11	.40	.56	.70	.99	.6	.85	.26
Year of/After PIL	.45	.20	.07	.21	.03	.02	.45	.23	.40	.13	.06	.01
Outcome Mean (SD)		.00 (1.00)			.00 (1.00)			-.06 (1.00)		-.06 (1.01)		
Student*Years	5,052,949	5,052,949	5,052,949	5,061,602	5,061,602	5,061,602	1,722,010	1,722,010	1,722,010	1,720,700	1,720,700	1,720,700
Students	1,671,839	1,671,839	1,671,839	1,660,139	1,660,139	1,660,139	826,920	826,920	826,920	823,829	823,829	823,829
School*Years	16,441	16,441	16,441	16,441	16,441	16,441	5,690	5,690	5,690	5,690	5,690	5,690
Schools	2,555	2,555	2,555	2,555	2,555	2,555	1,392	1,392	1,392	1,392	1,392	1,392

R^2	.396	.399	.399	.412	.414	.415	.413	.414	.416	.429	.430	.431
Panel B: VAM Teacher Sample												
4+ Years Before	-.029 (.055)	.064 (.121)	-.104** (.045)	-.054 (.040)	.035 (.093)	-.088* (.051)	-.061 (.076)	.054 (.149)	-.146 (.094)	-.045 (.056)	.027 (.121)	-.161** (.069)
3 Years Before	.026 (.031)	.048* (.027)	-.012 (.024)	-.002 (.025)	.017 (.020)	-.026 (.023)	-.003 (.045)	.029 (.035)	-.020 (.074)	.011 (.041)	.023 (.027)	-.020 (.053)
2 Years Before	-.000 (.018)	.000 (.012)	-.033** (.017)	.000 (.017)	-.007 (.012)	-.036** (.015)	-.006 (.030)	-.002 (.017)	-.037 (.038)	-.007 (.025)	-.011 (.016)	-.052* (.028)
Year of PIL	.022* (.011)	.020* (.011)	.033*** (.011)	.012 (.010)	-.005 (.009)	.013 (.009)	.028* (.016)	.042*** (.016)	.063** (.026)	.006 (.014)	.007 (.013)	.018 (.018)
1 Year After	.012 (.014)	.008 (.014)	.032** (.015)	-.002 (.012)	-.025** (.012)	.012 (.013)	.019 (.020)	.034 (.022)	.075* (.043)	-.013 (.016)	-.012 (.018)	.014 (.029)
2 Years After	.015 (.015)	.009 (.017)	.037* (.019)	-.002 (.014)	-.026* (.015)	.025 (.016)	.031 (.025)	.046 (.029)	.131** (.060)	-.012 (.020)	-.010 (.025)	.044 (.040)
3 Years after	.027 (.017)	.019 (.020)	.023 (.025)	.012 (.017)	-.018 (.019)	.022 (.020)	.046 (.032)	.055 (.036)	.168** (.081)	.000 (.026)	.000 (.031)	.064 (.054)
4+ Years After	.030 (.019)	.021 (.022)	.007 (.033)	-.010 (.020)	-.043* (.023)	.010 (.028)	.029 (.035)	.030 (.040)	.178* (.100)	-.026 (.031)	-.030 (.037)	.074 (.072)
P-Values from F-Statistic:												
Before PIL	.44	.14	.01	.43	.2	.00	.72	.55	.00	.58	.2	.00
Year of/After PIL	.61	.62	.68	.12	.1	.34	.43	.35	.05	.12	.1	.34
Outcome Mean (SD)		-.02 (1.01)			.00 (1.00)			-.08 (1.00)			-.05 (1.01)	
Student*Years	3,495,078	3,495,078	3,495,078	3,505,330	3,505,330	3,505,330	1,307,278	1,307,278	1,307,278	1,303,749	1,303,749	1,303,749
Students	1,585,785	1,585,785	1,585,785	1,575,287	1,575,287	1,575,287	715,799	715,799	715,799	712,143	712,143	712,143
School*Years	9,434	9,434	9,434	9,412	9,412	9,412	3,767	3,767	3,767	3,752	3,752	3,752
Schools	2,255	2,255	2,255	2,257	2,257	2,257	1,213	1,213	1,213	1,210	1,210	1,210
R^2	.422	.423	.424	.449	.450	.450	.432	.433	.434	.449	.449	.450
School FE	X		X	X		X	X		X	X		X
School*Principal FE		X			X			X			X	
School-Specific Linear Time Trends			X			X			X			X

Notes. Notes. Each column represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include district-year fixed effects. Sample includes all K-12 traditional and charter public schools present in any school year during the 2009-09 through 2015-16 school years. *All Schools* includes all school-year observations in the 2008-09 through 2015-16 school years; *PIL Eligible* includes schools with principals hired during or after the 2008-09 school year who did not complete World Class Schooling or Driving for Results as assistant principals. PIL is an indicator variable, defined at the school-year level, which equals one in the year (and all subsequent years) in which a principal has completed PIL induction in the same school (i.e., principal-school cells). P-Value from F-Statistic displays the p-values of F-tests of the joint significance of the pre-treatment effects (i.e., Before PIL) and of the post-treatment effects (i.e., Year of/After PIL). Coefficients statistically significant at *10% **5% and ***1% levels. *All Students* based on full sample of students. *VAM Teacher Sample* based on sample that includes school-grade-year observations with teacher VAM (i.e., *VAM Teacher Sample*). Coefficients statistically significant at *10% **5% and ***1% levels.

Table A7. Principal and School Characteristics, by Year Principal Completed PIL

		Year of PIL Completion				
	All Principals	**1**	**2**	**3**	**4**	**5**
Panel A. Principal Characteristics						
Age	42.15 (7.14)	40.91 (6.67)	40.84 (6.7)	42.97 (7.46)	41.64 (6.55)	43.23*** (7.58)
Female	.51	.57	.47	.56	.49	.5**
White	.87	.75	.85	.86	.94	.83***
Black	.11	.18	.12	.11	.05	.14***
Hispanic	.01	.04	0	.01	.01	.01***
Other Race	.02	.03	.03	.02	0	.01***
Experience	14.64 (7.47)	12.05 (6.42)	12.68 (6.79)	16.12 (7.60)	13.82 (7.06)	15.98*** (7.96)
Bachelor's Degree	.17	.14	.12	.22	.14	.20***
Advanced Degree	.83	.81	.88	.78	.86	.8***
Panel B. School Characteristics						
Enrollment	548.27 (339.08)	465.27 (143.37)	526.65 (268.08)	553.47 (264.79)	563.99 (375.58)	559.06* (491.91)
Female	.48 (.04)	.48 (.03)	.48 (.03)	.49 (.05)	.48 (.04)	.49 (.03)
Age	10.44 (3.21)	10.86 (2.86)	10.35 (3.13)	10.21 (3.12)	10.9 (3.38)	10.32*** (3.27)
White	.70 (.32)	.75 (.3)	.75 (.29)	.65 (.33)	.77 (.29)	.65*** (.33)
Minority	.23 (.3)	.21 (.29)	.19 (.27)	.28 (.32)	.18 (.27)	.28*** (.31)

FRPL	.47 (.27)	.51 (.25)	.45 (.27)	.50 (.28)	.46 (.24)	.47*** (.28)
IEP	.16 (.05)	.18 (.05)	.17 (.05)	.16 (.06)	.16 (.06)	.16** (.06)
ELL	.03 (.05)	.01 (.03)	.02 (.04)	.04 (.06)	.02 (.05)	.04*** (.06)
Gifted	.03 (.03)	.03 (.04)	.03 (.03)	.03 (.04)	.03 (.03)	.03 (.03)
Math Proficiency	.62 (.24)	.56 (.24)	.61 (.25)	.60 (.25)	.64 (.23)	.63** (.24)
ELA Proficiency	.65 (.19)	.62 (.18)	.65 (.19)	.63 (.19)	.67 (.17)	.64** (.2)
Charter	.03	.12	.01	.04	.03	.01***
City	.2	.17	.15	.28	.15	.22***
Suburban	.4	.38	.45	.39	.33	.47***
Rural	.29	.27	.33	.23	.42	.2***
Town	.10	.18	.08	.10	.10	.11**
Principals	534	28	142	184	111	69
Schools	635	30	173	226	143	91
School*Years	2,664	77	606	898	639	444

Notes. In Panel A, proportions are reported, except for age and experience, which report means (standard deviation). In Panel B, school-level means (standard deviation) reported, except for charter status and urbanicity, which report proportions. The sample includes K-12 traditional and charter public schools with PIL eligible principals (those hired during or after the 2008-09 school year) who completed PIL induction (World Class Schooling and Driving for Results) within the first five years of the principalship. *Year of PIL Completion* indicates the year during a principal's career that he/she completed PIL induction. Group-level differences are statistically significant at *10%, **5%, and ***1% (and are indicated in the *Year 5* column).

Table A8. Effect of PIL Induction Among PIL Completers

	VAM			Achievement (All Students)		Achievement (VAM Teacher Sample)	
	Math	ELA	Teacher Turnover	Math	ELA	Math	ELA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PIL	.018*** (.006)	.003 (.004)	.011* (.007)	.029* (.017)	.018 (.014)	.052*** (.022)	.036** (.018)
Outcome Mean (SD)	0.00 (.090)	0.00 (.060)	.11	0.00 (1.00)	0.00 (1.00)	-0.02 (1.01)	-0.02 (1.01)
Unit*Years	9,248	10,580	100,003	666,725	665,684	477,481	476,964
Units	4,570	5,174	31,265	344,674	343,498	270,228	268,864
School*Years	1,517	1,510	2,950	2,433	2,433	1,528	1,524
Schools	491	488	705	582	582	491	491
R^2	.260	.290	.071	.405	.418	.425	.439

Notes. Each column represents a separate regression. Coefficients are reported with robust standard errors (clustered at the school level). All regressions include school and district-year fixed effects. The sample includes K-12 traditional and charter public schools with PIL eligible principals (those hired during or after the 2008-09 school year) who completed PIL induction (World Class Schooling and Driving for Results) within the first five years of the principalship. In columns (1)-(3), the *Unit* is teachers; in columns (4)-(7), the *Unit* is students. Coefficients statistically significant at *10% **5% and ***1% levels.

Appendix B. Measuring Teacher Effectiveness

We measure teacher effectiveness based on a teacher's value-added contribution to student achievement. Although teacher value-added estimates only capture certain aspects of teaching practice and behaviors (Steinberg & Kraft, 2017; Grossman et al., 2013), students assigned to higher value-added teachers have been shown to have higher college-going rates, earn higher salaries later in life, and be less likely to have children as teenagers (Chetty, Friedman, & Rockoff, 2014b). And, while much has been written on the consequences of different modeling choices with respect to estimating teacher value-added, we follow Kraft (2019) and estimate teacher effectiveness using a restricted maximum likelihood approach. We specify the model as:

$$(1)\ Achievement_{ijst} = \beta_1 Ach_{ijs(t-1)} + \beta_2 Ach^{other}_{ijs(t-1)} + \beta_3 \boldsymbol{X_{it}} + \beta_4 \boldsymbol{C_{jt}} + \beta_5 \boldsymbol{Z_{st}} + \Omega_{jt} + \mu_{ijst}$$

where *Achievement* for student *i* with teacher *j* at school *s* in year *t* is modeled as a function of a student's prior year test score in the same subject ($Ach_{ijs(t-1)}$) and prior year test score in the other subject ($Ach^{other}_{ijs(t-1)}$). For example, if we are estimating teacher value-added for teacher *j* in math in school year *t*, $Ach_{ijs(t-1)}$ will be student *i*'s math test score from the prior school year, and $Ach^{other}_{ijs(t-1)}$ will be student *i*'s ELA test score from the prior school year. **X** is a vector of time-varying student characteristics, including: age, race, gender, grade level, free or reduced price-lunch eligibility status, special education status, English language learner (ELL) status, and gifted status. **C** is a vector of time-varying classroom characteristics, which are the

student characteristics aggregated to the classroom level; and **Z** is a vector of time-varying school characteristics, which are the student characteristics aggregated to the school-level.[22]

The parameter estimate $\widehat{\Omega}_{jt}$ is the teacher*year random effect, capturing teacher *j's* estimated value-added contribution to student achievement (in either math or ELA) in school year *t*. Given that we model teacher effectiveness as a function of lagged student test scores, teacher effectiveness measures will just be for teachers who, in a given school year, teach in grades 4-8.

There is ongoing debate about the most appropriate (i.e., least biased and most efficient) approach for estimating teacher effectiveness using student test scores – i.e., teacher value-added measures, or VAMs. Critics cite the lack of random assignment, potential unobserved confounders, and the lack of clear modeling guidelines as reasons to avoid using VAMs to assign teacher effectiveness (Morganstein & Wasserstein, 2014). Proponents argue that, on average, VAM results accurately predict long-term student outcomes (Chetty et al., 2014; Chetty et al., 2014c) and have been cross-validated with experimental evidence (Kane & Staiger, 2008). Particularly relevant to this paper is the finding that different methods estimate similar teacher effects, conditional on controlling for students' prior achievement (Chetty et al., 2014). Prior research has relied on a teacher fixed effects approach to estimate teacher VAMs in which the estimated teacher fixed effects are adjusted using the Empirical Bayes post-estimation shrinkage estimator (see e.g., Atteberry, Loeb, and Wyckoff, 2015; Grissom and Loeb, 2017). In other research, the random effects estimates recovered from restricted maximum likelihood is

[22] We test the sensitivity of our VAM estimates to the inclusion of classroom, school, and teacher-by-school random effects along with student leave-out estimates of classroom characteristics. Similar to Cohodes et al. (2021), our VAM estimates are robust to the inclusion of these additional terms. We also test the stability of our VAM estimates to the inclusion of higher-order polynomials (i.e., squared and cubic terms) in student achievement; our primary VAM estimates are robust to the inclusion of these higher order terms. Results available upon request.

preferred to a teacher fixed effect that has been shrunk via the Empirical Bayes method because it produces efficient and consistent estimators for the variance of true teacher effects (Kraft, 2019). While one concern related to the random effects estimator is that teacher assignment may be correlated with student characteristics, we control for prior student achievement and observable student and peer characteristics; these controls mitigate concern that the sorting of students to teachers is based on prior academic performance and other observable student characteristics. Second, as sample size increases, the random effects and fixed effects estimators converge (Guarino et al., 2015; Kane & Staiger, 2008).