Teachers College Record
2022, Vol. 124(12) 62–94

DOI: 10.1177/01614681221147738
journals.sagepub.com/home/tcz
SAGE

Supporting Professional Learning at Scale: Evidence from the District of Columbia Public Schools

Julie Cohen, PhD[1] and Emily Wiseman, PhD[2]

Abstract

Background/Context: Professional development (PD) programs have been the primary tool school districts have used to improve teachers' knowledge and skills, though the evidence is mixed on the degree to which these investments translate into improved outcomes for teachers and their students. Further, most research has tracked researcher-designed and researcher-implemented programs, meaning we know far less about the outcomes of PD designed and implemented by districts. Given that implementation and associated outcomes may look different without tight research parameters, we need more systematic research about district-designed and implemented PD. During early years of PD implementation, it is more likely to observe changes in more proximal outcomes, including an increased sense of trust and collaboration with colleagues, which could, in turn, support teacher retention. Any intervention, but especially those that necessitate substantial changes in instructional activities, likely takes time to promote changes to downstream outcomes like high-stakes assessments of teaching and student achievement.

Purpose/Objective/Research Question/Focus of Study: We analyze the relationship between the design and implementation of an ambitious PD/professional learning (PL) program, called Learning Together to Advance Our Practice (LEAP), and a range of outcomes across 3,000 teachers in the District of Columbia Public Schools (DCPS). We examine the extent to which teacher-reported frequency of participation in two specific PD structures—one-on-one coaching and team seminars—are each associated with improved outcomes of interest. Proximal outcomes include teacher perceptions of the PL program and peer culture at their school, as well as school- and

[1]University of Virginia, Charlottesville, VA, USA
[2]EY-Parthenon, New York, NY, USA

Corresponding Author:
Julie Cohen, University of Virginia, 218-G Bavaro Hall, 417 Emmet Street South, Charlottesville, VA 22903-1738, USA.
Email: jjc7f@virginia.edu

district-wide retention. More distal measures include teacher classroom practice and student achievement.

Research Design: We capitalize on researcher-designed and district-administered survey questions, along with rich administrative data, to understand the relationship between this at-scale, intensive PL program and a range of outcomes over two years, from 2016 to 2018. DCPS implemented LEAP simultaneously in every school at the beginning of the 2016–2017 school year. As a result, our ability to identify how our outcome variables would have changed in the absence of LEAP is limited. We address this issue by measuring differential implementation because the frequency of teacher participation in LEAP varied within schools, within LEAP teams within a year, or within a teacher across a two-year period. We hypothesize that more exposure to LEAP yields greater improvements in outcomes. In separate models, we attempt to limit competing explanations by controlling for: (1) observable attributes of teachers and time, and unobservable, time-invariant attributes of schools; (2) unobservable, time-invariant attributes of LEAP teams; and (3) unobservable, time-invariant attributes of teachers.

Conclusions/Recommendations: We find that greater reported engagement with LEAP is associated with improved teacher perceptions of LEAP and the peer culture at their school, as well as improved teacher retention, especially at the school level. This suggests that PL programs that center within-school connections and supports for teachers—in this case, vertically structured LEAP teams led by school-based LEAP leaders—may support positive school-level outcomes. However, we find little evidence of improved teacher skills and teacher contributions to student achievement, at least in the first two years of LEAP implementation. It will take more time and research to understand the degree to which and ways in which the district's investment in LEAP is associated with the range of desired outcomes.

Keywords

teacher professional development, teacher collaboration, teacher retention, teaching quality, urban schools

High-quality teaching is a key mechanism for improving student outcomes, but teachers vary enormously in their instructional effectiveness (Chetty et al., 2014; Rivkin et al., 2005). There are several established channels to improve teaching quality, including the recruitment and differential retention of more effective teachers as well as the development of less effective teachers. Professional development (PD) programs have been the primary tool school districts have used to improve teachers' knowledge and skills, though the evidence is mixed on the degree to which PD investments translate into improved outcomes for teachers and their students (Garet et al., 2008; Kennedy, 2016). Further, the majority of research has tracked

researcher-designed and researcher-implemented programs, meaning we know far less about the outcomes of PD designed and implemented by districts (Lynch et al., 2019). Given that PD implementation and associated outcomes may look different without tight research parameters, we need more systematic research of district-designed and implemented PD. To that end, we analyze the relationship between the design and implementation of an ambitious PD program and a range of outcomes across 3,000 teachers in the District of Columbia Public Schools (DCPS).

In 2016, DCPS developed a PD model called Learning Together to Advance Our Practice (referred to as LEAP) based on literature that indicates PD experiences are more effective at promoting positive outcomes when they are ongoing, school-embedded, and personalized for teachers (Blase & Blase, 2000; Desimone & Pak, 2017; Fullan & Knight, 2011). In contrast to traditional approaches characterized by large group lectures in a school auditorium, the LEAP approach is more akin to what scholars have termed professional *learning* (PL), rather than professional *development* (PD) (Webster-Wright, 2009). PL is characterized by the development of structures and systems for ongoing and sustained learning, rather than the content of a finite curricular program. Though PL models vary, research highlights some key features, including individualized coaching for teachers (Allen et al., 2015; Kraft et al., 2018) and interaction across teacher teams (Bruce et al., 2010; Zambo & Zambo, 2008). These approaches are associated with more collaborative, trusting workplaces, which, in turn, support teacher self-efficacy and job satisfaction, and ultimately teacher retention (Borman & Dowling, 2008; Carver-Thomas & Darling-Hammond, 2017; Loeb et al., 2005), teacher performance (Penuel et al., 2011), and student achievement (Heller et al., 2012; Kraft et al., 2018). As such, LEAP was designed as a content-focused, school-embedded learning program organized around cross-grade teams, led by coaches, termed "LEAP leaders," who facilitated ongoing group learning in team "seminars" and provided individualized coaching and feedback.

Unfortunately, even PL models that leverage "best practices" can fall short of improving high-stakes outcomes at scale (Garet et al., 2008, 2011, 2016). In contrast with demonstration programs implemented by researchers in more controlled settings, district- and state-wide programs present unique challenges (Borko, 2004; Hill & Grossman, 2013). In particular, implementation issues can mitigate positive, intended outcomes (Banerjee et al., 2017; Borman et al., 2007). During early years of PL implementation, it is more likely to observe changes in outcomes most proximal to the implementation process. PL is characterized by active engagement, so practice is often shared in group settings where teachers experience an increased sense of trust and collaboration with colleagues, both of which are also feasibly related to retention (Carver-Thomas & Darling-Hammond, 2017; Simon & Johnson, 2015).[1] As teachers and instructional leaders engage with the new PL models that require they shift many aspects of their teaching, they tend to adapt and modify components of centrally defined programs (McLaughlin, 1987). Any intervention, but especially those that necessitate substantial changes in PL activities, takes time to faithfully implement and promote changes to downstream outcomes like high-stakes assessments of teaching and student achievement (Kennedy, 2016; Lockwood et al., 2010).

This paper leverages data from DCPS to describe the extent to which teacher-reported frequency of participation in two specific PL structures—one-on-one coaching and team seminars—are each associated with improved outcomes of interest, including teacher perceptions of the PL program and peer culture, teacher classroom practice, teacher retention, and student achievement after the first two years of LEAP implementation. We capitalize on researcher-designed, district-administered surveys, along with rich administrative data, to understand the relationship between this at-scale, intensive PL program and a range of outcomes of interest over two years from 2016 to 2018. We explore the extent to which teacher participation in LEAP is associated with:

- Teacher perceptions of LEAP
- Teacher perceptions of the peer culture at their school
- Teacher retention
- Teaching practice
- Student achievement

We make several important contributions to the literature on professional learning at scale. First, we respond directly to Lockwood and colleagues' (2010) call for more rigorous, at-scale evaluations that disaggregate coaching from complementary components of PL programs. Parsing analyses of professional learning into the constituent components of the program can help districts better understand the "drivers" of associated outcomes and more strategically allocate resources. In addition, although some studies imply that coaching might influence intermediary outcomes such as teacher beliefs, dispositions, and knowledge (Campbell & Malkus, 2014; Kraft & Hill, 2019), few have empirically observed these outcomes across large and diverse populations of teachers. Understanding more clearly how PL opportunities might influence different outcomes – in both the short and long term – could help districts develop more realistic expectations around timetables for teacher development.

Finally, few studies have been able to disentangle how differences in PL frequency might be associated with differences in observed outcomes. Here we examine the relationship between PL structures and outcomes using a series of fixed effects regression models that allow us to examine differential frequency of coaching and seminar attendance within the same school, within teams of teachers, or within teachers over time. Such evidence can help districts move beyond a "what works" mindset about PL to a more nuanced understanding of how different levels of support within the same PL program may differentially influence outcomes across teachers.

Background and Framework

Our analytic approach stems from DCPS's conceptualization of high-quality professional learning experiences, which centers around the core idea that teachers benefit from dedicated time and space to focus on the teaching of disciplinary content (Kennedy, 2016); that is, rather than focus on management practices or other

content-generic skills, PL should center on subject-specific teaching practices that support students in engaging with rigorous academic content (Desimone & Garet, 2015). Importantly, PL is distinct from traditional, "one-off" PD sessions with outside providers in that it necessitates active, school-based work over a sustained period of time (Lynch et al., 2019; Yoon et al., 2007). In addition, district personnel theorized that teachers would benefit from a space for collaboration across grade levels, facilitated by a school-based leader with content expertise, coupled with individualized support and feedback from the same leader (Borko, 2004).

Hill, Beisiegel, and Jacob (2013) underscore the need for decomposing and studying distinct components of multifaceted PL programs. Empirically disentangling components of PL can provide insights about the degree to which teachers and instructional leaders implemented different learning activities and structural supports and then whether such differences in implementation are associated with differences in observed outcomes. Accordingly, we examine the distinct literatures focused on the use of teacher teams and individual coaching.

Teacher Teams

One common element of many PL programs, including LEAP, is group learning among teacher teams. In particular, teacher engagement with content-focused, cross-grade teams has been associated with increased peer collaboration (Kazemi & Franke, 2004; Little, 2012; McLaughlin & Talbert, 2001), which is associated with group analysis of student work, including plans to address student needs (Astuto et al., 1993; Stoll et al., 2006). Teachers working together to address challenges and facilitate skill-building can cultivate a sense of mastery, which has been associated with increased teacher self-efficacy (Bruce et al., 2010; Zambo & Zambo, 2008), motivation and persistence (Moolenaar et al., 2012), and willingness to experiment with new approaches (Puchner & Taylor, 2006). Some have argued that to facilitate effective mastery experiences in teams, teachers must first build trust and set norms to share about their teaching (Bruce et al., 2010; Supovitz, 2002; Tschannen-Moran, 2001; Vangrieken et al., 2017).

This process of moving from structured interaction to fruitful and sustained collaboration is likely to take time (Puchner & Taylor, 2006), particularly in environments that emphasize accountability and link teaching performance to consequential outcomes (Zambo & Zambo, 2008); however, such investments can pay dividends. Over time, the integration of high-frequency, structured, and content-focused collaboration is likely to enhance both teacher job satisfaction (Borman & Dowling, 2008; Simon & Johnson, 2015) and teacher retention (Ingersoll, 2001; Miller et al., 1999; Pounder, 1998). Sustained collaboration is also related to improved teacher practice and student achievement (Goddard et al., 2007; Ronfeldt et al., 2015; Saunders et al., 2009; Supovitz & Christman, 2003; Vescio et al., 2008). Though this research suggests the potential value of teaming and collaboration for improving outcomes, some of the studies cited are small-scale with specific samples that limit generalizability; others provide simple

correlations with little attempt to address potential biases (Bruce et al., 2010; Saunders et al., 2009; Supovitz & Christman, 2003). More recent work in Miami-Dade, Florida, has explored the value of collaboration with quasi-experimental methods and larger populations of teachers (Ronfeldt et al., 2015), demonstrating that teachers and schools reporting higher levels of collaboration also have greater impacts on student achievement. However, more work is needed in district contexts, particularly those focused on efforts to *improve* collaboration rather than to observe existing collaboration.

Teacher Coaching

Although team structures and collaboration may be helpful to realize certain goals, individual teachers may also have idiosyncratic needs that would be more effectively addressed individually. To that end, increasing attention has been paid to one-on-one teacher coaching that targets individual needs. A recent meta-analysis of 60 causal studies found that individual coaching programs, on average, improve observed teaching by 0.49 standard deviations (SDs) and student reading achievement by 0.16 SDs (Kraft et al., 2018). Programs that offered individualized coaching in combination with group learning (like LEAP) had an average effect size that was 0.31 SDs higher than the effect estimated for programs that included only individual coaching. Although coaching improves teaching and student outcomes in small-scale experiments, it has historically produced more mixed results at scale (Atteberry & Bryk, 2011; Biancarosa et al., 2010; Lockwood et al., 2010). Kraft and colleagues (2018) document a descriptive, negative relationship between program size and observed effects, calling for efficacy trials that identify the effective features of large-scale coaching programs.

The extant literature has provided insights about how coaching *can* support teacher development, but we know little about whether large-scale, district-implemented coaching actually translates into observable improvements in teacher and student outcomes. Given the increasing prevalence of coaching in districts around the country, the field needs more rigorous evidence about the success of district-designed and implemented coaching supports. It may well be that more idiosyncratic school- and teacher-based needs are more powerful levers for improving instructional quality and student outcomes. As such, investments in scalable programs across large and diverse populations of teachers working in a range of school contexts may not be the surest path to teacher development (Coburn, 2003). This underscores the need for more studies that focus on the implementation and associated outcomes of such programs to build a richer and more comprehensive evidence base for districts wanting to know more about such tradeoffs inherent in efforts to scale.

Learning Together to Advance Our Practice (LEAP)

Over the past decade, teachers, school leaders, and policymakers in DCPS have engaged in systematic and sustained efforts to improve instructional quality. These efforts began in 2009 with IMPACT, the multimeasure teacher evaluation system

linked to high-stakes outcomes for teachers. IMPACT was successful in improving the composition of the workforce and encouraging some teachers to develop their observable skills (Adnot et al., 2017; Dee & Wyckoff, 2015). However, with the adoption of Common Core State Standards (CCSS) in 2012 and the shift to the associated PARCC exam for students in 2015, district leaders recognized many teachers were not equipped with the knowledge and skills to engage in effective content-specific pedagogy.[2] LEAP was designed to support teacher development in service of supporting students in meeting the CCSS's ambitious learning goals.

LEAP embeds both group and individual learning structures into the daily school routine. Each week, a LEAP leader with demonstrated proficiency in English language arts (ELA) or mathematics convenes a cross-grade LEAP team (e.g., third- to fifth-grade math teachers) for 90-minute seminars during common release time created by district-level personnel. LEAP teams are organized at the school level, and district leaders encourage school-based adaptations to the content of LEAP. Seminars represent a dedicated time for teachers to reflect on student learning and teaching practice alongside colleagues teaching the same content at similar grade level. Seminar lesson plans were organized learning modules designed to cover a connected set of discipline-specific knowledge and skills to be developed over time. The district conceptualized the multigrade teams as a vehicle for enhancing teacher knowledge of the vertical trajectories of student learning over many years; that is, third-, fourth-, and fifth-grade math teachers all work on concepts related to proportional reasoning, though they rarely have opportunities to discuss the ways in which those concepts develop over time. LEAP seminars were thus conceived of as a space for making those temporal links explicit, so that students, too, might be better positioned to see the connections between material covered in different grades. Though LEAP leaders could customize content based on teachers' needs, the district provided detailed plans for the weekly meetings that theoretically supported teachers in examining data, learning new content, applying the content through planning and practice, and then monitoring and reflecting on progress to facilitate growth over time. The 90-minute seminars were built into common planning time for teachers, either before, during, or after the typical school day. The district provided schedule flexibility to accommodate either earlier arrival or later departure times for teachers.

LEAP leaders also observe teachers in their classrooms as part of the weekly LEAP cycle to follow up with a structured, one-on-one debrief, at times decided by the teacher–leader dyad. Coaching affords the opportunity for LEAP leaders to observe teachers enacting the practices discussed in seminars and provide individualized feedback. The district employed a multistep debriefs structure that included an affirmation of what's working in a teacher's classroom, a reflection on areas for improvement, a co-construction of next steps with target focus skill for the coming week, and finally, a plan for an upcoming lesson with that skill in mind. All core content teachers (ELA, math, and/or early childhood educators) in elementary and middle schools, or math, ELA, science, and social studies teachers in high schools, as well as special education

and English as a second language teachers, were placed into content-specific LEAP teams to engage in the weekly LEAP cycle.

Both coaching and seminar are conceptually linked to outcomes of interest, like increased school cohesion, more positive teacher perceptions of instructional supports, increased teacher self-efficacy and satisfaction, and improved teaching quality and increased student achievement.

Conceptual Model

Although the design of LEAP incorporates many features of effective PL, the program's effectiveness is far from inevitable. Just as setting aside dedicated time for teachers to work together is not sufficient to ensure high-quality collaboration (Little, 2003; Ronfeldt et al., 2015; Wei et al., 2010), designing an evidence-based PL program may not ensure the program will improve instruction. Indeed, many well-designed policies, successful in smaller pilots, fail to produce positive outcomes when implemented at scale (Fixsen et al., 2005).

Conceptually, LEAP's key structures are hypothesized to support teacher development through multiple mechanisms to influence a variety of proximal and more distal outcomes (see theory of change diagram, Figure 1). This figure undoubtedly oversimplifies the teacher development process—which is likely nonlinear, for example—but provides a high-level, research-based set of hypotheses to guide our empirical work. As shown, we hypothesize substantial overlap between group PL and individual coaching in supporting teacher mastery experiences.

Methods, Data, and Sample

DCPS implemented LEAP simultaneously in every school at the beginning of the 2016–2017 school year. As a result, our ability to identify how our outcome variables would have changed in the absence of LEAP is limited. We address this issue by measuring differential implementation as differences in the frequency of teacher participation in LEAP within schools, within LEAP teams within a year, or within a teacher across a two-year period. We hypothesize that more exposure to LEAP yields greater improvements in outcomes and limits competing explanations by controlling for observable attributes of teachers and time, and unobservable time-invariant attributes of schools (Equation 1), unobservable time-invariant attributes of LEAP teams (Equation 2), or unobservable time-invariant attributes of teachers (Equation 3).

$$Y_{ijst} = \alpha + \beta_1 \left(Coaching_{ijst} \right) + \beta_2 \left(Seminar_{ijst} \right) + X_{it}\vartheta + Y_t\delta + \gamma_s + \varepsilon_{ijst} \quad (1)$$

$$Y_{ijst} = \alpha + \beta_1 \left(Coaching_{ijst} \right) + \beta_2 \left(Seminar_{ijst} \right) + X_{it}\vartheta + Y_t\delta + \theta_j + \varepsilon_{ijst} \quad (2)$$

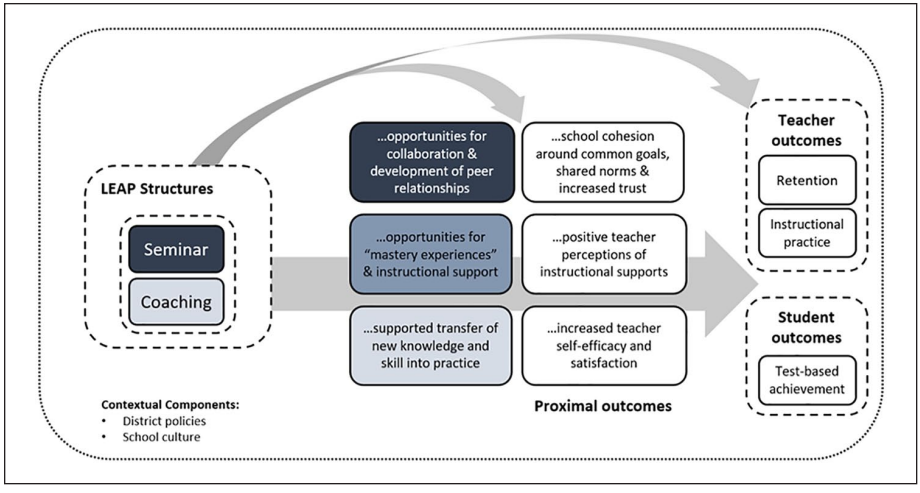


Figure 1. Theory of Change.

$$Y_{ijst} = \alpha + \beta_1 \left(Coaching_{ijst} \right) + \beta_2 \left(Seminar_{ijst} \right) + X_{it}\vartheta + Y_t\delta + \tau_i + \varepsilon_{ijst} \quad (3)$$

In these models, Y_{ist} represents the outcomes of interest (teacher perceptions of LEAP, teacher retention, and student achievement) for teacher i in school s in time t. Teacher perceptions are measured by a six-point agreement scale based on survey responses; teacher retention is observed via a dichotomous indicator for retain or attrit; and student achievement is estimated based on changes in a teacher-by-year value-added score rank. Each of these outcomes is described in more detail the following section. β_1 represents the estimated relationship between teacher-reported coaching session frequency (a frequency-per-month unit) in time t and the outcome of interest, whereas β_2 represents the estimated relationship between teacher-reported seminar session frequency (a frequency-per-month unit) in time t and the outcome of interest. ϑ represents the relationship between a vector of teacher characteristics including teacher experience and whether they taught a tested subject status in time t and outcomes of interest. This set of controls was selected purposely to retain the largest sample of teachers, while also incorporating meaningful sources of variation in teacher experiences. δ represents a control for year one ($t = 0$) or year two ($t = 1$) of LEAP, γ_s represents school fixed effects, θ_j represents LEAP team fixed effects, and τ_i represents teacher fixed effects.

In this framework, the inclusion of a school fixed effect allows for comparisons of teachers across LEAP teams within schools. This controls for all the time-invariant factors that differ across schools (e.g., principals and students). It does not account for differences that exist between LEAP teams within and across schools (e.g., more effective LEAP leaders or systematic differences between grade and content areas). Model 2 with LEAP team

fixed effects compares frequency differences between teachers within the same LEAP team. Finally, frequency could endogenously differ among teachers within LEAP teams; for example, teachers who are more motivated may seek out LEAP leaders who coach more frequently and simultaneously engage in other practices that improve outcomes. Conversely, teachers who are struggling to improve may receive more coaching, biasing the relationship with outcomes in the other direction. Models that include a teacher fixed effect identify the effect of LEAP solely by examining how year to year changes in LEAP frequency for a teacher are related to changes in outcomes for that teacher.

Each of these approaches has strengths and weaknesses. For example, models with teacher fixed effects provide the greatest control for potential confounds, and we therefore privilege these estimates in our discussion of results. The only exception is the teacher retention estimates, where we don't observe the retention decisions of teachers in t if they attrited in $t - 1$; however, we include the full set of estimates to explore the robustness of our results. Taken together, they provide credible approaches for limiting a causal interpretation between LEAP and the alternative outcomes. Given the limited examples of at-scale analyses of PL, we follow Conaway and Goldhaber's (2020) recommendation to report statistical significance at conventional levels as well as at the $p < 0.20$ level, particularly for pressing policy-relevant questions like how to build professional learning systems for large populations of teachers. Though doing so increases the risk of making a Type I error (detecting a statistically significant relationship where there is not one), we view this risk as tenable because of the potential reward of detecting relationships between components of professional learning and outcomes of interest. This work is correlational and not causal, and we view our findings as suggestive rather than conclusive. Districts and researchers alike would benefit from additional research about the relationships among facets of professional learning and a range of outcomes.

To understand the role of session frequency over time, we also specify a lagged model, which is summarized in Equation 4. This allows us to understand the separate contribution of both current- and prior-year participation in LEAP's structures during the second year of LEAP. We include a LEAP team fixed effect to address potential selection issues associated with schools and teams. Estimates that substitute school fixed effects for LEAP team effects are also included for comparison. Because these estimates rely on data for two years, samples are restricted for the reasons discussed previously.

$$\begin{aligned} Y_{ijt} = {} & \alpha + \beta_1 \left(Coaching_{it} \right) + \beta_2 \left(Coaching_{it-1} \right) + \beta_3 \left(Seminar_{it} \right) \\ & + \beta_4 \left(Seminar_{it-1} \right) + X_{it} \vartheta + \gamma_j + \varepsilon_{ijt} \end{aligned} \tag{4}$$

In this model, β_1 represents the difference in outcome Y_{ijt} associated with one additional coaching/seminar per month in the *current-year* frequency of coaching, among observationally similar teachers who received the same amount of coaching in the *prior* year. β_2 represents the coefficient of interest for *prior-year* frequency of coaching, β_3 represents the coefficient of interest for *current-year* frequency of seminar, and β_4 represents the coefficient of interest for *prior-year* frequency of seminar.[3]

Data and Sample

Our analysis follows the conceptual model presented in Figure 1. We examine LEAP during academic year (AY) 2016–2017 and AY 2017–2018, the first two years of its implementation. We employ district-wide surveys of teachers to better understand the relationship between engagement with LEAP and teacher perceptions, retention, observed teaching practice, and student achievement. As described in detail below, we map survey items onto specific constructs illustrated in Figure 1. We then create composite measures that aggregate conceptually related items. Finally, we merge survey data with administrative data about teachers, students, and schools.

Our analysis draws primarily on the end-of-year surveys DCPS administered to teachers. The instructional culture teacher survey called INSIGHT, designed and administered by The New Teacher Project (TNTP), asks teachers to reflect on school climate, opportunities for PD, and plans to continue teaching. Working with DCPS, we augmented the INSIGHT survey with LEAP-specific questions that prompted teachers to report their individual session frequency of LEAP coaching and seminar. The response rate to the end-of-year teacher survey in 2017 was 73% and in 2018 improved to 96%.

DCPS also provided annual LEAP rosters that match teachers to their LEAP team and leader. We merge survey data and the LEAP roster with administrative data that provide rich information about teachers and schools, including teacher performance, as measured through IMPACT, teacher experience, gender, and race, as well as student characteristics (e.g., racial demographics, percent of students eligible for free or reduced-price meals, percent of students in special education programs, etc.), school type (i.e., elementary, middle, high), and school poverty status.

Analytic Variables

We used survey data to create our primary independent variables. Our measures of coaching and seminar session frequency capture the frequency per month with which teachers reported engaging in these LEAP structures. Teachers indicated whether they received coaching and attended seminar never, rarely, about once per month, about once every other week, about once per week, or more than once per week. We converted this scale to a frequency measure to represent the number of coaching sessions per month.[4] For teachers who responded "never," we input 0 LEAP sessions per month, for "rarely" we input 0.5, for "once every other week" we input 2, and for "once per week" we input 4 sessions per month. For teachers who suggested they engage in LEAP "more than once per week," we input 5 sessions per month, a conservative estimate of average engagement. Across both years of LEAP, the mean session frequency of coaching was 2.44 sessions per month, with more than 40% of teachers reporting coaching no more than once a month (Table 1 and Appendix Figure 1). The mean teacher-reported seminar session frequency was 4.04 sessions per month, with relatively little variation (Table 1 and Appendix Figure 2).

Table 1. Variable Descriptions.

	LEAP Year 1	LEAP Year 2	Both Years of LEAP
	(1)	(2)	(2)
Coaching session frequency	2.50	2.37	2.44
	(1.71)	(1.71)	(1.71)
	1,838	1,569	3,407
Seminar session frequency	4.10	3.98	4.04
	(1.46)	(1.59)	(1.53)
	1,862	1,861	3,723
Perceptions of LEAP*	0.00	0.00	0.00
	(0.99)	(1.01)	(1.00)
	1,883	1,874	3,757
Perceptions of peer culture*	–0.04	0.04	0.00
	(1.01)	(0.98)	(1.00)
	2,446	2,390	4,836
School-level retention	0.78	0.83	0.81
	(0.42)	(0.37)	(0.40)
	3,166	3,229	6,395
District-level retention	0.86	0.91	0.89
	(0.35)	(0.29)	(0.32)
	3,166	3,229	6,395
Change in observational teaching score	0.13	0.15	0.14
	(0.83)	(0.77)	(0.80)
	2,644	2,681	5,325
Change in teacher value-added score	–0.03	0.04	0.01
	(1.07)	(1.00)	(1.03)
	312	340	652

Note: Numbers displayed represent simple means with standard deviations in parentheses followed by the number of teacher observations.
*Represent standardized variables.

Our proximal outcomes, teacher perceptions of LEAP and teacher perceptions of the peer culture at their school, are constructed using composite measures of survey items. Teacher perceptions of LEAP is a simple teacher-by-year average of five survey items that probe the value of LEAP. Specifically, teachers are asked to evaluate statements using a six-point scale (strongly disagree, disagree, somewhat disagree, somewhat agree, agree, strongly agree): (1) overall, participating in LEAP is a valuable use of my time; (2) my students will learn more in my classroom because of my participation in LEAP; (3) my teaching will improve because of my participation in LEAP; (4) participating in LEAP Seminar is valuable to improving my instructional practice; and (5) participating in LEAP Coaching is valuable to improving my instructional practice.

Together, these items have an alpha reliability coefficient of 0.97 and are composited using a teacher-by-year average with an across-year mean of 2.83 (SD = 1.55, N = 3,723).

Teacher perceptions of peer culture is a composite measure constructed using the following survey items that employ the same six-point agreement scale: (1) teachers at my school share a common vision of what effective teaching looks like; (2) there are many teachers at my school who set an example for me of what highly effective teaching looks like; (3) at my school, teachers use a common vocabulary to discuss effective teaching; (4) the time I spend collaborating with my colleagues is productive; and (5) there is a low tolerance for ineffective teaching at my school. Together, these items have an alpha reliability coefficient of 0.89 and are composited using a teacher-by-year average with an across-year mean of 3.49 (SD = 1.11, N = 3,720). Both composite variables—teacher perceptions of LEAP and of peer culture at their school—are then standardized to have a mean of zero and standard deviation of one (Table 1, Appendix Figures 3 and 4).

Our distal outcomes, teacher retention, change in observed teaching practice, and change in teacher contribution to student achievement, are constructed using administrative data. Teacher retention in t is calculated by assessing the status of that individual in $t + 1$. We create two measures of observed teacher retention: one at the school level and one at the district level. The school-level variable accounts for teacher retention in schools, either through remaining in teaching or transferring to another (nonteaching) position within that school. The district-level variable accounts for both of those possibilities, while also incorporating "retained" teachers who transfer to another school within DCPS.[5] Because we are interested in the relationship between LEAP session frequency and a teacher's decision to continue teaching in DCPS, we exclude teachers who were involuntarily dismissed through the IMPACT evaluation system (3–4% of teachers each year). Appendix Figure 5 displays the distribution of teacher retention across both years of LEAP, highlighting that about 80% of teachers are retained in their schools each year. In contrast, about 90% of teachers are retained in the district each year.

Our measure of change in observed teaching practice is based on the scores from the teacher observation component of the IMPACT evaluation system. DCPS observes and rates all teachers multiple times each year. In AY 2016–2017, the first year of LEAP, the observation rubric switched from the Teaching and Learning Framework (TLF) to the Essential Practices (EP). Both rubrics rate teaching on a four-point scale and have similar means and SDs (2015–2016 TLF mean = 3.28, SD = 0.39; 2016–2017 EP mean = 3.26, SD = 0.39). We standardize scores by year to create a measure of teaching comparable across years. The standardized observation scores are shown in Table 1; the mean change in observational scores is about 0.14 standardized rubric points (Appendix Figure 6).

Our final outcome of interest is a teacher's contribution to student achievement. Teachers who teach a tested grade and subject in DCPS (about 17% of teachers during

each year of LEAP)[6] are evaluated using teacher value-added measures of student achievement as part of the IMPACT system. This is a statistical measure of a teacher's contribution to student achievement on the PARCC exam that accounts for prior student achievement, as well as covariates typically associated with performance. Teacher value-added scores have been standardized (Table 1).

Sample

Our analytic sample is drawn from teachers working in DCPS during the 2016–2017 and 2017–2018 school years. DCPS employs approximately 3,300 teachers and 450 LEAP leaders each year. There were 110 schools in DCPS in 2017–2018, of which roughly 20% were identified by DCPS as low-poverty. The remaining schools were high-poverty schools, of which 40 had previously been identified as Targeted 40 schools (the lowest-performing schools in the district).

Our analytic sample is smaller than the population of teachers in DCPS because many of our measures come from teacher surveys, with associated nonresponse. Some variables are also lagged values from a prior year, requiring prior year data, missing for first-year teachers and those who did not respond to the survey in $t - 1$. As a result, depending on the formulation of models, some teacher observations are missing.

Table 2 compares the full sample of DCPS teachers to the sample of teachers who responded to the two survey frequency measures during either the first or second year of LEAP. There are several statistically significant differences between our survey respondents and the full district, but in most cases, these differences are not substantively meaningful. For example, whereas the final IMPACT evaluation score (scale of 100–400) represents a statistically significant difference, the difference between survey respondents and the full sample (2.47 points) is minimal. (The SD of IMPACT scores for all participants is 44 points.) Teachers in our survey sample also have about one less year of experience compared to the full DCPS sample. About 22% of survey respondents teach a tested grade and subject, compared to about 12% of nonrespondents (17% of the full sample of DCPS teachers in this period teach a tested grade and subject). Teachers in our sample are also somewhat more likely to teach in a school with more positive teacher-reported school climate. However, survey respondents are not more or less likely to teach in a high poverty or Targeted 40 (lowest-performing) school in the district. Survey respondents have somewhat higher scores on classroom observation scores, but these differences, too, are relatively small. Though these differences are substantively small, even for those that are statistically significant, readers should keep these differences in mind as they interpret our findings.

Limitations

DCPS implemented LEAP for all teachers simultaneously, limiting options to construct strong counterfactuals. As we designed our project, we emphasized the importance of differential LEAP implementation among teachers and an ability to limit

Table 2. Sample Comparison.

	Full District	Survey Respondents	
	(1)	(2)	
Final IMPACT score	328	2	*
	(44)	(1)	
Teaching experience	9.36	–0.75	**
	(7.62)	(0.20)	
Tested grade and subject	0.17	0.10	***
	(0.38)	(0.01)	
School climate average	3.23	0.08	***
	(0.50)	(0.01)	
High poverty school	0.44	0.02	
	(0.50)	(0.01)	
Targeted 40 school	0.34	0.00	
	(0.48)	(0.01)	
Change in observed teaching	0.14	0.07	***
	(0.80)	(0.02)	
Change in value-added	0.01	0.17	*
	(1.03)	(0.08)	
School-level teacher retention	0.81	0.04	***
	(0.40)	(0.01)	
District-level teacher retention	0.89	0.04	***
	(0.32)	(0.01)	
Teacher observations with IMPACT rating	6,596	3,402	

Note: Means and SDs from full district reported in column 1, differences and clustered SEs from separate regression of covariate on survey sample reported in column 2.
* $p < 0$.
** $p < 0.01$.
*** $p < 0.001$.

competing explanations for changes in outcomes of interest. To that end, we have employed a variety of variables that control for mechanisms other than LEAP that may have influenced the proximal and distal outcome measures, including teachers and school attributes, as well as school, LEAP team, and teacher fixed effects. Given the short duration of our panel, we find these controls credible. Even so, it is possible that our estimates of the relationship between outcomes and LEAP frequency do not address all potential confounds. Thus, we encourage readers to employ caution when interpreting our estimates as causal; for example, is LEAP coaching provided with more frequency to teachers who demonstrate comparably less growth in teaching skills? If so, our estimates may understate how LEAP influences these outcomes. Alternatively, an already strong teacher who also has a strong desire to improve may seek out additional LEAP coaching compared to other teachers or even that same teacher in another year. If so, our estimates may overstate the effect of LEAP.

More broadly, we are limited by the measures employed in this study. Specifically, we draw on teacher surveys administered at the end of the year to assess LEAP frequency, and policies like state testing may influence teaching outcomes at particular points in the year, as the school environment adapts to meet these time-specific goals (Spillane et al., 2011). We use teacher reports of LEAP session frequency, rather than observed frequency.[6] Researchers note that carefully designed survey measures often correlate with observed practice (Mayer, 1999), but differences occur. Finally, we are limited by small sample size in a few of our model specifications, specifically with regards to the models intended to estimate the relationship between LEAP frequency and teacher contribution to student achievement. Examining the relationship between LEAP and teacher value-added requires two consecutive years of value-added estimates for teachers, further limiting the sample.

We are additionally limited by the lack of a robust measure of coach quality. Prior literature suggests that the credentials of a coach are likely important (Campbell & Malkus, 2014; Coburn & Russell, 2008). We collected a content knowledge test that was used during the coach screening and hiring process; however, DCPS did not validate this measure, and we found little variation in scores. We also know that the extent to which the coach fosters trusting and positive relationships with teachers may influence outcomes (Ippolito, 2010; Woulfin & Jones, 2018). Unfortunately, we were unable to collect measures that shed light upon these constructs. However, analyses of the relationship between coaching and outcomes rarely differentiate coach quality, and frequency is often found to be related to outcomes (Kraft et al., 2018).

Results

In general, we find that greater reported engagement with LEAP is associated with improved teacher perceptions of LEAP and the peer culture at their school, as well as improved teacher retention, especially at the school level. However, we do not find consistent evidence of improved teacher skills and teacher contributions to student achievement. We turn to a more detailed examination of each of the research questions.

RQ 1a: To What Extent Is Participation in LEAP Associated with Teacher Perceptions of LEAP?

In a year when teachers participated in LEAP more frequently, they perceive LEAP more positively than in a year in which they participated in LEAP less frequently (Table 3, column 3). This is true for both seminar and coaching. A teacher who reports receiving one additional coaching session per month (representing about 0.60 of a SD of coaching frequency) and one additional seminar per month (representing about 0.70 of a SD of seminar frequency) is estimated to perceive the LEAP program about 0.22 SDs more positively compared to when that same teacher received a lower frequency of LEAP (Table 3: sum of rows 1 and 3, column 3). For teachers in their second year of LEAP, there is no evidence that the frequency of LEAP in the first year influences

Table 3. Teacher Perceptions of LEAP and LEAP Seminar and Coaching Session Frequency.

	Current Values Model			Lagged Values Model	
	(1)	(2)	(3)	(4)	(5)
Coaching session frequency	0.181***	0.170***	0.148***	0.167***	0.161***
	(0.010)	(0.013)	(0.025)	(0.021)	(0.030)
Coaching session frequency (last)	-	-	-	0.032++	0.029
	-	-	-	(0.022)	(0.031)
Seminar session frequency	0.127***	0.164***	0.071+	0.178***	0.158**
	(0.017)	(0.023)	(0.039)	(0.035)	(0.056)
Seminar session frequency (last)	-	-	-	0.035	0.054
	-	-	-	(0.041)	(0.058)
School fixed effects	*X*			*X*	
Team fixed effects		*X*			*X*
Teacher fixed effects			*X*		
Teacher-year observations	3,399	3,140	1,766	883	877

Note: Clustered standard errors in parentheses.
** $p < 0.01$.
*** $p < 0.001$.
+ $p < 0.1$.
++ $p < 0.2$.

their perceptions in the second year (Table 3, columns 4 and 5); however, we do see that for the retained teachers, their perceptions of LEAP are about 0.32 SDs greater compared to teachers on the same LEAP team who experience one fewer coaching session and LEAP seminar per month (Table 3: column 5, sum of coaching and seminar frequency). Taken together, these results suggest a strong, positive relationship between teacher participation in the LEAP program and teacher perceptions of the value of LEAP. This relationship is estimated to be stronger when teacher fixed effects are replaced by school or team fixed effects and teacher covariates. These models include cross-teacher variation that likely helps identify LEAP's effects, but they may also include variation that reflects selection into LEAP participation.

RQ 1b: To What Extent Is Participation in LEAP Associated with Teacher Perceptions of Peer Culture at Their School?

When teachers report participating in LEAP more frequently, they are estimated to perceive the school culture among their peers more positively compared to a year in which they report participating in LEAP less frequently (Table 4, column 3). Employing teacher fixed effects reduces the variability in coaching and seminar frequency, and, as a result, coefficient estimates are reduced, and seminar frequency is not statistically significant at traditional levels. However, a teacher who received one additional coaching and an additional seminar session each month perceives their school's peer culture

Table 4. Teacher Perceptions of Peer Culture by LEAP Seminar and Coaching Session Frequency.

	Current Values Model			Lagged Values Model	
	(1)	(2)	(3)	(4)	(5)
Coaching session frequency	0.131***	0.144***	0.077***	0.093***	0.111***
	(0.011)	(0.014)	(0.023)	(0.022)	(0.032)
Coaching session frequency (last)	-	-	-	0.037^{+}	0.029
	-	-	-	(0.022)	(0.030)
Seminar session frequency	0.119***	0.129***	0.085*	0.164***	0.184**
	(0.019)	(0.026)	(0.040)	(0.041)	(0.062)
Seminar session frequency (last)	-	-	-	0.075^{+}	0.106^{+}
	-	-	-	(0.045)	(0.062)
School fixed effects	*X*			*X*	
Team fixed effects		*X*			
Teacher fixed effects			*X*		*X*
Observations	3,396	3,137	1,765	882	876

Note: Clustered standard errors in parentheses.
*$p < 0.05$.
**$p < 0.01$.
***$p < 0.001$.
$^{+}p < 0.1$.

about 0.16 SDs more positively than that same teacher did prior to the increase in LEAP frequency (Table 4: sum of rows 1 and 3, column 3). When we examine these relationships for teachers only in the second year of LEAP, there is suggestive evidence ($p < 0.20$) that participation in seminar during the prior year of LEAP has a lagged influence on perceptions of peer culture in the second year. In total, we estimate that retained teachers who report experiencing an additional coaching and seminar session per month in the second year of LEAP, as well as an additional seminar session in the first year of LEAP, seem to perceive the peer culture at their school about 0.401 SDs higher than their LEAP team peers (Table 4: sum of rows 1, 3, and 4, column 5).

Our estimates suggest that exposure to LEAP is associated with improvement in teachers' perceptions of both LEAP and the peer culture of the school. These results are consistent with our conceptual model that the components of LEAP should be associated with positive shifts in these proximal outcomes. Our conceptual model also indicates that improved proximal outcomes would, in turn, support improvements in teacher retention, teaching skills, and student achievement. We now turn to examining these more distal outcomes.

RQ 1c: To What Extent Is Participation in LEAP Associated with Teacher Retention?

Table 5. Teacher Retention in School by LEAP Seminar and Coaching Session Frequency.

	Current Values Model		Lagged Values Model	
	(1)	(2)	(3)	(4)
Coaching session frequency	0.010*	0.013*	0.003	0.003
	(0.004)	(0.006)	(0.009)	(0.012)
Coaching session frequency (last)	-	-	0.014^{++}	0.016
	-	-	(0.009)	(0.013)
Seminar session frequency	0.026**	0.021*	0.043**	0.035^{++}
	(0.008)	(0.011)	(0.017)	(0.025)
Seminar session frequency (last)	-	-	–0.015	–0.021
	-	-	(0.017)	(0.026)
School fixed effects	*X*		*X*	
Team fixed effects		*X*		*X*
Observations	3,307	3,054	868	862

Note: Clustered standard errors in parentheses.
* $p < 0.05$.
** $p < 0.01$.
$^{++}p < 0.2$.

Given our findings regarding perceptions of LEAP and peer culture in the school, we estimate the relationship between LEAP frequency and retention separately for schools and the district. Overall, LEAP frequency has a mixed relationship with teacher retention. Here we privilege models with team, rather than teacher, fixed effects. For teachers who report receiving an additional coaching and seminar session per month, compared to their peers on the same LEAP team, school-based teacher retention is estimated to increase by 3.4 percentage points (Table 5: sum of rows 1 and 3, column 2). Most of this increase is attributable to seminar session frequency. Said differently, receiving an additional seminar and coaching session each month is associated with a decrease in school-level attrition by 17%, as about 80% of DCPS teachers remain at their schools. For retained teachers in the second year of LEAP, there is not a relationship between coaching session frequency in the current or prior year of LEAP; however, we do observe a second-year relationship between an additional seminar session and school-level retention that is statistically significant at $p < 0.20$ (Table 5, row 3, column 4).

Observed retention within the district is also positive, improving by about 1 percentage point for an additional monthly coaching session (Table 6, row 1, column 2). Increased seminar session frequency is predicted to improve district retention, but in our preferred model this relationship is not statistically significant (Table 6, row 3, column 2). Additionally, we do not observe any relationships between district retention and teacher participation in LEAP in our lagged models.

Taken together, and consistent with the peer culture results, we see increased participation in LEAP as moderately positively and statistically significantly associated

Table 6. Teacher Retention in District by LEAP Seminar and Coaching Session Frequency.

	Current Values Model		Lagged Values Model	
	(1)	(2)	(3)	(4)
Coaching session frequency	0.005++	0.009*	0.010++	0.006
	(0.004)	(0.005)	(0.006)	(0.010)
Coaching session frequency (last)	-	-	–0.006	0.000
	-	-	(0.007)	(0.010)
Seminar session frequency	0.013*	0.007	0.014	0.001
	(0.006)	(0.009)	(0.014)	(0.021)
Seminar session frequency (last)	-	-	–0.011	–0.017
	-	-	(0.013)	(0.022)
School fixed effects	*X*		*X*	
Team fixed effects		*X*		*X*
Observations	3,307	3,054	868	862

Note: Clustered standard errors in parentheses.
* $p < 0.05$.
++$p < 0.2$.

with improved school retention. We see less consistent of a relationship between teacher participation in LEAP and district retention, although there is some suggestive evidence of a positive relationship at the district level, too.

RQ 1d: To What Extent Is Participation in LEAP Associated with Observed Improvements in Teaching Practice?

LEAP participation is not associated with improvements on the high-stakes classroom observation measure employed by the IMPACT evaluation system. Additional coaching sessions are somewhat negatively associated with improvements to observed teaching, but these shifts are small and not statistically significant. We observe inconsistent and weak evidence that an additional seminar session per month is associated with decreased teaching performance as measured by IMPACT (Table 7, row 3, column 3). When we observe these teachers just in the second year of LEAP, we do not see evidence that current or prior participation in LEAP is associated with observed teaching in the second year.

RQ 1e: To What Extent Is Teacher Participation in LEAP Associated with Improvements in Teacher Contribution to Student Achievement?

Employing within-year standardized teacher value-added scores, we explore the extent to which engagement with LEAP is related to teachers' contributions to student achievement in the first two years of LEAP. Across models, we find inconsistent estimates that are typically insignificant (Table 8).

Table 7. Teacher Participation in LEAP and Change in Summative Teaching Scores.

	Current Values Model			Lagged Values Model	
	(1)	(2)	(3)	(4)	(5)
Coaching session frequency	0.021*	0.018	0.036	0.015	0.018
	(0.010)	(0.015)	(0.037)	–0.017	–0.027
Coaching session frequency (last)	-	-	-	–0.013	–0.018
	-	-	-	–0.018	–0.027
Seminar session frequency	0.008	–0.010	–0.130*	0.015	0.012
	(0.018)	(0.025)	(0.064)	–0.029	–0.046
Seminar session frequency (last)	-	-		–0.009	0.016
		-	-	–0.034	–0.058
Teacher covariates	*X*	*X*		*X*	*X*
School fixed effects	*X*			*X*	
Team fixed effects		*X*			*X*
Teacher fixed effects			*X*		
Observations	2,755	2,559	1,597	884	878

Note: Clustered standard errors in parentheses.
*$p < 0.05$.

Discussion

Improved teaching quality is the most potent school-based mechanism for improving student outcomes. Some small-scale experiments suggest that intensive opportunities for professional learning enhance teachers' skills, but these results often rely on research-implemented interventions. There is remarkably little evidence of PL interventions broadly implemented by school personnel in large, urban districts (Kraft et al., 2018; Lockwood et al., 2010). Based on our review of extant studies, the analyses presented here with 3,402 teacher-year observations represent the largest sample of coached teachers in the literature.

Employing differential implementation of LEAP across teachers, and within teachers over time, we hypothesize that if successful, LEAP will initially shift teachers' perceptions, and then, over time, influence their behavior in terms of decisions to remain teaching at the same school or in the district, as well as in terms of their instructional approach in the classroom. To explore this empirically, we examine the relationship between frequency of the two primary LEAP components (one-on-one teacher coaching and group learning in teacher seminars) and teacher perceptions of LEAP and the school culture among teachers, as well as teacher retention, teaching skills, and teachers' ability to improve student achievement.

Our most proximal outcomes, teacher perceptions of the program and of the peer culture at their school, are positively associated with participation in LEAP structures. Our estimates of more distal outcomes are considerably more mixed. Teacher retention is positively related to teacher participation in coaching and seminar, although the

Table 8. Teacher Participation in LEAP and Change in Contribution to Student Achievement.

	Current Values Model			Lagged Values Model	
	(1)	(2)	(3)	(4)	(5)
Coaching session frequency	0.006	–0.028	0.130	0.017	0.010
	(0.038)	(0.082)	(0.187)	(0.071)	(0.193)
Coaching session frequency (last)	-	-		–0.103	–0.108
	-	-		(0.127)	(0.251)
Seminar session frequency	0.114^{++}	–0.038	0.184	0.089	0.169
	(0.072)	(0.120)	(0.319)	(0.146)	(0.540)
Seminar session frequency (last)	-	-		–0.091	–0.066
	-	-		(0.150)	(0.323)
Teacher covariates	*X*	*X*		*X*	*X*
School fixed effects	*X*			*X*	
Team fixed effects		*X*			*X*
Teacher fixed effects			*X*		
Observations	398	382	232	149	148

Note: Clustered standard errors in parentheses.
$^{++}$ $p < 0.2$.

relationship is more substantive and consistently statistically significant for school-based teacher retention than district retention. Teacher classroom observation scores are not consistently related to either coaching session frequency or participation in seminar. LEAP coaching frequency appears unrelated to teacher value added, and seminar frequency is related in inconsistent directions, depending on model specification. While dissapointing given the investment in LEAP, these findings are in line with many prior studies of at-scale coaching (Biancarosa et al., 2010; Garet et al., 2008, 2011; Lockwood et al., 2010; Marsh et al., 2010).

We interpret these findings in light of three key contextual factors. First, the relationships we observe occurred in the first two years of the program. During this time, enabling conditions like trust and deprivatization of practice may still be forming, and therefore outcomes like teacher skill development and its influence on student achievement may just be emerging. Tracking LEAP's influence on the more distal outcomes over the next two years will be important. Second, the LEAP program represents an at-scale, fully implemented initiative without an identified control group. We approach this carefully in our analysis using comparisons of LEAP frequency within teachers across years and within teams throughout the two-year period, but we recognize that a lack of contrast may suppress observable differences. In a study of existing collaboration across schools, Ronfeldt and colleagues (2015) found that teacher teams produce a collectivist, supportive culture that benefits all students. From a policy lens, this is unequivocally positive in the sense that all teachers and students should benefit from the PL structures introduced by the LEAP program. From an analytic perspective, however, this may reduce our ability to detect changes in outcomes using differential PL frequency at the teacher level.

Third, we observe the complicated evolution of LEAP implementation from its inception to the period of study examined here in a separate mixed methods analysis (Boguslav et al., 2020). In that analysis, we document the ways in which the developers of LEAP balanced the competing desires for school-based autonomy with district-provided structure, materials, and guidance. This balancing resulted in an evolution of LEAP over time, with variable school-based adoption of materials and processes like district-created materials for seminar and district-endorsed structures for coaching conversations. Within this context, the assessment of LEAP frequency alone becomes particularly limiting, because the quality of LEAP was likely variable across school sites, and even among teams led by different LEAP leaders in the same school. Robust, validated measures of LEAP leader quality would have been instrumental in helping us detect differential shifts in our outcomes of interest. Moreover, our findings focus on macro-level structural aspects of PL, rather than more nuanced micro-processes occurring on each team. As such, there are many questions we cannot answer, including the degree to which different LEAP teams focused on specific instructional practices or pieces of content in seminar time. Variation in foci and heterogeneity in these micro-level processes may contribute to the results—or lack thereof—that we report here. We see attention to these micro-level processes as a critical direction for future research.

Programs hoping to build on this research at scale should also think carefully about measures of coaching quality and the coach–teacher relationship. Just as variation in seminar content might help explain variation in observed outcomes, so, too, might outcomes be explained by more micro-level interactions between a LEAP leader and coach during observations and debriefs. Experimental evidence suggests that individual coach assignment produces large variations in treatment effects (Blazar & Kraft, 2015), and in a conceptual synthesis of recent coaching studies, Blazar suggests "coaches *are* the intervention" (Blazar, 2020, p. 5). Despite this emphasis, the research has not yet coalesced around the attributes that characterize effective coaches or effective coaching. There is some suggestive evidence that credentials and experience are important (Campbell & Malkus, 2014; Coburn & Russell, 2008; Coburn & Woulfin, 2012), but we lack clarity about the specific credentials and experience that matter most.

Assessing the quality of a coach–teacher relationship is even more challenging, particularly at scale. One analysis examined differences between district-hired coaches and school-hired coaches, concluding that school-based coaches were more likely to form trusting relationships conducive to higher quality coaching conversations (Kane & Rosenquist, 2019). In the case of LEAP, however, almost all coaches were school-based, limiting our ability to make comparisons. A recent study used careful qualitative analysis to determine the extent to which coaches were viewed as legitimate sources of expertise within a school, and therefore became more or less "enmeshed" in the daily routines (Woulfin, 2020). Another used social-network data to systematically determine the extent to which coaches facilitated supportive communication with teachers who had varied interest in the reform goals of coaches (Hashim, 2020). Both of these approaches have limitations, but they underscore the potential for more innovative exploration of complicated coaching constructs at scale. Could coaches record videos of their debrief sessions, or submit their goals for teachers ahead of the observation? Could principals

evaluate coaches by attending coaching sessions, much like they observe classrooms to evaluate teachers? Future research to develop such measures of coaching quality would be very helpful.

Finally, we note that from an analytic perspective, we had originally hoped to disentangle the unique contribution of two PL structures and their related outcomes, per the framework set forth by Hill and colleagues (2013). For example, we hypothesized we might find suggestive evidence that coaching drives instructionally focused outcomes like improved scores on observational measures of teaching. In contrast, we hypothesized that teacher seminar would be more substantially related to teacher retention, because it intentionally creates a school-based community for teachers. However, we were ultimately dissuaded from drawing strong conclusions about the distinctions between the two structures of PL, due to the modest coefficient sizes and the inconsistent statistical significance of the estimates. These kinds of analyses seem particularly useful for future research, where the treatment–control contrast may be greater, or where there are additional years of data to observe the emergence of stronger relationships.

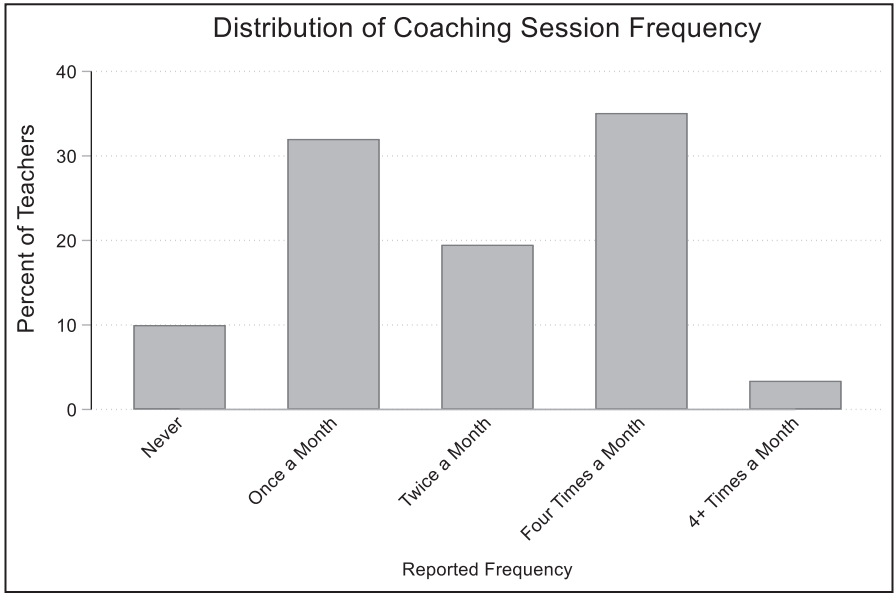
Conclusion

Overall, we find that a PL program that leverages one-on-one coaching and content-based teacher teams can be positively associated with outcomes of interests, even within the first two years of the program. In particular, teacher perceptions of the program, perceptions of the peer culture at their school, and school-based retention are positively related to teacher participation in the structures of LEAP. This suggests that PL programs that center within-school connections and supports for teachers—in this case, vertically structured LEAP teams led by school-based LEAP leaders—may support positive school-level outcomes. However, we do not observe statistically meaningful relationships between participation in LEAP and other important, teacher-level outcomes like observed teaching quality and teacher contributions to student achievement. Future studies should think critically about how to embed measures of coach and coaching quality to better understand mechanisms that may produce these mixed results at scale.

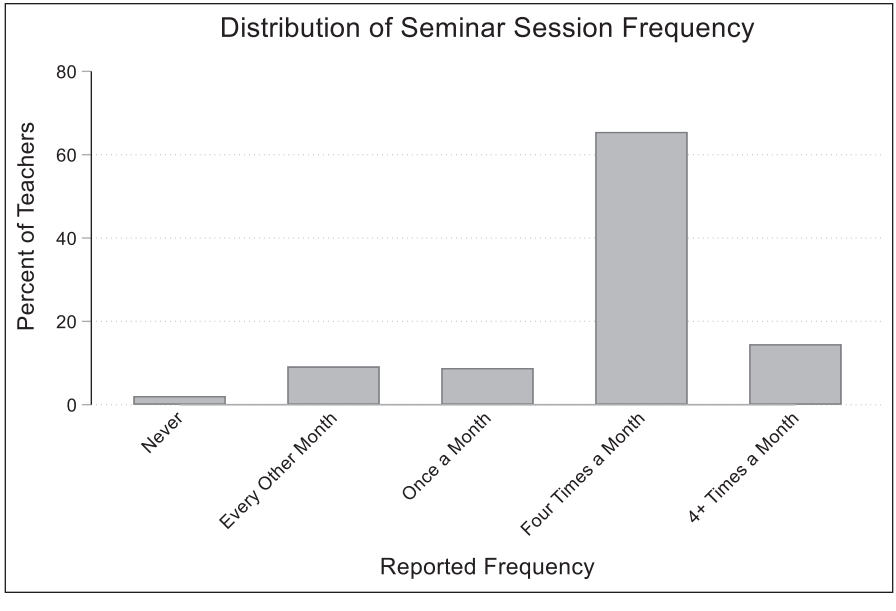
LEAP is a resource-intensive program. DCPS hired 450 LEAP leaders (coaches) who invested a substantial number of hours each week in group PL and individual coaching. Every DCPS teacher invested, on average, more than two hours every week in LEAP-related activities. Principals made significant ongoing investments in LEAP. It will take more time and research to understand the degree to which and ways in which these investments are associated with the range of desired outcomes.

In the meantime, the field should invest much greater effort in better understanding how to facilitate the development of teachers at-scale, alongside with building an evidence base about the tradeoffs of doing so. Students, teachers, administrators, and policymakers would meaningfully benefit from stronger evidence of ways in which teaching quality can be improved at scale. Professional learning and coaching provide a promising mechanism to realize improved teaching effectiveness, but much remains to be learned about how to design and implement such programs effectively.

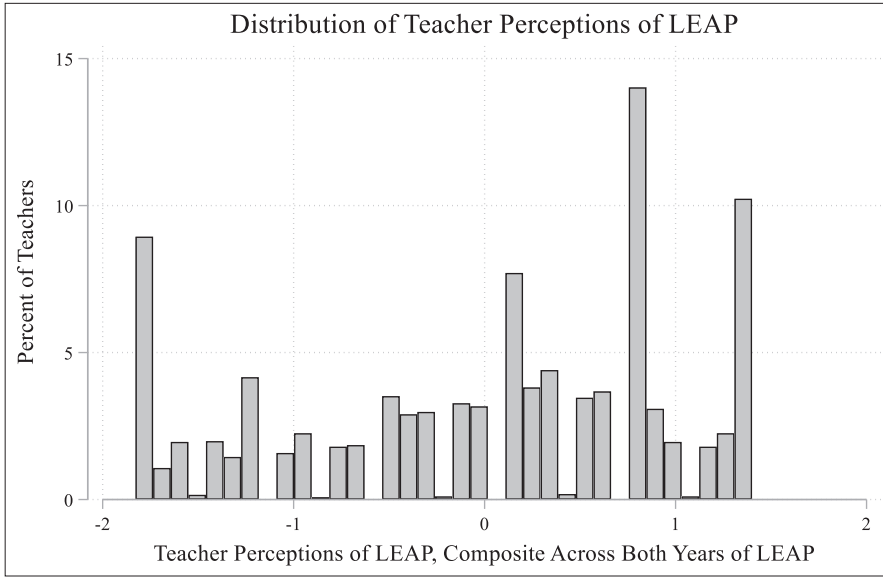
Appendix



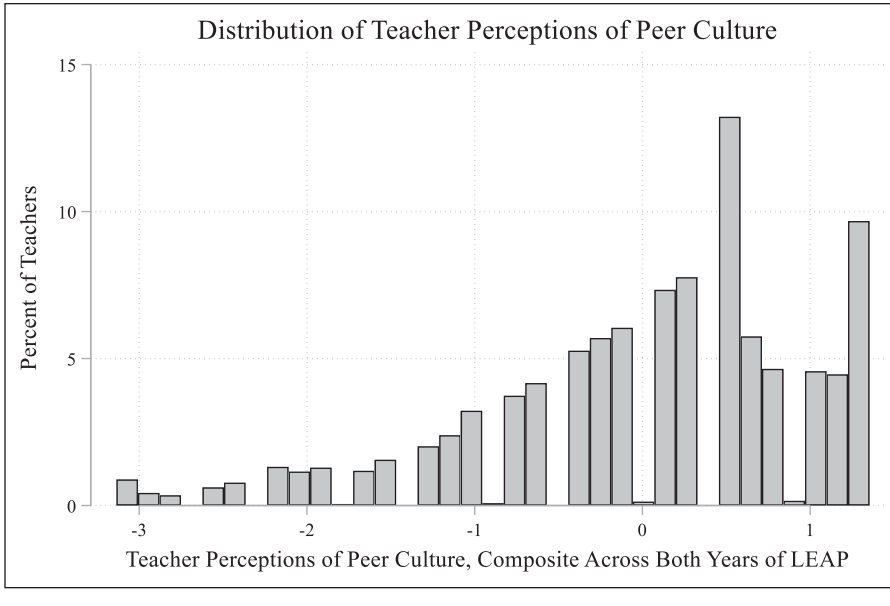
Appendix Figure 1.



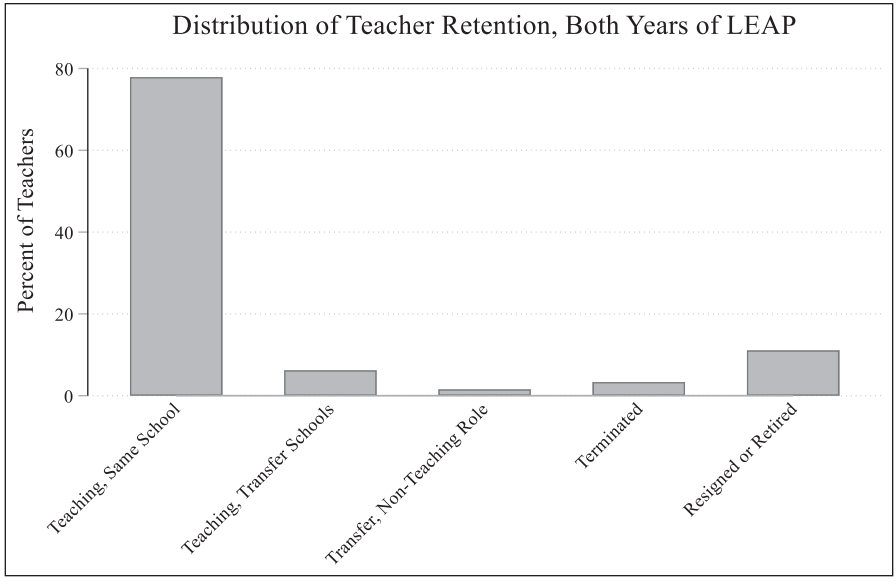
Appendix Figure 2.



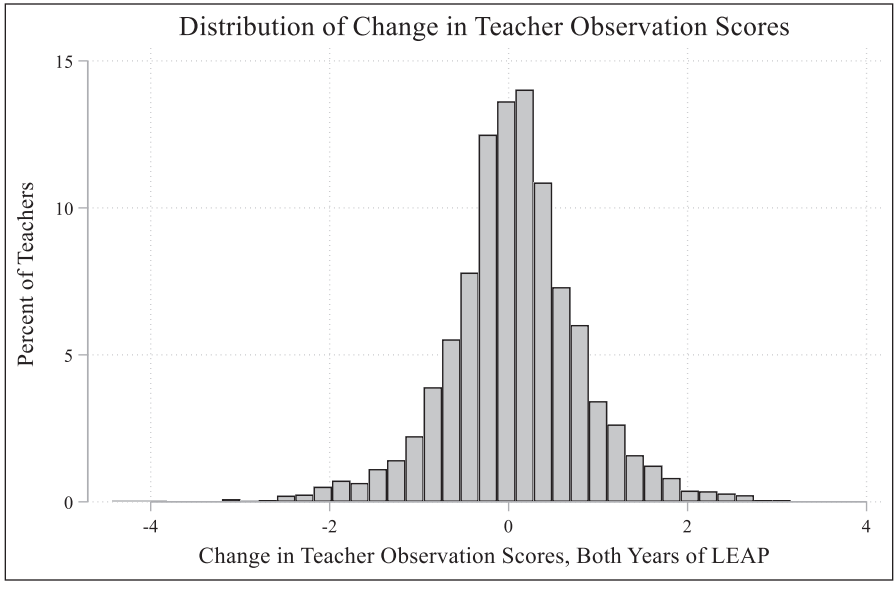
Appendix Figure 3.



Appendix Figure 4.



Appendix Figure 5.



Appendix Figure 6.

Declaration of Conflicting Interests

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

Funding

Authors are listed alphabetically. We are grateful to the District of Columbia Public Schools for supplying the data employed in this research and to Katie Burke, Alison Williams, Chris Miller, Astrid Atienza, Sooyon Stiller for answering our many questions. We appreciate TNTP making the INSIGHT survey data available to us. We received consistent mentorship on this work from Jim Wyckoff, and excellent research assistance from Gabriel Gassmann and Katherine Sadowski. We appreciate comments on an earlier draft from Allison Atteberry, Emanuele Bardelli, and Matt Ronfeldt. We are grateful for financial support from the Charles and Lynn Schusterman Family Philanthropies and the Overdeck Family Foundation. The research reported here was also supported by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305B140026 to the Rectors and Visitors of the University of Virginia. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Errors are attributable to the authors.

Notes

1. Again, we note a lack of rigorous evidence in this space and caveat that observations of increased retention in relation to collaboration may be confounded by school-level factors such as a principal who adopts high-quality curricula, for example, which fosters both increased collaboration as teachers make sense of the new materials and increased retention as teachers appreciate the enhanced curricular support.
2. "We found that the PD that they were receiving at the time, which was more traditional, was not really pushing our teachers from good to great" (2017 interview with the deputy chief of LEAP at that time).
3. Prior year frequency is set to zero for observations in LEAP year one ($t = 0$).
4. In the survey administered during the first year of LEAP, there were separate questions for coaching activities (observation and debrief), which were collapsed in the survey administered during year two of LEAP. We collapse the year-one questions by employing the mean of the relevant categories and then averaging the responses for the separate measures to form a single coaching frequency measure.
5. In each year of our study, about 6% of teachers transferred to another school in the district.
6. Tested grade and subject status is determined by the DCPS IMPACT Evaluation System, using "Group 1" General Education Teachers with Individual Value-Added student achievement data. The share of group 1 teachers is fairly consistent over time, as reported in other assessments of DCPS programming (see, for example, Dee & Wyckoff, 2015).
7. We also had access to coach-reported frequency through an online platform called Whetstone; however, we were limited by the scope of this reporting. First, Whetstone administrative reports were only available in the second year of the study. Second, this data was only recorded by a subset of coaches. For example, in comparison with survey data, we found that about 800 teachers reported engaging in one-on-one coaching in the second year of LEAP but do not appear in the Whetstone data. We therefore concluded that teacher reports, although limited in their own ways, provided a more comprehensive accounting of frequency in this context.

References

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement in DCPS. *Educational Evaluation and Policy Analysis*, *39*(1), 54–76.

Allen, J. P., Hafen, C. A., Gregory, A. C., Mikami, A. Y., & Pianta, R. (2015). Enhancing secondary school instruction and student achievement: Replication and extension of the My Teaching Partner-Secondary Intervention. *Journal of Research on Educational Effectiveness*, *8*(4), 475–489. https://doi.org/10.1080/19345747.2015.1017680

Astuto, T. A., Clark, D. L., Read, A. M., McGree, K., & Fernandez, L. (1993). *Challenges to dominant assumptions controlling educational reform*. Regional Laboratory for the Educational Improvement of the Northeast and Islands.

Atteberry, A., & Bryk, A. S. (2011). Analyzing teacher participation in literacy coaching activities. *The Elementary School Journal*, *112*(2), 356–382.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, *31*(4), 73–102.

Biancarosa, G., Bryk, A. S., & Dexter, E. R. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *The Elementary School Journal*, *111*(1), 7–34. https://doi.org/10.1086/653468

Blase, J., & Blase, J. (2000). Effective instructional leadership: Teachers' perspectives on how principals promote teaching and learning in schools. *Journal of Educational Administration*, *38*(2), 130–141.

Blazar, D. (2020). Teacher coaching to improve instruction at scale: opportunities and challenges in policy contexts. *Teachers College Record*, *122*(10), 1–9.

Blazar, D., & Kraft, M. A. (2015). Exploring mechanisms of effective teacher coaching: A tale of two cohorts from a randomized experiment. *Educational Evaluation and Policy Analysis*, *37*(4), 542–566.

Boguslav, A., Cohen, J., Katz, V., Sadowski, K., Wiseman, E., & Wyckoff, J. (2020). *Targeted professional development at scale: Evidence from DC public schools*. Paper presented at the annual meeting of the Association for Education Finance and Policy, Fort Worth, TX.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15. https://doi.org/10.3102/0013189X033008003

Borman, G. D., & Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research*, *78*(3), 367–409.

Borman, G. D., Slavin, R. E., Cheung, A. C. K., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal*, *44*(3), 701–731. https://doi.org/10.3102/0002831207306743

Bruce, C. D., Esmonde, I., Ross, J., Dookie, L., & Beatty, R. (2010). The effects of sustained classroom-embedded teacher professional learning on teacher efficacy and related student achievement. *Teaching and Teacher Education*, *26*(8), 1598–1608.

Campbell, P. F., & Malkus, N. N. (2014). The mathematical knowledge and beliefs of elementary mathematics specialist-coaches. *ZDM—The International Journal on Mathematics Education*, *46*(2), 213–225. https://doi.org/10.1007/s11858-013-0559-6

Carver-Thomas, D., & Darling-Hammond, L. (2017). *Teacher turnover: Why it matters and what we can do about it*. Learning Policy Institute.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–2679. https://doi.org/10.1257/aer.104.9.2633

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3–12.

Coburn, C. E., & Russell, J. L. (2008). Getting the most out of professional learning communities and coaching: Promoting interactions that support instructional improvement. *Learning Policy Brief*, *1*(3), 1–4.

Coburn, C. E., & Woulfin, S. L. (2012). Reading coaches and the relationship between policy and practice. *Reading Research Quarterly*, *47*(1), 5–30.

Conaway, C., & Goldhaber, D. (2020). Appropriate standards of evidence for education policy decision making. *Education Finance and Policy*, *15*(2), 383–396.

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, *34*(2), 267–297. https://doi.org/10.1002/pam.21818

Desimone, L. M., & Garet, M. S. (2015). Best practices in teachers' professional development in the United States. *Psychology, Society, & Education*, *7*(3), 252. https://doi.org/10.25115/psye.v7i3.515

Desimone, L. M., & Pak, K. (2017). Instructional coaching as high-quality professional development. *Theory into Practice*, *56*(1), 3–12. https://doi.org/10.1080/00405841.2016.1241947

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. University of South Florida.

Fullan, M., & Knight, J. (2011). Coaches as system leaders. *Educational Leadership*, *69*(2), 50–53.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H., Doolittle, F., Zhu, P., & Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement*. NCEE 2008-4030. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., Garrett, R., Yang, R., & Borman, G. D. (2016). *Focusing on mathematical knowledge: The impact of content-intensive teacher professional development*. NCEE 2016-4010. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., & Hurlburt, S. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation*. NCEE 2011-4024. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record*, *109*(4), 877–896.

Hashim, A. (2020). Coaching and districtwide improvement: Exploring the systemic leadership practices of instructional coaches. *Teachers College Record*, *122*(10), 1–44.

Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching*, *49*(3), 333–362.

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads, and challenges. *Educational Researcher*, *42*(9), 476–487. https://doi.org/10.3102/0013189X13512674

Hill, H. C., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, *83*(2), 371–384.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*(3), 499–534.

Ippolito, J. (2010). Three ways that literacy coaches balance responsive and directive relationships with teachers. *The Elementary School Journal*, *111*(1), 164–190. https://doi.org/10.1086/653474

Kane, B. D., & Rosenquist, B. (2019). Relationships between instructional coaches' time use and district- and school-level policies and expectations. *American Educational Research Journal*, *56*(5), 1718–1768.

Kazemi, E., & Franke, M. L. (2004). Teacher learning in mathematics: Using student work to promote collective inquiry. *Journal of Mathematics Teacher Education*, *7*(3), 203–235. https://doi.org/10.1023/B:JMTE.0000033084.26326.19

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*(4), 945–980.

Kraft, M., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, *88*(4), 547–588.

Kraft, M. A., & Hill, H. C. (2019). *Developing ambitious mathematics instruction through web-based coaching: A randomized field trial*. EdWorkingPaper No. 19-119. Annenberg Institute for School Reform at Brown University.

Little, J. W. (2003). Constructions of teacher leadership in three periods of policy and reform activism. *School Leadership and Management*, *23*(4), 401–419.

Little, J. W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education*, *118*(2), 143–166. https://doi.org/10.1086/663271

Lockwood, J. R., McCombs, J. S., & Marsh, J. A. (2010). Linking reading coaches and student achievement: Evidence from Florida middle schools. *Educational Evaluation and Policy Analysis*, *32*(3), 372–388. https://doi.org/10.3102/0162373710373388

Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education*, *80*(3), 44–70.

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, *41*(3), 260–293.

Marsh, J. A., Sloan McCombs, J., & Martorell, F. (2010). How instructional coaches support data-driven decision making: Policy implementation and effects in Florida middle schools. *Educational Policy*, *24*(6), 872–907.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis*, *21*(1), 29–45.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, *9*(2), 171–178.

McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. University of Chicago Press

Miller, M. D., Brownell, M. T., & Smith, S. W. (1999). Factors that predict teachers staying in, leaving, or transferring from the special education classroom. *Exceptional Children*, *65*(2), 201–218.

Moolenaar, N. M., Sleegers, P. J., & Daly, A. J. (2012). Teaming up: Linking collaboration networks, collective efficacy, and student achievement. *Teaching and Teacher Education*, *28*(2), 251–262.

Penuel, W. R., Gallagher, L. P., & Moorthy, S. (2011). Preparing teachers to design sequences of instruction in earth systems science: A comparison of three professional development programs. *American Educational Research Journal*, *48*(4), 996–1025.

Pounder, D. G. (1998). *Restructuring schools for collaboration*. SUNY Press.

Puchner, L. D., & Taylor, A. R. (2006). Lesson study, collaboration and teacher efficacy: Stories from two school-based math lesson study groups. *Teaching and Teacher Education*, *22*(7), 922–934.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458. https://doi.org/10.1111/j.1468-0262.2005.00584.x

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher collaboration in instructional teams and student achievement. *American Educational Research Journal*, *52*(3), 475–514. https://doi.org/10.3102/0002831215585562

Saunders, W. M., Goldenberg, C. N., & Gallimore, R. (2009). Increasing achievement by focusing grade-level teams on improving classroom learning: A prospective, quasi-experimental study of Title I schools. *American Educational Research Journal*, *46*(4), 1006–1033.

Simon, N. S., & Johnson, S. M. (2015). Teacher turnover in high-poverty schools: What we know and can do. *Teachers College Record*, *117*(3), 1–36.

Spillane, J. P., Parise, L. M., & Sherer, J. Z. (2011). Organizational routines as coupling mechanisms: Policy, school administration, and the technical core. *American Educational Research Journal*, *48*(3), 586–619.

Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change*, *7*(4), 221–258. https://doi.org/10.1007/s10833-006-0001-8

Supovitz, J. A. (2002). Developing communities of instructional practice. *Teachers College Record*, *104*(8), 1591–1626.

Supovitz, J. A., & Christman, J. B. (2003). Developing communities of instructional practice: Lessons from Cincinnati and Philadelphia. *CPRE Policy Briefs* (pp. 1–9). University of Pennsylvania.

Tschannen-Moran, M. (2001). Collaboration and the need for trust. *Journal of Educational Administration*, *39*(4), 308–331.

Vangrieken, K., Meredith, C., Packer, T., & Kyndt, E. (2017). Teacher communities as a context for professional development: A systematic review. *Teaching and Teacher Education*, *61*, 47–59.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, *24*(1), 80–91. https://doi.org/10.1016/j.tate.2007.01.004

Webster-Wright, A. (2009). Reframing professional development through understanding authentic professional learning. *Review of Educational Research*, *79*(2), 702–739.

Wei, R. C., Darling-Hammond, L., & Adamson, F. (2010). *Professional development in the United States: Trends and challenges*. National Staff Development Council.

Woulfin, S. (2020). Crystallizing coaching: An examination of the institutionalization of instructional coaching in three educational systems. *Teachers College Record*, *122*(10), 1–32.

Woulfin, S. L., & Jones, B. (2018). Rooted in relationships: An analysis of dimensions of social capital enabling instructional coaching. *Journal of Professional Capital and Community*, *3*(1), 25–38.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. (2007). Reviewing the evidence on how teacher professional development affects student achievement. *Issues & Answers*. Rel 2007, No. 033. Regional Educational Laboratory Southwest.

Zambo, R., & Zambo, D. (2008). The impact of professional development in mathematics on teachers' individual and collective efficacy: The stigma of underperforming. *Teacher Education Quarterly*, *35*(1), 159–168.

Author Biographies

Julie Cohen is an associate professor of curriculum, instruction, and special education in the School of Education and Human Development at the University of Virginia. She studies teachers and teaching, with a focus on policies and practices that support the development of effective instructional practices.

Emily Wiseman is a consultant in the education practice at EY-Parthenon in New York City. She earned her PhD in Education Policy from the University of Virginia, where she worked in a research-practice partnership with the District of Columbia Public Schools. Before completing her doctorate, Emily was a reading specialist for the DC Public Schools.