EdWorkingPaper No. 22-597

Does Teacher Professional Development Improve Student Learning? Evidence from Leading Educators' Teacher Fellowship Model

Ariana Audisio
Leading Educators

Rebecca Taylor-Perryman
Leading Educators

Tim Tasker
Leading Educators

Matthew P. Steinberg
George Mason University

Teachers are the most important school-specific factor in student learning. Yet, little evidence exists linking teacher professional learning programs and the various strategies or components that comprise them to student achievement. In this paper, we examine a teacher fellowship model for professional learning designed and implemented by Leading Educators, a national nonprofit organization that aims to bridge research and practice to improve instructional quality and accelerate learning across school systems. During the 2015-16 and 2016-17 school years, Leading Educators conducted its fellowship program for teachers and school leaders to provide educators ongoing, collaborative, job-embedded professional development and to improve student achievement. Relying on quasi-experimental methods, we find that a school's participation in the fellowship model increased student proficiency rates in math and English language arts on state achievement exams. Further, student achievement benefitted from a more sustained duration of teacher participation in the fellowship model, and the impact on student achievement varied depending on the share of a school's teachers who participated in the fellowship model and the extent to which teachers independently selected into the fellowship model or were appointed to participate by school leaders. Taken together, findings from this paper should inform professional learning organizations, schools, and policymakers on the design, implementation and impact of teacher professional learning.

VERSION: June 2022

Suggested citation: Audisio, Ariana, Rebecca Taylor-Perryman, Tim Tasker, and Matthew P. Steinberg. (2022). Does Teacher Professional Development Improve Student Learning? Evidence from Leading Educators' Teacher Fellowship Model. (EdWorkingPaper: 22-597). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/ah2f-z471

Does Teacher Professional Development Improve Student Learning?

Evidence from Leading Educators' Teacher Fellowship Model

Ariana Audisio
Rebecca Taylor-Perryman
Tim Tasker
Leading Educators

Matthew P. Steinberg
George Mason University

June 24, 2022

The authors thank Leading Educators for providing the proprietary program data for this paper and Natalie Truong for research assistance. The research reported here was supported in part by the Carnegie Corporation of New York. Corresponding author: M. Steinberg (msteinb6@gmu.edu).

Abstract

Teachers are the most important school-specific factor in student learning. Yet, little evidence exists linking teacher professional learning programs and the various strategies or components that comprise them to student achievement. In this paper, we examine a teacher fellowship model for professional learning designed and implemented by Leading Educators, a national nonprofit organization that aims to bridge research and practice to improve instructional quality and accelerate learning across school systems. During the 2015-16 and 2016-17 school years, Leading Educators conducted its fellowship program for teachers and school leaders to provide educators ongoing, collaborative, job-embedded professional development and to improve student achievement. Relying on quasi-experimental methods, we find that a school's participation in the fellowship model increased student proficiency rates in math and English language arts on state achievement exams. Further, student achievement benefitted from a more sustained duration of teacher participation in the fellowship model, and the impact on student achievement varied depending on the share of a school's teachers who participated in the fellowship model and the extent to which teachers independently selected into the fellowship model or were appointed to participate by school leaders. Taken together, findings from this paper should inform professional learning organizations, schools, and policymakers on the design, implementation and impact of teacher professional learning.

Keywords: Teacher professional development; student achievement; professional learning communities; teacher coaching; distributed leadership.

Introduction

After decades of research, there is consensus that teachers are the most important school-specific factor in student learning (Darling-Hammond, 2000; Nye et al., 2004; Wright et al., 1997). Students with more effective teachers not only score better on achievement tests (Aaronson et al., 2007; Rivkin et al., 2005; Rockoff, 2004), but they also graduate at higher rates (Koedel, 2008) and have higher rates of post-secondary enrollment (Jackson, 2014). In the longer run, students of more effective teachers attain higher rates of post-secondary degrees (Lee, 2018), experience greater success in the labor market (Chetty et al., 2014), and generate significantly higher lifetime earnings (Chetty et al., 2014; Hanushek, 2011).

If student success depends to such a considerable degree on teacher effectiveness, then teacher effectiveness undoubtedly rests on the opportunities teachers have for developing strong instructional skills. To paraphrase a common saying among educators: good teachers are not born, they are developed. To wit: a growing body of research demonstrates that teacher effectiveness increases dramatically during their first three years in the profession (Boyd et al., 2005; Hanushek et al., 2005; Henry et al., 2011), with more recent evidence suggesting that teacher effectiveness grows throughout the first decade of a teacher's tenure in the profession (Papay & Kraft, 2015). As any veteran teacher can attest, however, teachers improve not just in their first few years in the profession, but they are often required to continue to advance and grow their instructional practice over the course of their entire careers in response to a seemingly ever-evolving set of policies, standards, and expectations (Coolahan, 2002; Day, 1999).

Despite the importance of educator professional development for student learning, we still know relatively little about the effectiveness of professional learning programs and the various strategies or components that comprise them. In fact, scholars have observed that the role that individual teacher development and coaching programs play in improving both the quality of teaching and student achievement is both undertheorized and understudied (Desimone, 2009; Wayne et al., 2008). Moreover, relatively few teacher professional development programs have demonstrated strong evidence of increased outcomes at scale, across both a broad range of settings and student groups, in part due to a lack of generalizable knowledge about the specific features of such programs that effectively improve student learning (Hill et al., 2021).

Where specific knowledge of the program features and their connection to student learning is available, challenges arise when considering how to assess the impact of a differentiated sets of programs that often share some elements in common while still maintaining other unique features. This challenge is only exacerbated by the relatively large degree of variation in program features and the multitude of educational contexts in which professional learning programs are implemented. All this programmatic and contextual variation limits the utility of most professional learning intervention studies for local educational agencies (LEA) and the professional learning providers supporting them. This is because these groups are concerned about successful implementation of a particular program within a specific, local context rather than a generic intervention implemented without concern for context (Hill et al., 2021).

In this paper, we aim to address these important gaps and understudied questions by examining both the features and the impact of a teacher professional learning program implemented across a small but diverse set of urban school districts in the United States. The professional learning program examined herein was designed and implemented by Leading Educators, a national nonprofit organization that aims to bridge research and practice to improve instructional quality and accelerate learning across school systems. During the 2015-16 and

2016-17 school years, Leading Educators ran a fellowship program for teachers and school leaders across four different states that was designed to provide educators ongoing, collaborative, job-embedded professional development. Over the course of two school years, the fellowship program aimed to develop teachers' and leaders' beliefs, knowledge, and skills to lead ongoing, pedagogical development for the teachers they, in turn, supported back at their schools. The professional learning these teachers and leaders led within their schools was specific to mathematics and English language arts (ELA) curriculum standards, and it aimed to shift the school's instructional culture, the fellows' own instructional practices, and the instructional practices of their colleagues. The ultimate aim of all these adult learning efforts was to improve student learning in the schools supported by the program.

Examination of the subtle variations in approach across the multiple contexts in which the fellowship was implemented allows us to explore several unresolved questions regarding the impact of this type of professional learning program. For instance, program feature variation both across and within the fellowship regions creates an opportunity to evaluate the relative impact of those diverse features on student learning. In addition, by focusing on programming delivered over the course of two years, we have an opportunity to assess questions about the amount of time required to bring about considerable change in skills and behavior among educators as well as the subsequent impact of those changes on students' learning. Finally, access to data from beyond just the two years of active programming allows us to explore questions regarding the sustainability of impact. Specifically, comparing the effect before and after the conclusion of the fellowship will shed light on whether schools can continue to accelerate improvements in instruction even after support from an external partner ends.

Relying on quasi-experimental methods for causal inference, including difference-in-differences and event study strategies, we find that a school's participation in the fellowship model increased student proficiency rates on math and English language arts (ELA) state achievement exams by 8.5 and 5.3 percentage points, respectively. Our findings further suggest that student achievement benefitted from a more sustained duration of teacher participation in Leading Educators' fellowship model (i.e., two years rather than one year). We also show that specific program features generated differential benefits for student learning. Specifically, we show that student achievement varied depending on the share of a school's teachers who participated in the fellowship model and the extent to which teachers independently selected into the fellowship model or were appointed to participate in it by their school leaders.

By conducting this rigorous internal impact evaluation of its educator fellowship model, Leading Educators acknowledges the importance of understanding the aspects of its service provision – and the contexts in which those services are provided – that may be most effective for improving teacher professional practice and student performance. Thus, this evaluation aims to inform other organizations on the design and implementation features related to improving teacher professional development and, ultimately, student learning. Further, we offer this paper as a 'call to action' for other professional learning partner organizations to intentionally and rigorously examine their internal operations and impact. Ultimately, findings from this paper should illuminate the importance of organizational transparency as a means of understanding the elements of program design and implementation that are most effective at improving teacher practice and student learning.

Related Literature

In a recent landmark study, Kraft and colleagues (2018) conducted a meta-analysis of 60 primary reports on teacher coaching, which demonstrated that coaching significantly improves teachers' instructional practices as well as their students' learning outcomes. Notably, these researchers found that several common programmatic features were associated with improved teacher practice and student achievement: job-embedded practice; an intense and sustained duration of teacher coaching and professional development; a focus on discrete instructional skills; and active learning (Kraft et al., 2018). Their study also uncovered some important differences, however, corresponding to the type or focus of the coaching being provided. For instance, they found that content-specific coaching (e.g., coaching tailored to math instruction) to be marginally more effective at improving teachers' instructional practice and student achievement than was coaching focused more broadly on general pedagogical skills that are not tied to a specific content area. Furthermore, the pairing of coaching with other developmental strategies revealed some notable differences, with coaching being most effectively offered alongside group training or instructional resources (Kraft et al., 2018). Finally, the results revealed no significant differences according to the quantity of coaching received (Kraft et al., 2018), suggesting that coaching quality and the specific components – or combinations thereof – may be most important for its impact.

Beyond understanding the impact of coaching, the field lacks a robust body of evidence demonstrating that a diverse range of educator professional learning programs consistently lead to significant improvements in student learning. For instance, in summarizing the findings of all 67 evaluations of Investing in Innovation federally funded programs, Boulay et al. (2018) found that fully 53 of those offered teachers professional learning as a key program component. Of those 53 interventions, only 10 of those reported on an ELA or math student learning outcome, and only 6 of those (less than 10% of all evaluations) demonstrated evidence of at least 1 positive, statistically significant impact on student learning. Notably, the professional learning offered to teachers in five of these six studies showing a positive impact on students' ELA or math outcomes was content-specific, a finding further supporting the importance of content-specificity in the design and implementation of teacher professional development.

Other research examining the optimal duration of educator professional learning programs, a primary implementation feature, offers up a contradictory set of findings. Another recent meta-analysis shows that whereas teacher professional development improved student reading achievement, program duration did not explain any differences in the magnitude of the impact (Didion et al., 2020). Likewise, Lynch and colleagues (2019) found no association between the duration of professional learning and its impact. In keeping with these results, Kraft (2020) posits that professional learning programs will generally achieve smaller effects when they either require cumulative decisions or more sustained efforts over time, suggesting that simply extending the period of implementation is no guarantee for greater program impact.

Other studies suggest there might be a threshold for the minimum amount of time required before the effects of an intervention among educators can be detected among their students. For example, in a study of a large number of elementary schools, Timperley and Alton-Lee (2008) found that teachers needed two years to develop the skills to effectively lead inquiry cycles with their colleagues before those efforts produced significant improvements in student learning. Similarly, Dolfin and colleagues (2019) found that a professional learning program incorporating teacher institutes along with coaching support for content, pedagogy, and professional learning culture did significantly improve students' test scores, but only after two years of implementation. In view of this mixed pattern of findings regarding program duration, Hill and colleagues (2021) argue that further exploration of the relationship between the duration of professional learning and the magnitude of its impacts on learning and achievement is necessary.

In contrast to the attention paid to issues involving program duration, the question of which educator roles should participate in professional learning programs has been largely unexamined in the existing research literature. In particular, the question of whether teachers should participate in professional learning programs alongside school leaders (e.g., principals) or system leaders (e.g., principal managers, cross school instructional coaches, Directors of Literacy) has been paid remarkably little attention. One theoretical justification for involving school and system (or district-level) leaders in professional learning rests on the importance of having a clear vision for instructional improvement and coherent structures to support it (Fullan & Quinn, 2015; Cobb et al., 2019). Indeed, Hill and colleagues (2018) link the null results of a math professional learning program to the lack of clear messaging and alignment from district leaders, offering some support for the value of school and system leader involvement or participation. In light of this, it is reasonable to expect that involving school and system leaders in professional learning typically reserved for teachers could lead to additional benefits for student achievement.

A final, largely understudied aspect of teacher professional learning concerns the sustainability of its impact on longer-term changes to teachers' instructional practice and their students' learning outcomes (Antoniou & Kyriakides, 2013; Desimone & Stuckey, 2014; Kennedy, 2016). Several school-based interventions have demonstrated positive, long-term impacts on non-academic outcomes like student-teacher relationships (Okonofua et al., 2022) or attendance (Elias et al., 1991), and there is some evidence that improvements in instructional practice and student outcomes can also be maintained by a school after the program or intervention that produced them has been discontinued. For example, when following up with a random sub-sample of schools from a larger study of professional learning, Timperley and Alton-Lee (2008) found suggestive evidence that student learning gains were sustained for at least one year after the program facilitators withdrew their support; yet, to our knowledge, there are no studies finding causal evidence of the long-term impact of a professional development program on student learning. Other research looks beyond sustainability of program impacts and considers the lasting impact that teachers themselves have on their students' longer-term achievement. As an example, Kane and Staiger (2008) find that the influence a specific teacher has on student learning during any single school year is halved during each subsequent year. Specifically, teacher quality scores from the prior year explain roughly 50% of the variance in student achievement in the subsequent year, reduced to 25% of the variance in the year after that (Kane & Staiger, 2008). Taken together, these findings suggest that the achievement gains derived from professional learning programs can indeed be sustained beyond the program's term. Gains in student learning are likely to degrade over time, however, particularly when students transition into the classrooms of teachers in other schools who may not have participated in the program that brought about those gains.

Leading Educators' Fellowship Program

Leading Educators worked with school leaders and a subset of teachers and other educators in formal and informal leadership roles (including department chairs, Assistant Principals, and instructional coaches, collectively referred to as teacher leaders) to implement its fellowship model during the 2015-16 and 2016-17 school years. Alongside school leaders, teacher leaders participated in leadership development sessions, coaching focused on classroom instruction and leadership of cycles of professional learning, bias awareness and mitigation training, and sessions designed to improve pedagogical and content knowledge. In turn, Leading Educators supported these teacher leaders to take the knowledge and skills they had gained and translate them into cycles of professional learning that they would facilitate for their peer teachers back at their schools each week. Below, we describe the programmatic features and implementation of this fellowship model, and we offer initial evidence on changes to teachers' instructional practice among fellowship participants.

Program Model

The fellowship program aimed to improve teachers' knowledge and practice in content-specific pedagogy and rigorous, standards-based instruction by improving teacher leaders' skills for leading content-team professional learning, coaching others, and improving their school's learning culture. For these improvements to be sustained after the fellowship, strong structures that support teacher leadership and retain high-performing principals, teacher leaders, and teachers need to be developed within LEAs and schools (see outcomes in Figure 1).

Leading Educators' program logic model for achieving these outcomes integrated best-practices related to two distinct but complementary elements: 1) the structure or design of the professional learning approach; and 2) the topics on which the learning was focused. In terms of its structure, Leading Educators' model embodies five key principles of teacher professional learning (see Figures 1 and 2). First, the design exemplified the principle of distributed instructional leadership (e.g., Robinson et al., 2008; Seashore Louis et al., 2010), wherein responsibility for improving instruction lies not with a single individual, but rather is distributed across all the educators within a school and is only achieved through their collective efforts. Consistent with this principle, Leading Educators' logic model begins with ensuring that LEAs and school leaders are committed to investing in and supporting teacher leadership. Second, Leading Educators' model gave teachers access to the content experts who would help them expand or amend their existing knowledge in order to improve their instructional practice (e.g., Crandall, 1983; Timperley & Alton-Lee, 2008). Access to content expertise occurred at two different levels: Leading Educators' staff served as content experts that led learning sessions and provided coaching to teacher leaders throughout the year, who in turn also served as another set of experts, bringing skills and knowledge back to their peers. Third, the professional learning led by teacher leaders, an immediate output of the activities in the logic model, followed a cyclical approach wherein teachers explored a single, consistent topic across multiple sessions (e.g., Crow & Hirsh, 2015; Wiener & Pimentel, 2017), continuously returning to that same topic to examine it from distinct perspectives or process points (e.g., learning, planning, practicing). Fourth, teacher leaders led professional learning at their schools as a collaborative effort, with teachers in grade- and subject-area teams learning alongside one another, offering each other ideas and sharing feedback (e.g., Cohen, 2011; Lynch et al., 2019). Lastly, professional learning was integrated directly into the school day in a job-embedded manner, instead of being offered offsite or outside of normal school hours (Croft et al., 2010; Dana, 2010; Pacchiano et al., 2016).

< Figure 1 about here>

< Figure 2 about here>

The first year of the fellowship program was content agnostic and focused on developing the leadership skills of program participants; however, by the 2016-17 school year, all professional learning was grounded in the specific academic content areas in which teachers were working, which represents the first key aspect of focus within the programming model. This evolution reflected research demonstrating that teachers' instructional practice (e.g., Cohen

& Hill, 2001; Garet et al., 2016; National Academies of Sciences, Engineering, and Medicine, 2020) and impact on student learning (Kraft et al., 2018) improves more when professional learning is content-specific versus content-agnostic. A second aspect of focus involved fellows learning how teachers' biases and expectations can shape the quality of the instruction they provide students and, in turn, the outcomes students are able to achieve (Gershenson et al., 2016; Pajares, 1992). Besides learning about their impact, teachers and leaders also learned specific strategies for interrupting those biases and low expectations in both themselves and others. Third, components such as learning sessions, coaching, and the cycles of professional learning sought not only to increase teachers' content knowledge but also their pedagogical content knowledge, which is a critical competency for improving student learning (Baumert et al., 2010; Hill et al., 2008). Lastly, Leading Educators' model focused building a strong instructional culture within schools and aimed to create a culture of learning for both students and adults alike (e.g., Fullan, 1995; TNTP, 2012).

Program Implementation

Four regions were enrolled in Leading Educators' fellowship program between the 2015-16 and 2016-17 school years: the greater New Orleans and Baton Rouge areas; the greater Kansas City area; the Washington, DC metro area; and the greater Memphis area. Although Leading Educators began as a pilot program in New Orleans in 2008, Kansas City, Memphis, and Washington D.C. were identified as expansion sites based on potential interest, local funding opportunities, and student need.[1] Any district or charter school located within these four geographic areas was considered eligible to participate if at least 70% of the student population qualified for free or reduced price lunch. Leading Educators recruited applicants from schools

[1] Due to data limitations, Memphis has been excluded from the empirical analysis on the implementation and impact of the teacher fellowship model.

that met those criteria through targeted invitations and open information sessions and events. Given the organization's expectation that partnering with school teams would have a deeper impact than working with individuals alone, potential candidates were encouraged to consider the interest of others from their school and to apply as a team. To apply, prospective fellows were required to have at least two years' teaching experience and also to have received an effective or highly effective rating on their previous year's teaching evaluation. All candidates completed the same three phase competency-based selection process, regardless of whether they applied as individuals or teams. The phases included a written application, instructional assessment, and interview and role play. Leading Educators' staff and consultants were required to calibrate on the formal admissions rubric before being eligible to score each of the three application phases. The admissions team identified a minimum cut score for each of these three phases, and candidates were either invited to the next phase or rejected based on those. For teacher leaders applying as part of a team, the cut score was calculated across the team; however, members of a team could be denied individually if the assessors had significant concerns. See Table B5 in appendix B for additional details on the fellow selection process.

Once admitted to the fellowship program, teacher leaders and school leaders attended approximately two weeks of professional development in the summer and four workshops throughout the school year. In addition, teacher leaders met monthly with a leadership coach for one-on-one or team-based support. Some teacher leaders ultimately chose to enroll as individuals, and some school leaders chose to enroll several teacher leaders from their schools, often participating alongside them. In some cases, LEA leaders also participated in some of the programming, although there was not any specialized or differentiated programming for participants working at that level of the education system.

Leading Educators enrolled two separate cohorts of teachers and leaders from these four regions. For the first cohort, who began in 2015-16, much of their first year of programming focused on skills for coaching and leading teams. By the end of their first year in the fellowship, however, Leading Educators had launched its very first content-specific sessions focused on the ELA and math shifts for college and career-readiness standards. Whereas Cohort 2016-17 experienced two days of in-person, content-specific learning on the ELA and math shifts, the initial 2015-16 cohort experienced this same content through an online course. Later in the summer of 2016, both cohorts attended a regional institute where they prepared to design and lead content-focused cycles of professional learning back at their schools. Throughout the following year, both cohorts participated in content-specific workshops focused on evidence of teacher and student learning aligned to the standards and shifts. With respect to coaching, fellows who enrolled as individuals were provided one-on-one coaching; however, fellows who enrolled together as teams participated in group coaching alongside the other teacher leaders in their school.

Fellow Characteristics

There were notable differences in the backgrounds of teachers and leaders who enrolled as individuals compared to those who enrolled as part of a school team. Table 1 displays the share of participants who entered with graduate degrees, their average years of experience in K-12 education, and the summary scores on the three components of the application process (i.e., written application, instructional assessment, and interview) that participants submitted prior to enrolling. Compared to teachers enrolling as teams, participants enrolling as individuals were 1.5 times as likely to have a graduate degree, had one more year of teaching experience, on average, scored higher on each component of the selection process, and were more likely to identify as a person of color.

<Table 1 about here>

Teacher and leader completion of program activities was a challenge in all three regions. Participation in all program offerings was more common for fellows enrolling as individuals than participants who enrolled with their team. This may have been related to differences in recruitment. For instance, team teacher leaders may have been more likely to participate at the request of their principal, leading to more variability in teacher leader interest and confidence in the program. Individual teacher leaders, on the other hand, may have learned about the program on their own and invested significant time in finding funding to attend and securing principal support.

Leading Educators evaluated progress developing teachers and leaders towards its ultimate goal of increasing student achievement using an annual survey assessing intermediate improvement on the proximal indicators in its theory of change: pedagogical content knowledge; beliefs about equity; instructional practice; leadership competencies; and school culture (this annual diagnostic survey is included in Appendix B). Only fellows from Cohort 2016-17 participated in the diagnostic survey, but teachers and leaders from all three regions did complete it in both the spring of 2016 and 2017. Teacher leaders also submitted a performance assessment demonstrating their skills for leading adult learning as well as using student data. These assessments suggested improvement was occurring at all levels along the theory of change, as shown in Figures 3 and 4. Although these results are limited by the sample of teachers responding to both surveys and differ somewhat from the analytic sample included in this study, they do motivate and inform the current study. They are discussed below for the purposes of providing a detailed description of program implementation and its impact on the proximal outcomes of teacher skills, knowledge, beliefs, and culture.

Program Fellow Development

Data from the baseline annual diagnostic survey revealed that teachers were somewhat more advanced in their standards-aligned instructional knowledge and practices in mathematics compared to ELA. Specifically, teachers' average math score at baseline was 77% compared to 65% in ELA. Comparing results between the two survey waves provides strong evidence of changes in teachers' skills, knowledge, beliefs, and culture. For instance, at the individual level, teacher leaders' beliefs about racially and ethnically diverse students, on average, became more equitable following one year of programming (see Figure 3). Moreover, following just one year of programming, the average score on the pedagogical knowledge and practice items increased by approximately 8 percentage points in ELA and math (see Figure 4).

In terms of interpersonal or social outcomes, teacher leaders reported modest growth in the instructional cultures within their schools as well as how frequently they exercised specified leadership competencies. By drawing on those improved leadership skills, teacher leaders effectively transferred their own learning back to their schools through cycles of ongoing professional learning. As evidence of this, 95% of fellows submitted artifacts demonstrating they had led at least one cycle of professional learning, whereas 88% led two cycles of professional learning. More of these cycles focused on ELA versus math, with 54% focusing on ELA, 34% on math, and the remaining 12% were judged to be focused on content-agnostic topics. Principals and teacher leaders rated these cycles as valuable, with 73% of teacher leaders strongly agreeing or agreeing that these cycles were highly effective and enabled them to raise student achievement in their schools.

< Figure 3 about here>

< Figure 4 about here>

In sum, findings from the extant research suggest that the duration of professional learning programs, the specific components or strategies they employ, and how those are combined together into a coherent program all warrant greater consideration. Data regarding the implementation of Leading Educators' model and the changes observed among the proximal outcome indicators provide additional rationale for examining this program's impact on student achievement. Accordingly, in the next section, we describe the data and empirical approach for estimating the impact of participation in the teacher fellowship model on students' grade-level proficiency in ELA and math. By leveraging both program specific and publicly available data in the context of a differences-in-differences and event study approach, we examine the impact of Leading Educators' teacher professional learning program on student achievement both during and after the partnership.

Data & Sample

We construct a school*grade*year dataset for the 2009-10 through 2018-19 school years. The dataset contains grade-level administrative data for grades 3-8 from 1715 regular schools in Louisiana, Washington D.C. and Kansas City.[2] The data contain detailed information at both the school*grade*year and school*year levels. At the school*grade*year level, data include: student ethnicity, gender, enrollment count, and proficiency scores. At the school*year-level, data include: the number of students receiving free and reduced lunch status, magnet indicator, charter indicator, number of full-time equivalent staff, and the number of students in special

[2] We excluded special education schools, vocational education schools, alternative education schools and common core data reportable programs. Data from Memphis was collected but excluded from implementation and impact analyses because student proficiency data from one treatment year as well as a key student demographic data (i.e., ELL status) were missing for most years.

education programs and English language learners (ELL) programs. Except for student enrollment in special education and ELL programs, all other demographic indicators were obtained from the U.S. Department of Education's Common Core of Data (CCD).

State assessment data and ELL and special education data sources varied by region, as follows. For Washington DC, we obtained math and ELA proficiency rates at the school*grade*year level from 2012 to 2019 from the Comprehensive Assessment System (CAS) and the Partnership for Assessment of Readiness for College and Careers (PARCC) District of Columbia assessment. Schools transitioned from CAS to PARCC in the school year 2014-2015. The number students enrolled in special education programs and English learners' programs that took the state assessment each year was used as a proxy for the total number of students in those programs. Schools with less than 10 students in those programs were reported as n<10 and we replaced those with zeros in the analytical file. For Louisiana, we obtained math and ELA proficiency rates at the school*grade*year level from 2009 to 2019 from the Louisiana Educational Assessment Program (LEAP) state assessment. The 2012-2013 scores were reported only at the LEA level during a transition to a Common Core aligned test; therefore, the scores from those years are missing from the analytical file. The Louisiana Department of Education reports special education enrollment for preschool students only. The number of preschool students enrolled in special education programs was used as a proxy for total special education enrollment. For Kansas City, we obtained math and ELA proficiency rates at the school*grade*year level for the school years 2010-2019 from the Northwest Evaluation Association (NWEA) MAP assessment and the Missouri School Improvement Program (MSIP) state assessment[3].

[3] We thank Dr. Cameron Anglum from St. Louis University in Missouri for sharing the Kansas City data.

Primary sourced data (i.e., data from Leading Educators) included a list of schools enrolled in the fellowship model during school years 2015-16 to 2016-17, the number of teachers enrolled by school and program year, number of years enrolled, an indicator for schools that had one or more LEA leaders enrolled in the program, and an indicator that differentiates teachers that enrolled individually in the fellowship program from teachers that enrolled as a school-based team alongside their school principals. We excluded from the analysis file schools with teachers that received support from Leading Educators and districts that partnered with Leading Educators to receive district level support before school years 2015-16 and during school years 2017-18 to 2018-19. Finally, and for descriptive purposes, we collected data about teachers' level of education, years of experience in K-12 education, and summary scores on the three components of the application process participants completed to enroll, which consisted of an online application, an interview day, and an instructional assessment that included a classroom observation and submission of student data. The application process varied across regions and years and scores were missing for some individuals.

School Characteristics

Complete data for all 10 years covered by this study were not available for any school, so the characteristics and number of schools from each year included in the analyses varied according to the question being examined. The mean characteristics of schools in the treatment and comparison samples across the 10 years of the study are displayed in Table 2. Because the analytical sample used for the ELA outcomes analysis had only 39 more cases (+ 0.5%) than the math sample, the characteristics of the ELA sample are used to describe both samples. On average, there were 529 schools in the comparison group and 29 schools in treatment group.

Across these three regions, the treatment schools had rates of student proficiency similar to the comparison schools in both math and ELA across the 10 years. During the 6 years that preceded the treatment, the average proficiency of the schools in the fellowship program was 2 percentage points higher in both math and ELA. In contrast with the comparison schools, the treatment schools had a statistically significantly higher proportion of students who identified as Hispanic, students who identified as Black, students who were English language learners, and students who were identified as neurodiverse learners participating in special education programs. It also has a statistically significantly lower proportion of students who identified as White or Asian and no statistically significant difference in the percentage of students who qualified for free and reduced lunch. The treatment sample had a statistically significantly higher proportion of charter schools, a statistically significantly lower proportion of magnet schools, and a statistically significantly lower student-teacher ratio.

< Table 2 about here>

School characteristics varied modestly across regions and to a relatively smaller degree between the treatment and comparison schools within regions (see Table A1). In terms of race and ethnicity, schools in Louisiana and Washington DC had considerably higher percentages of students identifying as Black compared to Kansas City. Instead, Kansas City had the highest percentage of students identifying as Hispanic. Schools in Louisiana had the fewest students qualifying for free or reduced price lunch, receiving special education services, or designated as English learners. The proportion of schools functioning as charters varied considerably both across and within regions, with all treatment and comparison Washington DC schools organized as charters, whereas only 12% of comparison schools in Kansas City were designated as charters. In terms of school staffing, treatment schools in both Louisiana and Washington DC had higher numbers of full-time teachers than did their respective comparison schools, whereas schools in Kansas City were much more comparable on this particular dimension. Finally, the ratio of students to teachers was fairly consistent across and within regions, ranging from roughly 12 to 16 pupils per teacher.

Program Characteristics

In the Leading Educators' Teacher Fellowship Model section above we described the vision, program components and focus of the fellowship model that were common in all schools and region. In this section we highlight four program components that varied at the teacher and school levels. We exploit this variation to explore heterogeneity in the impact of the fellowship model on student achievement across the following four program components (see Figure 5).

Saturation. This is the proportion of a school's full-time teachers who participated in the fellowship model during the 2015-16 and 2016-17 school years. A higher proportion of teachers enrolled in the program could lead to more students exposed to trained teachers during the fellowship years and higher outcomes for students. At the same time, the large time commitment required of teachers and the leadership component of the program could mean that a smaller and more experienced group of teachers would benefit more from the fellowship model than a larger group. A larger group of teachers also translated into less individualized coaching time in the context of this fellowship model, which would reduce the effect on teacher and students.

Figure 5 shows the distribution of saturation across school-grades. While the mean proportion of teachers supported in a school across the two years was 6%, the figure shows a high proportion of schools with saturation levels of around 2% to 3%. Because LEA leaders enrolled in the fellowship program are not included in the calculation of saturation, schools with zero saturation represent schools in which only LEA leaders enrolled and schools with low saturation could include schools with additional participation of LEA leaders.

Enrollment type. This variable indicates how the teachers decided to apply for the fellowship program, which are likely correlated with unobservable teacher characteristics like motivation and leadership skills. One type of enrollment involved teachers independently learning about the program, applying for funds, and enrolling as individuals. In the other type of enrolment, school leaders learned about the program and secured funds and then appointed a group of teachers from their schools to participate in the program. It could be that teachers in this second group enrolled in part to comply with their principal's request and had less motivation and readiness than teachers who applied as individuals. Figure 5 shows that more teachers enrolled as individuals compared to teachers who enrolled as part of a school team. There was a total of 67 school*grades across 17 schools where teachers enrolled as individuals and 47 school*grades across 12 schools where teachers enrolled as teams. As explained in the program saturation section above, program type and program saturation cannot be separated, but results from the analysis of the effect of these two program components could help us formulate stronger hypotheses about the effect of program saturation and teacher selection.

Duration. This is the amount of time that teachers were enrolled in the program. While the program was designed to have exactly one or two years of duration, depending on the region, some teachers withdrew before the end of the program. Using data about the exact enrollment and withdrawal days we created a continuous variable representing the amount of time in years and fractions of years that teachers remained enrolled in the program. While the ideal amount of time teachers should be exposed to continuous professional development has important budgeting and practical implications, to our knowledge, this is the first study examining variation in student achievement with varying duration of external support. Figure 5 shows the distribution of program duration across school-grades in the fellowship schools. As expected based on program design, the majority of the participants were enrolled between 1 and 2 years. Program duration of less than one year or between more than one and less than two years is the result of teachers withdrawing from the program, which could be correlated with teacher selection into the program and could be a source of bias if the reasons for withdrawal were not random (e.g., due to unobserved teacher skills and attitudes). On average, teachers enrolled in the program for 1 year and three months[4].

Local educational agency leader participation. While the fellowship model had no specialized programing or coaching for LEA leaders, some district leaders decided to enroll alongside teachers and school leaders, and a few enrolled instead of the teachers. This study did not collect information about how the LEA leaders implemented what they learned back in their educational agency, but we believe LEA leaders enrolling alongside teachers may have facilitated the creation of an aligned vision for instruction, an important system condition connected to gains in student achievement (Fullan & Quinn, 2015; Cobb et al., 2019). We defined the level of participation on the part of LEA leaders by three categories. The first category represents schools where only LEA leaders and no teachers enrolled in the program, with only 6 school*grades across 2 schools represented in this category as shown in Figure 5. The second category represents schools where LEA leaders enrolled alongside teachers, with 17 school*grades across 7 schools represented in this category. Finally, in most schools, only teachers enrolled. There were 82 school*grades across 20 schools in this category.

<Figure 5 about here>

[4] See Figure A1 in the appendix for a distribution of Duration by region

Empirical Approach

We rely on difference-in-differences and event study strategies to estimate the causal effect of Leading Educators' intervention during the 2015-16 and 2016-17 school years. We also examine heterogeneity by program saturation and duration and the potentially differential effect for schools where LEA leaders participated alongside teachers and for teachers that enrolled as teams. We describe our main and heterogeneity models, below.

Difference-in-Differences and Event Study

We rely on a two-way fixed effects (TWFE) difference-in-differences approach to estimate the average effects of the teacher fellowship model. This approach compares the change in student proficiency (in math and ELA) among schools participating in Leading Educators' fellowship model to changes in student proficiency among non-participating schools in the same district and grades in the years prior to the start of the program (2009-10 through 2014-15) and the years during and after the program (2015-16 through 2018-19). We also present results from a non-parametric event study model, which enables an assessment of the parallel-trends assumption underlying the difference-in-differences approach. In the context of the event study approach, the year-specific effects following the introduction of the teacher fellowship model will provide insight into whether any program effects accumulate or fade over time. Further, the year-specific effects prior to the intervention provide a test of the parallel-trends assumption; indeed, any statistically significant effects before the treatment period will signal pre-treatment differences in outcome trends, a violation of an important assumption of the parallel-trends assumption.

We specify the difference-in-differences model as follows:

$$Y_{gst} = \alpha + \gamma(Treat_{st}) + X_{st} + Z_{gst} + \theta_{gdt} + \delta_s + \varepsilon_{gst} \quad (1)$$

where Y denotes the percentage of students that are proficient or advanced in ELA (or math) for grade g in school s during school year t. The coefficient γ of $Treat$ is the estimate of the main treatment effect. $Treat$ denotes whether school s received treatment in school year t. X is a vector of school level characteristics, Z is a vector of grade level characteristics, θ_{gdt} and δ_s denote grade*LEA*year fixed effects and school fixed effects, respectively, and ε_{gst} denotes the random error term. We cluster the standard errors at the school level (to account for the correlation among students attending the same school).

We specify the event study model as follows:

$$Y_{gst} = \sum_{j=-6}^{j=3} \gamma_j \left(Treat_{st+j}\right) + X_{st} + Z_{gst} + \theta_{gdt} + \delta_s + \varepsilon_{gst} \tag{2}$$

where the γ_j coefficients of $Treat$ denote the pre- and post-treatment effects for students supported by teachers enrolled in the program. All other variables are defined as in Eq. (1), and we cluster the standard errors at the school level.

In both models, and to make sure the groups we are comparing are equivalent, we start by including in the regression equation all observable characteristics of the schools and grades in the sample believed to be correlated with the intervention and the outcomes. Data on these observable characteristics include the percentage of students receiving free or reduced price lunch, special education, and English language learner programs; the percentage of each ethnicity category, a charter school indicator, a magnet school indicator, and the school's student-teacher ratio. There may be many other characteristics that are not easy to observe but we know commonly vary across levels, including within: grade (e.g., neurotypical age maturity), schools (e.g., school culture, professional development conditions), LEAs (e.g., district policies, system-level conditions), and school year (e.g., catastrophic events, policy changes, test changes, etc.). We can account for these group- and time-specific differences using fixed effects. The school fixed effects account for all unobservable differences between schools, and the grade*LEA*year fixed effects account for all unobserved differences among grades in the same year and the same school district (i.e., LEA).

Exploring Heterogeneity by Program Implementation Characteristics

To further understand the influence of different components of the fellowship model on student outcomes, we took advantage of natural variation in four characteristics of the program model. Specifically, we estimated the differential effect for schools where teachers remained enrolled in the program above the average time and for specific lengths of enrollment time, for schools with above average proportion of teachers in the fellowship program (or above average saturation) and for specific levels of saturation, for schools that had LEA leaders enrolled in the program alongside teachers compared to schools that did not, and finally for schools where teachers enrolled as a team compared to teachers that enrolled as individuals. Equations (3) to (10) present models to estimate differential effects by program characteristics by adding an interaction term for each program characteristic to equations (1) and (2), as follows:

$$Y_{gst} = \alpha + \gamma(Treat_{st}) + \beta(Treat_{st})(Saturation_s) + X_{st} + Z_{gst} + \theta_{gdt} + \delta_s + \varepsilon_{gst} \tag{3}$$

$$Y_{gst} = \sum_{j=-6}^{j=3} \gamma_j\left(Treat_{st+j}\right) + \sum_{j=-6}^{j=3} \beta_j\left(Treat_{st+j}\right)(Saturation_s) + X_{st} + Z_{gst} + \theta_{gdt}$$
$$+\delta_s + \varepsilon_{gst} \tag{4}$$

where $Saturation$ denotes the average proportion of teachers enrolled in the program between the 2015-16 and 2016-17 school years (where the proportion is the count of program participants in school s divided by the number of full-time equivalent teachers in school s). In equations (3) and (4), the effect of the program on treated schools is a linear combination of $\hat{\gamma} +$

$\hat{\beta}(Saturation_s)$. All other variables are defined as in Eq. (1), and we cluster the standard errors at the school level.

$$Y_{gst} = \alpha + \gamma(Treat_{st}) + \beta(Treat_{st})(Team_s) + X_{st} + Z_{gst} + \theta_{gdt} + \delta_s + \varepsilon_{gst} \tag{5}$$

$$Y_{gst} = \sum_{j=-6}^{j=3} \gamma_j\left(Treat_{st+j}\right) + \sum_{j=-6}^{j=3} \beta_j\left(Treat_{st+j}\right)(Team_s) + X_{st} + Z_{gst} + \theta_{gdt} + \delta_s + \varepsilon_{gst} \tag{6}$$

where $Team$ is an indicator that equals 1 if teachers enrolled as a team and 0 if teachers enrolled as individuals. In equations (5) and (6), the effect of the program on treated schools equals $\hat{\gamma} + \hat{\beta}$ when $Team = 1$, and $\hat{\gamma}$ when $Teams = 0$.

$$Y_{gst} = \alpha + \gamma(Treat_{st}) + \beta(Treat_{st})(TimeEnrolled_s) + X_{st} + Z_{gst} + \theta_{gdt} + \delta_s + \varepsilon_{gst} \tag{7}$$

$$Y_{gst} = \sum_{j=-6}^{j=3} \gamma_j\left(Treat_{st+j}\right) + \sum_{j=-6}^{j=3} \beta_j\left(Treat_{st+j}\right)(TimeEnrolled_s) + X_{st} + Z_{gst} + \theta_{gdt} + \delta_s + \varepsilon_{gst} \tag{8}$$

where $TimeEnrolled$ denotes the fractions of years (maximum 2 years) during which teachers remained enrolled in the program. In equations (7) and (8), the effect of the program on treated schools is a linear combination of $\hat{\gamma} + \hat{\beta}(TimeEnrolled_s)$. For each quartile of enrollment time in the sample, we use the corresponding mean of enrollment time by interacting the treatment indicator with the mean enrollment time for each quartile.

$$Y_{gst} = \alpha + \gamma(Treat_{st}) + \beta(Treat_{st})(TeachLEA_s) + X_{st} + Z_{gst} + \theta_{gdt} + \delta_s + \varepsilon_{gst} \tag{9}$$

$$Y_{gst} = \sum_{j=-6}^{j=3} \gamma_j\left(Treat_{st+j}\right) + \sum_{j=-6}^{j=3} \beta_j\left(Treat_{st+j}\right)(TeachLEA_s) + X_{st} + Z_{gst} + \theta_{gdt} + \delta_s + \varepsilon_{gst} \tag{10}$$

where $TeachLEA$ is an indicator that equals 1 if both school teachers and LEA leaders were enrolled in the program and 0 if only teachers or only leaders enrolled. In equations (9) and (10), the effect of the program on treated schools equals $\hat{\gamma} + \hat{\beta}$ when $TeachLEAs = 1$, and $\hat{\gamma}$ when $TeachLEAs = 0$. These leaders attended the same sessions that teachers attended, and we believe that they could have provided additional coaching to teachers, or that it provided them with an opportunity for aligning priorities and setting common expectations. On the other hand, schools in the same LEA that did not enroll in the program could have received some spillover treatment from their enrolled LEA leaders.

Results

Main Effects

Table 3 presents the main effects of the teacher fellowship model on student math and ELA proficiency. On average, pooled difference-in-differences results indicate that student proficiency in math and ELA improved by 8.5 and 5.3 percentage points, respectively, as a result of a school's participation in the teacher fellowship model. Event study results provide evidence of no pre-treatment effects in any of the six pre-treatment years (see Table 3 and Figure 6). Further, the post-treatment effects in math show positive and statistically significant differences in the two years immediately after starting the treatment, marginally significant effects three years after and not discernible effects four years after. The post-treatment effects in ELA show marginally significant effects one and two years after initial treatment and no discernible effects three and four years after.

< Figure 6 about here>

< Table 3 about here>

Table 4 summarizes these results in four ways: (a) percentage point change in student achievement; (b) percent change in the share of proficient students; (c) change in standard deviation units; and (d) relative magnitude of effect sizes benchmarked against the effects of other education interventions with strong evidence of positive effects (Kraft, 2020; Lipsey et al., 2012). Based on these different characterizations of the main effects, there is strong and consistent evidence of a positive impact of the teacher fellowship model on student proficiency in both ELA and math across the partnership years and in the years subsequent to fellowship participation. The size of the standardized effects range between 0.2 and 0.6 standard deviations; according to IES benchmarks for interventions that use broad scope standardized tests as the outcome measure (Lipsey et al., 2012), these effects are considerably greater than the average effect size of 0.08 for educational interventions at the elementary school levels and above the average effect size of 0.15 found in middle school interventions. In Kraft (2020) framework for interpreting effect sizes, an effect size above 0.20 should be considered large. The medium to large, but not statistically significant effects found one and two years after the conclusion of the fellowship are consistent with research that have found that effects tend to be modest or not significant the further away the assessments are from the intervention (Angrist et al., 2016, Bailey et al., 2017; Kraft, 2020; Ruiz-Primo et al., 2002).

The smaller effects found in ELA are also consistent with findings from many empirical studies showing stronger effects on math scores than reading scores (Hansen et al., 2018; Cronin et al., 2005). Program evaluation data suggested a similar increase in teacher leaders' knowledge and practice scores in ELA and math, but teacher leaders' baseline math knowledge and practice was also higher (see Figure 4). This could suggest that math teachers enrolled in the program were more effective or experienced teachers and more likely able to benefit from the program without compromising planning time.

< Table 4 about here>

Effects by Program Saturation

Table 5 summarizes the effects at each of four quartiles of saturation. Effects are larger at the lowest level of saturation with the largest effects in the second year of treatment (0.15 in ELA and 0.21 in math). Effects are slightly smaller in the second quartile when 4% of teachers in a school are enrolled, with second year effects of 0.09 for ELA and 0.16 for math. The effect becomes non-statistically significant when saturation increases to 6% for all years except for the second program year in math when the effect 0.11 and highly significant. Effect becomes either significantly negative or non-significant in the fourth quartile when the saturation mean is 12%. The largest negative effects in the fourth quartile are in the post treatment years when all support had been withdrawn (-0.25 for ELA one year after treatment and -0.22 for math two years after treatment).

The negative results for teacher saturation in the fourth quartile may seem counterintuitive but may be explained by differences in coaching dosage and enrollment method and has important implications for program design. The coaching time received by teachers in lower saturation schools was larger and more personalized as individuals received one-on-one coaching and teams received team coaching. These findings may support the importance of differentiating programming supports according to the baseline experience of participants, as well as the importance of building investment at all levels. Further, the two program models operating simultaneously differed in recruitment and enrollment method which are both theoretically and empirically correlated with program saturation[5]. Whereas individually enrolling teachers chose to apply as individuals and were responsible for locating funding for the program on their own, school leaders chose to enroll a team from their school, found the funding, and then selected and recruited teacher leaders to participate. It may be that in the individual teacher leader model, teacher leaders felt a stronger sense of investment and agency. In the school team model, some teacher leaders may have felt obligated to attend and even resentful of the additional demands on their time, and as a result may have been less likely to participate or engage. It may also be that the greater number of teacher leaders a school leader chose to send reduces the likelihood that those teachers are prepared to engage in the program. Baseline skills, as measured by years of experience, graduate degrees, and scores on the application, interviews, and classroom observations were notably lower for these larger groups than it was for individuals. This may have led to a range in teacher readiness to adopt new practices and share them with their peers. Another potential explanation is that with lower levels of skills and experience, team-enrolled teacher leaders were more negatively impacted by time taken away from instructional planning and preparation. Given the lack of access to aligned curricular materials in particular, teachers would have been pulled between creating or adapting their own materials, improving their practice, and leading their peers. Findings from the heterogeneity analysis of program type support these hypotheses and are presented next.

< Table 5 about here>

Effects by Enrollment Type

Table 6 shows the effects for the two types of program enrollment. The results show no clear differential relationships between the type of enrollment and the program effects. In

[5] Pearson's correlation between program saturation and program type is 0.5 and statistically significant at the 1% level.

general, we do find that there are statistically significantly greater effects from the individual type schools compared to the team type schools. The effects across the four posttreatment years are positive and significant for the individual enrollment teachers in both ELA (0.09) and math (0.1) and non-statistically significant for the team enrollment teachers. The year-specific effects show no clear pattern between the two groups with some post treatment years with non-significant effects and some with positive effect in both groups. The somewhat larger results for the individual enrollment teachers provide support for the hypothesis that the positive results at lower saturation levels could be influenced by teacher selection. Teachers that enrolled as individuals could have been more motivated and prepared than teachers selected by school leaders to participate in a team. The fact that the effects for individual enrollment are not consistently higher than the effects when teachers enrolled as teams is also an indication that teacher selection is not the only program component that matters, and most likely individualized coaching and teacher collaboration play an important role in driving student outcomes.

<Table 6 about here>

Effects by Program Duration

Table 7 shows the effects for different durations of teacher enrollment in the fellowship program. Each duration level represents the mean value at each of four quartiles of program duration. Effects when teachers enrolled for 15 months or less are mostly non-statistically significant with some statistically significant negative effects when teachers enroll for 5 months. Notably, schools where teachers enrolled for the maximum of two years experienced positive and significant effects that were sustained two years after completions of the fellowship for math (0.16) and one year after for ELA (0.1). Most of the differences in program duration are the product of program design. The fellowship model enrolled two cohorts of teachers, the first cohort for two years and the second cohort for one year. Teachers that enrolled for less than a year or for a period longer than a year but shorter than two years, withdrew before completing the program. Comparing schools with participants that completed the intended amount of time (one or two full years) help us make conclusions about the most effective type of program. Participants who received two years of programming received one year of content-focused development and one year of leadership development, while those who received one year of programming received only the content-focus development. The difference in the effects of the program that was designed for two years compared to one year is shown in the last column of Table 7. Schools where teachers enrolled for two years increased student proficiency by 0.16 percentage points more compared to schools where teachers enrolled for one year across the two years of programing and the two follow up years combined in both ELA and math. The year specific effects are also significantly higher for the two-year program in all years except for year one in ELA and years one and two in math. This means that the two-year program had a larger lasting effect on teachers. A non-statistically significant difference in year one between one- and two-year program duration, which include teachers that stayed in the program for the entirety of the intended program duration, provides some support to the assumption that the greater effect was related to the increased time and programing and not to differences in teacher selection. These results suggest teachers benefit from ongoing, sustained development and for the combination of content-focused pedagogical and general leadership development.

<Insert **Table 7** Heterogeneous Effects of Fellowship Model, by Program Duration about here>

Effects by LEA Leader Participation

Table 8 shows the effects for the schools in the sample where LEA leaders enrolled alongside teachers and for the schools in the sample where either teachers or only LEA leaders enrolled. The results show that only when LEA leaders enrolled alongside teachers, the effects are positive and statistically significant across the four post-treatment years (0.07 for ELA and 0.1 for math), with larger significant effects in the second year of treatment (0.11 for ELA and 0.16 for math). These results strongly support the idea that instructional alignment between teachers, school leaders and teacher leaders matter and could be key for the sustainability of program effects. As with teacher selection, we cannot separate LEA leader selection into the program from the program effects, but selecting and training a few highly effective system leaders could be a much easier task than selecting and training hundreds of skilled teachers. More research is needed to understand the specific mechanisms by which LEA leader participation could be related to improvements in student achievement outcomes.

< Table 8 about here>

Conclusion

School districts nationwide invest heavily in developing their educators, annually allocating 3-5 percent of their budgets to fund teacher professional development initiatives (Kraft et al., 2018). For example, in the 2016-17 school year, expenditures on public elementary and secondary schools in the United States totaled $739 billion, or $14,439 per public school student (National Center for Education Statistics, U.S. Department of Education); thus, the annual cost of teacher professional development to districts was $22-37 billion, or $433-722 per public school student. Yet, teacher professional development programs have historically failed to improve teachers' instructional practice or student achievement (Kraft et al., 2018), and there is limited evidence on the specific features of professional learning programs that are most salient for improving student achievement across contexts (Hill et al., 2021). Not only have recent meta-analyses been unable to shed light on the optimal duration of professional learning (Didion et al., 2020, Lynch et al., 2019), but the lack of clear information on the resources and factors necessary for effective implementation may account for a large part of the difficulty in scaling research-based strategies (Hollands et al. 2016). This paper adds to the literature by describing in detail the implementation of a teacher professional learning effort and by identifying characteristics that influence the effect on student achievement and the extent to which that effect is sustained over time.

Findings from this paper are consistent with other rigorous evaluations of educator professional learning programs that have shown a positive impact on student learning through the development of both teachers (e.g., Dolfin et al., 2019; Young et al., 2017) as well as school leaders (e.g., Gates et al., 2019). It provides strong evidence that Leading Educators' teacher leader fellowship model increased student achievement in math with different levels of support and program characteristics. Additionally, these effects are larger and are sustained for at least two years after the program ends, particularly when the program targets a small proportion of selected teachers or leaders in each school to implement its professional learning intervention, when teachers remain in the program for two years compared to shorter durations, or when LEA leaders participate alongside teachers compared to teachers participating without LEA leaders. The ELA results follow the same pattern, though the overall ELA results were marginally significant. Additionally, as these effects are at the school level, these findings suggest improvements in teacher effectiveness were achieved across the school. As Chetty and colleagues (2014) found, improving teacher effectiveness has long-term outcomes on students' college attendance and lifetime earnings even after the initial test score gains fade out.

The effects were achieved with a relatively low-cost investment of $75 per student per year, primarily covering personnel costs for designing and leading the professional development and coaching (for additional details on cost, see Table A6). For example, Kraft (2020) gathered per-pupil cost information for 68 education interventions to propose per-pupil cost benchmarks where less than $500 is considered low, $500 to less than $4,000 is considered moderate, and greater than $4,000 is considered large. Kraft contrasts a higher-cost tutoring intervention (effect size 0.23 SD with an annual cost of more than $2,500 per student) with universal free breakfast (effect size 0.09 SD with a cost of $50-200 per student) and argues that the impact of universal free breakfast may be considered more impressive given the relatively low cost. Foster and colleagues (2013) also reviewed the cost effectiveness of teacher professional development relative to other interventions, and reported similar ranges in cost, with one professional development program producing a modest effect (0.03 SD) on math achievement at $44 per student and a master's degree program at $702 per student producing larger effects (0.22 SD) on math achievement. The Leading Educators Fellowship's approach to job-embedded, ongoing learning resulted in whole school effects at a low cost by targeting a subset of teachers and leaders per school, enabling cost savings compared to a direct to teacher or student approach. When compared with these frameworks, this program produced and, in some cases, sustained, medium to large effects at low cost.

While this study provides new evidence on the importance of assessing the readiness of schools and leaders to engage in an initiative for ongoing, school-based learning, there are remaining limitations that the analysis was unable to address. First, the study was unable to control for unobserved differences that led some leaders and schools to seek out the program. Additionally, the results of the event study by quartile of saturation suggested strong, enduring effects could be achieved by a small group of teacher leaders, but a likely higher baseline knowledge and skills of these teacher leaders may indicate these effects are not entirely due to the program. A better understanding of the role of baseline knowledge, skills, and conditions could support the development of differentiated approaches that address all schools' readiness for shifting teaching and learning. As the heterogeneous effects of saturation level may in part be explained by the importance of teacher and leader motivation in seeking out learning opportunities, this study also supports further investigation into how teacher motivation influences the rate at which schools implement and sustain new practices. Further understanding of these factors can help a variety of school systems and nonprofit partners make strategic design choices that maximize the value of professional learning initiatives for all schools in an effort to more widely distribute the benefits of rigorous, job-embedded teacher professional development.

References

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

Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Pathak, P. A., & Walters, C. R. (2016). Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics, 34*(2), 275-318.

Antoniou, P., & Kyriakides, L. (2013). A dynamic integrated approach to teacher professional development: Impact and sustainability of the effects on improving teacher behaviour and student outcomes. *Teaching and Teacher Education*, *29*(1), 1-12.

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness, 10*(1), 7-39.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., ... & Tsai, Y. M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, *47*(1), 133-180.

Boulay, B., Goodson, B., Olsen, R., McCormick, R., Darrow, C., Frye, M., ... & Sarna, M. (2018). *The Investing in Innovation Fund: Summary of 67 evaluations. Final Report. NCEE 2018-4013.* Washington, DC: U.S. Department of Education.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). Explaining the short careers of high-achieving teachers in schools with low-performing students. *American Economic Review*, *95*(2), 166-171.

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.

Cobb, P., Jackson, K., Henrick, E., & Smith, T. M. (2020). *Systems for instructional improvement: Creating coherence from the classroom to the district office*. Cambridge, MA: Harvard Education Press.

Cohen, D. K. (2011). *Teaching and its predicaments.* Cambridge, MA: Harvard University Press.

Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works.* New Haven, CT: Yale University Press.

Coolahan, J. (2002). Teacher education and the teaching career in an era of lifelong learning. *OECD Education Working Papers, 2*(1), 1-39.

Crandall, D. P. (1983). The teacher's role in school improvement. *Educational Leadership, 41*(3), 6-9.

Croft, A., Coggshall, J. G., Dolan, M., & Powers, E. (2010). *Job-embedded professional development: What it is, who is responsible, and how to get it done well.* Washington, DC: National Comprehensive Center for Teacher Quality.

Cronin, J., Kingsbury, G. G., McCall, M. S., & Bowe, B. (2005). *The impact of the No Child Left Behind Act on student achievement and growth (2005 edition).* Portland, OR: Northwest Evaluation Association.

Crow, T., & Hirsh, S. (2015). Learning team cycle of continuous improvement. *Tools for Learning Schools, 19*(1), 1-7.

Dana, N. F. (2010). Teacher quality, job-embedded professional development, and school-university partnerships. *Teacher Education and Practice*, *23*(3), 321-325.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1), 1-44.

Day, C. (1999). *Developing teachers: The challenges of lifelong learning*. London: Routledge.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181-199.

Desimone, L. M., & Stuckey, D. (2014). Sustaining teacher professional development. In L. E. Martin, S. Kragler, D. J. Quatroche, & K. L. Bauserman (Eds.), *Handbook of professional development in education: Successful models and practices, PreK–12* (pp. 467–482). New York, NY: Guilford.

Didion, L., Toste, J. R., & Filderman, M. J. (2020). Teacher professional development and student reading achievement: A meta-analytic review of the effects. *Journal of Research on Educational Effectiveness, 13*(1), 29-66.

Dolfin, S., Richman, S., Choi, J., Streke, A., DeSaw, C., Demers, A., & Poznyak, D. (2019). *Evaluation of the Teacher Potential Project*. Washington, DC: Mathematica.

Elias, M. J., Gara, M. A., Schuyler, T. F., Branden-Muller, L. R., & Sayette, M. A. (1991). The promotion of social competence: Longitudinal study of a preventive school-based program. *American Journal of Orthopsychiatry, 61*(3), 409–417.

Foster, J. M., Toma, E. F., & Troske, S. P. (2013). Does teacher professional development improve math and science outcomes and is it cost effective?. *Journal of Education Finance*, 255-275.

Fullan, M. (1995). The school as a learning organization: Distant dreams. *Theory into Practice, 34*(4), 230-235.

Fullan, M., & Quinn, J. (2015). *Coherence: The right drivers in action for schools, districts, and systems*. Thousand Oaks, CA: Corwin Press.

Garet, M. S., Heppen, J. B., Walters, K., Smith, T. M., & Yang, R. (2016). *Does content-focused teacher professional development work? Findings from three Institute of Education Sciences Studies*. Washington, DC: U.S. Department of Education.

Gates, S. M., Baird, M. D., Doss, C. J., Hamilton, L. S., Opper, I. M., Master, B. K., ... & Zabar, M. A. (2019). *Preparing school leaders for success: Evaluation of New Leaders' Aspiring Principals Program, 2012-2017*. Santa Monica, CA: RAND Corporation.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, *52*, 209-224.

Hansen, M., Levesque, E. M., Valant, J., & Quintero, D. (2018). *Brown Center report on American education: Trends in NAEP math, reading, and civics scores.* Washington, DC: Brookings Institution.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 466-479.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (No. w11154). Cambridge, MA: National Bureau of Economic Research.

Hawley, W., Jordan, J. I. & Landa, M. (n.d.). *Common beliefs survey: Teaching racially and ethnically diverse students.* Montgomery, AL: Learning for Justice. Retrieved from www.learningforjustice.org on October 28, 2021.

Henry, G. T., Bastian, K. C., & Fortner, C. K. (2011). Stayers and leavers: Early-career teacher effectiveness and attrition. *Educational Researcher*, *40*(6), 271-280.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, *39*(4), 372-400.

Hill, H. C., Corey, D. L., & Jacob, R. T. (2018). Dividing by zero: Exploring null results in a mathematics professional development program. *Teachers College Record*, *120*(6), 1-42.

Hill, H., Papay, J., Schwartz, N., Johnson, S., Freitag, E., Donohue, K., ... & Williamson-Zerwic, B. (2021). *A learning agenda for improving teacher professional learning at scale.* Providence, RI: Research Partnership for Professional Learning.

Hollands, F. M., Kieffer, M. J., Shand, R., Pan, Y., Cheng, H., & Levin, H. M. (2016). Cost-effectiveness analysis of early reading programs: A demonstration with recommendations for future research. *Journal of Research on Educational Effectiveness*, *9*(1), 30-53.

Jackson, C. K. (2014). Teacher quality at the high school level: The importance of accounting for tracks. *Journal of Labor Economics*, *32*(4), 645-684.

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

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

Koedel, C. (2008). Teacher quality and dropout outcomes in a large, urban school district. *Journal of Urban Economics*, *64*(3), 560-572.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241-253.

Kraft, M. A., 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.

Lee, S. W. (2018). Pulling back the curtain: Revealing the cumulative importance of high-performing, highly qualified teachers on students' educational outcome. *Educational Evaluation and Policy Analysis*, *40*(3), 359-381.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., ... & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms.* Washington, DC: National Center for Special Education Research.

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.

National Academies of Sciences, Engineering, and Medicine. (2020). *Changing expectations for the K–12 teacher workforce: Policies, preservice education, professional development, and the workplace.* Washington, DC: National Academies Press.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257.

Okonofua, J. A., Goyer, J. P., Lindsay, C. A., Haugabrook, J., & Walton, G. M. (2022). A scalable empathic-mindset intervention reduces group disparities in school suspensions. *Science Advances*, *8*(12), 1-10.

Pacchiano, D. M., Whalen, S. P., Horsley, H. L., & Parkinson, K. (2016). *Efficacy study of a professional development intervention to strengthen organizational conditions and effective teaching in early education settings.* Evanston, IL: Society for Research on Educational Effectiveness.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, *62*(3), 307-332.

Papay, J. P. & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*(1), 105-119.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.

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

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*(2), 247-252.

Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching, 39*(5), 369-393.

Seashore Louis, K., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School Effectiveness and School Improvement, 21*(3), 315-336.

Timperley, H., & Alton-Lee, A. (2008). Reframing teacher professional learning: An alternative policy approach to strengthening valued outcomes for diverse learners. *Review of Research in Education, 32*(1), 328-369.

TNTP. (2012). *The irreplaceables: Understanding the real retention crisis in America's urban schools.* New York, NY: Author.

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher*, *37*(8), 469-479.

Wiener, R., & Pimentel, S. (2017). Practice what you teach: Connecting curriculum & professional learning in schools. Washington, DC: *Aspen Institute*.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*(1), 57–67.

Young, V. M., Schmidt, R., Wang, H., Cassidy, L., & Laguarda, K. (2017). *A comprehensive model of teacher induction: Implementation and impact on teachers and students.* Menlo Park, CA: SRI International.

Tables & Figures

Table 1. Characteristics of Teacher Fellowship Program Participants

	Kansas City		Louisiana		Washington D.C.		Overall	
	Individual	Team	Individual	Team	Individual	Team	Individual	Team
Total Enrolled	7.00	2.0	6.00	14.00	8.00	10.00	21.00	26.00
% Completing program	0.86	0.5	0.67	0.57	1.00	0.80	0.86	0.65
Average Years Experience	6.00	NA	6.33	1.67	16.00	5.89	7.38	4.83
% with graduate degrees	0.86	1.0	0.50	0.58	0.88	0.10	0.76	0.42
Average Application Score	0.82	NA	0.54	0.50	0.57	0.55	0.62	0.53
Average Interview Score Score	0.83	NA	0.42	0.54	0.70	0.72	0.63	0.63
Average Instructional Assessment Score	0.58	NA	0.33	0.49	0.76	0.54	0.56	0.51
% Identifying as a Person of Color	0.14	0.0	0.67	0.50	0.67	0.89	0.47	0.61

Notes: Table shows data collected as part of the participant application process during spring 2015 and 2016. Sample includes 47 of the 49 teachers directly supported in the final analytical sample of schools. Memphis was excluded from the analytical sample due to data limitations. Percentage completing the program represents the percent of teachers that enrolled in the spring of 2015 or 2016 and remained enrolled until the spring of 2017. Average application score from Leading Educators Teacher Leader Application Rubric, average classroom observations scores from TNTP Core Teaching, and average interview score from Leading Educators interview scoring rubric were scaled to 0-1, where 0 is the minimum number of points in each rubric and 1 is the maximum.

Table 2. Characteristics of Schools, by Treatment Status

	Comparison Schools	Treatment Schools	p_value of difference
Proficiency ELA	0.36	0.36	0.39
Proficiency Math	0.32	0.31	0.11
Hispanic	0.12	0.18	0.00
Black	0.56	0.62	0.00
White	0.27	0.16	0.00
Asian	0.02	0.01	0.00
FRL	0.53	0.52	0.66
ELL	0.05	0.08	0.00
Special_Ed	0.05	0.09	0.00
Magnet	0.07	0.04	0.02
Charter	0.27	0.79	0.00
Full_time_teachers	32.97	37.34	0.00
Student_teacher_ratio	15.57	13.56	0.00
Schools	529.00	29.00	NA
N	8237.00	527.00	NA

Notes: Based on school by grade by year analytic sample for ELA achievement outcome (there are 8198 cases in comparison group and 510 cases in treatment group in the analytical sample for Math). In the school years 2009-10 to 2014-15 before the treatment the pooled and comparison group percentage proficiency mean (standard deviation) in ELA was 0.31 (0.22) and 0.30 (0.23) in Math. The treatment mean (standard deviation) was 0.33 (0.21) for ELA and 0.32 (0.22) for Math.

Table 3. Effect of Fellowship Model on Student Achievement

	Pooled DD ELA (1)	Pooled DD Math (2)	Event study ELA (3)	Event study Math (4)
Effect	0.053*	0.085**		
	(0.031)	(0.037)		
Effect 6 years before			-0.003	0.031
			(0.082)	(0.045)
Effect 5 years before			-0.001	0.035
			(0.099)	(0.041)
Effect 4 years before			-0.041	0.008
			(0.069)	(0.052)
Effect 3 years before			0.044	-0.009
			(0.057)	(0.044)
Effect 2 years before			0.012	0.029
			(0.035)	(0.047)
Effect 1st year			0.059*	0.075***
			(0.034)	(0.020)
Effect 2nd year			0.072*	0.145***
			(0.043)	(0.040)
Effect 1 year after			0.042	0.091*
			(0.046)	(0.050)
Effect 2 years after			0.033	0.085
			(0.048)	(0.061)
Observations	8,764	8,708	8,764	8,708
R^2	0.91307	0.88021	0.91316	0.88036
Within R^2	0.01667	0.01544	0.01764	0.01668

Notes: Each column represents a separate regression. Coefficients with robust standard errors (clustered at the school level). Data are from 2009-10 through 2018-19 school years. The effect represents the change in proportion of students that are proficient or advanced between Spring of the year immediately before the first year of treatment (2015-2016) and the specified number of years after the first year of treatment, between LE schools and the comparison schools. Teachers enrolled the summer of 2015-2016 or the summer of 2016-2017 for 1 or two years and some withdrew earlier. All regressions include grade by year by LEA fixed effects, school fixed effects, and controls for charter, magnet, student-teacher ratio, school level percentage FRL, ELL, and special education, and grade level percentage of white students (white ethnicity was chosen as reference so the effect refers to students of color). Coefficients statistically significant at the 10% (*), 5% (**), and 1% (***) levels.

Table 4. Contextualizing the Effect of Fellowship Model on Student Achievement

	Effects expressed as percentage point increase in students that are proficient or advanced	Effects expressed as percentage increase in the percentage of students that are proficient or advanced	Effects expressed in standard deviation units	Effects magnitude benchmarked against similar interventions
Math				
Effect across 4 years	8.5%**	28%**	0.37**	Large
Effect 1st year	7.5%***	25%***	0.33***	Large
Effect 2nd year	14.5%***	48%***	0.63***	Large
Effect 1 year after	9.1%*	30%*	0.40*	Large
Effect 2 years after	9%	28%	0.37	Large
ELA				
Effect across 4 years	5.3%*	17%*	0.24*	Large
Effect 1st year	5.9%*	19%*	0.27*	Large
Effect 2nd year	7.2%*	23%*	0.33*	Large
Effect 1 year after	4%	14%	0.19	Medium
Effect 2 years after	3%	11%	0.15	Medium

Notes: The effects across 4 years represent the difference between treated and the comparison schools on the change in the percentage points of proficient or advanced students across school years 2015-16 through 2018-19 compared to the average proficiency across school years 2009-10 through 2014-15. The year specific effects represent the same difference on the year specified but the comparison is the year immediately before treatment 2014-2015. We used the comparison group percentage proficiency mean in ELA (0.31) and math (0.30) in the baseline years 2009-10 through 2014-15, and the pooled standard deviation in ELA (0.22) and math (0.23) in the same years to estimate the total projected increase as a percentage of the baseline scores and to convert to standard deviation units. To report a magnitude as it compares to other interventions, we used Kraft (2020) and Lipsey et al (2012). Coefficients statistically significant at the 10% (*), 5% (**), and 1% (***) levels.

Table 5. Heterogeneous Effects of Fellowship Model, by Program Saturation

	2% Saturation	4% Saturation	6% Saturation	12% Saturation	Difference in 2% to 12% Saturation
ELA					
Effect across 4 years	0.122 (0.030)***	0.065 (0.020)***	0.009 (0.018)	-0.160 (0.050)***	-0.282 (0.071)***
Effect 1st year	0.106 (0.044)*	0.071 (0.035)	0.036 (0.033)	-0.068 (0.067)	-0.174 (0.088)
Effect 2nd year	0.149 (0.043)***	0.089 (0.031)**	0.029 (0.026)	-0.152 (0.056)**	-0.301 (0.084)***
Effect 1 year after	0.147 (0.050)**	0.068 (0.034)	-0.011 (0.025)	-0.247 (0.061)***	-0.394 (0.099)***
Effect 2 years after	0.099 (0.048)	0.050 (0.033)	0.001 (0.036)	-0.146 (0.101)	-0.245 (0.131)
Math					
Effect across 4 years	0.155 (0.036)***	0.098 (0.026)***	0.039 (0.023)	-0.134 (0.058)*	-0.290 (0.082)***
Effect 1st year	0.122 (0.062)	0.089 (0.049)	0.055 (0.043)	-0.046 (0.076)	-0.168 (0.332)
Effect 2nd year	0.211 (0.049)***	0.162 (0.034)***	0.113 (0.030)***	0.033 (0.076)	-0.245 (0.109)*
Effect 1 year after	0.191 (0.057)***	0.117 (0.037)***	0.043 (0.027)	-0.180 (0.078)*	-0.371 (0.009)***
Effect 2 years after	0.192 (0.046)***	0.111 (0.033)***	0.029 (0.029)	-0.216 (0.068) ***	-0.409 (0.097)***

Notes: Each cell shows estimates (in percentage points) for treated schools based on the linear combination of $\hat{\gamma} + \hat{\beta}(Saturation_s)$ for different values of saturation. The mean (standard deviation) of saturation is 0.058 (0.048) percentage points. See Table A2 for estimates of full specification of equations 3 and 4 upon which these estimates are based. Coefficients statistically significant at the 10% (*), 5% (**), and 1% (***) levels.

Table 6. Heterogeneous Effects of Fellowship Model, by Enrollment Type

	Individual	Team
ELA		
Effect across 4 years	0.092 (0.031)***	0.029 (0.021)
Effect 1st year	0.016 (0.037)	0.093 (0.052)
Effect 2nd year	0.027 (0.040)	0.102 (0.041)**
Effect 1 year after	0.016 (0.045)	0.065 (0.041)
Effect 2 years after	0.014 (0.039)	0.044 (0.044)
Math		
Effect across 4 years	0.13 (0.038)***	0.055 (0.029)
Effect 1st year	0.076 (0.063)	0.083 (0.074)
Effect 2nd year	0.143 (0.063)*	0.144 (0.046)***
Effect 1 year after	0.130 (0.063)	0.062 (0.037)
Effect 2 years after	0.126 (0.053)*	0.052 (0.042)

Notes: Each cell shows estimates (in percentage points) for treated schools based on the linear combination of $\hat{\gamma} + \hat{\beta}(EnrollmentType)$ for EnrollmentType $=$ 1 if LEA leaders enrolled alongside teachers and 0 otherwise. See Table A3 for estimates of full specification of equations 7 and 8 upon which these estimates are based. Coefficients statistically significant at the 10% (*), 5% (**), and 1% (***) levels.

Table 7. Heterogeneous Effects of Fellowship Model, by Program Duration

	˜5 months	1 year	˜15 months	2 years	Difference in 1 to 2 years
ELA					
Effect across 4 years	-0.171 (0.057)**	-0.071 (0.034)	-0.028 (0.026)	0.094 (0.024)***	0.165 (0.042)***
Effect 1st year	-0.067 (0.084)	-0.012 (0.055)	0.012 (0.044)	0.078 (0.036)	0.091 (0.056)
Effect 2nd year	-0.127 (0.089)	-0.039 (0.054)	-0.002 (0.041)	0.105 (0.035)**	0.144 (0.064)*
Effect 1 year after	-0.238 (0.078) ***	-0.111 (0.046)*	-0.055 (0.036)	0.099 (0.039)**	0.021 (0.060)***
Effect 2 years after	-0.198 (0.087)**	-0.092 (0.044)	-0.047 (0.034)	0.082 (0.043)	0.174 (0.062)**
Math					
Effect across 4 years	-0.140 (0.074)	-0.040 (0.046)	0.004 (0.035)	0.126 (0.030)***	0.166 (0.052)***
Effect 1st year	-0.042 (0.119)	0.008 (0.080)	0.030 (0.065)	0.092 (0.050)	0.084 (0.074)
Effect 2nd year	0.004 (0.123)	0.066 (0.074)	0.093 (0.055)	0.168 (0.042)***	0.102 (0.087)
Effect 1 year after	-0.135 (0.088)	-0.033 (0.053)	0.011 (0.041)	0.135 (0.045)***	0.168 (0.047)**
Effect 2 years after	-0.245 (0.091)**	-0.097 (0.053)	-0.028 (0.040)	0.163 (0.045)***	0.260 (0.071)***

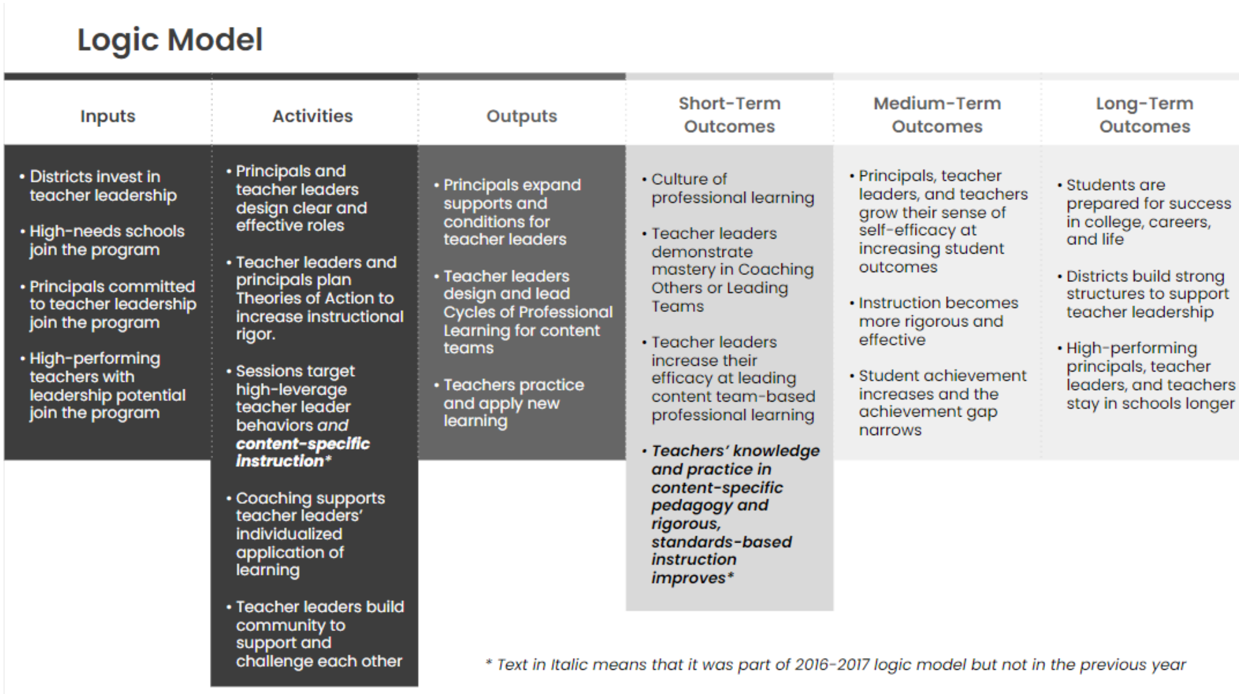
Notes: Each cell shows estimates (in percentage points) for treated schools based on the linear combination of $\hat{\gamma} + \hat{\beta}(TimeEnrolled_s)$ for different values of enrollment time. The mean (standard deviation) of time enrolled is 1.286 (0.529) years. See Table A4 for estimates of full specification of equations 5 and 6 upon which these estimates are based. Coefficients statistically significant at the 10% (*), 5% (**), and 1% (***) levels.

Table 8. Heterogeneous Effects of Fellowship Model, by LEA Leader Participation

	Teacher and LEA	Teacher or LEA only
ELA		
Effect across 4 years	0.074 (0.021)***	-0.030 (0.029)
Effect 1st year	0.077 (0.042)	0.006 (0.038)
Effect 2nd year	0.114 (0.035)***	-0.030 (0.033)
Effect 1 year after	0.084 (0.039)*	-0.062 (0.035)
Effect 2 years after	0.046 (0.037)	-0.017 (0.051)
Math		
Effect across 4 years	0.098 (0.029)***	0.031 (0.040)
Effect 1st year	0.089 (0.061)	0.033 (0.050)
Effect 2nd year	0.159 (0.041)***	0.100 (0.045)*
Effect 1 year after	0.116 (0.042)**	0.024 (0.046)
Effect 2 years after	0.125 (0.035)***	-0.008 (0.055)

Notes: Each cell shows estimates (in percentage points) for treated schools based on the linear combination of $\hat{\gamma} + \hat{\beta}(TeachLEA)$ for TeachLEA $= 1$ if LEA leaders enrolled alongside teachers and 0 otherwise. See Table A5 for estimates of full specification of equations 7 and 8 upon which these estimates are based. Coefficients statistically significant at the 10% (*), 5% (**), and 1% (***) levels.

Figure 1. Teacher Fellowship Model Logic Model

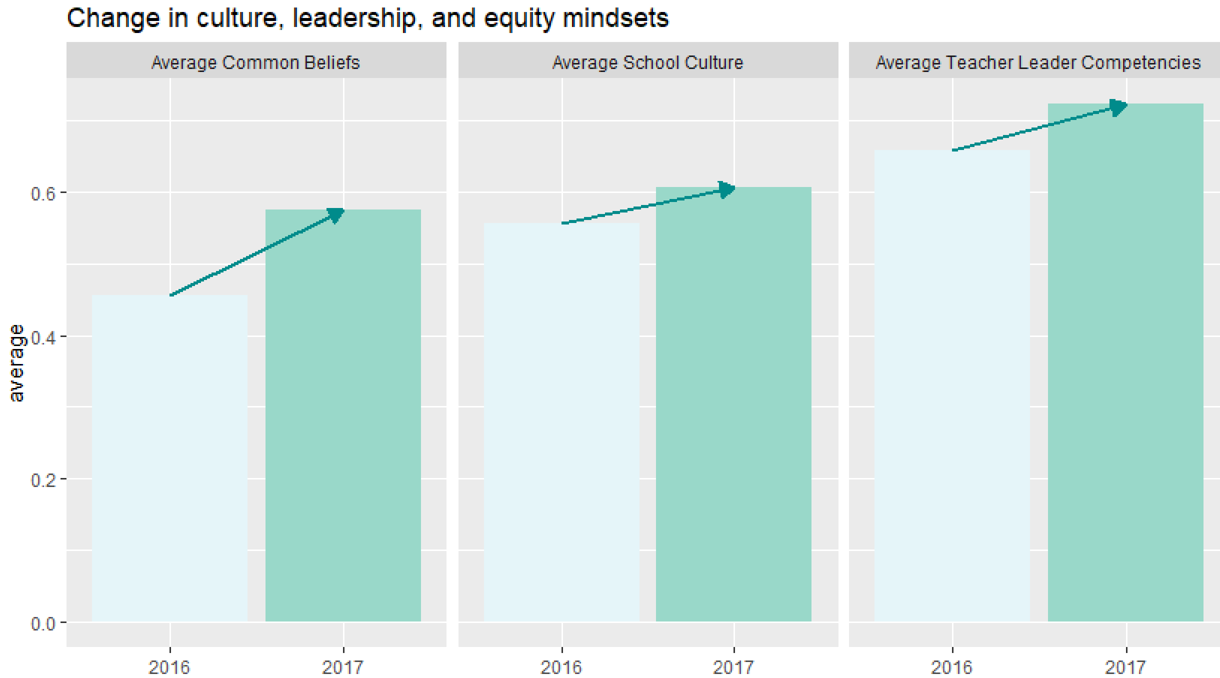


Notes: The logic model describes implementation of Leading Educators' fellowships during the 2015-16 and 2016-17 school years. Based on organizational experience and research demonstrating the value of content-specific professional learning, Leading Educators integrated content-specific sessions and coaching aligned to mathematics and English language arts standards for the 2016-17 school year. © Leading Educators 2017

Figure 2. Teacher Fellowship Model Theory of Change

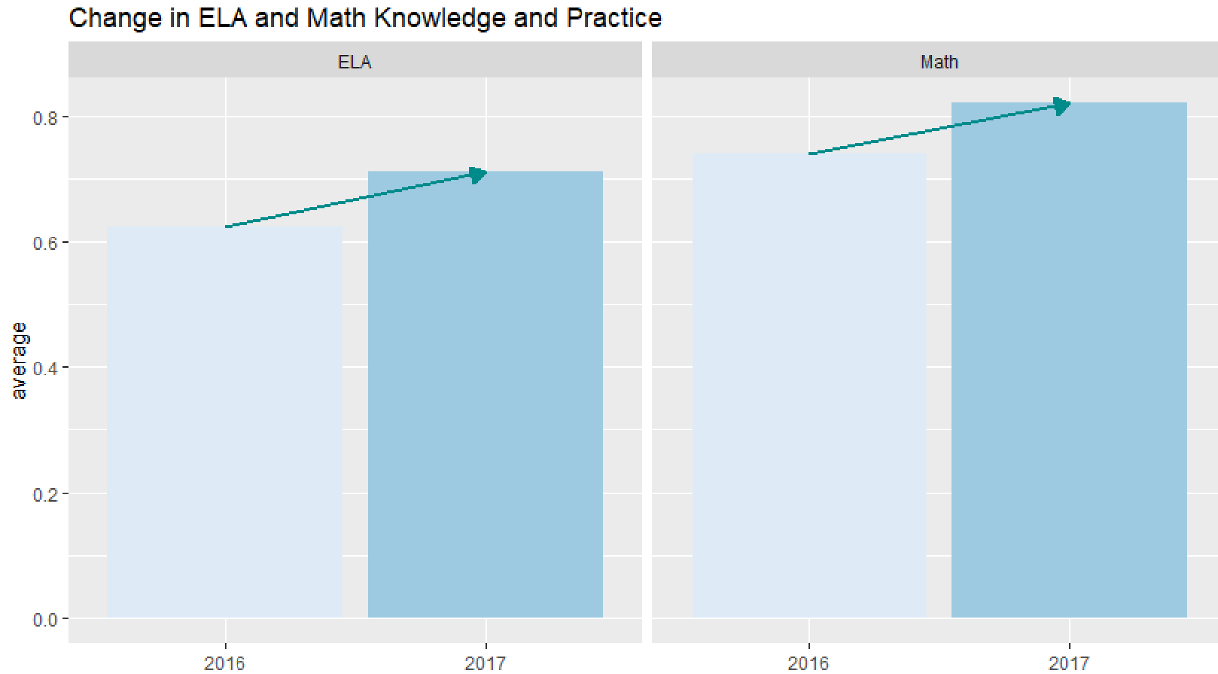
Notes: Figure shows in green the vision that Leading Educators was trying to achieve during the school years 2015-16 through 2016-17. The programmatic focus of the professional development is shown in yellow. In blue, it shows the mechanism by which the program tried to accomplish its vision.

Figure 3. Teacher Leadership, Equity Mindsets and School Culture



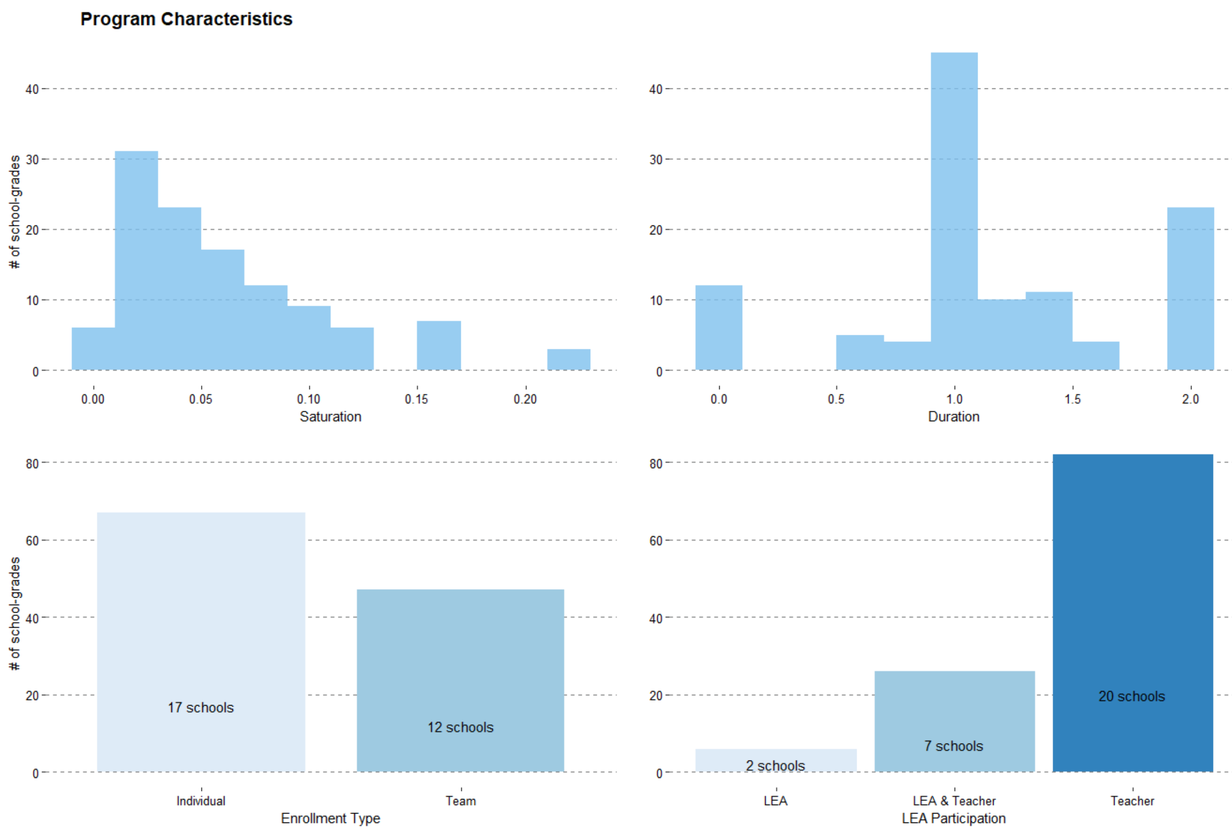
Notes: Data from Leading Educators diagnostic survey administered in the Spring of school year 2015-16 and 2016-17. The sample includes 62 fellows from Cohort 2016-17 who took the pre and posttests from a population of 112 teachers. The average Common Beliefs is an indicator created by scaling and averaging eight prioritized equitable beliefs from the Learning for Justice's Common (formerly Teaching Tolerance) Beliefs survey (Hawley et al., n.d.). The beliefs, found in Table B1 in the appendix, were scaled to 0-1 with 1 representing more equitable beliefs and 0 representing less equitable beliefs. Teacher Leader Competencies and School Culture are indicators creating by scaling and averaging the survey items in Tables B2 and B3 correspondingly in the appendix. The survey questions were also scaled from 0-1, with 0 representing no frequency and 1 representing high frequency.

Figure 4. Teacher Leader Knowledge and Practice, by ELA and Math



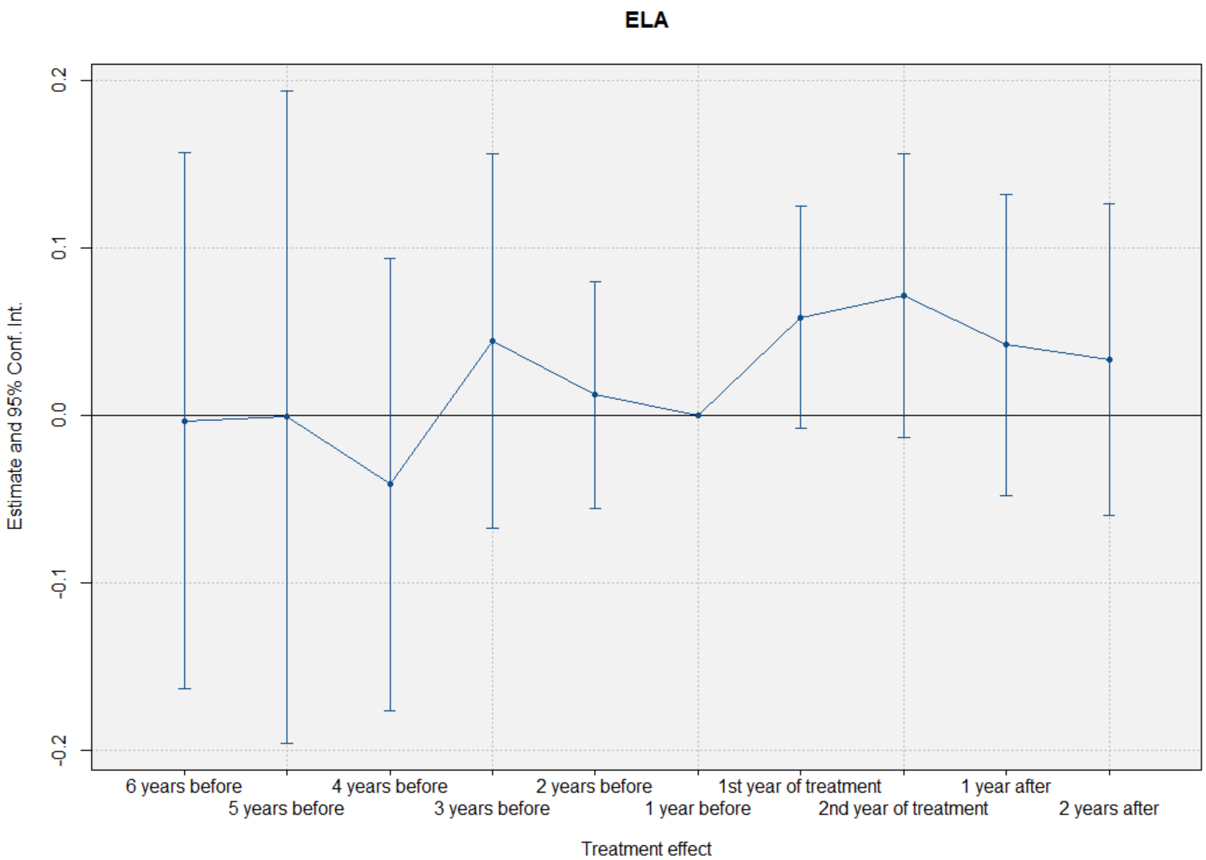
Notes: Data from Leading Educators diagnostic survey administered in the Spring of school year 2015-16 and 2016-17. The sample includes 62 fellows from Cohort 2016-17 who took the pre and posttests from a population of 112 teachers. Each column represents a subject specific (ELA/math) survey item from the knowledge and practice assessment survey found in Table B4 in the appendix. For each knowledge item (columns 1 to 2 for ELA and 4 for math), teachers received a score between 0 and 1, with 1 being completely correct and 0 being completely not correct. For each ELA practice item (columns 3 to 5) participants received a score from 0 to 1, with 0 representing no frequency and 1 representing high frequency. For each math practice item (columns 1 to 3), participants received a score from 0 to 1, with 0 representing strong disagreement and 1 representing strong agreement.

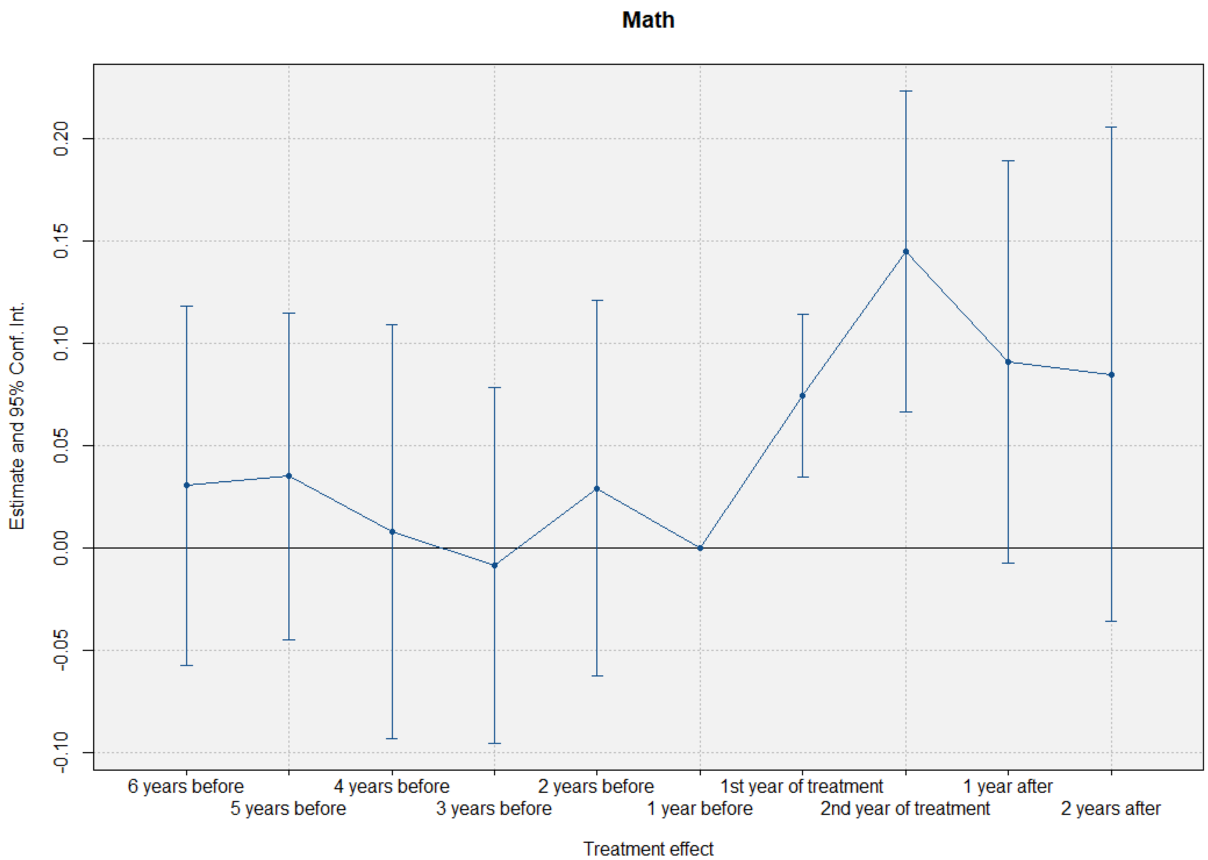
Figure 5. Distribution of Program Characteristics



Notes: Figure shows the distributions of four characteristics that varied across schools in the fellowship model for the school by grade analytic sample in the program years 2015-2017, for ELA achievement outcome (there were an average of 70 school-grades in the analytical sample for Math achievement outcome, for years 2015-2017): Saturation (the proportion of a school's full-time teachers who participated in the program), duration (years teachers remained enrolled in the program), enrollment type (if teachers enrolled in the program as individuals or as a school team) and LEA leader participation (if LEA leaders enrolled in the program alongside teachers, without any teacher enrolling, or did not enrolled (only teachers enrolled).

Figure 6. Event Study Estimates of Fellowship Model Effects on Student Achievement





Notes: Figure shows the average effect of the Fellowship model (with 95% confidence intervals) on the proportion of students proficient or advanced in ELA and Math in the pre- and post- treatment periods (relative to the year immediately before the first year of treatment). See Table 3 for complete summary of estimates upon which these estimates are based.

Appendix A: Tables & Figures

Table A1. Characteristics of Schools, by Treatment Status and Region

	Kansas City		Louisiana		Washington D.C.	
	Comparison	Treat	Comparison	Treat	Comparison	Treat
Proficiency ELA	0.47	0.37	0.31	0.36	0.30	0.36
Proficiency Math	0.43	0.34	0.26	0.27	0.29	0.33
Hispanic	0.17	0.24	0.10	0.16	0.14	0.15
Black	0.31	0.57	0.65	0.53	0.79	0.73
White	0.46	0.16	0.21	0.25	0.04	0.09
Asian	0.02	0.01	0.03	0.02	0.01	0.01
FRL	0.56	0.69	0.48	0.35	0.68	0.55
ELL	0.09	0.17	0.02	0.02	0.06	0.06
Special_Ed	0.11	0.10	0.02	0.01	0.11	0.16
Magnet	0.02	0.14	0.10	0.01	0.00	0.00
Charter	0.12	0.56	0.21	0.73	1.00	1.00
Full_time_teachers	31.08	30.49	35.05	44.54	26.89	36.46
Student_teacher_ratio	15.31	12.58	15.71	15.17	15.57	12.94
Schools	132.00	6.00	320.00	10.00	77.00	13.00
N	2532.00	154.00	4834.00	171.00	871.00	202.00

Notes: Based on school by grade by year analytic sample for ELA achievement outcome (there are 2,493 cases in Kansas City control group and 137 cases in Kansas City treatment group in the analytical sample for Math and the number of cases in Louisiana and Washington D.C. are identical in both samples).

Table A2. Effects of Fellowship Model, by Program Saturation

	Pooled DD ELA (1)	Pooled DD Math (2)	Event study ELA (3)	Event study Math (4)
Effect	0.178***	0.214***		
	(0.036)	(0.036)		
Effect:Saturation	-2.818***	-2.899***		
	(0.576)	(0.841)		
Effect 6 years before			0.013	0.051
			(0.084)	(0.042)
Effect 5 years before			0.016	0.055
			(0.102)	(0.036)
Effect 4 years before			-0.024	0.029
			(0.070)	(0.054)
Effect 3 years before			0.055	0.007
			(0.060)	(0.043)
Effect 2 years before			0.014	0.033
			(0.043)	(0.043)
Effect 1st year			0.141**	0.156***
			(0.062)	(0.030)
Effect 2nd year			0.210***	0.260***
			(0.066)	(0.053)
Effect 1 year after			0.226***	0.265***
			(0.070)	(0.035)
Effect 2 years after			0.148**	0.274***
			(0.058)	(0.057)
Effect:Saturation 1st year			-1.739***	-1.681***
			(0.594)	(0.515)
Effect:Saturation 2nd year			-3.011***	-2.447***
			(0.615)	(0.723)
Effect:Saturation 1 year after			-3.941***	-3.709***
			(0.853)	(0.888)
Effect:Saturation 2 years after			-2.449***	-4.086***
			(0.657)	(1.202)
Observations	8,764	8,708	8,764	8,708
R^2	0.91333	0.88047	0.91348	0.88076
Within R^2	0.01955	0.01764	0.02132	0.01997

Notes: Each column represents a separate regression. Coefficients with robust standard errors (clustered at the school level). Data are from 2009-10 through 2018-19 school years. The effect represents the change in proportion of students that are proficient or advanced between Spring of the year immediately before the first year of treatment (2015-2016) and the specified number of years after the first year of treatment, between LE schools and the comparison schools. Teachers enrolled the summer of 2015-2016 or the summer of 2016-2017 for 1 or two years and some withdrew earlier. The saturation effect represents the same change for schools above the average saturation of 6%. Saturation is the school's average number of teachers enrolled between 2015-16 and 2016-17 divided by the average number of full-time employees in the same years. All regressions include grade by year by LEA fixed effects, school fixed effects, and controls for charter, magnet, student-teacher ratio, school level percentage FRL, ELL, and special education, and grade level percentage of white students (white ethnicity was chosen as reference so the effect refers to students of color).

Table A3. Effects of Fellowship Model, by Enrollment Type

	Pooled DD ELA (1)	Pooled DD Math (2)	Event study ELA (3)	Event study Math (4)
Effect	0.092**	0.132**		
	(0.040)	(0.057)		
Effect:Team Enrollment	-0.063	-0.077		
	(0.059)	(0.069)		
Effect 6 years before			-0.242***	-0.105**
			(0.060)	(0.051)
Effect 5 years before			-0.241***	-0.083
			(0.060)	(0.051)
Effect 4 years before			-0.237***	-0.187***
			(0.056)	(0.049)
Effect 3 years before			0.076	1.797
			(0.065)	(4,238.795)
Effect 2 years before			0.053	0.216***
			(0.047)	(0.052)
Effect 1st year			0.016	0.076***
			(0.018)	(0.022)
Effect 2nd year			0.027	0.143***
			(0.039)	(0.028)
Effect 1 year after			0.016	0.130**
			(0.051)	(0.060)
Effect 2 years after			0.014	0.126*
			(0.037)	(0.069)
Team × Effect 6 years before			0.292***	0.138**
			(0.078)	(0.064)
Team × Effect 5 years before			0.293***	0.116*
			(0.098)	(0.063)
Team × Effect 4 years before			0.235***	0.220***
			(0.071)	(0.063)
Team × Effect 2 years before			-0.027	-0.233***
			(0.072)	(0.080)
Team × Effect 1st year			0.077	0.007
			(0.057)	(0.039)
Team × Effect 2nd year			0.075	0.001
			(0.082)	(0.074)
Team × Effect 1 year after			0.049	-0.067
			(0.087)	(0.077)
Team × Effect 2 years after			0.030	-0.075
			(0.076)	(0.097)
Team × Effect 3 years before				-1.813
				(4,238.816)
Observations	8,764	8,708	8,764	8,708
R^2	0.91313	0.88030	0.91366	0.88089
Within R^2	0.01734	0.01618	0.02335	0.02109

Notes: Each column represents a separate regression. Coefficients with robust standard errors (clustered at the school level). Data are from 2009-10 through 2018-19 school years. The effect represents the change in proportion of students that are proficient or advanced between Spring of the year immediately before the first year of treatment (2015-2016) and the specified number of years after the first year of treatment, between LE schools and the comparison schools. Teachers enrolled the summer of 2015-2016 or the summer of 2016-2017 for 1 or two years and some withdrew earlier. The team effect is the same effect for schools where principals and teachers enrolled as a team. All regressions include grade by year by LEA fixed effects, school fixed effects, and controls for charter, magnet, student-teacher ratio, school level percentage FRL, ELL, and special education, and grade level percentage of white students (white ethnicity was chosen as reference so the effect refers to students of color).

Table A4. Effects of Fellowship Model, by Program Duration

	Pooled DD ELA (1)	Pooled DD Math (2)	Event study ELA (3)	Event study Math (4)
Effect	-0.239***	-0.209*		
	(0.091)	(0.115)		
Effect:Time Enrolled	0.167***	0.167**		
	(0.053)	(0.071)		
Effect 6 years before			0.003	0.035
			(0.082)	(0.052)
Effect 5 years before			0.006	0.039
			(0.098)	(0.048)
Effect 4 years before			-0.034	0.013
			(0.069)	(0.059)
Effect 3 years before			0.047	-0.008
			(0.059)	(0.052)
Effect 2 years before			0.003	0.018
			(0.048)	(0.053)
Effect 1st year			-0.105	-0.077
			(0.086)	(0.087)
Effect 2nd year			-0.187*	-0.038
			(0.112)	(0.134)
Effect 1 year after			-0.325**	-0.204
			(0.132)	(0.172)
Effect 2 years after			-0.270*	-0.361**
			(0.139)	(0.158)
Effect:Time Enrolled 1st year			0.092**	0.085*
			(0.043)	(0.046)
Effect:Time Enrolled 2nd year			0.146**	0.103
			(0.067)	(0.077)
Effect:Time Enrolled 1 year after			0.212***	0.170
			(0.081)	(0.111)
Effect:Time Enrolled 2 years after			0.176**	0.262**
			(0.073)	(0.103)
Observations	8,764	8,708	8,764	8,708
R^2	0.91327	0.88041	0.91339	0.88067
Within R^2	0.01897	0.01711	0.02022	0.01929

Notes: Each column represents a separate regression. Coefficients with robust standard errors (clustered at the school level). Data are from 2009-10 through 2018-19 school years. The effect represents the change in proportion of students that are proficient or advanced between Spring of the year immediately before the first year of treatment (2015-2016) and the specified number of years after the first year of treatment, between LE schools and the comparison schools. Teachers enrolled the summer of 2015-2016 or the summer of 2016-2017 for 1 or two years and some withdrew earlier. The time enrolled effect represents the same effect for schools above the average enrollment duration of 1.3 years. All regressions include grade by year by LEA fixed effects, school fixed effects, and controls for charter, magnet, student-teacher ratio, school level percentage FRL, ELL, and special education, and grade level percentage of white students (white ethnicity was chosen as reference so the effect refers to students of color).

Table A5. Effects of Fellowship Model, by LEA Leader Participation

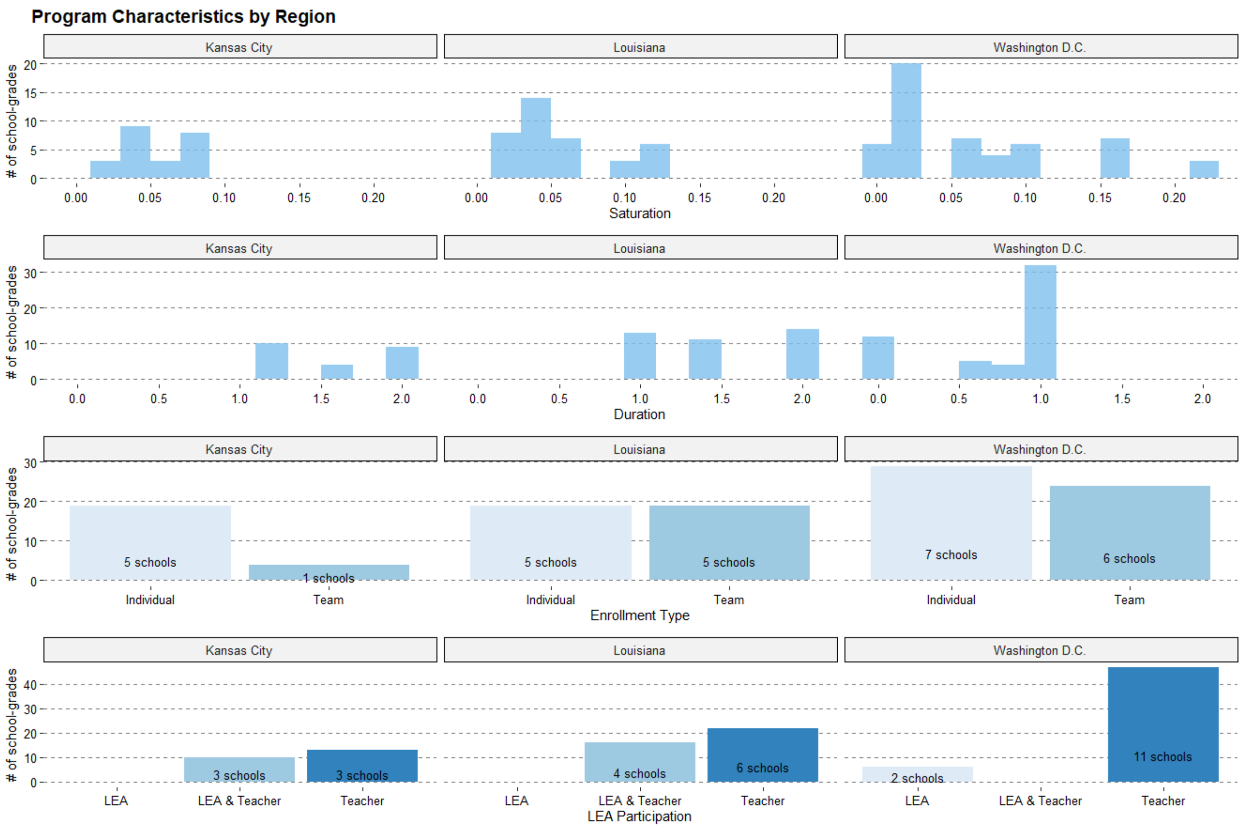
	Pooled DD ELA (1)	Pooled DD Math (2)	Event study ELA (3)	Event study Math (4)
Effect	-0.030	0.031		
	(0.028)	(0.044)		
Effect:LEA Leader	0.103***	0.067		
	(0.036)	(0.057)		
Effect 6 years before			0.017	0.045
			(0.084)	(0.043)
Effect 5 years before			0.019	0.049
			(0.102)	(0.039)
Effect 4 years before			-0.021	0.022
			(0.071)	(0.051)
Effect 3 years before			0.060	0.004
			(0.058)	(0.041)
Effect 2 years before			0.025	0.040
			(0.037)	(0.042)
Effect 1st year			0.006	0.033
			(0.017)	(0.029)
Effect 2nd year			-0.030	0.100
			(0.030)	(0.065)
Effect 1 year after			-0.062*	0.024
			(0.036)	(0.059)
Effect 2 years after			-0.017	-0.008
			(0.051)	(0.062)
TeachLEA × Effect 1st year			0.071	0.056
			(0.043)	(0.034)
TeachLEA × Effect 2nd year			0.144***	0.059
			(0.053)	(0.074)
TeachLEA × Effect 1 year after			0.146**	0.093
			(0.057)	(0.078)
TeachLEA × Effect 2 years after			0.063	0.133
			(0.068)	(0.087)
Observations	8,764	8,708	8,764	8,708
R^2	0.91315	0.88024	0.91329	0.88045
Within R^2	0.01751	0.01570	0.01918	0.01740

Notes: Each column represents a separate regression. Coefficients with robust standard errors (clustered at the school level). Data are from 2009-10 through 2018-19 school years. The effect represents the change in proportion of students that are proficient or advanced between Spring of the year immediately before the first year of treatment (2015-2016) and the specified number of years after the first year of treatment, between LE schools and the comparison schools. Teachers enrolled the summer of 2015-2016 or the summer of 2016-2017 for 1 or two years and some withdrew earlier. The TeachLEA effect represent the same effect for schools where one or more LEA leaders enrolled alongside teachers. All regressions include grade by year by LEA fixed effects, school fixed effects, and controls for charter, magnet, student-teacher ratio, school level percentage FRL, ELL, and special education, and grade level percentage of white students (white ethnicity was chosen as reference so the effect refers to students of color).

Table A6. Costs of Leading Educators Fellowship Model

- **Equipment and materials costs:** The program used Google Drive and Canvas to manage documents and share program information. Google Drive products can be used free of cost, and Canvas memberships were 14.95 per participant per year. Total operating expenses added to about $430 per teacher.
- **Personnel costs:** Each region was staffed by two coaches, one program manager, and one program director. A central design and facilitation team of four created sessions, assessments, and tools. Participants were offered 160 hours of professional development and 20 hours of coaching over the two year period, both within and outside of the school day. Participants were encouraged to engage in ongoing learning during planning time for 2-3 hours per month during the school day. Costs per teacher $7,800
- **Facilities cost:** Most activities were hosted in school or district buildings, which some summer spaces at conference centers. Total cost was about $2,300 per teacher.
- **Overhead expenses:** About $2500 per teacher.
- **Cost paid by students or families:** There was no charge to students or parents.
- **Sources of funding:** The Fellowship was covered by a mix of local and national philanthropy and direct fees from schools.

Figure A1. Distribution of Program Characteristics by region



Notes: Figure shows the distributions of three characteristics that varied across schools in the fellowship model for the school by grade analytic sample in the program years 2015-2017, for ELA achievement outcome by region (there were an average of 13 school-grades in the analytical sample for Math achievement outcome in Kansas City, 27 in Louisiana and 31 in Washington D.C., for years 2015-2017): Saturation (the proportion of a school's full-time teachers who participated in the program), duration (years teachers remained enrolled in the program), enrollment type (if teachers enrolled in the program as individuals or as a school team) and LEA leader participation (if LEA leaders enrolled in the program alongside teachers, without any teacher enrolling, or did not enrolled (only teachers enrolled).

Appendix B: Surveys and Rubrics

Table B1. Common Beliefs Survey

(Adapted, with permission, from a self-reflection tool created by Learning for Justice (formerly Teaching Tolerance; Hawley et al., n.d.))

Beliefs about Equity: Please rate your agreement with the following statements.

	Strongly Agree	Agree	Somewhat Agree	Neutral	Somewhat Disagree	Disagree	Strongly Disagree
I don't think of my students in terms of their race or ethnicity. I am color blind when it comes to my teaching.							
The gap in achievement among students of different races is about poverty not race.							
Teachers should adapt their teaching to the distinctive cultures of African American, Latino, Asian, and Native American students.							
In some cultures, students are embarrassed to speak in front of others, and so I take this into account and don't call on those students in class.							
When students come from homes where educational achievement is not a high priority, they often don't do their homework and their parents don't come to school events. This lack of parental support							

undermines my efforts to teach these students.							
It is not fair to ask students who are struggling with English to take on challenging academic assignments.							
I believe I should reward students who try hard, even if they are not doing well in school. Building their self-esteem is important.							
I try to keep in mind the limits of my students' ability and give them assignments I know they can do so they do not become discouraged.							
Students of different races and abilities often have different learning styles, and good teachers match their instruction to those styles.							
Grouping students of different levels of achievement for instruction may benefit some students, but it can undermine the progress that could otherwise be made by higher achieving students.							
Before students are asked to engage in complex learning tasks, they need to							

have a solid grasp of basic skills.							
With all the pressures to raise student achievement, finding and using examples for the cultural, historic and everyday lived experiences of my students takes away (or could take away) valuable time from teaching and learning what matters most.							
Talking about race with my colleagues could open up a can of worms; little good is likely to come from it.							

Table B2. Teacher leader competencies survey
(Developed by Leading Educators)

Teacher leader competencies: How often do you carry out the following actions?

Statement	**Almost Never**	**Sometimes**	**Frequently**	**Almost Always**	**Always**	**Not Applicable/ Do Not Know**
Models the belief that, regardless of circumstances, all children can master rigorous material						
Acknowledges and confronts racial and other biases in self and others						
Accurately senses and seeks to understand colleagues' preferences, emotions, and perspectives						

Accepts criticism and feedback and takes responsibility for actions						
Adjusts behaviors to respect colleagues' preferences, emotions, and perspectives						
Creates a sense of urgency by ensuring team members see the need for change and the importance of immediate action						
Makes decisions to drive high-quality results while respecting the values and capacity of teammates and school (potential subtract)						
Invites diverse stakeholders (considering students, community members, school leaders, and colleagues) to provide input on and participate in project planning and implementation						
Shifts cognitive lift to teammates by asking open-ended non-rhetorical questions and including think time						
When giving feedback, makes specific, non-judgmental, and factual statements						
Encourages unfiltered discussion to explore differing opinions and reach shared decisions						
Analyzes evidence of teaching and learning with team, celebrating success and making adjustments as needed						

Develops strong relationships and actively renews and repairs relationships as necessary.						

Table B3. School culture survey
(Developed by Leading Educators)

School culture How often do the following actions occur?

Statement	**Almost Never**	**Sometimes**	**Frequently**	**Almost Always**	**Always**	**Not Applicable/ Do Not Know**
Teachers are focused on creating equity for all students through rigorous content.						
Teachers tailor their instructional practices to meet their students' individual learning needs.						
Teachers share norms and values.						
Teachers are honest about their growth areas and ask for help when needed.						
Teachers continually deepen their content knowledge and pedagogy.						
Teachers practice and get feedback on new instructional and planning skills.						
Professional learning opportunities include opportunities to experience, reflect, build shared language, and apply new knowledge.						
Meetings and professional learning opportunities target specific goals to improve teacher and student learning.						
Teachers trust and respect one another.						
Teachers talk to each other and gather ideas about the specific challenges they face in their own classrooms.						
Teachers stick to evidence and data when making statements and decisions.						

Teachers share and give feedback on each other's lesson plans.						
Teachers share and give feedback on each other's teaching practice.						
Teachers consider evidence of student learning when planning for instruction.						
Teachers rely on evidence when deciding whether to adopt new instructional materials or practices.						
Teachers provide students with many opportunities to participate in classroom discussions.						
Teachers consider student input or preferences when planning instructional units.						
Teachers seek out strategies for making classroom content engaging for all students.						

Table B4. Knowledge and practice assessment survey
(Adapted from an assessment created by Student Achievement Partners in 2014)

MATH KNOWLEDGE AND PRACTICE [ECE AND K-2]

This section is focused on standards-based instructional planning and practice. The purpose of this section is to get a sense of of what teachers know and are able to do before planning professional learning for the year. These questions are designed to get accurate data of strengths, preconceived notions, misunderstandings, misconceptions, and knowledge gaps. If you have resources you regularly use in your instructional planning, feel free to consult them.

This section is intentionally challenging. Do not be discouraged if you do not know all or most of the answers. If "I don't know" is the most accurate answer, please use it. Your answers here will not be used to evaluate or judge you, but instead to get an accurate gauge of planning and practice of standards-based instruction in your school. This information will ensure professional learning cycles are aligned to your practice.
Which of the following belongs to the major work of the indicated grade?

Grade	**Select all that apply**				
K	Compare numbers	Tell and write time from analog and digital clocks to the nearest five	Understand meaning of addition and subtraction	Develop understanding of fractions as numbers	I don't know

		minutes using a.m. and p.m.			
1	Add and subtract within 20	Measure lengths indirectly and by iterating length units	Extend understanding of fraction equivalence and ordering	Identify arithmetic patterns (including patterns in the addition or multiplication tables) and explain them using properties of operations	I don't know
2	Identify line of symmetry in two dimensional figures	Understand place value	Apply and extend previous understandings of multiplication and division to multiply and divide fractions	Represent and solve problems involving addition	I don't know

Please rate your agreement with the following statements.

Statement	**Strongly Agree**	**Agree**	**Somewhat Agree**	**Neutral**	**Somewhat Disagree**	**Disagree**	**Strongly Disagree**
I regularly apply rigor in standards and assessment in my instruction.							
I regularly apply deep knowledge of mathematical content and pedagogy in my grade band in my instruction.							
I regularly apply the standards for mathematical practice in my instruction.							

ELA KNOWLEDGE AND PRACTICE [ECE AND LOWER ELEMENTARY K-2 ONLY]

This section is focused on standards-based instructional planning and practice. The purpose of this section is to get a sense of of what teachers know and are able to do before planning professional learning for the year. These questions are designed to get accurate data of strengths, preconceived notions, misunderstandings, misconceptions, and knowledge gaps. If you have resources you regularly use in your instructional planning, feel free to consult them.

This section is intentionally challenging. Do not be discouraged if you do not know all or most of the answers. If "I don't know" is the most accurate answer, please use it. Your answers here will not be used to evaluate or judge you, but instead to get an accurate gauge of planning and practice of standards-based instruction in your school. This information will ensure professional learning cycles are aligned to your practice.

In a typical lesson, please respond about the percentage of your time you are engaging students in the following activities during class. Percentages do not need to add up to 100, as some items may overlap.

	Never use	**1-25% of a typical lesson**	**26-50% of a typical lesson**	**51-75% of a typical lesson**	**76-100% of a typical lesson**
Use of a single grade-level text for whole-class reading, writing, and/or discussion					
Use of leveled readers to support struggling students in place of the grade-level text other students are reading in class					
In-class writing assignments in response to or about texts					

Which of the following approaches for selecting texts for reading aligns with your state's English language arts and literacy standards? Check all that apply.

1. Using abridged or adapted versions of complex texts for struggling readers
2. Assigning complex texts that all students in a class are required to read
3. Selecting texts for individual students based on their reading level
4. Selecting texts for a class based on qualitative factors like knowledge demands, as well as quantitative factors like word and sentence length
5. Other approach (please describe):
6. I don't know

Mr. Jones is developing a lesson plan to go with the text, "Lost Penguin Back in his Natural Habitat." How could Mr. Jones provide the appropriate scaffolds so that all students - including those who read below grade-level - have opportunities to engage in the work of the lesson in a way that best aligns with your state's English language arts and literacy standards?

1. He could rewrite the text and substitute more complex text and difficult vocabulary with easier words and phrases.

2. He could create a podcast or audio recording of the passage for students to listen to as they read along.
3. He could build background knowledge by providing a summary of the text.
4. He could group students homogeneously and give the English Language Learners a simpler text on the same subject.
5. I don't know

Table B5. Leading Educators Selection Process

Application Process Overview

In order to apply, teachers must have at least two years experience and have received a rating of Effective or Highly Effective on last year's Evaluation. All candidates complete the same process, and each phase is scored by assessors using rubrics who pass a norming and training exercise.

1. **Written Application:** applicants submit a short application answering two questions:
    0. The Leading Educators Fellowship is a rigorous and challenging two-year program that empowers teacher leaders to expand their impact. What do you hope to learn and develop through your participation in the Fellowship?
    1. Please describe your teacher leader responsibilities for next year. How do you hope to impact the teachers and students at your school in this role?
    2. Describe a time in your past two years as an educator when you set a goal with your students and struggled to meet it. How did you respond, and what do you think contributed to your success or failure?
2. **Instructional Assessment:** applicants submit a student data essay describing results with students and are observed teaching a lesson in the content area and submit an evidence-based reflection on what students learned and how it prepared them for mastery of grade-level content.
3. **Interview and Role Play:** applicants participate in an interview where they answer questions about their content area and participate in a short role play. Sample questions include:
    0. What are the key priorities of the school in this content area? How would you address these priorities as a teacher leader?
    1. In the last year, how have you deliberately grown in your instructional practice? How do you know?
    2. The applicant participates in a short role play giving feedback to a teacher to assess their content expertise and how they manage the relationship.

	Written Application	Instructional Assessment	Interview and Role Play
Equity: Models the belief that all children can master rigorous material			
Growth: Actively seeks opportunities to leverage strengths and develop growth areas			
Opportunities to Lead: Clear charge from the principal in terms of a leadership role or responsibility for the next two years. Responsible for leading at least two adults.			
Results: Works diligently and purposefully to reach results without lowering expectations			
Vision: Clearly communicates vision of success for students and teachers			

Assess: Analyzes and reflects on student achievement data against benchmarks towards end of year goals			
Instructional Expertise: Achieves strong results with students			
Relationship Management Appropriately matches leadership styles to individual and contextual needs by identifying the skill level and motivation of colleagues			
Community Collaborates with colleagues to increase the collective impact on student success Supports, celebrates, and challenges colleagues			
Self-Management Identifies emotional triggers and manages reactions to conflict and stressful situations			
Plan: Analyzes context to identify the highest-need annual and interim priorities with clear links to vision of success for students and teachers			