# Pathways to Instructional Leadership

## Implementation and Outcomes from a Job-Embedded School Leader Training Program

BENJAMIN K. MASTER, ELIZABETH D. STEINER,
CHRISTOPHER JOSEPH DOSS, HANNAH ACHESON-FIELD

**Sponsored by TNTP**

For more information on this publication, visit www.rand.org/t/RRA255-1

Published by the RAND Corporation, Santa Monica, Calif.

# Preface

The contributions of assistant principals (APs) toward improving student and staff outcomes has not been a focus of much empirical research to date. In recent years, as the increased focus on improving student outcomes has shifted expectations, APs are now expected to assume some instructional leadership tasks. To date, there have been few studies of AP professional development opportunities, and little is known about the extent to which APs are trained to undertake instructional leadership activities. In the accompanying report, *Pathways to Instructional Leadership: Implementation and Outcomes from a Job-Embedded School Leader Training Program*, the authors present findings about implementation and impacts on student outcomes from a four-year study of the Pathway to Leadership in Urban Schools (PLUS) program in a large, urban public school district. PLUS was developed by TNTP (formerly The New Teacher Project) to train APs in instructional leadership. The findings in this report could help policymakers, district staff, and training program providers understand participants' perceptions of programs that focus on training APs in instructional leadership and the impact that such a program can have on student and staff outcomes. The findings could also help program providers and district staff understand some of the benefits and challenges associated with training school leaders in the context of district-provider partnerships. This technical appendix provides more detail about data, samples, methodology, and results from additional exploratory analyses conducted by the authors.

This study was undertaken by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through postsecondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking. TNTP sponsored the report, with funding from the U.S. Department of Education's Investing in Innovation (i3) grant program, grant number U411C140039.

More information about RAND can be found at www.rand.org. Questions about this technical appendix should be directed to bmaster@rand.org, and questions about RAND Education and Labor should be directed to educationandlabor@rand.org.

# Contents

Preface ..... iii
Figures ..... v
Tables ..... vi
Abbreviations ..... vii
Technical Appendix ..... 1
References ..... 35

# Figures

Figure A.1. Unadjusted ELA Achievement Levels of PLUS-Exposed Schools Centered on the Year Before PLUS Exposure ..... 29
Figure A.2. Unadjusted Mathematics Achievement Levels of PLUS-Exposed Schools Centered on the Year Before PLUS Exposure ..... 30
Figure A.3. Differential Schoolwide Mathematics Performance over Time for Schools That Hired a Leader Compared with Schools That Never Hired a Leader ..... 32
Figure A.4. Differential Schoolwide ELA Performance over Time for Schools That Hired a Leader Compared with Schools That Never Hired a Leader ..... 32

# Tables

Table A.1. Number of Interviews Conducted, by Cohort and Interview Year ..... 2
Table A.2. Interview Invitations, Replacements, and Refusals, by Cohort and Interview Year ..... 3
Table A.3. Number of Interviews Conducted, by School Grade-Level Configuration ..... 3
Table A.4. Number of Interviews Conducted, by Cohort, Year, and Role ..... 4
Table A.5. Example Leader Interview Questions ..... 5
Table A.6. Characteristics of New PLUS and New Non-PLUS Assistant Principals ..... 10
Table A.7. Characteristics of Treated and Untreated Students in Pooled ELA Coaching Analytical Sample, With and Without Weighting ..... 15
Table A.8. Characteristics of Treated and Untreated Students in Pooled Math Coaching Analytical Sample, With and Without Weighting ..... 16
Table A.9. Estimates of the Impact of Leader Coaching on Student Achievement for a Subsample of PLUS-Coached and Comparison Teachers, With and Without Controls for Teachers' Prior-Year Value-Added ..... 18
Table A.10. Characteristics of Coached and Non-Coached Teachers in Pooled Retention Sample, With and Without Weighting ..... 23
Table A.11. Characteristics of Coached and Non-Coached Teachers in Pooled Value-Added Model Retention Sample ..... 24
Table A.12. Cumulative Rates of Novice APs No Longer Working as School Administrators in the District One, Two, or Three Years After First Becoming an AP ..... 25
Table A.13. Cumulative Rates of Novice APs Being Promoted to Principal Roles in the District One, Two, or Three Years After First Becoming an AP ..... 25
Table A.14. Adjusted Student Achievement of Leader-Coached Teachers Relative to Comparison Teachers Across All PLUS Cohorts and Years ..... 25
Table A.15. Cumulative Rates of Teachers Leaving the District One, Two, or Three Years Following Receipt or Nonreceipt of Leader Coaching ..... 26
Table A.16. Cumulative Rates of Teachers Leaving the District One, Two, or Three Years Following First Receipt or Nonreceipt of Leader Coaching, by Teachers' Student Achievement Gains ..... 26
Table A.17. Characteristics of Ever-PLUS Schools and Never-PLUS Schools ..... 29
Table A.18. Comparative Interrupted Time Series Analysis of Changes in Level and Trends in School Average Achievement Before and After Exposure to a Leader ..... 33
Table A.19. Thresholds and Results for Fidelity of Implementation Tracking ..... 34

# Abbreviations

| | |
|---|---|
| AP | assistant principal |
| CEM | coarsened exact matching |
| CITS | comparative interrupted time series |
| ELA | English language arts |
| ELL | English-language learner |
| i3 | Investing in Innovation Fund |
| PD | professional development |
| PLUS | Pathway to Leadership in Urban Schools |
| SY | school year |

# Technical Appendix

In this technical appendix, we include additional details about the data, methodology, and limitations of our analyses of both the implementation and impacts of the Pathway to Leadership in Urban Schools (PLUS) program. In addition, we report on supplemental findings from exploratory analyses that we conducted related to the impacts of PLUS on schoolwide student-achievement outcomes. Finally, at the end of this appendix, we include a descriptive summary of implementation fidelity metrics gathered by the program implementer, TNTP. These fidelity metrics were prespecified at the start of the PLUS grant, and we report on them to fulfill a requirement of the grant.

## Sample and Methods for Implementation Analyses

We drew on three data sources to understand implementation of the PLUS program: program data and materials provided by TNTP, background interviews with PLUS staff, and interviews with Leaders (assistant principals [APs] and principals in the PLUS program). We describe each of these in this section.

### *Program Data Provided by TNTP*

As the program implementer, TNTP tracked fidelity of implementation and collected data about teachers who were coached by Leaders during their participation in the program. TNTP also gathered information on Leader performance through evaluations of Leaders during the program. TNTP shared all these data with us, along with extensive background materials related to PLUS program components (e.g., the selection process, guidance for host principals, the Leader-evaluation process) and performance standards.

### *Background Interviews with PLUS Staff*

We conducted background telephone interviews with the TNTP staff who managed the program multiple times per year over the four years of the study. Overall, we conducted 24 interviews with TNTP staff (about six interviews per year over four years). Interviews ranged from 30 to 45 minutes in length and included one to four TNTP staff. All the TNTP staff we contacted agreed to participate in the interviews in all years. The interviews were designed to elicit an overall understanding of the program's design and intended implementation and to explore the evolution of the program over time. We also used these interviews to clarify questions that arose in our review of program data. The protocol was semistructured, and the main questions covered key program components. Questions were constantly revised to elicit information about program updates and changes over time.

### *Leader Interviews*

We conducted 45–60-minute telephone interviews with Leaders annually in the spring, beginning in 2016 and ending in 2019, to gather information about program implementation and understand their experiences in the PLUS program. Our goal was to conduct longitudinal interviews with the same leaders over the course of three years: during their first and second years in PLUS and as PLUS alums one year after their completion of the program. However, conducting longitudinal interviews was not always possible because, over time, several participants did not respond to scheduling emails or, in rare cases, declined to participate.

We interviewed about half of the leaders in Cohorts I–III in each year of the project, for a total of 51 interviews over the four years of the study. The number of interviewees in each cohort in a given year ranged from four to eight (Table A.1). Individuals in Cohort III were slightly under-sampled in both 2018 and 2019. Over the four years of the study, we interviewed 22 total unique Leaders, 16 of whom participated in interviews in all applicable years (12 leaders in Cohorts I and II participated for three years, and four leaders in Cohort II participated for two years). We also interviewed one Cohort I Leader twice and five Leaders, across all cohorts, in only one year. In each year, a small number of Leaders refused to participate (zero to two Leaders per year) or did not respond to our request (zero to three Leaders per year; Table A.2).

**Table A.1. Number of Interviews Conducted, by Cohort and Interview Year**

| | Cohort I | | Cohort II | | Cohort III | | |
|---|---|---|---|---|---|---|---|
| Interview Year | Number in Cohort | Number interviewed | Number in Cohort | Number Interviewed | Number in Cohort | Number Interviewed | Total Interviews |
| 2016 | 13 | 8 | – | – | – | – | 8 |
| 2017 | 12 | 7 | 13 | 6 | – | – | 13 |
| 2018 | 12 | 7 | 13 | 6 | 12 | 4 | 17 |
| 2019 | – | – | 13 | 8 | 11 | 5 | 13 |

NOTES: Although we tried to interview the same members in each cohort, there were several cases in which we were unable to reach interviewees from previous years or interviewees were unable to participate. Table A.2 shows these cases.

In year 1 of the study, we randomly sampled half the leaders in the program (Cohort I) to participate in the interviews. After the initial randomization, we adjusted the sample as needed to achieve representation on important characteristics, such as school configuration (i.e., elementary, middle, and high schools) (Table A.3) and role (Table A.4). We did not achieve perfect representation in all years because of occasional participant refusals or lack of response and the limited number of Leaders who fit a given set of characteristics. In year 2 of the study, we invited the same Cohort I Leaders to participate in an interview, replacing those who did not respond to invitation emails or who declined to participate with a Leader who had similar characteristics (e.g., school configuration and role). In year 2 of the study, we also randomly

**Table A.2. Interview Invitations, Replacements, and Refusals, by Cohort and Interview Year**

| Cohort | Interview Year | Number Interviewed | Number Invited | Number Replaced | Number Refused | Number of No Response |
|---|---|---|---|---|---|---|
| Cohort 1 | 2016 | 8 | 8 | – | 0 | 0 |
| | 2017 | 7 | 8 | 1 | 1 | 0 |
| | 2018 | 7 | 7 | 0 | 0 | 0 |
| Cohort 2 | 2017 | 6 | 9 | – | 1 | 2 |
| | 2018 | 6 | 7 | 0 | 0 | 1 |
| | 2019 | 8 | 11 | 0 | 0 | 3 |
| Cohort 3 | 2018 | 4 | 7 | – | 1 | 2 |
| | 2019 | 5 | 10 | 0 | 2 | 3 |

NOTES: In each cohort and year, the number interviewed is equal to the number invited minus the number who refused and provided no response.

**Table A.3. Number of Interviews Conducted, by School Grade-Level Configuration**

| Cohort | Grade Level | Cohort 1 | | Cohort 2 | | Cohort 3 | |
|---|---|---|---|---|---|---|---|
| | | Number in Cohort | Number Interviewed | Number in Cohort | Number Interviewed | Number in Cohort | Number Interviewed |
| 2016 | Elementary | 4 | 2 | – | – | – | – |
| 2016 | K–8 | 1 | 1 | – | – | – | – |
| 2016 | Middle school | 4 | 3 | – | – | – | – |
| 2016 | High school | 2 | 2 | – | – | – | – |
| 2017 | Elementary | 5 | 1 | 3 | 2 | – | – |
| 2017 | K–8 | 1 | 1 | 2 | 1 | – | – |
| 2017 | Middle school | 4 | 3 | 4 | 2 | – | – |
| 2017 | High school | 2 | 2 | 4 | 1 | – | – |
| 2018 | Elementary | 7 | 3 | 2 | 1 | 3 | 1 |
| 2018 | K–8 | 1 | 1 | 2 | 1 | 1 | 0 |
| 2018 | Middle school | 2 | 1 | 5 | 2 | 4 | 1 |
| 2018 | High school | 2 | 2 | 3 | 1 | 4 | 2 |
| 2019 | Elementary | – | – | 2 | 1 | 2 | 1 |
| 2019 | K–8 | – | – | 3 | 3 | 1 | 0 |
| 2019 | Middle school | – | – | 5 | 3 | 4 | 1 |
| 2019 | High School | – | – | 3 | 1 | 3 | 2 |

NOTES: This table does not include the one leader working in the central office in 2018 and 2019. A – indicates that the cohort was not interviewed in a given year. There is one leader in Cohort I in 2016 that is omitted because we lacked school grade level.

sampled half the Leaders in Cohort II to participate in interviews and adjusted the sample after randomization, using the approach used for Cohort I, to achieve representation on important characteristics. Recruitment in years 3 and 4 of the study followed the same pattern, as shown in Tables A.1 and A.2.

In 2016 and 2017, nearly all interviewees were APs. However, in 2018 and 2019, several interviewees had moved into principal roles, and two were serving in other administrator positions (Table A.4). In 2017, we had data on whether Leaders were internal or external hires and their AP role (general, pupil services, or curriculum and instruction) and adjusted the interview sample after randomization to achieve representation on these characteristics in addition to school grade level and role. In 2017, nine out of the 13 interviewees were internal hires. About half (seven out of 13) were serving in general AP roles, four served in pupil-services roles and two served in curriculum and instruction roles.

**Table A.4. Number of Interviews Conducted, by Cohort, Year, and Role**

| Cohort | Grade Level | Cohort I | | Cohort II | | Cohort III | |
|---|---|---|---|---|---|---|---|
| | | Number in Cohort | Number Interviewed | Number in Cohort | Number Interviewed | Number in Cohort | Number Interviewed |
| 2016 | Assistant principal | 13 | 8 | – | – | – | – |
| 2016 | Principal | 0 | 0 | – | – | – | – |
| 2016 | Other administrator | 0 | 0 | – | – | – | – |
| 2017 | Assistant principal | 10 | 7 | 13 | 6 | – | – |
| 2017 | Principal | 2 | 0 | 0 | 0 | – | – |
| 2017 | Other administrator | 0 | 0 | 0 | 0 | – | – |
| 2018 | Assistant principal | 6 | 3 | 10 | 5 | 11 | 3 |
| 2018 | Principal | 6 | 4 | 2 | 0 | 1 | 1 |
| 2018 | Other administrator | 0 | 0 | 1 | 1 | 0 | 0 |
| 2019 | Assistant principal | – | – | 8 | 5 | 9 | 3 |
| 2019 | Principal | – | – | 5 | 3 | 1 | 1 |
| 2019 | Other administrator | – | – | 0 | 0 | 1 | 1 |

NOTES: – indicates that cohort was not interviewed in a given year.

We used a semistructured interview protocol, which was refined slightly in each year to reflect the Leader's tenure in the program (i.e., first year, second year, or alum), included new aspects of the program (e.g., transition of the program to district management), and probed key themes that emerged in the previous year. The interviews were led by an analyst who was familiar with PLUS program components and goals. The semistructured protocol allowed for

balance and consistency in the questions asked and ensured coverage of important content while allowing respondents to elaborate or offer unsolicited input.

The protocol covered a range of topics related to program implementation and Leader experience, such as career history and background; school leadership and on the job training; and experience with PLUS professional development, PLUS coaching, and cohort interactions. Table A.5 provides a list of sample questions. These examples are not exhaustive but are intended to illustrate the nature of the information we gathered from Leaders.

**Table A.5. Example Leader Interview Questions**

| Interview Topic | Example Question |
|---|---|
| Background | Has your understanding of school leadership changed or evolved now that you have another year of experience? Please explain. |
| School leadership and on the job training | Are you able to devote as much time to observation and feedback (coaching) cycles with teachers as you would like? |
| | Tell me about your work coaching teachers this year. How many teachers do you coach, and how often do you interact with them? |
| PLUS professional development (PD) | Overall, do you think your training with PLUS is adequately preparing you for the school leadership roles and responsibilities that you anticipate taking on? |
| | To what extent does your cohort serve as a professional learning community and source of support? |

### *Analytic Approach*

We analyzed the implementation data, which included program data provided by TNTP, interviews with TNTP staff, and interviews with Leaders to address research questions one and two. We first describe our approach to analyzing the interview data and then describe our approach to analyzing the program data.

Prior to our analysis, we compared the interview notes with the audio recording to check for clarity and accuracy and cleaned them to serve as a near-transcript of the conversation. We engaged in member checking as appropriate to ensure data accuracy. We loaded the cleaned interview notes into the qualitative software (NVivo10 in 2016–2018 and Dedoose in 2019) and coded them using a thematic codebook developed by the evaluation team. We structured the codebook to enable the interview data to be coded according to PLUS program components to capture implementation details (e.g., number of teachers coached by Leaders) and Leaders' experiences (e.g., their perceptions of PLUS coaching) and to enable analysis of similarities and differences over time, across cohorts, and across Leader characteristics. We updated the codebook in each year to account for changes in the protocol; in the early stages of coding, we refined the codebook to capture themes that emerged from the data. In each year, one analyst coded the interviews, which avoided introducing bias from multiple coders.

We first coded the interview data according to PLUS program components. We then conducted a second round of coding, analyzing the data accordingly to capture descriptions of program components (e.g., content of monthly PD sessions), the extent of implementation (e.g., how many teachers did Leaders coach during their participation in PLUS), and Leaders' experiences (e.g., to what extent did Leaders view PLUS coaching as helpful?). Finally, we summarized the interview data for each program component within each year and for each cohort across years. To address research question 1, these summaries included a description of each program component according to TNTP staff and Leaders, descriptions of changes over time, and descriptions of current district practices; we drew primarily on the interviews with TNTP staff for this information. To address research question 2, these summaries captured TNTP staffs' and Leaders' accounts of implementation and Leaders' experiences with the program. We also analyzed the Leader interview data according to a limited number of Leader characteristics (i.e., school grade level and focus of the AP role on pupil services, curriculum or instruction, or general). However, because the numbers of interviewees across these characteristics were small (Tables A.4 and A.5), it was difficult to draw meaningful conclusions. The summaries included common themes across study years and PLUS cohorts as well as instances of disagreement and a discussion of similarities and differences across cohorts and years of the study. We did not find many differences across cohorts or across Leader characteristics, but, when we did, we described them in the text.

We analyzed the program data provided by TNTP by reviewing the materials and broadly categorizing them according to program components (e.g., Leader evaluation process, PD, host principal training) and then reviewing them in detail for content. To address research question 1, we triangulated the information describing program components derived from these materials with the summaries of the interviews to generate descriptions of PLUS program components, changes over time, and descriptions of district practices. To address research question 2, we triangulated the implementation fidelity metrics provided by TNTP with our interview summaries to gauge the extent to which program components were implemented as intended.

### *Limitations*

Our ability to thoroughly evaluate implementation of all PLUS program components was limited by our available data and resources. Specifically, the research budget did not allow us to conduct interviews beyond those with Leaders and selected PLUS staff. We therefore did not interview several key groups of stakeholders, including non-PLUS APs, host principals, and district staff, which hampered our ability to fully document the extent to which some aspects of the program were implemented. In addition, our research budget required us to limit the length of our interviews with Leaders and PLUS staff, and, because of individual variation in the length of responses, we were not always able to ask every question of every respondent. In addition, the implementation fidelity data provided by TNTP are limited to the metrics that TNTP was required to track as part of its grant and only provide a high-level view of adherence to a limited

number of implementation goals. Because of these various limitations, the data available did not allow us to evaluate the extent of implementation at a similar depth for all program components or to rigorously assess implementation quality.

The interview data used in the implementation analysis provide a rich picture of implementation of key program elements and illustrate Leaders' experiences during and after their tenure with the PLUS program. Although we did triangulate Leaders' responses with those provided by PLUS staff and program documentation, note that interview data are self-reports and thus subject to various biases, such as social desirability. Moreover, although the interview data are crucial for providing context and include approximately half the Leaders in each PLUS cohort, the perceptions of the Leaders in this sample may not be representative of the full population of Leaders.

## Impact Analysis Methods and Samples

In this section, we describe each of the analytic models referenced in our main report, beginning with analyses of Leaders' professional outcomes, then effects on the student-achievement gains of coached teachers, and then effects on the retention rates of coached teachers. In each case, we discuss our methodology and samples and then detail limitations that were specific to the analysis. For each model, we also include descriptive tables summarizing the baseline characteristics of individual students, teachers, or Leaders in our samples. Finally, we conclude this section by discussing the methodology for and sharing the results of our supplemental exploratory analysis of Leaders' effects on schoolwide student-achievement outcomes.

### *Evaluating Leader Retention and Promotion*

To evaluate the career trajectories of Leaders, we examined whether they were more likely than other newly hired APs to remain working in the district for a longer period of time or to be promoted to principal roles. With respect to retention, we focused on retention in any role in the district because some APs were promoted to principal roles within their first few years on the job, and we consider this to be a positive outcome with respect to within-district retention. However, we were unable to distinguish between voluntary and involuntary retention in our data.

With respect to principal promotion, we focused on whether APs were promoted to be principals within the school district. However, our data did not identify individuals who may have left the district to take on a principal role in another district. As a consequence, this analysis does not speak to the overall rate at which PLUS or non-PLUS APs were promoted to the principalship inclusive of promotions outside of the district.

**Methods and sample:** To assess PLUS AP retention in the district, we compared the rate at which newly hired Leaders were still working in the district in the year or years after they were first hired with the rate at which other newly hired APs in the district from the same hiring-year

cohort were similarly retained in the district. We excluded from this analysis any PLUS participants who had prior experience as APs in the district so that we could compare just new APs in both PLUS and the comparison group. We also excluded one Leader who was already a principal at the start of the program.

As a sensitivity check, we also ran a version of the model that did not restrict APs based on their experience levels but instead relied on controls for prior experience. Results were similar and are available upon request, but our preferred approach was to focus on the comparison between cohorts of APs that were identical with respect to being new to the AP role in the district.

In addition to restricting our comparison sample to just newly hired APs who were not part of the PLUS program, we also controlled for a small number of AP and school characteristics that we theorized might be most relevant to retention outcomes in our models. We restricted our models to this limited number of variables to avoid over-specification given our limited sample size.

We followed a similar approach when assessing PLUS AP promotion. Here, we used the same sample as in the retention analysis and compared the rate at which newly hired Leaders were both still working in the district and promoted to principal roles with the rate at which other newly hired APs from the same hiring-year cohort were promoted to principal roles in the district. We also ran sensitivity checks on a sample that controlled for experience rather than restricting our sample to novice APs, but the results were not meaningfully different.

In our linear probability model for analyzing AP retention or promotion, we employed models of the following form:

$$Y_{ict} = \beta_0 + PLUS_{ict}\beta_1 + \boldsymbol{X_{ict}\beta_2} + \varepsilon_{ict}$$

In this analysis, we created cohorts of individual APs, *i*, which started in the district in the same cohort, *c* (i.e., same school year [SY]). $PLUS_{ict}$ is an indicator for being a Leader when joining the district in a particular cohort of new APs. $\boldsymbol{X_{ict}}$ is a vector of AP characteristics including gender, an indicator for being white, enrollment size of the school, grade band of school, the school's prior year average English language arts (ELA) and math scores, and indicators for cohort. $\varepsilon_{ict}$ is an AP-level idiosyncratic error term. In our retention analyses, the outcome of interest is $Y_{ict}$, which is an indicator if an AP left *t* years (i.e., one, two, or three years) after beginning their AP job in the district. If an AP left, that person was retained in the data and marked every year thereafter as having left. Thus, the outcome is a cumulative measure of leaving. Similarly, in our promotion analysis, the outcome of interest is an indicator if the AP was promoted within the district one, two, or three years after starting in the AP role in the district.

Note that the number of years of retention outcomes that we could observe varied for each cohort. We could observe whether APs hired in all three cohorts were still in the district as of the start of their second year. Then, for Cohort II, we could observe retention through the start of their third year, and, for Cohort I, we could observe retention through the start of their fourth year in the district. A descriptive comparison of PLUS and non-PLUS APs in our sample panels is shown in Table A.6.

**Limitations:** Our analysis of Leader retention and promotion was primarily limited by the small number of both PLUS and comparison APs first hired in each SY. This limited our statistical power to gauge whether small or moderate differences were statistically significant rather than potentially resulting from random chance. It also limited the number of theoretically relevant covariate controls that we could usefully account for in our models.

Another limitation of our analysis of APs' retention and promotion outcomes is that we could not observe whether Leaders were systematically different from other new APs in terms of their roles and responsibilities in schools. The AP role in this district typically focused on one of three areas: pupil services, curriculum and instruction, or general. APs in the curriculum and instruction role were primarily responsible for overseeing the instructional program and curriculum materials and coaching and evaluating teachers, among other building and staff management tasks. APs in the pupil services role were primarily responsible for addressing student discipline issues and interacting with families and community members, as well as some building-management tasks. APs in the general role were expected to focus on pupil services and on curriculum and instruction. Although we know from our interview data that different Leaders served in all three areas in their schools, we do not know if, on average, Leaders were more likely to work in the area of curriculum and instruction. If Leaders differed in this way, and if APs' role assignments were associated with differences in job retention or promotion, our analyses could not control for this directly and could potentially be biased as a result.

A separate limitation of our analyses of retention and promotion was a lack of information about transitions to roles that were not related to school-building administration. Individuals who moved to the district central office or another district role not included in the data on school administrators were not counted as promotions and were counted as departures (in both PLUS and comparison groups) simply because we could not observe these outcomes among the comparison sample. We know that one Leader moved to the central office in this way but do not know if any comparison APs did so during the time frame of the study.

A final limitation pertained to our analysis of AP promotions. We could not compare the rate at which APs who left the district were promoted into principal roles. Although Leaders who left the district usually went on to principal roles, we could not confirm the rates at which promotions occurred for comparison APs who left the district. As a result, although most Leaders who left the district did so to take on principal roles in other districts, we could not evaluate whether this was similar to or different from promotion rates for comparison APs who left the district.

**Table A.6. Characteristics of New PLUS and New Non-PLUS Assistant Principals**

| | PLUS APs | | Non-Plus APs | |
|---|---|---|---|---|
| **Panel A: One- and Two-Year Retention** | **Mean** | **Standard Deviation** | **Mean** | **Standard Deviation** |
| AP characteristics | | | | |
| Female | 0.484 | 0.508 | 0.673 | 0.474 |
| White | 0.484 | 0.508 | 0.404 | 0.495 |
| Experience in district | 6.258 | 6.537 | 9.654 | 8.595 |
| School characteristics | | | | |
| In elementary school | 0.258 | 0.445 | 0.308 | 0.466 |
| In middle school | 0.290 | 0.461 | 0.212 | 0.412 |
| In high school | 0.290 | 0.461 | 0.385 | 0.491 |
| In K–8 school | 0.129 | 0.341 | 0.077 | 0.269 |
| In other type of school | 0.032 | 0.180 | 0.019 | 0.139 |
| School lagged ELA score | –0.213 | 0.529 | –0.072 | 0.586 |
| School lagged mathematics score | –0.228 | 0.524 | –0.129 | 0.586 |
| Percentage minority of student body | 0.912 | 0.075 | 0.894 | 0.098 |
| School enrollment | 757.045 | 505.934 | 765.887 | 549.681 |
| *N* (APs) | 31 | | 51 | |
| **Panel B: Three-Year Retention** | | | | |
| AP characteristics | | | | |
| Female | 0.524 | 0.512 | 0.730 | 0.450 |
| White | 0.476 | 0.512 | 0.459 | 0.505 |
| Experience in district | 6.190 | 6.690 | 10.568 | 8.811 |
| In elementary school | 0.238 | 0.436 | 0.324 | 0.475 |
| In middle school | 0.286 | 0.463 | 0.216 | 0.417 |
| In high school | 0.286 | 0.463 | 0.378 | 0.492 |
| In K–8 school | 0.143 | 0.359 | 0.081 | 0.277 |
| In other type of school | 0.048 | 0.218 | 0.000 | 0.000 |
| School lagged ELA score | –0.266 | 0.419 | –0.142 | 0.538 |
| School lagged mathematics score | –0.301 | 0.429 | –0.182 | 0.562 |
| Percentage minority of student body | 0.911 | 0.084 | 0.898 | 0.084 |
| School enrollment | 742.781 | 470.171 | 834.574 | 575.056 |
| *N* (APs) | 21 | | 37 | |
| **Panel C: Four-Year Retention** | | | | |
| AP characteristics | | | | |
| Female | 0.545 | 0.522 | 0.625 | 0.500 |
| White | 0.545 | 0.522 | 0.438 | 0.512 |
| Experience in district | 5.636 | 6.038 | 9.938 | 8.037 |
| School characteristics | | | | |
| In elementary school | 0.273 | 0.467 | 0.375 | 0.500 |
| In middle school | 0.364 | 0.505 | 0.188 | 0.403 |
| In high school | 0.182 | 0.405 | 0.375 | 0.500 |
| In K–8 school | 0.091 | 0.302 | 0.063 | 0.250 |
| In other type of school | 0.091 | 0.302 | 0.000 | 0.000 |
| School lagged ELA score | –0.235 | 0.341 | –0.209 | 0.529 |
| School lagged mathematics score | –0.272 | 0.423 | –0.266 | 0.564 |
| Percentage minority of student body | 0.924 | 0.070 | 0.917 | 0.060 |
| School enrollment | 706.490 | 414.015 | 807.078 | 618.542 |
| *N* (APs) | 11 | | 16 | |

### *Evaluating PLUS Effects on Coached Teachers' Student Achievement*

**Methods**: Coaching may improve individual teachers' instructional effectiveness in the year in which coaching is delivered, and it may also improve teachers' effectiveness in subsequent years and with future students that they teach. However, in our evaluation, we limited our focus to the effects of Leader coaching on students in the year in which a teacher actually received coaching.

We focused our analysis on the contemporaneous effects of coaching for several reasons. First, we hypothesized that focusing on the effects of coaching in the year(s) when it is delivered to teachers would allow us to detect a mix of impacts that might stem from either permanent improvement in teachers' practices or from any short-term changes in teachers' practices associated with Leaders' influence in the year that coaching was delivered. Separately, measuring longer-term effects on teachers was complicated by a variety of factors, including the inability of our data to show whether some teachers continued to receive coaching after their Leader graduated from PLUS and our inability to measure post-coaching effects for later PLUS cohorts within the time frame of our study. Finally, changes in teaching assignments over time moved some coached teachers into and out of tested grades and subjects from one year to the next. The instability of the sample of teachers that taught tested subjects and grades over time represents a type of attrition that would have reduced our effective sample size and could potentially bias any results. For all of these reasons, we limited our focus to testing for the contemporaneous effects of coaching. This test afforded us the best statistical power for measuring effects that we expected might be small or moderate in size.

To assess the contemporaneous impacts of PLUS on academic outcomes for the students of coached teachers, we compared outcomes for students taught by PLUS-coached teachers (the "treatment group") with outcomes for comparable students taught by teachers who taught in the same school and year but who were taught only by teachers who have never received PLUS coaching ("comparison group"). We used coarsened exact matching (CEM) to match students who were similar on a few key dimensions (details below), but our primary method for ensuring a fair comparison was the control variables included in our statistical models. When conducting the matching, we created separate matched samples by PLUS cohort and year. Thus, we created separate samples and weights for each PLUS cohort year combination. We then pooled our data on treated and comparison teachers across SYs and cohorts to estimate an overall average effect of coaching during the period of the study. Our pooled sample included coached teachers in each of three SYs and across all three coached cohorts, but we lacked records of which teachers were coached in SY 2015–2016, the first year in which the first PLUS cohort was in the program.

Our model for estimating the impacts of PLUS on student-achievement and attendance outcomes is described in the following equation:

$$Y_{itsy} = \beta_0 + \beta_1 PLUS_{itsy} + \boldsymbol{X}_{itsy}\beta_2 + \boldsymbol{\theta}_{ts}\beta_3 + \boldsymbol{\phi}_{sy}\beta_4 + \gamma_s + \alpha_y + \varepsilon_{itsy}$$

where

- $Y_{itsy}$ = outcome for student ($i$) taught by teacher ($t$) in school ($s$) in cohort-year ($y$). Our two achievement outcomes are student-standardized test scores in grades 4–8 in ELA and mathematics, with test scores standardized within SY and grade level.
- $PLUS_{itsy}$ = an indicator for a student in a classroom of a coached teacher in a given year
- $\boldsymbol{X}_{itsy}$ = a vector of student characteristics, including prior-year ELA and mathematics scores and a quadratic term for those scores,[1] prior SY attendance, student ethnicity, home language, an indicator for the student's mother having a college degree or higher, whether the student was retained in grade last year, English-language learner (ELL) status, special education status, and grade-level fixed effects
- $\boldsymbol{\theta}_{ts}$ = a vector of teacher-level characteristics, including an indicator for having a multi-subject teaching credential and a single-subject teaching credential, as well as the teacher's race, gender, age, and years of experience in the district. Additional covariates include indicators for other types of credentials, indicators for being in a teaching role related to ELL courses or to special education courses, and an indicator for being a first-year teacher. If a student is matched to more than one teacher, then the above characteristics are averaged between teachers
- $\boldsymbol{\phi}_{sy}$ = indicators for the cohorts of PLUS administrators that a school houses in a cohort year
- $\gamma_s$ = a set of school fixed effects
- $\alpha_y$ = a set of cohort-year fixed effects
- $\varepsilon_{itsy}$ = a student-level idiosyncratic error term. Standard errors are clustered at the teacher level because treatment assignment was determined at the level of teachers within individual schools.

We included school fixed effects to ensure a fairer contrast between PLUS-coached and non-coached teachers. The presence of the school fixed effect essentially means that our model is estimating the differential performance of students taught by PLUS-coached teachers relative to other students at the same school. This allowed us to more precisely assess whether the Leader coaching contributed to student learning specifically in coached teachers' classrooms. However, if Leaders were also contributing to overall schoolwide student academic performance, our model would explicitly filter that out. We included data from teachers and students in non-PLUS schools in this analysis solely to facilitate more-accurate modeling of the influence of our teacher and student covariates. We included cohort-year fixed effects to account for the pooling nature of the sample and the fact that matches were made within cohort year. We included indicators for having more than one PLUS administrator from different cohorts in a school in any cohort year to account for possible dosage effects of having more than one PLUS program cohort represented.

[1] Although some students in SY 2017–2018 would have been "treated" by PLUS in SY 2016–2017, we assume that any benefits of PLUS in the prior year would be adequately reflected in their SY 2016–2017 test scores.

As a sensitivity check, we also ran versions of our models with controls for school characteristics in lieu of school fixed effects. Results were similar but somewhat attenuated relative to our preferred specification using school fixed effects. These results are available upon request.

**Weights:** Our primary means of ensuring a fair comparison between students potentially impacted by Leader coaching of teachers and other district students was to include a robust set of controls for student and teacher characteristics, as well as school fixed effects. All of these factors might be related to student-achievement outcomes. In addition, we used sampling weights generated by CEM (Iacus, King, and Porro, 2012).

The goal of weighting observations in our comparison sample was to ensure that greater weight was given to comparison students who were similar on a few key dimensions to our treated students. This helped ensure that our comparison was conducted within a sample of reasonably similar students such that the estimated coefficients on the control variables in our model were appropriate. Weighting our sample this way also helped ensure that our weighted treatment and comparison samples were equivalent at baseline in terms of key outcomes measures, which is a requirement for quasi-experimental studies to meet What Works Clearinghouse evidence standards with reservations.

We used CEM to create strata (or *bins*) of students corresponding to all possible permutations of a few key characteristics that may have been relevant to the outcome measures and that reflect differences in underlying student learning contexts. These characteristics were "coarsened" by defining researcher-selected thresholds that we theorized might represent meaningfully different types of students. Every student in our treatment sample was then matched "exactly" to any number of comparison students in identical bins for every characteristic included in our matching. We focused the matching on a limited set of criteria to minimize loss of the treatment group sample (which can occur if no comparable student is available in a bin) while still usefully narrowing the comparison sample included in our analysis.

More specifically, we weighted students according to the following four characteristics: ELA and mathematics scores in the year prior to our outcome measures, student rates of absence from school in the year prior, and student grade levels. For all baseline outcomes, we defined four bins corresponding to quartiles of the distribution. Therefore, students were matched based on where they fell in each of the four bins for each of the matching variables. The weight of any control students in each bin corresponded to the ratio between the number of treatment and control students in that bin. For baseline student grade levels, we bucketed students into their specific grade level group.

**Samples and sample restrictions:** Our analysis spanned three SYs and three PLUS cohorts. We focused only on students in grades 4–8 and excluded from our analysis any students for whom we did not have both current and prior-year test scores in both subjects. We were able to match (and generate CEM weights) for more than 99 percent of students in our treatment group. However, some potential comparison students were excluded from the analysis if they were not

sufficiently similar on our matching variables. A total of 73 percent of Leaders had at least one coached teacher included in this analysis.

We provide descriptive statistics about our initial and CEM-matched analytic samples in in Table A.7 for ELA and Table A.8 for mathematics, which show weighted and unweighted samples for each SY sample and separately for ELA versus mathematics. Overall, as expected, we achieved a close balance on those student-level measures that we explicitly matched and weighted against, with all treatment and control groups well within What Works Clearinghouse requirements on our baseline outcome measures. For all other differences between the groups, we relied on our covariate controls to ensure a fair comparison. We also report in these tables on a variety of student- and teacher-level characteristics that we do not include in our matching but are controlled for in our models. For these characteristics, there are, at times, meaningful differences between our treatment and control groups, and we rely on our statistical model to adjust for any associations between these teacher and student characteristics and our outcome measures that might otherwise bias our results.

When linking PLUS-coached teachers to either ELA or mathematics outcomes, we considered only students with a PLUS-coached math teacher as the "treated" group for analysis of mathematics outcomes. Similarly, for ELA outcomes, we focused only on students of coached teachers who actually teach ELA in cases in which students had different teachers in the two subject areas.

Finally, our analysis restricted the pool of potential comparison students by excluding any students with current teachers whom we identified as ever having received Leader coaching (regardless of subject area). This helped ensure that our results were not biased by comparing the students of currently coached teachers with students of previously coached teachers. Unfortunately, we lack a record of which teachers that PLUS Cohort I Leaders may have coached in SY 2015–2016. As a result, we were unable to exclude students of teachers in Leaders' schools who may have had prior exposure to Leader coaching but who were not still receiving coaching in SY 2016–2017. This may have biased effect estimates downward slightly to the extent that some of the control students may have been benefiting from any persistent advantages of the coaching that was delivered in SY 2015–2016. In other words, the presence of previously coached teachers in the same schools may have made for a slightly more challenging within-school reference point in our analyses of the relative performance of coached teachers in one of the three cohorts that contributes to this analysis.

**Table A.7. Characteristics of Treated and Untreated Students in Pooled ELA Coaching Analytical Sample, With and Without Weighting**

| | Analytical Sample | | | | Weighted Analytical Sample | | | |
|---|---|---|---|---|---|---|---|---|
| | PLUS | | Comparison Group | | PLUS | | Comparison Group | |
| Lagged student characteristics | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation |
| ELA achievement* | –0.049 | 0.900 | –0.027 | 0.944 | –0.048 | 0.900 | –0.032 | 0.926 |
| Math achievement* | –0.071 | 0.896 | 0.007 | 0.898 | –0.072 | 0.896 | –0.055 | 0.912 |
| Percentage absent* | 3.103 | 4.077 | 3.004 | 3.817 | 3.107 | 4.080 | 3.066 | 3.924 |
| Special education | 0.105 | 0.306 | 0.110 | 0.312 | 0.105 | 0.306 | 0.133 | 0.339 |
| ELL | 0.230 | 0.421 | 0.225 | 0.418 | 0.230 | 0.421 | 0.213 | 0.410 |
| Retained | 0.000 | 0.017 | 0.001 | 0.038 | 0.000 | 0.017 | 0.000 | 0.021 |
| Black | 0.054 | 0.226 | 0.061 | 0.239 | 0.054 | 0.226 | 0.071 | 0.257 |
| White | 0.118 | 0.323 | 0.165 | 0.371 | 0.118 | 0.323 | 0.156 | 0.363 |
| Asian | 0.450 | 0.498 | 0.442 | 0.497 | 0.450 | 0.498 | 0.438 | 0.496 |
| Hispanic | 0.288 | 0.453 | 0.230 | 0.421 | 0.288 | 0.453 | 0.246 | 0.431 |
| Other | 0.068 | 0.251 | 0.073 | 0.260 | 0.068 | 0.251 | 0.068 | 0.251 |
| Home language English | 0.423 | 0.494 | 0.489 | 0.500 | 0.424 | 0.494 | 0.474 | 0.499 |
| Home language Spanish | 0.220 | 0.414 | 0.160 | 0.366 | 0.220 | 0.414 | 0.175 | 0.380 |
| Home language Chinese | 0.256 | 0.437 | 0.271 | 0.445 | 0.256 | 0.436 | 0.266 | 0.442 |
| Home language other | 0.101 | 0.301 | 0.080 | 0.272 | 0.101 | 0.301 | 0.085 | 0.278 |
| Mother's education: college or higher | 0.154 | 0.361 | 0.126 | 0.332 | 0.155 | 0.362 | 0.169 | 0.375 |
| Grade level* | 5.489 | 1.094 | 4.809 | 1.451 | 5.486 | 1.091 | 5.435 | 1.082 |
| Teacher characteristics | | | | | | | | |
| Age | 38.575 | 9.891 | 41.330 | 10.786 | 38.580 | 9.892 | 40.762 | 10.205 |
| Years of experience in district | 7.062 | 6.281 | 9.562 | 7.808 | 7.065 | 6.283 | 9.245 | 7.222 |
| Female | 0.594 | 0.360 | 0.666 | 0.415 | 0.594 | 0.360 | 0.676 | 0.376 |
| Black | 0.041 | 0.161 | 0.063 | 0.207 | 0.041 | 0.161 | 0.072 | 0.203 |
| Hispanic | 0.145 | 0.293 | 0.110 | 0.288 | 0.145 | 0.294 | 0.085 | 0.232 |
| White | 0.480 | 0.384 | 0.483 | 0.431 | 0.480 | 0.384 | 0.520 | 0.382 |
| Asian | 0.179 | 0.328 | 0.219 | 0.365 | 0.179 | 0.328 | 0.187 | 0.297 |
| Other | 0.155 | 0.277 | 0.125 | 0.285 | 0.155 | 0.277 | 0.137 | 0.263 |
| Special education teacher | 0.008 | 0.080 | 0.005 | 0.068 | 0.008 | 0.080 | 0.008 | 0.085 |
| ELL teacher | 0.156 | 0.342 | 0.515 | 0.485 | 0.156 | 0.342 | 0.244 | 0.403 |
| Single subject credential | 0.643 | 0.383 | 0.342 | 0.413 | 0.643 | 0.383 | 0.515 | 0.399 |
| Multiple subject credential | 0.465 | 0.408 | 0.640 | 0.423 | 0.465 | 0.408 | 0.465 | 0.404 |
| Administrative credential | 0.000 | 0.000 | 0.016 | 0.110 | 0.000 | 0.000 | 0.012 | 0.083 |
| Other credential | 0.097 | 0.255 | 0.112 | 0.272 | 0.097 | 0.254 | 0.143 | 0.285 |
| New teacher | 0.160 | 0.367 | 0.136 | 0.343 | 0.160 | 0.367 | 0.121 | 0.327 |
| *N* of students | 3,402 | | 56,202 | | 3,396 | | 45,665 | |
| *N* of schools | 20 | | 89 | | 19 | | 80 | |

NOTES: *Denotes inclusion in the CEM weighting. Models control for all variables shown.

**Table A.8. Characteristics of Treated and Untreated Students in Pooled Math Coaching Analytical Sample, With and Without Weighting**

| | Analytical Sample | | | | Weighted Analytical Sample | | | |
|---|---|---|---|---|---|---|---|---|
| | PLUS | | Comparison Group | | PLUS | | Comparison Group | |
| Lagged student characteristics | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation |
| ELA achievement* | –0.227 | 0.945 | –0.030 | 0.943 | –0.228 | 0.945 | –0.247 | 0.962 |
| Math achievement* | –0.251 | 0.939 | 0.003 | 0.897 | –0.251 | 0.940 | –0.253 | 0.939 |
| Percentage absent* | 3.627 | 4.357 | 3.022 | 3.839 | 3.625 | 4.359 | 3.494 | 4.185 |
| Special education | 0.116 | 0.320 | 0.110 | 0.313 | 0.116 | 0.320 | 0.155 | 0.362 |
| ELL | 0.278 | 0.448 | 0.226 | 0.419 | 0.279 | 0.448 | 0.257 | 0.437 |
| Retained | 0.001 | 0.030 | 0.001 | 0.038 | 0.000 | 0.022 | 0.001 | 0.024 |
| Black | 0.085 | 0.279 | 0.062 | 0.241 | 0.085 | 0.280 | 0.091 | 0.288 |
| White | 0.104 | 0.306 | 0.164 | 0.370 | 0.104 | 0.305 | 0.144 | 0.351 |
| Asian | 0.358 | 0.479 | 0.441 | 0.496 | 0.358 | 0.480 | 0.386 | 0.487 |
| Hispanic | 0.374 | 0.484 | 0.231 | 0.421 | 0.375 | 0.484 | 0.285 | 0.451 |
| Other | 0.060 | 0.237 | 0.074 | 0.261 | 0.060 | 0.237 | 0.067 | 0.249 |
| Home language English | 0.432 | 0.495 | 0.489 | 0.500 | 0.432 | 0.495 | 0.482 | 0.500 |
| Home language Spanish | 0.296 | 0.456 | 0.160 | 0.367 | 0.296 | 0.456 | 0.210 | 0.407 |
| Home language Chinese | 0.184 | 0.388 | 0.270 | 0.444 | 0.184 | 0.388 | 0.226 | 0.418 |
| Home language other | 0.088 | 0.283 | 0.080 | 0.272 | 0.088 | 0.283 | 0.082 | 0.274 |
| Mother's education: college or higher | 0.131 | 0.337 | 0.122 | 0.327 | 0.130 | 0.336 | 0.165 | 0.371 |
| Grade level* | 5.267 | 1.186 | 4.802 | 1.445 | 5.264 | 1.180 | 5.172 | 1.205 |
| Teacher characteristics | | | | | | | | |
| Age | 36.756 | 7.519 | 41.248 | 10.751 | 36.761 | 7.524 | 40.037 | 10.257 |
| Years of experience in district | 5.006 | 4.610 | 9.486 | 7.784 | 5.007 | 4.613 | 8.713 | 7.265 |
| Female | 0.586 | 0.369 | 0.672 | 0.410 | 0.587 | 0.369 | 0.695 | 0.383 |
| Black | 0.091 | 0.220 | 0.067 | 0.216 | 0.091 | 0.220 | 0.086 | 0.237 |
| Hispanic | 0.121 | 0.304 | 0.107 | 0.283 | 0.121 | 0.304 | 0.097 | 0.255 |
| White | 0.555 | 0.378 | 0.482 | 0.429 | 0.555 | 0.379 | 0.497 | 0.401 |
| Asian | 0.160 | 0.279 | 0.218 | 0.364 | 0.160 | 0.279 | 0.189 | 0.311 |
| Other | 0.073 | 0.205 | 0.126 | 0.284 | 0.073 | 0.205 | 0.131 | 0.265 |
| Special education teacher | 0.004 | 0.063 | 0.005 | 0.066 | 0.004 | 0.063 | 0.008 | 0.084 |
| ELL teacher | 0.237 | 0.404 | 0.504 | 0.485 | 0.238 | 0.404 | 0.311 | 0.442 |
| Single subject credential | 0.642 | 0.423 | 0.337 | 0.408 | 0.641 | 0.423 | 0.435 | 0.405 |
| Multiple subject credential | 0.374 | 0.395 | 0.646 | 0.420 | 0.374 | 0.395 | 0.515 | 0.420 |
| Administrative credential | 0.013 | 0.080 | 0.016 | 0.111 | 0.013 | 0.080 | 0.011 | 0.086 |
| Other credential | 0.124 | 0.270 | 0.112 | 0.273 | 0.124 | 0.270 | 0.175 | 0.318 |
| New teacher | 0.320 | 0.466 | 0.133 | 0.340 | 0.320 | 0.467 | 0.150 | 0.357 |
| *N* of students | 2,158 | | 56,068 | | 21,54 | | 36,150 | |
| *N* of schools | 20 | | 89 | | 19 | | 80 | |

NOTES: *Denotes inclusion in the CEM weighting. Models control for all variables shown.

**Limitations:** One limitation of our method for estimating coaching effects on student achievement in coached teachers' classrooms was the possibility that Leaders may have influenced school-wide outcomes in ways other than their effects on the students of coached teachers. To the extent that Leaders were able to improve student outcomes school-wide, this would have biased our estimates of their contributions to students of coached teachers downward. Our approach therefore should be interpreted as measuring the additional impacts of coaching, over and above any school-wide impacts of Leaders. However, as described later in this technical appendix, we did not find detectable effects of Leaders on school-wide academic achievement outcomes.

A second and more concerning limitation of our analysis of the impacts of Leaders' coaching was that assignment to receive PLUS coaching was nonrandom and was instead determined by the Leaders and principals at schools. In fact, Leaders were encouraged to identify teachers who were not receiving other formal coaching support from the district to coach. In this district, teachers most likely to be receiving other district coaching would have been novice teachers who participated in a district-wide induction program that included coaching supports. More-experienced teachers would typically not have received coaching. Teachers were also selected for PLUS coaching in part based on the match between each teacher's subject area and the AP's areas of instructional expertise. The intentional sorting of coaching to teachers without alternative sources of coaching support explicitly influences our interpretation of any PLUS effects that we identify. Instead of measuring the overall impact of Leader coaching, we interpret our estimates of the impact of Leader coaching as the relative effects of Leader coaching to business-as-usual instructional supports that were likely provided to comparison teachers in the district.

In addition to the intentional sorting of coaching assignments to teachers with fewer other instructional supports, it is plausible that Leaders and principals may have tended to assign coaching to teachers based on their perceived prior effectiveness. If this occurred, it would most likely bias our estimates downward if we assume that coaching was more likely to be targeted to teachers who were perceived as needing more support and that principals' perceptions of teachers' effectiveness were accurate. Unfortunately, for most coached teachers, we lacked historical data that would allow us to determine their prior value-added to their students' achievement in years before receiving coaching. We account for this in our models. This was in part because many coached teachers were new to teaching or had not previously taught in tested grades and subjects and because we lacked sufficient historical data linking teachers to students to evaluate prior-year value-added estimates for teachers in the first year of our coaching analysis, SY 2016–2017.

As a specification check, we examined coaching effects for the minority of coached teachers for whom we could estimate prior-year value-added. Although this sample differed substantially from our larger sample of interest, for this group, we were able to estimate Leader coaching

effects in two specifications.[2] First, we estimated PLUS effects without controlling for prior-year value-added, and then we estimated PLUS effects when controlling for prior-year value-added. Results of this specification check are provided in Table A.9. Details on how we estimated teacher value-added is provided later in this technical appendix.

**Table A.9. Estimates of the Impact of Leader Coaching on Student Achievement for a Subsample of PLUS-Coached and Comparison Teachers, With and Without Controls for Teachers' Prior-Year Value-Added**

| | ELA | | | | Mathematics | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1: No Prior-Year Value-Added | | Model 2: With Prior-Year Value-Added | | Model 1: No Prior-Year Value-Added | | Model 2: With Prior-Year Value-Added | |
| | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error |
| PLUS coached | 0.039* | (0.019) | 0.039* | (0.019) | –0.105** | (0.031) | –0.070* | (0.031) |
| Control for prior-year value-added | No | | Yes | | No | | Yes | |
| *N* (student observations) | 19,251 | | 19,251 | | 10,824 | | 10,824 | |
| *N* (coached teachers) | 25 | | 25 | | 11 | | 11 | |
| *N* (comparison teachers) | 324 | | 324 | | 289 | | 289 | |

NOTES: Samples consist of students taught by teachers who taught ELA or mathematics in grades 4–8 and for whom prior-year value-added performance metrics could be estimated. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

For this subgroup of coached teachers, estimated effects of Leader coaching were significant and positive in ELA and significant and negative in math, regardless of whether we controlled for prior-year value-added. Our estimates in ELA were almost identical regardless of whether or not we controlled for prior-year value-added, which would suggest that our other model controls, including controls for teacher experience, may have been sufficient to ensure a fair comparison. However, in mathematics, our estimates when not controlling for prior-year value-added were directionally more negative, which suggests that, in the absence of controls for prior-year value-added, our estimates of PLUS effects might have been negatively biased. It is possible that this apparent bias, if also present in our overall sample, might have contributed to our lack of findings of a PLUS coaching effect on students' mathematics achievement. Overall, both the evidence from this specification check and our uncertainty about Leaders' priorities when

[2] We also directly examined the prior-year value-added estimates of PLUS-coached and comparison teachers. We found that PLUS-coached teachers had, on average, directionally lower estimates. However, this descriptive comparison does not adjust for teacher experience levels, which is a key control variable included in our models because PLUS-coached teachers were less experienced on average than comparison teachers.

identifying teachers to coach point to the possibility that our estimates of PLUS effects on student achievement in our full sample may have understated the impacts of PLUS. As we discuss in the conclusion section of the main report, it is notable that we identified an overall positive effect of Leader coaching given both this and other factors that most likely would have led to a conservative bias in our estimates of Leader coaching effects on student achievement.

Finally, as previously discussed, we focused exclusively on the effects of Leaders' coaching in the year in which it was provided. We believed this analysis provided the best statistical power to detect any effects and was therefore the clearest way to test for evidence of any impacts of Leader coaching on students. We did not look for and do not provide an estimate of the extent to which coached teachers' instructional performance may have been permanently affected by their exposure to Leaders' coaching.

### *Evaluating PLUS Effects on Coached Teachers' Retention*

**Methods**: Coaching by Leaders could in theory improve support for teachers and improve experience on the job and thus encourage higher rates of teacher retention in the district. At the same time, Leader coaching involved frequent cycles of teacher observation and evaluation, which may have called attention to teachers who were consistently performing poorly. This could have led to lower retention rates for lower-performing teachers in particular.

Our methods for evaluating PLUS effects on coached teacher retention were similar to those we used to evaluate effects on coached teachers' student-achievement outcomes. To assess whether PLUS affected retention rates, we compared the cumulative rate at which coached teachers left the district over time with the departure rate of other teachers who did not receive PLUS coaching.

Our model for estimating the impacts of PLUS on coached teacher-retention rates is shown in the following equation:

$$Y_{isy} = \beta_0 + \beta_1 PLUS_{isy} + \boldsymbol{X}_{isy}\beta_2 + \gamma_s + \boldsymbol{\phi}_{sy} + \alpha_y + \varepsilon_{isy}$$

where

- $Y_{isy}$ = an indicator for whether teacher ($i$) in school ($s$) has left by year ($y$). $Y_{isy}$ is a cumulative measure marked as having left in all future years, such that a teacher who left in a previous year is retained in the sample
- $PLUS_{isy}$ = an indicator for a teacher who was coached in a given year
- $\boldsymbol{X}_{isy}$ = a vector of teacher characteristics, including race or ethnicity, gender, age, years of experience in the district, indicators for being in a teaching role related to ELL courses or special education courses
- $\gamma_s$ = a set of school fixed effects
- $\boldsymbol{\phi}_{sy}$ = time-varying school-level characteristics, such as proportion of the student body by race or ethnicity, home language, classified as ELL, and reclassified as English-

language proficient; the proportion of the student body with mothers with a college education is also included

- $\alpha_y$ = a set of year fixed effects
- $\varepsilon_{isy}$ = a student-level idiosyncratic error term. Standard errors are clustered at the school level to account for correlation in teacher retention decisions driven by school-wide factors.

We included school fixed effects in our models that estimate the effects on teacher retention. This effectively means coached teachers' retention rates are compared with those of other teachers in the same school. We believed that this was most appropriate because individual schools likely varied substantially in the rate at which they were successful in retaining teachers. We controlled for time-varying school characteristics to account for changes in school composition that could have affected teacher-retention decisions. In addition, to account for the pooled-by-year nature of the sample and yearly shocks that could contribute to retention decisions, we also included year fixed effects in our model.

**Weights:** Our analysis of PLUS impacts on student-achievement gains relied on both our covariate controls in our model and on weights to help ensure as fair of a comparison as possible. For the analysis of teacher retention rates, we used the same CEM matching technique to generate weights based on the following teacher characteristics: teacher age, gender, experience in the district, ethnicity, an indicator for special education teacher roles, and an indicator for ELL teacher roles.

**Sample:** Our analysis spanned three SYs and three PLUS cohorts, although different cohorts were represented in our analyses of year 1, 2, and 3 retention. Using the CEM weights, we were able to find suitable teacher matches for 99 percent of teachers in our PLUS-treated group. We did, however, exclude more potential comparison teachers in our analysis when they did not match according to the variables used in the CEM bins. A descriptive comparison of coached and non-coached samples before and after weighting is provided in Table A.10.

**Limitations:** Many of the limitations that we previously discussed of evaluating PLUS effects on coached teachers' student-achievement gains also apply here. The most important limitation is the fact that our estimates of the effects of Leader coaching on teacher retention rates may be biased if teachers were systematically assigned to receive Leader coaching based on unobserved characteristics that also made them more or less likely to remain teaching in the district. This could be the case, for example, if coaching tended to be allocated to teachers that principals viewed as lower performing and less likely to remain teaching in the district.

**Retention effects for teacher performance subgroups:** Increasing teacher-retention rates was a goal of Leaders in most cases but not when teachers were lower performing. Leaders intentionally sought to retain more-effective teachers and dismiss or counsel out ineffective teachers. To explore this distinction, we examined whether there was any evidence of differential PLUS effects on coached teacher retention for subgroups of teachers whose students had particularly low achievement gains.

In this analysis, we focused on teachers whose student-achievement gains could be observed (i.e., teachers who taught in tested subjects in grades 4–8). From this group, we distinguished between teachers whose average student-achievement gains were in the bottom 20 percent of the distribution of student-achievement gains, when controlling for observable student and classroom factors and school fixed effects, and the top 80 percent of performers. We chose this low threshold of performance because estimates of teachers' contributions to student achievement in any one year tend to be imprecise. However, prior research suggests that teachers whose students perform in the bottom quintile in one year are substantially more likely to also produce subpar student-achievement gains in future years as well (Loeb and Candelaria, 2012). Thus, this subgroup likely mostly represented teachers who were truly less effective than the average teacher at raising student achievement.

We used the following model to calculate teacher value-added measures:

$$Y_{itsy} = \beta_0 + \boldsymbol{X}_{itsy}\beta_1 + \boldsymbol{W}_{tsy}\beta_2 + \boldsymbol{C}_{tsy}\beta_3 + \boldsymbol{B}_{sy}\beta_3 + \varepsilon_{itsy}$$

where

- $Y_{itsy}$ = either the standardized reading or math school for student ($i$) assigned to teacher ($t$) in school ($s$) in year ($y$)
- $\boldsymbol{X}_{itsy}$ = a vector of student characteristics, including indicators for grade, classified as an ELL, reclassified as English-language proficient, classified as a special education student, ethnicity, home language, and mother having at least a college education; additional controls include a quadratic of prior-year standardized ELA and math scores
- $\boldsymbol{W}_{tsy}$ = teacher-level characteristics, including age, gender, experience in district, ethnicity, and role related to ELL or special-education instruction
- $\boldsymbol{C}_{tsy}$ = classroom-level characteristics of the variables contained in $\boldsymbol{X}_{itsy}$; this aggregated vector of characteristics includes all students across all classes taught by a teacher
- $\boldsymbol{B}_{sy}$ = school-level characteristics of the variables contained in $\boldsymbol{X}_{itsy}$
- $\varepsilon_{itsy}$ = a student-level idiosyncratic error term.

Per this equation, we predicted student achievement in the current year with student characteristics, classroom characteristics, school characteristics, and teacher characteristics. The residuals (the difference between a student's observed achievement and predicted achievement) were attributed to the teacher. We then averaged the student residuals by teacher and standardized the average by year. This standardized average became our estimate of a teacher's value-added measure. We created two value-added measures—one for ELA and one for math. In looking at teachers at the bottom quintile of value-added measures, we averaged the ELA and math value-added measures (where appropriate) and used the distribution of that average.

When estimating teacher value-added for these teachers, we used student achievement from the year in which each teacher was coached. We did so in part because the performances of the teachers in the year they were coached would have likely informed how Leaders managed them. In addition, because some coached teachers were in their first year of teaching and some experienced teachers were not teaching in grades 4–8 in a tested subject in the prior year, it was

not possible to estimate teacher value-added using historical data for a substantial portion of our sample. As we discuss below, using coached-year value-added can potentially influence the interpretation of any associations between teacher performance and Leader coaching.

**Subgroup sample:** As previously discussed, when evaluating possible retention effects among teachers with different levels of student achievement gains, our sample was restricted to teachers who taught reading or mathematics to student in grades 4–8. These were the only teachers for whom impacts on student achievement could be calculated. As a result, our sample size for this analysis was much smaller than for our overall retention analysis, and this subgroup analysis is not necessarily representative of impacts on the full sample of coached teachers. We did not employ CEM matching for analysis of this subgroup sample because our limited sample size made that infeasible. Descriptive information for this subsample is provided in Table A.11.

**Strengths and additional limitations of the subgroup analysis:** In contrast to our estimates of teacher retention among all coached teachers, for this subgroup analysis, we were less concerned about sorting of teachers to Leader coaching as a function of teacher performance. Instead, this analysis directly evaluates an impact that we believe may be dependent on teacher performance. As such, our estimates of this subgroup retention effect are more robust and less subject to bias than our estimates of overall PLUS effects on teacher retention.

In addition, however, a limitation of this subgroup analysis was that Leaders may have influenced both their coached teachers' performance with students in a given year and their likelihood of subsequently continuing to teach in the district. For example, if a Leader had a negative effect on a teacher's performance as a coach, this might lower the performance of students in that teacher's class that year while also influencing the teacher's future propensity to remain in the district. Unfortunately, we cannot rule out this potential interpretation of any PLUS effects on retention in this subgroup. However, it is doubtful that this potential bias had much influence on our estimates. We would not typically expect Leaders to have had a large enough effect on coached teachers' performance to shift the average achievement profile of his or her students into or out of the bottom quintile.

**Table A.10. Characteristics of Coached and Non-Coached Teachers in Pooled Retention Sample, With and Without Weighting**

| | Analytical Sample | | | | Weighted Analytical Sample | | | |
|---|---|---|---|---|---|---|---|---|
| | Coached | | Non-Coached | | Coached | | Non-Coached | |
| Year 1 retention | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation |
| Age | 36.827 | 11.115 | 42.358 | 12.230 | 36.776 | 11.162 | 36.447 | 10.441 |
| Female | 0.605 | 0.490 | 0.679 | 0.467 | 0.611 | 0.489 | 0.679 | 0.467 |
| Experience in district | 5.338 | 7.576 | 11.174 | 9.644 | 5.389 | 7.599 | 5.438 | 7.223 |
| Black | 0.056 | 0.231 | 0.049 | 0.215 | 0.057 | 0.232 | 0.056 | 0.230 |
| Hispanic | 0.241 | 0.429 | 0.146 | 0.353 | 0.238 | 0.427 | 0.171 | 0.377 |
| White | 0.431 | 0.496 | 0.442 | 0.497 | 0.430 | 0.496 | 0.445 | 0.497 |
| Asian | 0.179 | 0.385 | 0.243 | 0.429 | 0.181 | 0.386 | 0.230 | 0.421 |
| Other | 0.092 | 0.290 | 0.121 | 0.326 | 0.093 | 0.292 | 0.097 | 0.296 |
| Special education teacher | 0.092 | 0.290 | 0.075 | 0.263 | 0.083 | 0.276 | 0.077 | 0.266 |
| ELL teacher | 0.344 | 0.476 | 0.438 | 0.496 | 0.347 | 0.477 | 0.430 | 0.495 |
| *N* (sample) | 195 | | 7,695 | | 193 | | 6,352 | |
| *N* (unique teachers) | 195 | | 3,604 | | 193 | | 3,404 | |
| Year 2 retention | | | | | | | | |
| Age | 36.968 | 11.384 | 42.582 | 12.226 | 36.915 | 11.437 | 36.891 | 11.301 |
| Female | 0.608 | 0.490 | 0.683 | 0.465 | 0.615 | 0.488 | 0.688 | 0.463 |
| Experience in district | 5.486 | 7.748 | 11.418 | 9.651 | 5.542 | 7.773 | 5.874 | 7.767 |
| Black | 0.055 | 0.229 | 0.048 | 0.214 | 0.056 | 0.230 | 0.057 | 0.232 |
| Hispanic | 0.243 | 0.430 | 0.141 | 0.348 | 0.240 | 0.428 | 0.165 | 0.371 |
| White | 0.420 | 0.495 | 0.444 | 0.497 | 0.419 | 0.495 | 0.448 | 0.497 |
| Asian | 0.193 | 0.396 | 0.248 | 0.432 | 0.196 | 0.398 | 0.236 | 0.424 |
| Other | 0.088 | 0.285 | 0.119 | 0.324 | 0.089 | 0.286 | 0.094 | 0.293 |
| Special education teacher | 0.094 | 0.293 | 0.062 | 0.240 | 0.084 | 0.278 | 0.067 | 0.249 |
| ELL teacher | 0.348 | 0.478 | 0.450 | 0.498 | 0.352 | 0.479 | 0.443 | 0.497 |
| *N* (sample) | 181 | | 5,052 | | 179 | | 4,875 | |
| *N* (unique teachers) | 181 | | 3,155 | | 179 | | 3,059 | |
| Year 3 retention | | | | | | | | |
| Age | 36.713 | 12.250 | 42.886 | 12.394 | 36.605 | 12.362 | 36.520 | 11.726 |
| Female | 0.638 | 0.483 | 0.674 | 0.469 | 0.652 | 0.479 | 0.683 | 0.465 |
| Experience in district | 5.149 | 7.913 | 11.676 | 9.709 | 5.250 | 7.969 | 5.468 | 8.028 |
| Black | 0.043 | 0.203 | 0.047 | 0.211 | 0.043 | 0.205 | 0.055 | 0.228 |
| Hispanic | 0.223 | 0.419 | 0.142 | 0.349 | 0.217 | 0.415 | 0.168 | 0.374 |
| White | 0.479 | 0.502 | 0.437 | 0.496 | 0.478 | 0.502 | 0.446 | 0.497 |
| Asian | 0.213 | 0.411 | 0.279 | 0.449 | 0.217 | 0.415 | 0.265 | 0.442 |
| Other | 0.043 | 0.203 | 0.095 | 0.293 | 0.043 | 0.205 | 0.065 | 0.246 |
| Special education teacher | 0.021 | 0.145 | 0.005 | 0.073 | 0.000 | 0.000 | 0.000 | 0.000 |
| ELL teacher | 0.394 | 0.491 | 0.502 | 0.500 | 0.402 | 0.493 | 0.491 | 0.500 |
| *N* (sample) | 94 | | 2,229 | | 92 | | 2,160 | |
| *N* (unique teachers) | 94 | | 2,229 | | 92 | | 2,160 | |

**Table A.11. Characteristics of Coached and Non-Coached Teachers in Pooled Value-Added Model Retention Sample**

| | Coached | | Non-Coached | |
|---|---|---|---|---|
| Panel A: Year 1 retention | Mean | Standard Deviation | Mean | Standard Deviation |
| Age | 35.862 | 11.489 | 40.555 | 12.149 |
| Female | 0.584 | 0.496 | 0.684 | 0.465 |
| Experience in district | 4.455 | 6.214 | 8.962 | 8.639 |
| Black | 0.078 | 0.270 | 0.066 | 0.249 |
| Hispanic | 0.169 | 0.377 | 0.146 | 0.353 |
| White | 0.468 | 0.502 | 0.464 | 0.499 |
| Asian | 0.169 | 0.377 | 0.206 | 0.404 |
| Other | 0.117 | 0.323 | 0.119 | 0.324 |
| Special education teacher | 0.078 | 0.270 | 0.053 | 0.223 |
| ELL teacher | 0.312 | 0.466 | 0.558 | 0.497 |
| *N* (sample) | 77 | | 1,676 | |
| *N* (unique teachers) | 77 | | 873 | |
| Panel B: Year 2 retention | | | | |
| Age | 36.225 | 11.983 | 40.926 | 12.295 |
| Female | 0.580 | 0.497 | 0.675 | 0.469 |
| Experience in district | 4.594 | 6.404 | 9.099 | 8.640 |
| Black | 0.072 | 0.261 | 0.068 | 0.252 |
| Hispanic | 0.159 | 0.369 | 0.134 | 0.341 |
| White | 0.464 | 0.502 | 0.475 | 0.500 |
| Asian | 0.188 | 0.394 | 0.204 | 0.403 |
| Other | 0.116 | 0.323 | 0.119 | 0.324 |
| Special education teacher | 0.072 | 0.261 | 0.042 | 0.200 |
| ELL teacher | 0.304 | 0.464 | 0.562 | 0.496 |
| *N* (sample) | 69 | | 1,106 | |
| *N* (unique teachers) | 69 | | 719 | |
| Panel C: Year 3 retention | | | | |
| Age | 34.303 | 11.809 | 41.368 | 12.349 |
| Female | 0.595 | 0.498 | 0.665 | 0.472 |
| Experience in district | 3.865 | 6.156 | 9.398 | 8.573 |
| Black | 0.054 | 0.229 | 0.062 | 0.241 |
| Hispanic | 0.162 | 0.374 | 0.129 | 0.335 |
| White | 0.486 | 0.507 | 0.485 | 0.500 |
| Asian | 0.243 | 0.435 | 0.217 | 0.413 |
| Other | 0.054 | 0.229 | 0.108 | 0.310 |
| Special education teacher | 0.000 | 0.000 | 0.004 | 0.062 |
| ELL teacher | 0.378 | 0.492 | 0.588 | 0.493 |
| *N* (sample) | 37 | | 520 | |
| *N* (unique teachers) | 37 | | 520 | |

### *Impact Analysis Results Tables*

In the following tables, we detail the results of our primary analyses. These tables correspond to Figures 2 through 6 in the main report. They provide additional detail about each analysis, including sample sizes.

**Table A.12. Cumulative Rates of Novice APs No Longer Working as School Administrators in the District One, Two, or Three Years After First Becoming an AP**

| | Left Within 1 Year | | Left Within 2 Years | | Left Within 3 Years | | Left Within 4 Years | |
|---|---|---|---|---|---|---|---|---|
| | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error |
| PLUS APs | –0.074 | (0.054) | –0.049 | (0.083) | –0.135 | (0.109) | –0.150 | (0.162) |
| PLUS cohorts | Cohorts 1, 2, and 3 | | Cohorts 1, 2, and 3 | | Cohorts 1 and 2 | | Cohort 1 | |
| *N* (PLUS) | 31 | | 31 | | 21 | | 11 | |
| N (comparison) | 51 | | 51 | | 37 | | 16 | |

NOTES: Samples include only Leaders and comparison APs who were initially brand new to the role in the district. No results are statistically significant at the $p < 0.05$ threshold.

**Table A.13. Cumulative Rates of Novice APs Being Promoted to Principal Roles in the District One, Two, or Three Years After First Becoming an AP**

| | Promoted by Year 1 | | Promoted by Year 2 | | Promoted by Year 3 | | Promoted by Year 4 | |
|---|---|---|---|---|---|---|---|---|
| | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error |
| PLUS APs | n/a | n/a | –0.044 | (0.084) | 0.020 | (0.111) | 0.110 | (0.200) |
| PLUS cohorts | Cohorts 1, 2, and 3 | | Cohorts 1, 2, and 3 | | Cohorts 1 and 2 | | Cohort 1 | |
| *N* (PLUS) | 31 | | 31 | | 21 | | 11 | |
| *N* (comparison) | 51 | | 51 | | 37 | | 16 | |

NOTES: Samples include only Leaders and comparison APs who were initially brand new to the role in the district. No results are statistically significant at the $p < 0.05$ threshold. n/a = not applicable.

**Table A.14. Adjusted Student Achievement of Leader-Coached Teachers Relative to Comparison Teachers Across All PLUS Cohorts and Years**

| | ELA | | Mathematics | |
|---|---|---|---|---|
| | Coefficient | Standard Error | Coefficient | Standard Error |
| PLUS coached | 0.040* | (0.018) | –0.014 | (0.032) |
| *N* (student observations) | 49,061 | | 3,304 | |
| *N* (coached teachers) | 65 | | 43 | |
| *N* (comparison teachers) | 590 | | 531 | |

NOTES: Samples consist of students taught by teachers in grades 4–8 who taught ELA or mathematics between SYs 2016–2017 and 2018–2019. * = $p < 0.05$.

**Table A.15. Cumulative Rates of Teachers Leaving the District One, Two, or Three Years Following Receipt or Nonreceipt of Leader Coaching**

| | Left Within 1 Year | | Left Within 2 Years | | Left Within 3 Years | |
|---|---|---|---|---|---|---|
| | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error |
| PLUS coached | –0.035 | (0.025) | –0.042 | (0.037) | –0.015 | (0.063) |
| PLUS cohorts | Cohorts 1, 2, and 3 | | Cohorts 1, 2, and 3 | | Cohorts 1 and 2 | |
| *N* (coached) | 193 | | 179 | | 92 | |
| *N* (comparison) | 3,404 | | 3,059 | | 2,160 | |

NOTES: No results are statistically significant at the $p < 0.05$ threshold.

**Table A.16. Cumulative Rates of Teachers Leaving the District One, Two, or Three Years Following First Receipt or Nonreceipt of Leader Coaching, by Teachers' Student Achievement Gains**

| | Left Within 1 Year | | Left Within 2 Years | | Left Within 3 Years | |
|---|---|---|---|---|---|---|
| | Coefficient | Standard Error | Coefficient | Standard Error | Coefficient | Standard Error |
| Lower achievement gains sample | | | | | | |
| PLUS coached | 0.150 | (0.221) | 0.855** | (0.221) | 0.412 | (0.446) |
| PLUS cohorts | Cohorts 1, 2, and 3 | | Cohorts 1, 2, and 3 | | Cohorts 1 and 2 | |
| *N* (coached) | 9 | | 8 | | 7 | |
| *N* (comparison) | 288 | | 208 | | 104 | |
| Higher achievement gains sample | | | | | | |
| PLUS coached | 0.010 | (0.048) | 0.029 | (0.057) | 0.028 | (0.119) |
| PLUS cohorts | Cohorts 1, 2, and 3 | | Cohorts 1, 2, and 3 | | Cohorts 1 and 2 | |
| *N* (coached) | 68 | | 61 | | 30 | |
| *N* (comparison) | 746 | | 606 | | 416 | |

NOTES: Lower achievement gains sample is defined as the bottom quintile of teachers' student-achievement gains. Higher achievement gains sample is defined as the top four quintiles of teachers' student-achievement gains. Both samples consist only of teachers in grades 4–8 who taught ELA or mathematics. ** = $p < 0.01$.

## Exploratory Evaluation of Leader Effects on Schoolwide Achievement

In this section, we share our methodology and findings related to our exploratory analysis of the impacts of PLUS on school-wide achievement trajectories. In these analyses, we used panel data that spans SYs 2014–2015 through 2018–2019. We compared schools that eventually

adopted Leaders with schools that never adopted Leaders. A descriptive summary of these two groups of schools is provided in Table A.17.

### *Methods for Estimating Impacts of Leaders on Schoolwide Student Achievement*

A key concern in these analyses was the comparability of schools that hired Leaders and schools that did not. If these two types of schools differed on key unobserved dimensions, then the non-PLUS schools would not be a valid counterfactual for PLUS-led schools. Thus, we examined effects using two different methodologies: (1) an event study that non-parametrically estimates the effect of Leaders each year after adoption and (2) a comparative interrupted time series (CITS) approach that models the trends among schools that adopted Leaders and those that never did. The first approach assumed, and partially tested the assumption, that adopting and non-adopting schools had common pre-trends in student achievement. The second approach explicitly models those trends.

The following model illustrates our event-study approach:

$$Y_{ist} = \sum_{\tau=-x,\tau\neq 0}^{y} \delta_\tau \, PLUS_{st}^{\tau} + \boldsymbol{X_{ist}\beta_1} + \boldsymbol{S_{st}\beta_2} + \alpha_s + \gamma_t + \varepsilon_{ist}$$

where

- $Y_{ist}$ = either the standardized reading or math score for student ($i$) in school ($s$) in year ($t$)
- $\tau$ = time centered on the year before a school adopts a Leader; for schools that never adopt a Leader, $\tau = 0$, $\forall$ t
- $PLUS_{st}^{\tau}$ = indicator variables for the distance in years from centered time
- $\boldsymbol{X_{ist}}$ = a parsimonious vector of student characteristics that include race, an indicator for whether the mother has at least a college education, and grade-fixed effects
- $\boldsymbol{S_{st}}$ is a parsimonious vector of school characteristics, including the number of APs in a school, an indicator for getting any new AP in a given year, and an indicator for getting a new principal in a given year
- $\alpha_s$ = school-fixed effects
- $\gamma_t$ = year-fixed effects
- $\varepsilon_{ist}$ = a student-level idiosyncratic error term; standard errors are clustered at the school level.

In this model, we estimated student outcomes between students in schools that adopted Leaders $\tau$ quarters prior to $t$ and students in schools that did not adopt Leaders, minus stable differences between schools and years. This model provided two key pieces of information. First, it provided a test of any differential pretreatment trends between schools that adopted a Leader and schools that did not. If these schools were on a common trend, pre-trend would expect $\delta_\tau = 0$, $\forall\ \tau < 0$. Second, it estimated the effect of adopting a Leader on student outcomes separately

for each year after adoption. This allowed us to understand whether achievement increased or decreased as more time passed from the time of adoption.

Our results indicated that schools that adopted Leaders were trending downward in student achievement, relative to schools that never adopted a Leader, prior to adoption. Given these results, we employed the CITS approach as a second strategy for identifying the effects of Leaders on student outcomes. The following model illustrates this approach:

$$Y_{ist} = \beta_0 + year_t\beta_1 + year_t^2\beta_2 + PLUS_{st} * \tau_{st}\,\beta_3 + Post_{st} * PLUS_{st}\beta_4 + Post_{st} * PLUS_{st} * \tau_{st}\beta_5 + Post_{st} * PLUS_{st} * \tau_{st}^2\beta_6 + \boldsymbol{X_{ist}}\beta_7 + \alpha_s + \varepsilon_{ist}$$

where

- $Y_{ist}$ = either the standardized reading or math score for student ($i$) in school ($s$) in year ($t$)
- $year_t$ = a variable representing calendar time
- $PLUS_{st}$ = an indicator variable representing whether a school ever adopted a Leader
- $\tau_{st}$ = time centered on the year before a school adopts a Leader; for schools that never adopt a Leader, $\tau = 0, \forall$ t
- $Post_{st}$ = indicator for a school having adopted a Leader in a given year; it is an indicator for being in the "post-period" of Leader adoption
- $\boldsymbol{X_{ist}}$ = the same parsimonious vector of student characteristics used in the event study
- $\alpha_s$ = school-fixed effects
- $\varepsilon_{ist}$ = a student-level idiosyncratic error term; standard errors are clustered at the school level.

In this model, $\beta_1$ and $\beta_2$ control for a quadratic in the secular trend in student achievement for schools that never adopted Leaders. $\beta_3$ allows for a linear pre-trend in achievement for schools that adopt Leaders. $\beta_4$ allows for a one-time jump (or drop) in achievement when schools adopt a Leader. $\beta_4$ and $\beta_5$ allow for a quadratic post-trend in student achievement for school that adopted a Leader. In this model, the effect of Leaders on student achievement is determined by the difference in the quadratic post-trend, compared with the linear pre-trend of PLUS adopters and the secular trend for schools that never adopted Leaders, allowing for a one-time change in achievement in the year of PLUS adoption. School fixed effects control for average school differences among all schools in the sample. Figures A.1 and A.2 suggest that the functional forms used to model the secular trend and PLUS adopter time pre/post time trends are appropriate. Whereas the event study took a nonparametric approach to estimating the effect of Leaders on school achievement, the CITS approach explicitly modeled any difference in trends between schools that ultimately adopted Leaders and schools that never adopted Leaders.

**Table A.17. Characteristics of Ever-PLUS Schools and Never-PLUS Schools**

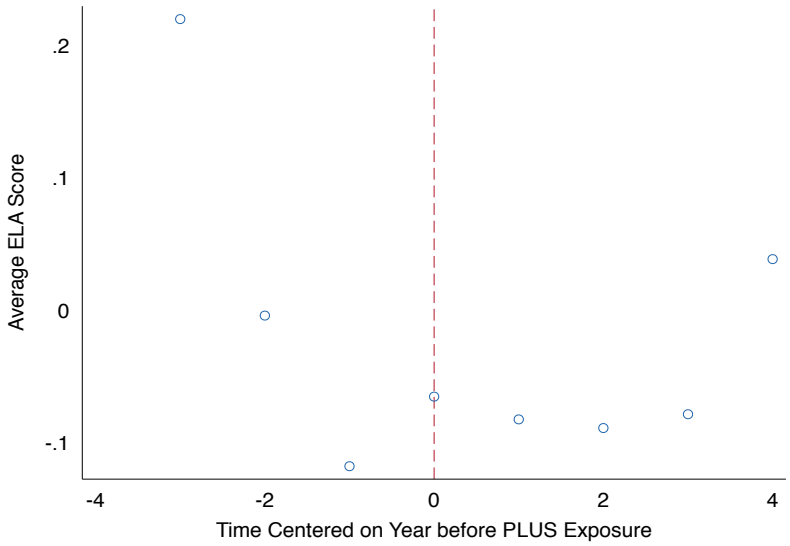
| | Ever PLUS Schools | | Never Plus Schools | |
|---|---|---|---|---|
| Panel A: 1- and 2-year retention | Mean | Standard Deviation | Mean | Standard Deviation |
| Black | 0.087 | 0.282 | 0.078 | 0.268 |
| White | 0.118 | 0.322 | 0.157 | 0.363 |
| Asian | 0.446 | 0.497 | 0.486 | 0.500 |
| Hispanic | 0.291 | 0.454 | 0.228 | 0.419 |
| Other | 0.059 | 0.235 | 0.052 | 0.223 |
| Home language English | 0.427 | 0.495 | 0.459 | 0.498 |
| Home language Spanish | 0.239 | 0.427 | 0.172 | 0.378 |
| Home language Chinese | 0.262 | 0.440 | 0.295 | 0.456 |
| Home Language other | 0.071 | 0.257 | 0.074 | 0.261 |
| ELL | 0.204 | 0.403 | 0.209 | 0.407 |
| Reclassified English-language proficient | 0.351 | 0.477 | 0.315 | 0.465 |
| Special education | 0.120 | 0.325 | 0.106 | 0.307 |
| Mother has college degree or higher | 0.194 | 0.396 | 0.207 | 0.405 |
| Percentage absent | 2.809 | 3.972 | 3.016 | 3.747 |
| Standardized mathematics score | –0.040 | 1.004 | 0.068 | 0.979 |
| Standardized ELA score | –0.057 | 0.994 | 0.065 | 0.989 |
| *N* (students) | 10,576 | | 14,066 | |
| *N* (schools) | 33 | | 67 | |

### *Estimated Impacts of Leaders on School-Wide Student Achievement*

Using the two approaches described previously, we explored whether hiring Leaders was associated with changes in school-performance trajectories. We did not find any clear indication that Leaders contributed in a significant way to changes in school-wide student achievement during the period of this study. Because we had limited historical achievement data for our first cohort of potentially impacted schools and limited post-PLUS data for our final cohort of schools, our results were sensitive to which cohorts we focused on and which models we used to examine them. Overall, our analyses of PLUS impacts on school-wide achievement were inconclusive.

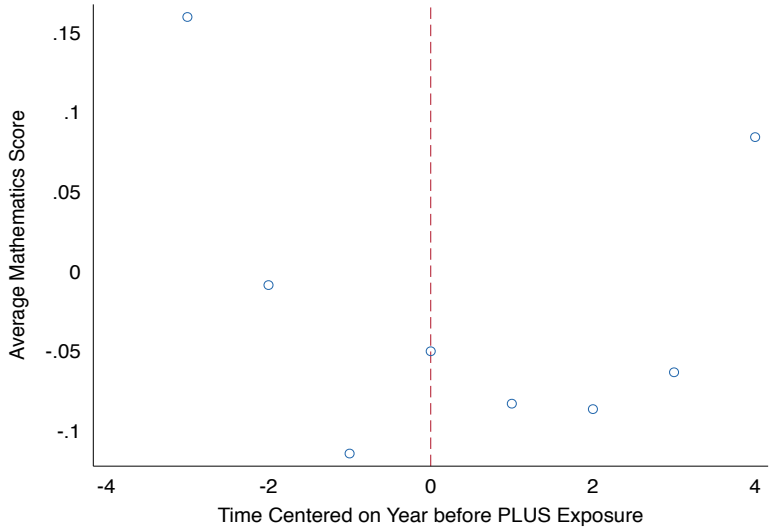
**Schools where Leaders worked were trending lower in achievement levels before Leaders arrived, and this negative trend continued initially after their arrival.** As illustrated in Figures A.3 (for mathematics) and A.4 (for ELA), Leader-led schools' achievement levels were trending downward relative to comparison schools in the rest of the district in the years immediately prior to a Leader arriving. In the first two years following the arrival of a Leader, this trajectory continued. By the second year with a Leader, the schools were on average significantly lower achieving than district schools that did not have a Leader. This downward

**Figure A.1. Unadjusted ELA Achievement Levels of PLUS-Exposed Schools Centered on the Year Before PLUS Exposure**




NOTES: Sample includes schools that were first led by Leaders any time between SYs 2015–2016 and 2018–2019. For each school in our sample, *Year 0* refers to the year prior to first hiring a Leader. Exposed schools were observed at least one year and up to four years prior to exposure to PLUS and at least one year and up to four years after exposure to PLUS.

**Figure A.2. Unadjusted Mathematics Achievement Levels of PLUS-Exposed Schools Centered on the Year Before PLUS Exposure**




NOTES: Sample includes schools that were first led by Leaders anytime between SYs 2015–2016 and 2018–2019. For each school in our sample, *Year 0* refers to the year prior to first hiring a Leader. Exposed schools were observed at least one year and up to four years prior to exposure to PLUS and at least one year and up to four years after exposure to PLUS.

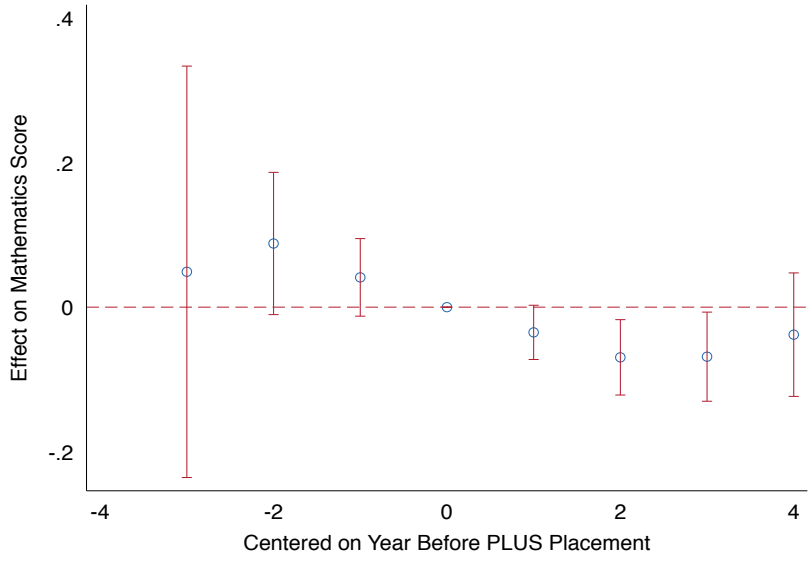
trend appears to have halted by the third and fourth years following PLUS exposure, however.

This achievement trend pattern (both before and after PLUS placement) was similar to what had been observed in prior research as associated with the hiring of a new school principal (Beteille, Kalogrides, and Loeb, 2012). Although we controlled for the typical average effect of hiring a novice AP in our model, it is possible that Leaders tended to be hired into a different kind of leadership position than the other newly hired APs in our data.[3] Whatever the reason, the results suggest a possible relationship between declining school test scores and the hiring of a new Leader in the school.

**When accounting for trends that predated PLUS exposure, we found no significant effect of Leaders on their schools' subsequent trajectories.** Using a more rigorous methodology that took into account preexisting achievement trends in schools with Leaders and those without, we explored whether schools experienced significant changes in their level or trajectory of achievement following the arrival of a Leader. Estimates of PLUS-induced changes in levels and trends in achievement are shown in Table A.18. Using this approach, which we expect would be more accurate with respect to causal impacts, we found no significant changes in school achievement levels or trajectories associated with the first arrival of a Leader in a school. Overall, it does not appear that exposure to a Leader caused significant changes in overall school-achievement trajectories or levels in the near term.

[3] As an additional specification check, we explored whether receiving a novice AP in the year prior to PLUS was associated with negative prior performance trends. We found that it was.

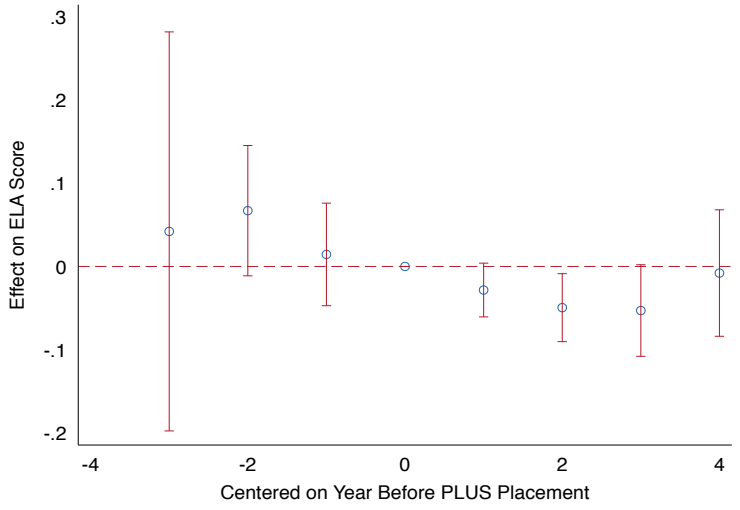
**Figure A.3. Differential Schoolwide Mathematics Performance over Time for Schools That Hired a Leader Compared with Schools That Never Hired a Leader**




.

NOTES: Estimated effects shown for schools that were first led by Leaders anytime between SYs 2015–2016 and 2018–2019. For each school in our sample, *Year 0* refers to the year prior to first hiring a Leader. Exposed schools were observed at least one year and up to four years prior to exposure to PLUS and at least one year and up to four years after exposure to PLUS.

**Figure A.4. Differential Schoolwide ELA Performance over Time for Schools That Hired a Leader Compared with Schools That Never Hired a Leader**




NOTES: Estimated effects shown for schools that were first led by Leaders anytime between SY 2015–16 and SY 2018–19. For each school in our sample, *year 0* refers to the year prior to first hiring a Leader. Exposed schools were observed at least one year and up to four years prior to exposure to PLUS and at least one year and up to four years after exposure to PLUS.

**Table A.18. Comparative Interrupted Time Series Analysis of Changes in Level and Trends in School Average Achievement Before and After Exposure to a Leader**

| | ELA | | Mathematics | |
|---|---|---|---|---|
| | Coefficient | Standard Error | Coefficient | Standard Error |
| PLUS pre-trend | –0.024 | (0.020) | –0.034 | (0.022) |
| PLUS immediate effect | 0.008 | (0.040) | –0.017 | (0.029) |
| PLUS post-trend (linear) | –0.025 | (0.043) | –0.010 | (0.040) |
| PLUS post-trend (quadratic) | 0.011 | (0.008) | 0.009 | (0.007) |
| *N* (schools) | 100 | | 100 | |

NOTE: No results are statistically significant at the $p < 0.05$ threshold. Schools are considered PLUS-exposed after their first Leader is hired to work there. All models control for a quadratic in comparison school achievement trends.

## Documenting Fidelity of Implementation for the i3 Grant

As a condition of the Investing in Innovation Fund (i3) grant, TNTP committed to rigorous measurement of the fidelity with which PLUS was implemented in the district over at least two years of program implementation. TNTP identified a series of measures and thresholds that aligned with the core components of the intervention and collected the data necessary to document whether each threshold was met. For each threshold and measure, TNTP collected data from at least two SYs of the PLUS program's implementation and tracked this data in accordance with i3 grant guidelines. The results are shown in Table A.19.

Overall, the PLUS program met all of its implementation goals. These goals were substantively related to the program's core components, and the results indicate that the program was implemented as intended and with high fidelity in the district.

**Table A.19. Thresholds and Results for Fidelity of Implementation Tracking**

| PLUS Component | Measure | Threshold Goals | Results | SYs Documented |
|---|---|---|---|---|
| Leaders receive preservice training | Attendance | 90% of Leaders attend ≥ 80% of the time | 100% in year 1<br>92% in year 2<br>met thresholds | 2015–2016 and 2016–2017 |
| Leaders receive coaching | Participation | ≥ 2 sessions per month, with feedback ≥ 80% of the time | 100% met thresholds in all years | 2015–2016, 2016–2017, and 2017–2018 |
| Leaders manage teacher teams | Caseload of teachers managed | 80% of Leaders manage an assigned team | 100% met thresholds in all years | 2015–2016, 2016–2017, and 2017–2018 |
| Leaders receive PD | Attendance | 90% of Leaders attend > 80% of PD sessions | 100% in year 1<br>92% in year 2<br>met thresholds | 2015–2016 and 2016–2017 |
| Leaders receive evaluations | Evaluations | 90% of Leaders receive evaluations | 100% met thresholds in all years | 2015–2016, 2016–2017, and 2017–2018 |

# References

Béteille, Tara, Demetra Kalogrides, and Susanna Loeb, "Stepping Stones: Principal Career Paths and School Outcomes," *Social Science Research*, Vol. 41, No. 4, 2012, pp. 904–919.

Iacus, Stefano M., Gary King, and Giuseppe Porro, "Causal Inference Without Balance Checking: Coarsened Exact Matching," *Political Analysis*, Vol. 20, No. 1, 2012, pp. 1–24.

Loeb, Susanna, and Christopher A. Candelaria, *How Stable Are Value-Added Estimates Across Years, Subjects and Student Groups?* Princeton, N.J.: Carnegie Foundation for the Advancement of Teaching, What We Know Series: Value-Added Methods and Applications, Knowledge Brief 3, October 2012.