Partners to Lead Evaluation

Final Summative Report

Melissa Arellanes | Melissa Brown-Sims | Marina Castro | Elisabeth Davis | Patricia Garcia-Arena | Eric Larsen | Bradley Salvato

APRIL 2023

American Institutes for Research® | AIR.ORG

AIR®
Advancing Evidence.
Improving Lives.

Acknowledgments

American Institutes for Research and the authors would like to acknowledge the contributions of Matthew Clifford, Feng Liu, and Cris Price to this research.

This work was funded by DuPage Regional Office of Education (ROE) through a 2017 U.S. Department of Education, Office of Elementary and Secondary Education, Education Innovation and Research Early-Phase grant (U411C170142) to DuPage ROE.

Abstract

Partners to Lead (PtL) is a school leadership professional development (PD) project funded by a 5-year Education Innovation and Research grant and implemented by the DuPage, Illinois, Regional Office of Education (ROE) in 37 public elementary, middle, and high schools in four Illinois ROEs. The American Institutes for Research® (AIR®), the independent evaluator of PtL, has completed an implementation and impact study of the program. This final report summarizes the PtL program and AIR's evaluation methods; the extent to which PtL was implemented as designed; and findings on PtL's impact on student achievement, changes to principal effectiveness as measured by the Illinois 5Essentials Survey, and principal retention. All 37 schools recruited to participate in PtL remained in the program across all 3 years of the intervention, and principals generally expressed satisfaction with the components of PtL. However, analyses indicate that PtL did not have a statistically significant positive impact on principal leadership effectiveness, student achievement, or principal retention. From March 2020 through the end of the intervention, some PtL principals may have become somewhat disengaged from the full PtL PD and coaching due to competing professional and personal priorities resulting from the COVID-19 pandemic, which may have attenuated the impact of PtL. Nonetheless, students in intervention schools scored 0.104 standard deviation higher in math in spring of the final year of the intervention, and the *p* value for the estimated positive impact on student math achievement was .06.

Contents

Tables

Figures

Executive Summary

The American Institutes for Research® (AIR®) has conducted an independent evaluation of the implementation and impact of the Partners to Lead (PtL) program, a leadership professional development (PD) project implemented by the DuPage, Illinois, Regional Office of Education (DuPage ROE). Public elementary, middle, and high schools were recruited from four Illinois ROEs (ROEs 1, 17, 19, and 28) to participate in PtL. Recruitment of rural schools was emphasized, and 18 of the 37 schools that participated in PTL are rural schools. The AIR study addressed the following research questions, which align with the program's logic model:

1. To what degree was PtL implemented with fidelity across participating PtL schools?
2. To what extent did school leadership quality and school culture improve in schools that participated in PtL in comparison to similar schools that did not participate in PtL?
3. To what extent did PtL participation have an impact on student learning in English language arts (ELA) and math in comparison to similar schools that did not participate in PtL?
4. To what extent did PtL participation have an impact on principal retention in comparison to similar schools that did not participate in PtL?

The PtL program was a significant investment by the U.S. Department of Education (ED) in three interrelated and promising principal PD processes, which include interventions to improve individual principal practice and distributed leadership in schools (Box ES1). DuPage ROE, ED, and AIR agreed to the program framework, research questions, and performance indicators. This final summative evaluation presents AIR's findings on program implementation and impact based on data from the 2018–19, 2019–20, and 2020–21 school years.

BOX ES1. PTL KEY COMPONENTS

- Time Utilization Analysis PD aims to support principals' efficient use of available time for instructional leadership, with the goals of developing distributed leadership and school-based collaboration practices and fostering continuous improvement of personal professional practice.
- The Leadership Framework supports principals with establishing organizational routines and processes for the regular and focused meetings of instructional leadership team, grade-level, and content-area teacher teams in order to continually improve instructional quality throughout the school. The Leadership Framework includes ongoing PD, technical assistance, and feedback loops that inform schoolwide, grade-level, and classroom-level continuous improvement. It includes the organizational mechanisms that enable principals to act as a multiplier by distributing effective practice throughout the school.
- Cycles of Inquiry provide schools with a systematic, evidence-driven improvement approach that can be applied to organizational, leadership, teaching, and learning problems.

There are three interconnected components to PtL: Time Utilization Analysis (TUA), the Leadership Framework (LF), and Cycles of Inquiry (COI) (Box ES1). AIR and DuPage ROE devised a fidelity-of-implementation matrix based on the PtL framework to gauge the extent to which the PtL program was implemented as designed. Fidelity of implementation was assessed separately for each of the three program components for each of the 3 years of the program. DuPage ROE and AIR decided that fidelity of implementation would be met at the program level if 90% of schools participating in the program met the fidelity-of-implementation threshold for that component for each year. For example, DuPage ROE and AIR decided that to meet fidelity of implementation for TUA at the program level in 2018–19, 90% of schools participating in the intervention would need to meet the threshold for adequate TUA implementation in 2018–19.

AIR also evaluated the impact of PtL program participation on three sets of outcomes: (a) principal leadership effectiveness as captured by the Illinois 5Essentials Survey, (b) schoolwide student achievement, and (c) principal retention. AIR used quasi-experimental designs to compare the outcomes of schools participating in the PtL program with those of nonparticipating (i.e., comparison) schools with similar characteristics. The primary contrast compares spring 2021 outcomes in PtL schools with spring 2021 outcomes of schools that were similar to PtL schools in the baseline year (2017–18), before PtL implementation began in the participating schools. Student outcomes were measured for all students in tested grades who took the regular state assessments.

Findings related to program impact and implementation can be summarized as follows:

- **All 37 schools recruited to participate in PtL remained in the program across all 3 years of the intervention.**
- **Overall, none of the 37 schools participating in PtL met the threshold for overall implementation of the TUA component in any year, although 100% of schools met the threshold for the TUA training component each year.**
  - Only 11% of principals participated in the TUA calendaring process in 2019–20. Principals found the task of coding their calendars to be overly burdensome, which was counterproductive to the ultimate goals of the TUA process (i.e., increased principal time allocated to instructional leadership).
  - No principals participated in an average of at least 30 minutes of TUA coaching per month in 2018–19, 2019–20, and 2020–21. Because PtL is a new program, the minimum thresholds for TUA coaching were based not on prior research but on educated guesses by DuPage ROE and AIR about what threshold might be needed to be able to effect change. Because no principals participated in at least 30 minutes of TUA coaching per month, on average, the threshold for TUA coaching set by DuPage ROE and AIR may

have been set too high. Moreover, coaches may have had difficulty distinguishing among TUA coaching, COI coaching, and other coaching, resulting in coaches underreporting the true number of minutes spent on TUA coaching.

- **Implementation of the LF component was mixed.**
  - Data on LF implementation are not available for 2018–19.
  - Ninety-two percent of PtL schools met the minimum threshold for LF implementation in 2019–20, exceeding the 90% threshold set by DuPage ROE and AIR.
  - However, only 70% of PtL schools met the minimum threshold in 2020–21, falling below the 90% goal set by DuPage ROE and AIR.
  - Principals in 97% of schools had identified an instructional leadership team (ILT) by 2019–20.
  - Principals or their designees met with the school's ILT at least 35 minutes per month, on average, in 73% of schools in 2019–20 and in 46% of schools in 2020–21.
  - ILT members met with teacher teams at least 35 minutes per month, on average, in 89% of schools in 2019–20 and in 73% of schools in 2020–21.
- **The percentage of PtL schools meeting the minimum threshold for implementation of the COI component increased from 59% in 2018–19 to 86% in 2019–20 and fell to 76% in 2020–21. These rates are all below the 90% threshold set by DuPage ROE and AIR.**
  - The percentage of principals who had completed all six initial COI trainings increased from 59% in 2018–19 to 62% in 2019–20 and 65% in 2020–21.
  - The percentage of principals who participated in at least 30 minutes of COI coaching per month, on average, increased from 11% in 2018–19 to 86% in 2019–20 and fell to 59% in 2020–21. Minimum thresholds for COI coaching were based not on prior research but on educated guesses by DuPage ROE and AIR about what threshold might be needed to be able to effect change. Because no principals participated in at least 30 minutes of COI coaching per month, on average, the threshold for COI coaching set by DuPage ROE and AIR may have been set too high. Moreover, coaches may have had difficulty distinguishing among TUA coaching, COI coaching, and other coaching, resulting in coaches underreporting the true number of minutes spent on COI coaching.
- **In general, principals were satisfied with the components of PtL**.
  - In each of the 3 program years, at least 87% of principals reported they were “satisfied” or “very satisfied” with the quality, intensity, practicality, and utility of the coaching they received, and more than 80% of principals reported they were “satisfied” or “very satisfied” with the COI components of the intervention.

  - In 2019–20, 66% of principals reported they were "satisfied" or "very satisfied" with the information and advice they received from their coaches about how to lead schools through the COVID-19 pandemic; the corresponding percentage in 2020–21 was 77%.
  - The samples of principals who were interviewed in 2018–19, 2019–20, and 2020–21 shared the benefits and positive experiences of participating in PtL.
- **Analyses indicate that PtL did not have a statistically significant positive impact on principal leadership effectiveness, student achievement, or principal retention.**
  - Accounting for baseline scores, school level, and student characteristics, schools that participated in PtL had lower 5Essentials Effective Leaders scores, on average, than comparison schools in spring of the final year of the intervention. The difference is equivalent to an effect size of −0.064 and is not statistically significant ($p < .05$).
  - After controlling for other factors included in the statistical model, we estimate that students in intervention schools scored 0.065 standard deviation (SD) higher in ELA and 0.104 SD higher in math in spring of the final year of the intervention. These differences, which are equivalent to percentile rank increases of 2.6 in ELA and 4.1 in math, are not statistically significant ($p < .05$), although the p value for the estimated positive impact on student math achievement is .06.
  - After controlling for baseline school and principal characteristics, we estimate that principal retention was 12.7 percentage points higher in intervention schools than in comparison schools, an effect size of 0.33 SD. However, this difference is not statistically significant ($p < .05$).

In March 2020, the middle year of the PtL program, all Illinois schools were closed because of the COVID-19 pandemic. For all Illinois public schools, in-person teaching and learning abruptly ceased. School administrators shifted gears and priorities to ensure that students and staff had access to food, technology, and resources at home as well as to transition to remote/virtual learning. From March 2020 through the end of the intervention, some PtL principals may have become somewhat disengaged from the full PtL PD and coaching due to competing professional and personal priorities, which may have attenuated the impact of PtL.

Background

School principals exert strong, although indirect, influence on student performance (Clifford et al., 2012; Grissom et al., 2021; Leithwood et al., 2004), and principals' influence strengthens with practice and professional support (Bartenen, 2019). Although principal professional development (PD) may be important for advancing principal practice, few rigorous studies of effective principal PD programs have been conducted (George W. Bush Institute, 2016; Herman

et al., 2017; Steele et al., 2021). Without evidence, significant questions remain about principal PD efficacy, design, and conditions for learning.

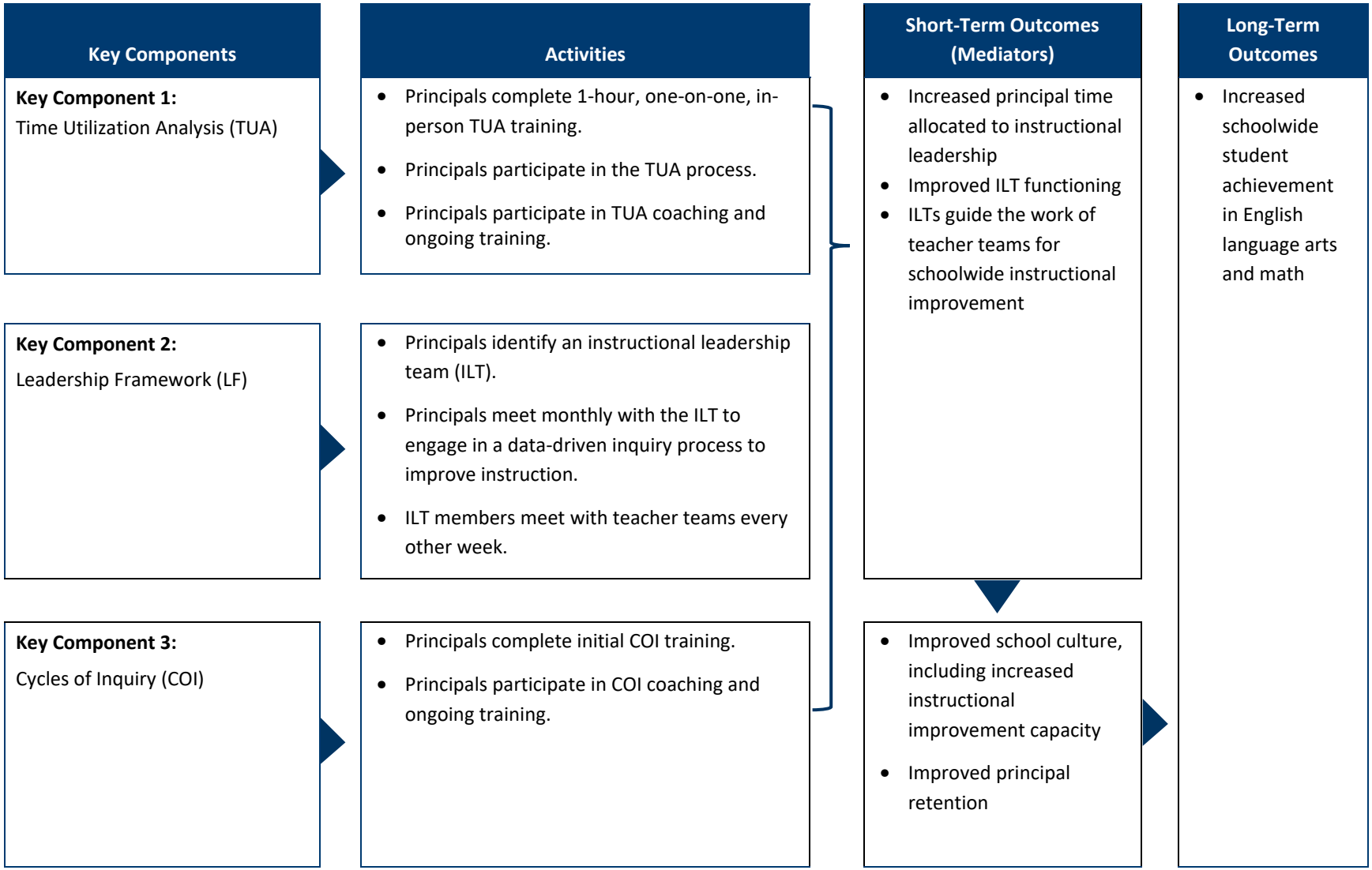
Partners to Lead (PtL) is an innovative, research-informed 3-year principal PD program funded by the U.S. Department of Education (ED) to address specific leadership learning needs and inform the field about what works in principal PD. Funded by a 5-year Education Innovation and Research (EIR) grant, the PtL program has been designed to be rigorous, replicable, and sustainable.

The American Institutes for Research® (AIR®), the independent evaluator of PtL, has completed an implementation and impact study of the program. This final summative report begins by briefly summarizing the PtL program and AIR's evaluation methods, including the study's research questions and activities. Next, the report presents AIR's findings on the extent to which the program was implemented as designed. The implementation evaluation section also describes adjustments that were made to the program over the course of the grant and trends in principal satisfaction with program components. Finally, the report presents findings on PtL's impact on student achievement, changes to principal effectiveness as measured by the Illinois 5Essentials Survey, and principal retention. The presentation of impact evaluation findings is designed to provide all information necessary for a What Works Clearinghouse (WWC) evidence review.

PtL PD Design

The PtL logic model, depicted in Figure 1, posits that schools can improve student outcomes by (a) aligning principals' and others' time to identified priority improvement areas, (b) establishing and maintaining systems of organizational routines that promote collaborative instructional improvement efforts by instructional leadership teams (ILTs) and teacher teams in ongoing instructional improvement efforts using a (c) Cycles of Inquiry (COI) process to develop and implement responsive instructional strategies that address priority student learning problems. To better align their time to priority improvement areas, principals receive tools and coaching support to delegate administrative tasks, which allows principals to prioritize their own time and organize the time of others to focus on instructional improvement.

Figure 1. Partners to Lead (PtL) Logic Model



PtL Components

There are three interconnected components to PtL: Time Utilization Analysis (TUA), Leadership Framework (LF), and COI. The remainder of this section provides details for each component.

TUA coaching. During the first 2 years of the intervention, PtL expected principals and coaches to examine how principals' time was allocated and take steps to improve their use of time for instructional support, which might include collaborating with groups of teachers to better understand which students were struggling, where these students were struggling, and what adults in the school could do differently to better support students.

To support TUA coaching, principals were asked to code at least 65% of their time and submit calendar data. A monthly TUA report was then delivered to principals and coaches for feedback on progress and planning. PtL school principals were expected to receive, on average, at least 105 minutes each month of in-person or virtual coaching during the academic year (e.g., between September and May of each school year). Of this time, 45 minutes per month were intended to be devoted specifically to coaching on TUA, and the remaining 60 minutes were intended to focus on COI.

Principals, however, found the task of coding their calendars to be overly burdensome, which was counterproductive to the ultimate goals of the TUA process (i.e., increased principal time allocated to instructional leadership). Because principals often did not have enough time to maintain and code their calendars, the time-use reports based on these calendars were often poor representations of how principals spent their days, which further reduced principals' incentives to maintain and code their calendars. By the third year of the intervention, calendar coding was dropped from the list of required program activities. However, the focus on principal time management was not abandoned and continued to be an area of both training and coaching support.

LF. The LF includes the following school-based organizational practices:

- Monthly coaching sessions and weekly touchpoints between principals and leadership coaches
- Monthly ILT meetings focused on instructional improvement and facilitated by the principal or the principal's designee
- Biweekly grade-level and/or content-area meetings focused on instructional improvement and facilitated by an ILT representative
- Ongoing job-embedded coaching
- Technical assistance
- Multiple feedback loops to inform a structured continuous improvement process

By implementing the LF, principals and teachers collaboratively established organizational routines designed to institutionalize effective practices and policies.

COI. COI provide schools with a systematic, evidence-driven improvement approach that can be applied to organizational, leadership, teaching, and learning problems. Cosner and colleagues describe COI as a rigorous, nonlinear improvement process that systematically identifies and tests localized innovations in schools (Cosner & Jones, 2016; Cosner et al., 2016). PtL supports COI learning through a formalized workshop series and individualized principal coaching. Each principal is expected to participate fully in workshops and coaching.

The COI PD engages principals in learning to continuously improve school-level instructional support systems, build their school's capacity, and improve educator learning. According to PtL, COI starts with an analysis of a wide variety of student performance data to identify a well-defined student learning problem. ILTs and teacher teams then explore instructional data to determine the root causes of the well-defined student learning problem. They then develop and implement high-leverage, research-based, responsive strategies aligned to the specific student learning problem. These results are achieved through goal setting, action planning, implementation, progress assessment, collaboration among the principal and ILT members to identify any adjustments that might be needed, and helping teachers attain the knowledge and skills necessary to implement a responsive strategy.

PtL project staff and partners offered six foundational COI PD training sessions, which all principals were required to complete. Coordinators or trainers offered makeup sessions throughout the year to accommodate principals' scheduling conflicts; they also offered repeat sessions to principals new to schools that already were participating in PtL. Additional, supplemental sessions beyond the six foundational sessions also were offered to principals. Project staff expected that all continuing principals would be exposed to the content in these supplemental sessions, either through in-group trainings or one-on-one site delivery.

Characteristics of PtL Schools[1]

Table 1 presents the number of elementary, middle, and high schools that participated in PtL by the school's rural status. Nearly half of the intervention schools were in rural areas. The majority of middle schools were in nonrural areas, and the majority of high schools were in

[1] PtL engaged a total of 55 schools across the following three types of schools: intervention schools (n = 37), pilot schools (n = 8), and demonstration schools (n = 10). Of the 55 schools engaged in PtL, 36 were in rural areas, as all 10 demonstration schools and all eight pilot schools were located in rural areas. Pilot schools were early implementers of the professional learning model, and their principals' experiences informed the PtL improvements. Pilot schools were exposed to all aspects of the full intervention. Demonstration schools were the early implementers of key aspects of the professional learning model, and their school experiences informed the PtL design. Demonstration schools were exposed to aspects of the full intervention.

rural areas. Among all schools participating in PtL, 16 schools were elementary schools, 10 were middle schools, and 11 were high schools.

Table 1. Number of Schools by School Level and Rural Status

School level	Nonrural	Rural	Total
Elementary	8	8	16
Middle	7	3	10
High	4	7	11
Total	19	18	37

Table 2 presents the number of elementary, middle, and high schools that participated in PtL in each Illinois Regional Office of Education (ROE). ROE 17 had the largest number of participating schools (20 schools), followed by ROE 19 (11 schools), ROE 28 (four schools), and ROE 1 (two schools).

Table 2. Number of Schools by School Level and Regional Office of Education (ROE)

School level	ROE 1	ROE 17	ROE 19	ROE 28	Total
Elementary	1	10	4	1	16
Middle	0	5	4	1	10
High	1	5	3	2	11
Total	2	20	11	4	37

Table 3 presents the average school enrollment, student demographics, and assessment participation and proficiency rates for the study years among the 37 PtL intervention schools. On average, schools enrolled slightly fewer than 600 students. Students who were non-Hispanic White constituted the largest racial/ethnic group enrolled, with an average of approximately 70% of students in schools in each of the 3 years. Fourteen percent of students in schools were Hispanic, on average; about 7% were Black; and 5% were Asian. About 40% of students were eligible for the National School Lunch Program, on average, across schools. Approximately 5% of students in schools were English learners (ELs), and about 15% had a disability. On average, almost all students who were eligible participated in the annual state assessment in 2018–19, and, on average, about 85% participated in 2020–21. Average English language arts (ELA) proficiency across schools was 38% in 2018–19 and 29% in 2020–21, and math proficiency was 33% in 2018–19 and 25% in 2020–21.

Table 3. Characteristics of Schools and Students, by Year

Characteristics	2018–19		2019–20		2020–21	
	Mean	SD	Mean	SD	Mean	SD
Total student enrollment	593.6	548.96	591.4	547.06	570.4	536.83
Percentage of students who are White	70.1%	0.23	69.4%	0.22	68.3%	0.23
Percentage of students who are Hispanic	13.7%	0.16	13.9%	0.16	13.8%	0.16
Percentage of students who are Black	7.1%	0.09	7.5%	0.09	7.1%	0.10
Percentage of students who are Asian	4.7%	0.06	4.8%	0.06	4.8%	0.06
Percentage of students eligible for the National School Lunch Program	41.6%	0.19	41.0%	0.19	40.3%	0.18
Percentage of students who are English learners	5.1%	0.07	5.8%	0.07	6.1%	0.08
Percentage of students with a disability	15.6%	0.04	14.5%	0.04	14.6%	0.04
School ELA participation rate	98.9%	0.02	NA	NA	85.7%	0.15
School math participation rate	98.2%	0.03	NA	NA	85.3%	0.16
School ELA proficiency rate	38.5%	0.12	NA	NA	29.0%	0.11
School math proficiency rate	33.0%	0.15	NA	NA	25.2%	0.14

Note. ELA = English language arts; NA = not applicable; SD = standard deviation. Table statistics represent all 37 schools participating in Partners to Lead.

Source. Annual Illinois Report Card data released by the Illinois State Board of Education: https://www.isbe.net/Pages/IL-Report-Card.aspx.

Overview of the Evaluation

The primary purpose of AIR's independent evaluation of PtL was to provide the DuPage, Illinois, ROE (DuPage ROE) with both formative and summative data about fidelity of program implementation and PtL's impact on principal practice, school conditions, principal retention, and student performance. This summative evaluation report provides DuPage ROE, its partners, and ED with results on PtL implementation and attainment of intended outcomes. AIR used surveys and participant data for the implementation analyses and student achievement, the Illinois 5Essentials Survey, and principal retention data for the impact analyses.

Box 1 presents the research questions (RQs) that AIR's PtL study was designed to answer. RQ 1 addresses implementation and is primarily descriptive. To address RQs 2 through 4, the study team employed quasi-experimental research designs. RQs 2 through 4 were registered in the Registry of Efficacy and Effectiveness Studies.[2]

BOX 1. PTL RESEARCH QUESTIONS

AIR's evaluation was guided by four RQs:

- **RQ 1:** To what degree was PtL implemented with fidelity across participating PtL schools?
- **RQ 2:** To what extent did school leadership quality and school culture improve in schools that participated in PtL in comparison to similar schools that did not participate in PtL?
- **RQ 3:** To what extent did PtL participation have an impact on student learning in English language arts and math in comparison to similar schools that did not participate in PtL?
- **RQ4:** To what extent did PtL participation have an impact on principal retention in comparison to similar schools that did not participate in PtL?

Data Sources and Analytic Approach

AIR's data collection and analytic approach was developed with input from PtL and Abt Associates, the technical assistance provider assigned to support EIR evaluators. The study called for extensive data collection to thoroughly document principal PD participation and determine its effects. As the external evaluator, AIR was responsible for receiving, managing, and analyzing all data. Each data collection activity and corresponding analytic approach is outlined in Table 4.

[2] The registry entries are #1832.1v3 Partners to Lead (Early03) Student Outcomes and #1832.2v1 Partners to Lead (Early03) Principal Outcomes; see https://sreereg.icpsr.umich.edu/sreereg/.

Table 4. Data Sources and Analytic Approaches

Data source	Research question	Data collected	Analytic approach
Formative Data Collection Activities			
Cycles of Inquiry (COI) professional development (PD) training participant data	1	Attendance data for the six foundational COI trainings	Descriptive analysis of principal attendance rates
Coaching logs	1	Self-reported amount of time spent by coaches on coaching activities with their assigned principals during the 2018–19, 2019–20, and 2020–21 school years	Descriptive analysis of coaching time allocations, in average minutes per month
Coaching survey[a]	1	Data gauging the change in school-level team use of COI, including instructional leadership team and teacher team activities during the 2018–19, 2019–20, and 2020–21 school years	Descriptive analysis of coaches' interactions with school administrators, disaggregated by school level, locale, and school year
Annual principal survey	1	An annual online survey administered to PtL principals during the 2018–19, 2019–20, and 2020–21 school years about their perceptions and satisfaction with the Partners to Lead (PtL) program PD	Descriptive analysis of principal responses by school year
Interviews with principals	1	Interviews with respondents who represent principals participating in PtL were conducted during the 2018–19, 2019–20, and 2020–21 school years. The purpose of the interviews was to learn the perspectives of these principals about program goals, policies, and practices as well as the organizational structures that affect implementation of PtL. Respondents also were asked to discuss the supports provided through PtL and to identify what additional supports are needed to sustain this program.	All interviews were audio-recorded with participant permission and later transcribed. The American Institutes for Research (AIR) evaluation team analyzed the transcripts using NVivo software using a priori codes aligned with program objectives as a preliminary means of organizing the data, and then they used a grounded approach to identify more fine-grained codes (see Glaser, 1965).
Principal calendars	1	Monthly calendars coded and submitted by principals during the 2019–20 school year. AIR used seven different categories to code events in principals' schedules. Principals received monthly reports detailing their time-use information.	Calendars were converted into data files, which we used to tabulate the percentage of total possible hours that each principal spent on certain tasks or categories each year.

Data source	Research question	Data collected	Analytic approach
Summative Data Collection Activities			
Effective Leaders measure from the Illinois 5Essentials Survey	2	This measure is based on teachers' responses to questions about the extent to which the principal works with teachers to implement a clear and strategic vision for school success. The measure has demonstrated reliability, with a Cronbach's alpha greater than 0.8 (Klugman et al., 2015). AIR collected Illinois 5Essentials Survey data from the publicly available resource on the Illinois State Board of Education (ISBE) website.[b]	Difference-in-differences design to evaluate the impact of PtL participation on leadership effectiveness
Student achievement data	3	Administrative student English language arts and math Partnership for Assessment of Readiness for College and Careers, Illinois Assessment of Readiness, and SAT® data for students in Grades 3–8 and 11	Comparative interrupted time series design to evaluate the impact of PtL participation on student achievement
Principal retention data	4	Personnel data, provided by ISBE, on individuals who were principals in the treatment and comparison schools from the 2013–14 school year through the 2020–21 school year, including records for other schools where the individual was a principal Publicly available, school-level student enrollment and demographic data from the Illinois Report Card[c]	Matched-comparison group design to evaluate the impact of PtL participation on principal retention

[a] The coaching survey was administered by AIR and completed by PtL coaches on a monthly basis.

[b] https://www.isbe.net/Pages/5Essentials-Survey.aspx.

[c] https://www.isbe.net/Pages/Illinois-State-Report-Card-Data.aspx.

Limitations

As with any evaluation, the PtL evaluation faced challenges that may affect inferences about the program's fidelity of implementation and outcomes. The primary limitations are as follows:

- **Impact of the COVID-19 pandemic on participant engagement in PtL PD.** In March 2020, the middle year of the PtL program, all Illinois schools were closed because of the COVID-19 pandemic. For all Illinois public schools, in-person teaching and learning abruptly ceased for the remainder of the 2019–20 school year. School administrators shifted gears and priorities to ensure that students and staff had access to food, technology, and resources at home as well as to transition to remote/virtual learning. From March 2020 through the end of the intervention, some PtL principals may have become somewhat disengaged from the

full PtL PD due to competing professional and personal priorities, which may have attenuated the impact of PtL. A number of factors could explain the lack of program impact on outcomes of interest. At this point, we cannot determine whether the program would have achieved its desired impact if it had been implemented with higher levels of fidelity. However, it reasonable to assume that the COVID-19 pandemic influenced program impact.

- **Impact of the COVID-19 pandemic on student outcome analysis.** In spring 2020, Illinois canceled its 2019–20 academic year student testing in response to the COVID-19 pandemic. As a result, AIR was unable to analyze the impact of PtL on spring 2020 student achievement. In addition, participation in testing during spring 2021 was optional. As a result, there was a decrease in the percentage of students who took the Illinois Assessment of Readiness (IAR) or SAT® in spring 2021 compared with previous years. To minimize bias, AIR excluded one school from the impact study that tested less than 50% of students across the IAR or SAT® in ELA or math.
- **Completeness and accuracy of the information reported in the coaching logs.** Coaching log entries were subject to human error as well as bias. To help ensure the accuracy of coaching data, AIR implemented a system in which coaches periodically received a record of their submitted logs, which they were asked to review. If these records indicated that some coaching sessions were not logged successfully, coaches were asked to enter the missing sessions into the coaching log system.
- **Reliance on self-reported survey data.** To track principal participation in and satisfaction with PD, AIR analyzed survey data completed by principals. Self-reported survey data can be prone to inaccuracies and biases, which may skew findings.
- **Focus on short-term effects.** Leadership impact on teaching and instruction is indirect; therefore, interventions aimed at improving leadership may take multiple years to display impact. The study was limited to reporting impact in the first 3 years of the intervention, given the 5-year duration of the grant and the timing of annual test score certification by the state.
- **Inconsistent survey data collection across study years.** Changes to questions in the annual principal survey across years limited the survey results we were able to present in the current report. We note in the survey figures when the wording to survey questions slightly differed across years.
- **Limited generalizability.** The study occurs within distinct state and regional contexts. PtL implementation occurs in Illinois only, within four regions of the state. Although implementing schools are diverse in many ways, state and regional policy contexts may limit generalizability.

Stakeholders should take these limitations into consideration when reading, interpreting, or generalizing findings in this report.

Fidelity of PtL Implementation

This section summarizes the implementation findings related to TUA, COI, and LF implementation. Data used to inform this section include PD attendance data, coaching logs, coaching surveys, and principal calendars. For each key component (TUA, LF, and COI), we present the indicators of implementation fidelity; the thresholds for high, moderate, and low implementation fidelity; and the threshold overall program-level fidelity of implementation for each year. The thresholds at the indicator and program levels were set by DuPage ROE staff and AIR. Because PtL is a new program, the thresholds for high, moderate, and low implementation fidelity were based not on prior research but on educated guesses by DuPage ROE staff and AIR about what threshold might be needed to be able to effect change.

TUA Implementation

One assumption of the PtL logic model is that if principals understand and document how they spend their time during the school day, they will be able to use these data to identify the tasks and barriers that are preventing them from spending time as instructional managers and leaders. AIR researchers and PtL leaders identified three indicators of TUA implementation: initial TUA training, participation in the TUA process, and regular TUA coaching. Table 5 details each indicator; the data source used to measure it; and how high, moderate, and low implementation fidelity were defined. Schools that implemented at moderate or high levels met the threshold for adequate implementation.

Table 5. Indicators of Time Utilization Analysis (TUA) Implementation

Indicator	Data source	High, moderate, and low implementation fidelity
Principals complete 1-hour, one-on-one, in-person TUA training (2018–19 and 2019–20).	Coaching survey	High = Completed the 1-hour training. Low = Did not complete the 1-hour training.
Principals participate in the TUA process (2018–19 and 2019–20).	Principal calendars	High = Principals code 65% or more of their calendar time with TUA codes. Moderate = Principals code 50% to less than 65% of their calendar time using TUA codes. Low = Principals code less than 50% of their calendar time with TUA codes.
Principals participate in 45 minutes per month of TUA coaching while school is in session (2018–19, 2019–20, and 2020–21).	Coaching logs	High = 45 minutes or more per month, on average Moderate = 30 to less than 45 minutes per month, on average Low = Less than 30 minutes per month, on average

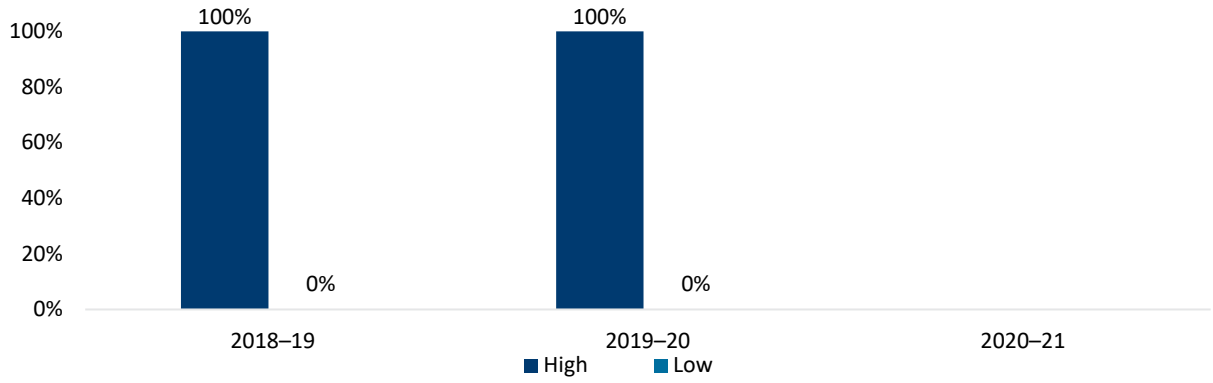
The "TUA process" is the method by which principals code their online calendars according to research-based task categories, submit calendars, and receive monthly TUA reports for coaching conversations. TUA coaching conversations were provided by PtL coaches during a regularly scheduled meeting, which also could include COI and LF content.

Principals found the task of coding every hour on their calendar to be overly burdensome, and the burden of coding calendars interfered with the goal of the TUA process: increased principal time allocated to instructional leadership. Therefore, DuPage ROE staff dropped the TUA process indicator for the third year of the intervention. Although principals continued to receive coaching focused on increasing principal time devoted to instructional leadership and building schoolwide capacity for instructional improvement, we no longer asked them to code their calendars and did not try to institute another way of systematically gathering information on principals' time allocation.

This section presents our findings on annual TUA fidelity of implementation. Additional findings are presented in Appendix A.

Participation in TUA training. Figure 2 reports the percentages of principals who completed the 1-hour, one-on-one, in-person TUA training. Principals were only required to complete this training once, and Figure 2 shows that in 2018–19 and 2019–20, all principals completed the training, either during the current year (for new principals) or in prior years (for continuing principals). Participation in TUA training was not required in 2020–21.

Figure 2. Time Utilization Analysis Training Participation: Percentage of Principals With High and Low Implementation Fidelity, by Year

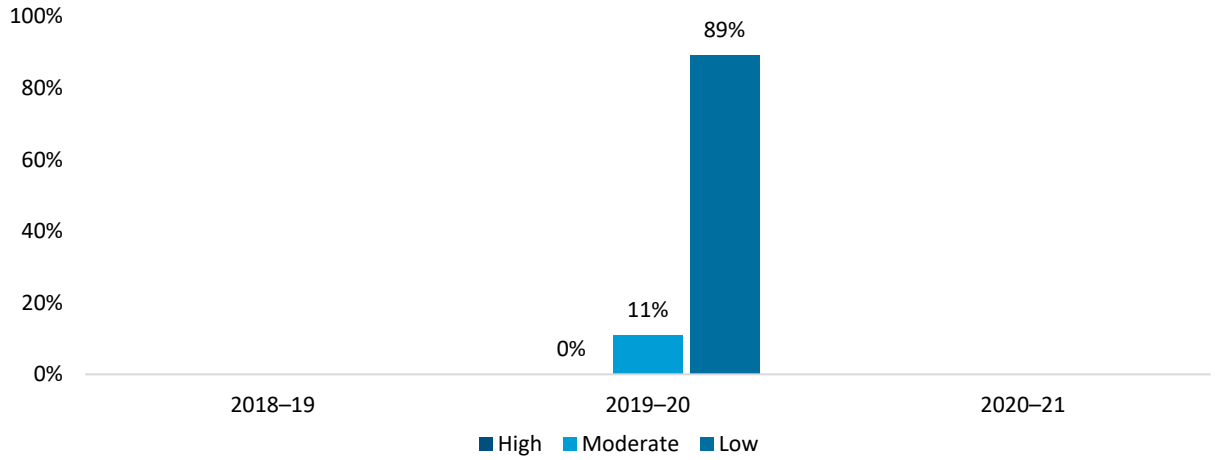


Note. This figure is based on 37 schools each year. See Table 4 and Table 5 for a description of the analytic approach. Data are based on coaching logs.

Participation in the TUA process. Figure 3 illustrates the degree to which principals participated in the TUA process. To be considered to have implemented the process with high fidelity, a principal had to have coded at least 65% of their calendar time with codes indicating TUA

activities, at least 50% but less than 65% of their time for moderate fidelity, and less than 50% of their time for low fidelity. Data needed to analyze TUA implementation were not available for 2018–19. In 2019–20, 11% of principals met the criteria for moderate fidelity, coding at least half of their calendar time with TUA activities. The remaining 89% of principals coded less than half of their time with TUA codes, indicating low fidelity. As noted earlier, principals were not asked to code their calendars in 2020–21.

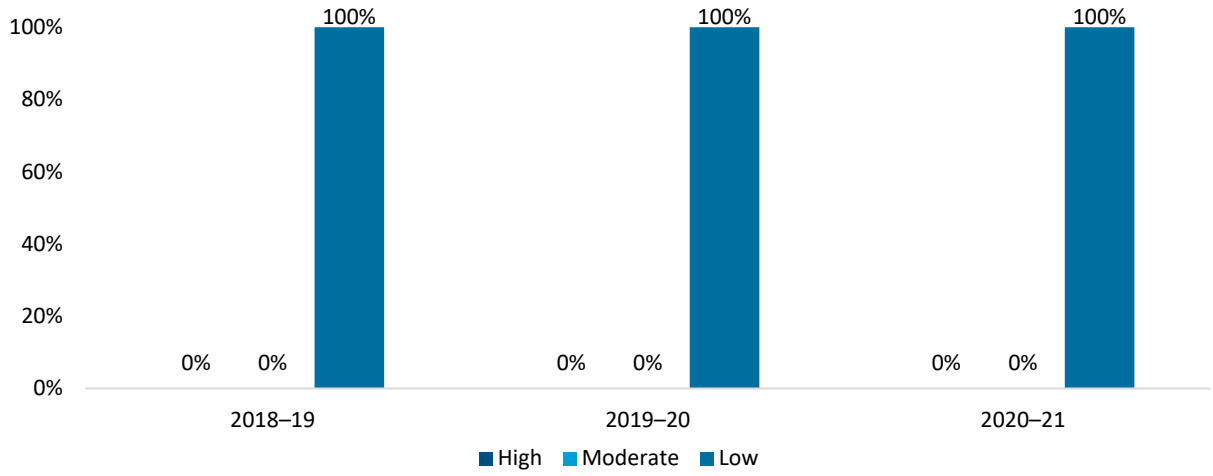
Figure 3. Time Utilization Analysis Process Participation: Percentage of Principals With High, Moderate, and Low Implementation Fidelity, by Year



Note. This figure is based on 37 schools each year. See Table 4 and Table 5 for a description of the analytic approach. Data are based on calendar files submitted to AIR by principals.

Intensity of TUA coaching. Participating PtL principals were asked to participate in TUA coaching in months when school was in session. Fidelity of implementation of this indicator was defined by the average length of time (in minutes per month) that principals engaged in TUA coaching during those months. For high fidelity, principals had to average at least 45 minutes per month engaged in TUA coaching; for moderate fidelity, at least 30 minutes but less than 45 minutes per month; and for low fidelity, less than 30 minutes per month. Across all 3 years, principals implemented with low fidelity (Figure 4).

Figure 4. Monthly Time Utilization Analysis Coaching: Percentage of Principals With High, Moderate, and Low Implementation Fidelity, by Year



Note. This figure is based on 37 schools each year. See Table 4 and Table 5 for a description of the analytic approach. Data are based on coaching logs.

Because PtL is a new program, the minimum thresholds for TUA coaching were based not on prior research but on educated guesses by DuPage ROE and AIR about what threshold might be needed to be able to effect change. Because no principals participated in at least 30 minutes of TUA coaching per month, on average, the threshold for TUA coaching set by DuPage ROE and AIR may have been set too high. Moreover, coaches may have had difficulty distinguishing among TUA coaching, COI coaching, and other coaching, resulting in coaches underreporting the true number of minutes spent on TUA coaching.

Overall Fidelity of Implementation

To assess whether PtL was implemented with fidelity across all TUA indicators each year, we assigned a school points for each applicable indicator based on the level of implementation fidelity reached. A school received 2 points if it implemented with high fidelity, 1 point for moderate fidelity, or 0 points for low fidelity. To reach the minimum threshold for adequate implementation, a school had to receive a predetermined number of points, which varied by year and by component, based on the applicable indicators in each program component in each year. DuPage ROE and AIR decided that overall fidelity of implementation would be met if at least 90% of schools met the minimum threshold for the component.

Table 6 documents the indicators associated with TUA and the minimum threshold required to demonstrate adequate fidelity of implementation. For the 2018–19 and 2019–20 school years, a school had to receive at least 4 points out of a possible 6 to demonstrate adequate fidelity of implementation. The indicators assessed include principals completing a 1-hour, one-on-one TUA training, participating in the TUA process, and engaging in at least 45 minutes of TUA

coaching during the school year. For the 2020–21 school year, only the TUA coaching indicator is assessed, and the minimum threshold is 1 point out of a possible 2.

Table 6. Time Utilization Analysis (TUA) Fidelity-of-Implementation Matrix

Program component	Expectations for fidelity of implementation	Indicators
TUA 6 points total in 2018–19 and 2019–20 2 points total in 2020–21	Principals complete 1-hour, one-on-one, in-person TUA training (2018–19 and 2019–20 only).	High = Completed the 1-hour training (2 points). Low = Did not complete the 1-hour training (0 points).
	Principals participate in the TUA process (2019–20 only).	High = Principals coded 65% or more of their calendar time with TUA codes (2 points). Moderate = Principals coded 50% to less than 65% of their calendar time using TUA codes (1 point). Low = Principals coded less than 50% of their calendar time with TUA codes (0 points).
	Principals participate in 45 minutes per month of TUA coaching while school is in session (2018–19, 2019–20, and 2020–21).	High = 45 minutes or more per month (2 points) Moderate = 30 to less than 45 minutes (1 point) Low = Less than 30 minutes (0 points)
	Adequate implementation at the school/principal level: 2018–19: 4 points 2019–20: 4 points 2020–21: 1 point	Adequate implementation at the sample level: 2018–19: 90% of schools meeting the threshold for adequate implementation 2019–20: 90% of schools meeting the threshold for adequate implementation 2020–21: 90% of schools meeting the threshold for adequate implementation

In each year, fewer than 90% of schools met the threshold for adequate implementation at the program level. In fact, none of the 37 schools participating in PtL met the threshold for overall TUA implementation in any year. This result suggests that the thresholds for TUA fidelity, and for TUA coaching in particular, may have been set too high by DuPage ROE and AIR.

Leadership Framework

PtL views ILTs and teacher teams as important for instructional improvement. The PtL logic model posits that school performance will improve as instructional leadership is distributed within schools. This section reports the percentage of schools in which principals identified an ILT each year, and the total duration of ILT and teacher team meetings each month. Table 7 details each indicator; the data source used to measure it; and the definitions of high, moderate, and low implementation fidelity. Schools that implemented at moderate or high levels met the threshold for adequate implementation.

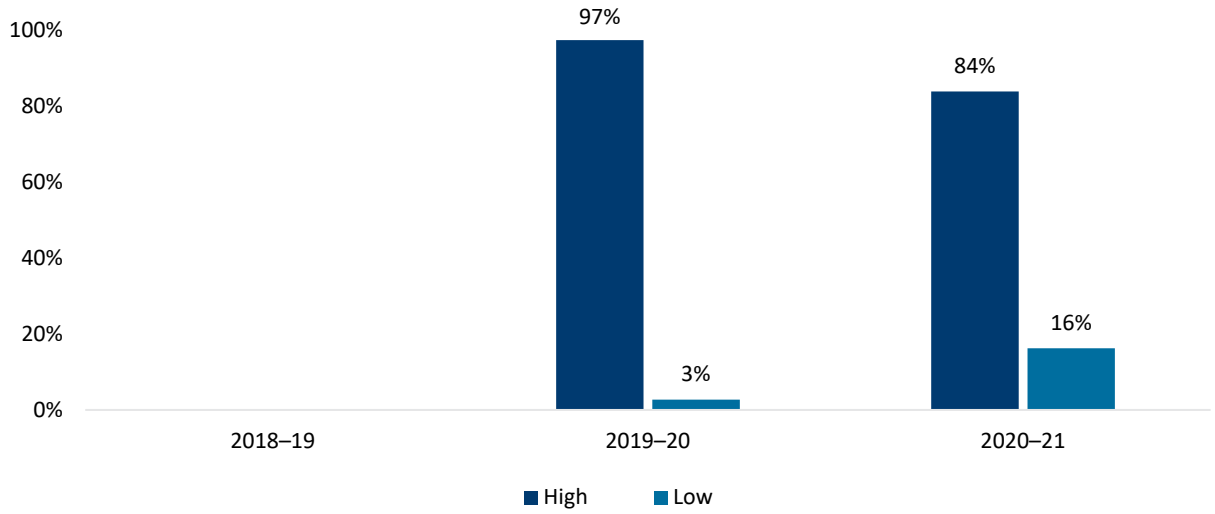
Table 7. Indicators of Leadership Framework Implementation

Indicator	Data source	High, moderate, and low implementation fidelity
Principals identify an instructional leadership team (ILT) (2018–19, 2019–20, and 2020–21).	Coaching survey	High = Yes Low = No
Principals meet with the ILT 60 minutes per month while school is in session (2018–19, 2019–20, and 2020–21).	Coaching survey	High = 60 minutes or more per month, on average Moderate = 35 to less than 60 minutes Low = Less than 35 minutes
ILT members meet with teacher teams every other week while school is in session (2018–19, 2019–20, and 2020–21).	Coaching survey	High = 60 minutes or more per month, on average Moderate = 35 to less than 60 minutes Low = Less than 35 minutes

This section presents our findings on annual LF fidelity of implementation. Unfortunately, LF implementation data for 2018–19 are not available, so we report on LF implementation for 2019–20 and 2020–21 only. Additional results and tables also are reported in Appendix A.

Schools that have an ILT. Nearly all principals (97%) identified an ILT in the 2019–20 school year, but only 84% identified an ILT in 2020–21 (Figure 5). Implementation data for the 2018–19 school year are not available.

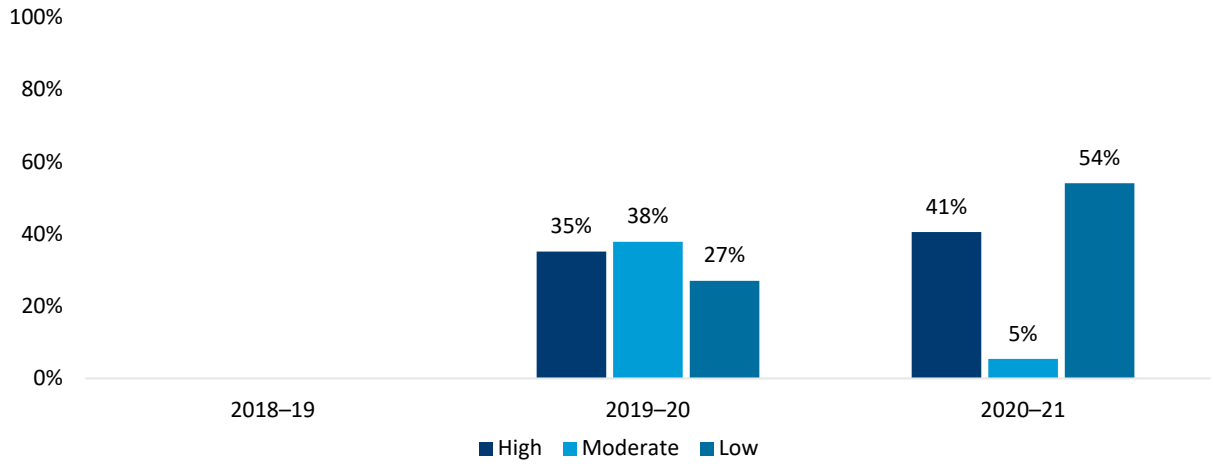
Figure 5. Instructional Leadership Team Identification: Percentage of Principals With High and Low Implementation Fidelity, by Year



Note. This figure is based on 37 schools each year. See Table 4 and Table 7 for a description of the analytic approach. Data are based on coaching surveys. Implementation data for the 2018–19 school year are not available.

Frequency of meetings between PtL principals and their ILT members. Figure 6 reports the proportion of PtL principals in each school year who met with the ILT for a given length of time per month, on average. A principal must meet with the ILT for an average of at least 60 minutes per month to achieve high implementation fidelity, at least 35 minutes but less than 60 minutes to implement with moderate fidelity, and less than 35 minutes to implement with low fidelity. Implementation data for the 2018–19 school year are not available. In 2019–20, 73% of principals met with their ILTs for at least 35 minutes per month, demonstrating either moderate or high fidelity. However, only 46% of principals achieved high or moderate fidelity in 2020–21.

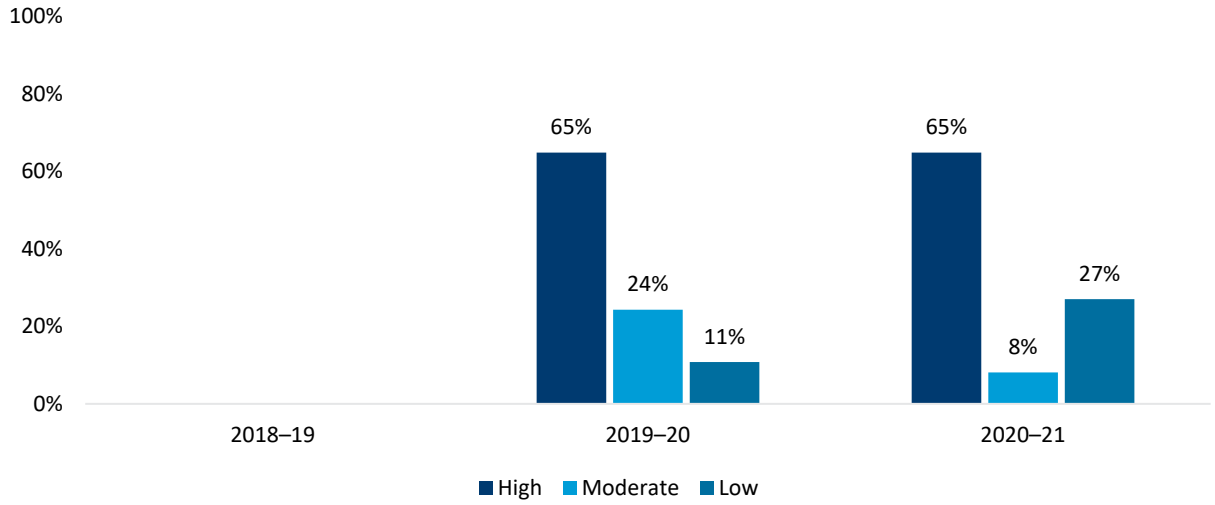
Figure 6. Principal–Instructional Leadership Team Meetings: Percentage of Principals Implementing With High, Moderate, and Low Implementation Fidelity, by Year



Note. This figure is based on 37 schools each year. See Table 4 and Table 7 for a description of the analytic approach. Data are based on coaching surveys. Implementation data for the 2018–19 school year are not available.

Frequency of teacher team meetings. Coaches also were asked to report the frequency and duration of teacher team meetings while school was in session. To implement with high fidelity, ILT members must meet with teacher teams for at least 60 minutes per month. To implement with moderate fidelity, ILT members must meet with teacher teams for at least 35 minutes but less than 60 minutes per month. To implement with low fidelity, ILT members must meet with teacher teams for less than 35 minutes per month. Implementation data for the 2018–19 school year are not available. In both 2019–20 and 2020–21, in 65% of schools, ILT members met with teacher teams for at least 60 minutes per month on average, demonstrating high implementation fidelity (Figure 7).

Figure 7. Teacher Team Meetings: Percentage of Schools Implementing With High, Moderate, and Low Implementation Fidelity, by Year



Note. This figure is based on 37 schools each year. See Table 4 and Table 7 for a description of the analytic approach. Data are based on coaching surveys. Implementation data for the 2018–19 school year are not available.

Overall Fidelity of Implementation

To assess whether PtL was implemented with fidelity across all LF indicators each year, AIR followed the same process as was followed for TUA. First, each school was assigned points for each applicable indicator based on the level of implementation fidelity reached. Second, AIR compared the total number of points assigned to a school with the threshold for adequate implementation for the LF component, which varied by year. DuPage ROE and AIR decided that overall fidelity of LF implementation would be met if at least 90% of schools met the minimum threshold for the component. Because implementation data are not available for 2018–19, AIR evaluated overall fidelity of LF implementation for 2019–20 and 2020–21 only.

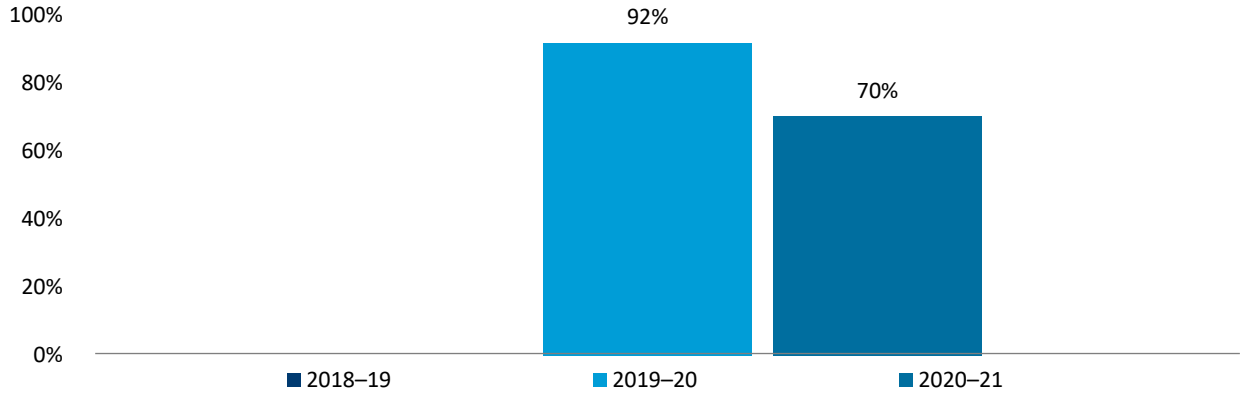
Table 8 defines the minimum threshold required to demonstrate adequate fidelity for the LF indicators. For the 2019–20 and 2020–21 school years, these indicators include principals identifying an ILT, principals meeting with the ILT for 60 minutes per month during the school year, and ILT members meeting with teacher teams every other week during the school year. To reach the minimum threshold in 2019–20 and 2020–21, a school had to earn at least 4 points out of a possible 6.

Table 8. Leadership Framework Fidelity-of-Implementation Matrix

Program component	Expectations for fidelity of implementation	Indicators
Leadership Framework Implementation data are not available for 2018–19 6 total points in 2019–20 and 2020–21	Principals identify an instructional leadership team (ILT).	High = Yes (2 points) Low = No (0 points)
	Principals meet with the ILT 60 minutes per month while school is in session.	High = 60 minutes or more per month, on average (2 points) Moderate = 35 to less than 60 minutes, on average (1 point) Low = Less than 35 minutes, on average (0 points)
	ILT members meet with teacher teams every other week while school is in session.	High = 60 minutes or more per month, on average (2 points) Moderate = 35 to less than 60 minutes, on average (1 point) Low = Less than 35 minutes, on average (0 points)
	Adequate implementation at the school/principal level: 2019–20: 4 points 2020–21: 4 points	Adequate implementation at the sample level: 2018–19: Not applicable 2019–20: 90% of schools meeting the threshold for adequate implementation 2020–21: 90% of schools meeting the threshold for adequate implementation

Figure 8 reports the percentage of schools that achieved the minimum threshold for fidelity of LF implementation each year. Implementation data for 2018–19 are not available. In the 2019–20 school year, 92% of schools met the minimum threshold for adequate implementation for the LF component, exceeding the target of 90% set by DuPage ROE and AIR. In the 2020–21 school year, 70% of schools met the minimum threshold for LF implementation, despite the competing priorities presented by the pandemic.

Figure 8. Percentage of Schools That Met the Threshold for Implementation Fidelity for the Leadership Framework (LF), by Year



Note. Data are based on authors' analyses of LF implementation data.

COI Implementation

To determine whether the PtL COI component was implemented with fidelity, AIR and PtL leaders formulated two indicators of COI implementation: participation in COI PD and engagement in regular COI coaching sessions. Table 9 details both indicators; the data source by which each indicator was measured; and the definitions of high, moderate, and low implementation fidelity. Schools that implemented at moderate or high levels met the threshold for adequate implementation.

Table 9. Indicators of Cycles of Inquiry (COI) Implementation

Indicator	Data source	High, moderate, and low implementation fidelity
Principals complete six initial professional development sessions (2018–19, 2019–20, and 2020–21).	Sign-in sheets	High = Six sessions Moderate = Four or five sessions Low = Three or fewer sessions
Principals participate in 60 minutes per month of COI coaching while school is in session (September–May 2018–19, 2019–20, 2020–21).	Coaching logs	High = 60 minutes or more per month, on average Moderate = 30 to less than 60 minutes Low = Less than 30 minutes

This section presents our findings on annual COI fidelity of implementation. Additional findings are presented in Appendix A.

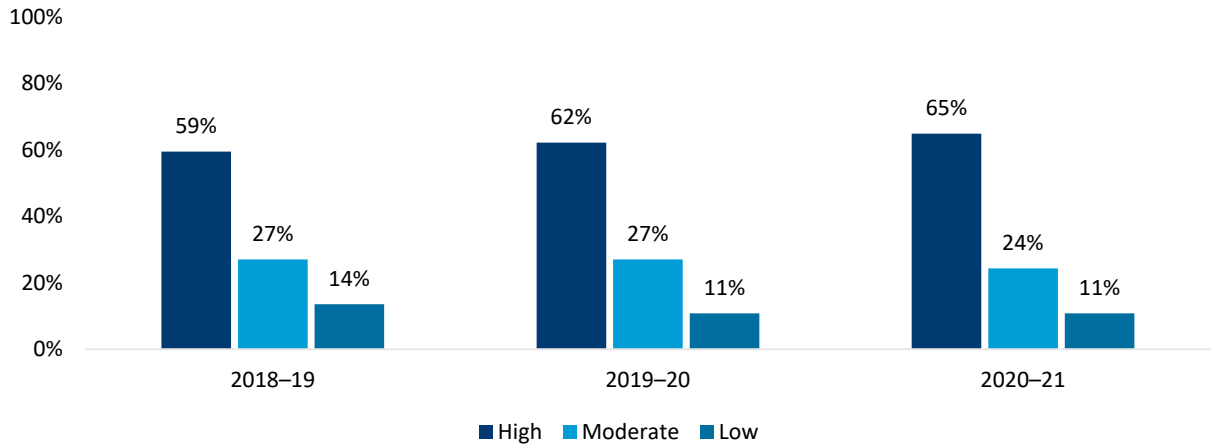
Attendance at COI PD sessions. Figure 9 illustrates the extent to which principals in each school completed the six initial COI trainings each school year. Continuing principals who had already

completed the trainings were not expected to repeat the trainings but are included in our counts of the number of schools implementing with high, moderate, or low fidelity.

In each of the 3 years, the majority of PtL schools implemented the COI attendance indicator with moderate or high fidelity: the percentage of principals who completed four or more of the six required trainings was 86% in 2018–19, 89% in 2019–20, and 89% in 2020–21. The percentage of principals who completed all six required trainings increased from 59% in 2018–19 to 62% in 2019–20 and 65% in 2020–21.

In some cases, coaches provided one-on-one COI trainings to principals who missed the initial and makeup COI training sessions. Because these one-on-one training sessions were not documented in the COI attendance data, the figures presented here may understate the true level of implementation of COI PD.

Figure 9. Cycles of Inquiry Professional Development Attendance: Percentage of Schools With High, Moderate, and Low Implementation Fidelity, by Year

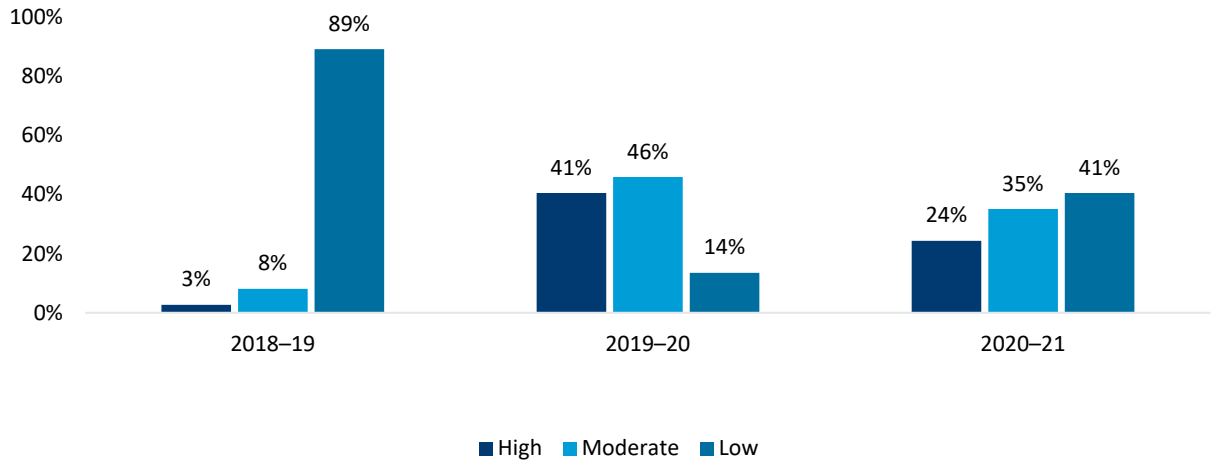


Note. This figure is based on 37 schools each year. See Table 4 and Table 9 for a description of the analytic approach. Data are based on sign-in sheets.

COI coaching intensity. Figure 10 reports implementation of COI coaching by school year. Principals were expected to engage in COI coaching only during months when school was in session. To meet the criteria for high fidelity, principals needed to engage in COI coaching for at least 60 minutes per month, on average.

In 2018–19, most principals (89%) participated in less than 30 minutes of COI coaching per month, on average, and only 11% of principals achieved moderate or high fidelity. However, the majority of principals were able to implement this component with moderate or high fidelity in subsequent years (87% in 2019–20 and 59% in 2020–21).

Figure 10. Cycles of Inquiry Coaching Intensity: Percentage of Principals With High, Moderate, and Low Implementation Fidelity, by Year



Note. This figure is based on 37 schools each year. See Table 4 and Table 9 for a description of the analytic approach. Data are based on coaching logs.

Overall Fidelity of Implementation

To assess whether PtL was implemented with fidelity across all COI indicators each year, AIR followed the same process as was followed for TUA and LF. First, each school was assigned points for each applicable indicator based on the level of implementation fidelity reached. Second, AIR compared the total number of points assigned to a school with the threshold for adequate implementation for the LF component, which varied by year. DuPage ROE and AIR decided that overall fidelity of LF implementation would be met if at least 90% of schools met the minimum threshold for the component.

COI had two indicators (principals' attendance at initial training sessions and participation in COI coaching) in each of the 3 years. Each indicator was worth up to 2 points, so a total of 4 points were possible each year. In each year, a school had to receive at least 2 of the 4 points to meet the fidelity-of-implementation standard (Table 10).

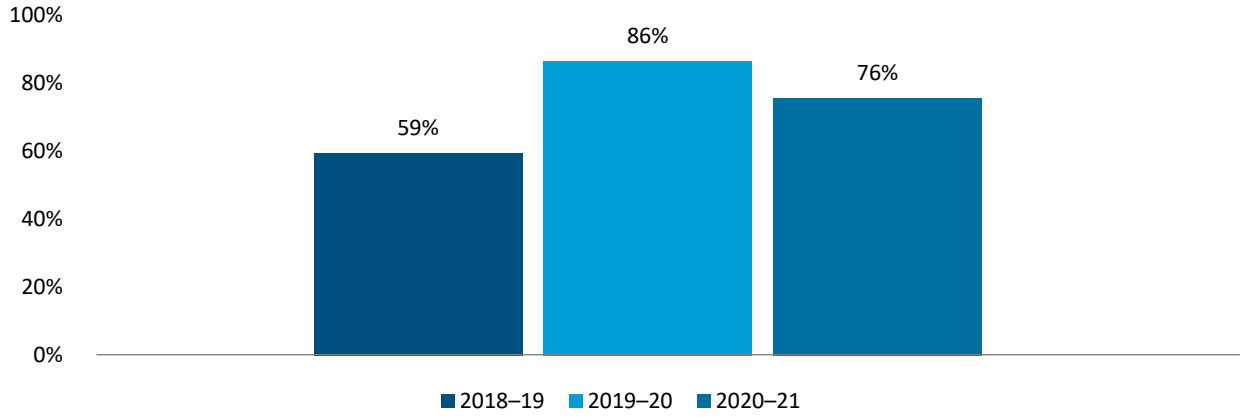
Table 10. Cycles of Inquiry (COI) Fidelity-of-Implementation Matrix

Program component	Expectations for fidelity of implementation	Indicators
COI 4 points total in 2018–19, 2019–20, and 2020–21	Principals complete six foundational training sessions (attendance).[a]	High = Six sessions (2 points) Moderate = Four or five sessions (1 point) Low = Three or fewer sessions (0 points)
	Principals participate in 60 minutes per month of COI coaching while school is in session (September–May).	High = 60 minutes or more per month, on average (2 points) Moderate = 30 to less than 60 minutes, on average (1 point) Low = Less than 30 minutes, on average (0 points)
	Adequate implementation at the school/principal level: 2018–19: 2 points 2019–20: 2 points 2020–21: 2 points	Adequate implementation at the program/sample level: 2018–19: 90% of schools meeting the threshold for adequate implementation 2019–20: 90% of schools meeting the threshold for adequate implementation 2020–21: 90% of schools meeting the threshold for adequate implementation

[a] Some foundational sessions were delivered by coaches, so the percentages reported for this indicator may be an underestimate.

Figure 11 reports the percentage of schools that achieved the minimum threshold for fidelity of COI implementation each year. In the 2018–19 school year, 59% of schools achieved the minimum threshold for fidelity of implementation. COI fidelity of implementation increased to 86% during the 2019–20 school year and then fell slightly to 76% of schools during the 2020–21 school year. Therefore, PtL schools did not attain the 90% target for fidelity of COI implementation in any of the 3 years.

Figure 11. Percentage of Schools That Met the Threshold for Implementation Fidelity for Cycles of Inquiry (COI), by Year



Note. Data are based on authors' analyses of COI implementation data.

Thus, for most components and in most years, the sample as a whole did not meet the minimum threshold of 90% of schools implementing with fidelity. The only exception was the LF component in 2019–20 (92%).

Fidelity of PtL implementation more than likely was greatly influenced by state and district responses to the COVID-19 pandemic. In March 2020, the state governor closed schools for in-person instruction. Because the elements of the TUA and LF components changed each year, it is not possible to track the pandemic's impact on TUA and LF implementation fidelity. However, the share of principals implementing COI components with fidelity increased from 2018–19 to 2019–20, but fell from 2019–20 to 2020–21.

Summary of Implementation Findings

Despite school closures and disruptions associated with the COVID-19 pandemic, PtL continued to provide PD and other supports to school leaders with moderate to high fidelity on many indicators, but, at the program level, fidelity of implementation was low across all 3 years.

- **Overall, none of the 37 schools participating in PtL met the threshold for overall implementation of the TUA component in any year, although 100% schools met the threshold for the TUA training component each year.**
  - Only 11% of principals participated in the TUA calendaring process in 2019–20. Principals found the task of coding their calendars to be overly burdensome, which was counterproductive to the ultimate goals of the TUA process (i.e., increased principal time allocated to instructional leadership).

  - No principals participated in an average of at least 30 minutes of TUA coaching per month in 2018–19, 2019–20, and 2020–21. Because PtL is a new program, the minimum thresholds for TUA coaching were based not on prior research but on educated guesses by DuPage ROE and AIR about what threshold might be needed to be able to effect change. Because no principals participated in at least 30 minutes of TUA coaching per month, on average, the threshold for TUA coaching set by DuPage ROE and AIR may have been set too high. Moreover, coaches may have had difficulty distinguishing among TUA coaching, COI coaching, and other coaching, resulting in coaches underreporting the true number of minutes spent on TUA coaching.

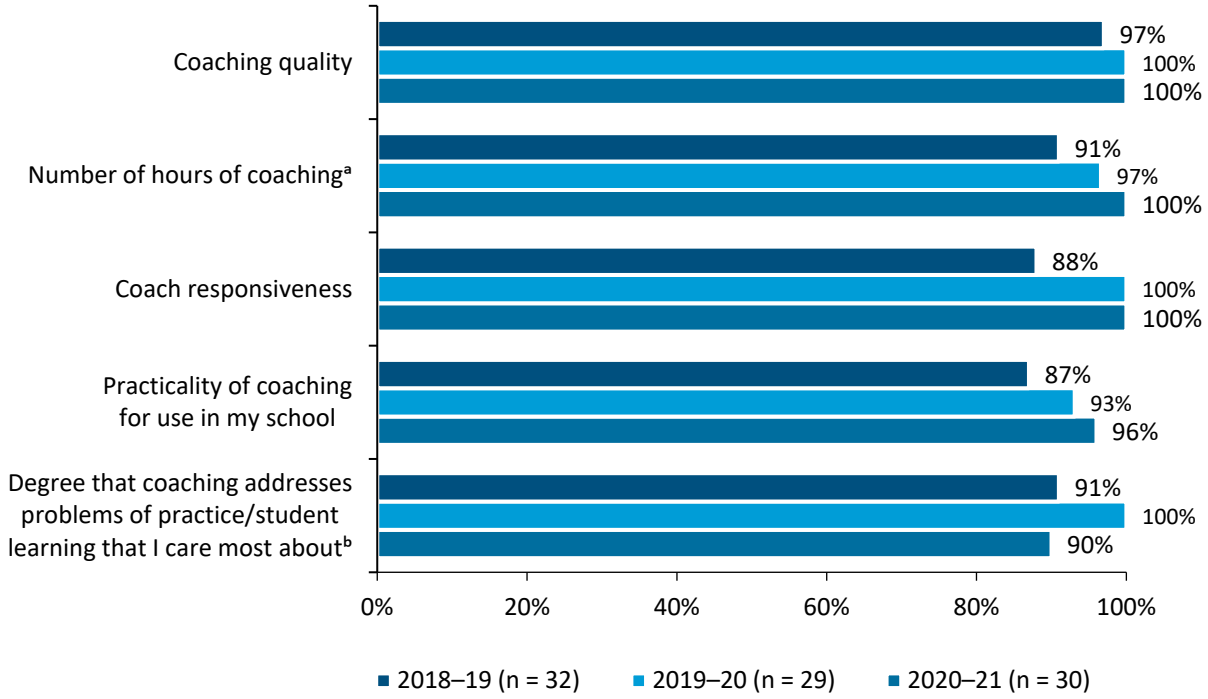
- **Implementation of the LF component was mixed.**
  - Data on LF implementation are not available for 2018–19.
  - Ninety-two percent of PtL schools met the minimum threshold for LF implementation in 2019–20, exceeding the 90% threshold set by DuPage ROE and AIR.
  - However, only 70% of PtL schools met the minimum threshold in 2020–21, falling below the 90% goal set by DuPage ROE and AIR.
  - Principals in 97% of schools had identified an ILT by 2019–20.
  - Principals or their designees met with the school's ILT at least 35 minutes per month, on average, in 73% of schools in 2019–20 and in 46% of schools in 2020–21.
  - ILT members met with teacher teams at least 35 minutes per month, on average, in 89% of schools in 2019–20 and in 73% of schools in 2020–21.
- **The percentage of PtL schools meeting the minimum threshold for implementation of the COI component increased from 59% in 2018–19 to 86% in 2019–20 and fell to 76% in 2020–21. These rates are all below the 90% threshold set by DuPage ROE and AIR.**
  - The percentage of principals who had completed all six initial COI trainings increased from 59% in 2018–19 to 62% in 2019–20 and 65% in 2020–21.
  - The percentage of principals who participated in at least 30 minutes of COI coaching per month, on average, increased from 11% in 2018–19 to 86% in 2019–20 and fell to 59% in 2020–21. Minimum thresholds for COI coaching were based not on prior research but on educated guesses by DuPage ROE and AIR about what threshold might be needed to be able to effect change. Because no principals participated in at least 30 minutes of COI coaching per month, on average, the threshold for COI coaching set by DuPage ROE and AIR may have been set too high. Moreover, coaches may have had difficulty distinguishing among TUA coaching, COI coaching, and other coaching, resulting in coaches underreporting the true number of minutes spent on COI coaching.

Perceptions of and Satisfaction With PtL

To gather insight into PtL principals' satisfaction with coaching, COI, and other program supports, AIR administered an annual survey to all PtL principals and interviewed a sample of principals in the spring of each school year from 2018–19 through 2020–21. The 2020 and 2021 surveys also included a question to gauge principals' satisfaction with the information and advice they received about how to lead their schools during the COVID-19 pandemic. AIR received survey responses from 86% of principals (*n* = 32) in 2018–19, 78% of principals (*n* = 29) in 2019–20, and 81% of principals (*n* = 30) in 2020–21. See Appendix B for additional information about the principals who responded to the survey.

Satisfaction with coaching supports. Figure 12 reports the percentage of survey respondents who reported that they were "satisfied" or "very satisfied" with the PtL coaching supports. Each year, most PtL principals reported satisfaction with the coaching supports they received. Specifically, between 88% and 100% of responding principals reported that they were either "very satisfied" or "satisfied" with each of the following: (a) coaching quality, (b) number of hours of coaching, (c) coach responsiveness, (d) practicality of coaching for use in their school, and (e) a focus on problems of practice or student learning that they care about. The percentage of principals who reported satisfaction with coaching number of hours, responsiveness, and practicality was 6 to 12 percentage points higher in 2020 and 2021 than in 2019. The percentage of principals who reported satisfaction with the degree that coaching addresses the problems of practice or student learning that they care about was 9 to 10 percentage points higher in 2020 than in 2019 or 2021.

Figure 12. Percentage of Principals Who Reported Being "Very Satisfied" or "Satisfied" With Each of the Five Aspects of PtL Coaching



Note. Data are based on the Annual Principal Survey, 2019, 2020, and 2021.

[a] In 2020–21, the survey text was "Total coaching hours."

[b] In 2018–19 and 2019–20, the survey text was "Degree that coaching addresses problems of practice that I care most about," and in 2020–21, the survey text was "Degree that coaching addresses problems of student learning that I care most about."

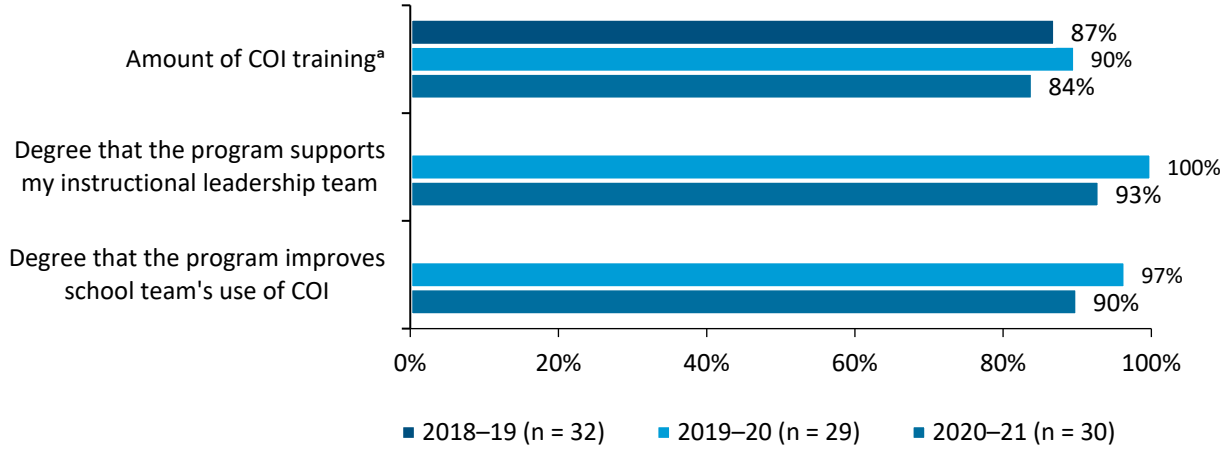
Principals provided additional feedback about the coaching support they received during interviews. In 2020–21, one principal stated, "The coach is great . . . that's been probably one of the bright spots is just having that intentional time to speak to someone and help navigate some of these movements." Another principal who was interviewed in 2019–20 elaborated on the nature of their relationship with their coach: "I have a one-on-one relationship with my coach. She's honest and she pushes me to do things and try things that I might otherwise put off or say, 'Oh, that's a great idea,' but then never do it. So, it's sort of an accountability piece for me but definitely the relationship piece." One principal who was interviewed in 2018–19 also had positive comments about their coach:

> I really enjoy working with my coach. I think being in this grant and allowing me to have this time with the coach has helped me transition into being a principal more than my school district has. I think that the conversation that [my coach and I are] having is what I need, especially because I'm here by myself as the administrator. So I [can] talk through all those processes with her, and [she is] coaching me through the cycle of the

> "What are the steps?" Or "What are you even thinking about doing?" And so, I can talk through, "I have these three ideas," and it's nice to just talk through that and be able to have someone to listen to what I'm thinking.

Satisfaction with COI. Figure 13 shows the percentage of survey respondents who reported that they were "satisfied" or "very satisfied" with the COI components. At least 80% of principal respondents reported in each of the 3 years that they were "satisfied" or "very satisfied" with the amount of COI training. In 2019–20 and 2020–21, at least 90% of principals reported that they were "satisfied" or "very satisfied" with the degree to which PtL supports their teacher teams and the degree to which PtL improved schools' use of COI.[3] The percentage of principals who reported satisfaction with the three COI components was between 6 and 7 percentage points higher in 2020 than in 2021. Additional annual survey findings are presented in Appendix B.

Figure 13. Percentage of Principals Who Reported Being "Very Satisfied" or "Satisfied" With Each of the Three Cycles of Inquiry (COI) Aspects of PtL



Note. Data are based on the Annual Principal Survey, 2019, 2020, and 2021.

[a] In 2020–21, the survey text was "Amount of training for COI."

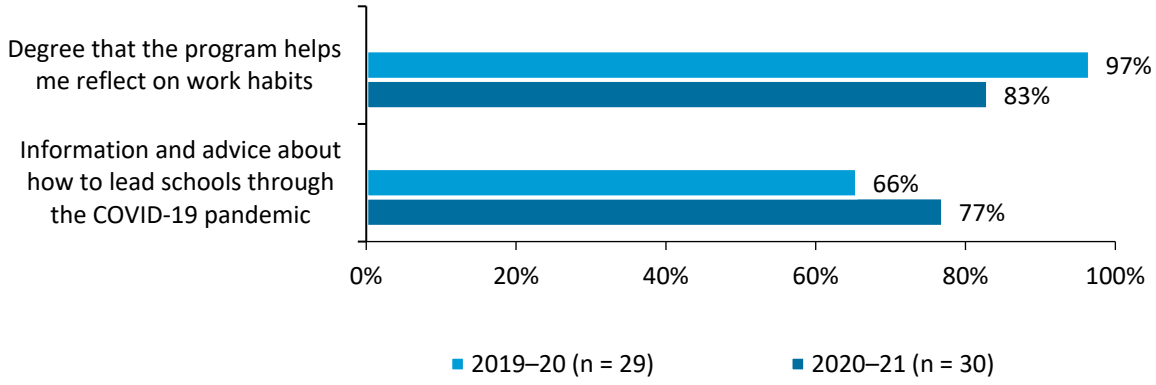
Principals who participated in interviews discussed how they valued the applicability of the training to their buildings. In particular, one principal who was interviewed in 2020–21 discussed how the process was applied to school improvement planning, saying, "It allows us to make it a living, breathing, working document [rather] than a template that ends up on the shelf." Another principal who was interviewed in 2020–21 explained how leadership teams are independently implementing the process: "The beauty of the cycle of inquiry as we've gone through this process is it's very much a collaborative conversation with me and the instructional team. A lot of times I will actually sit back and let them kind of lead." Another principal who was

[3] In 2018–19, AIR did not ask principals to report their level of satisfaction with the "Degree that the program improves school team's use of COI" or the "Degree that the program supports my instructional leadership team."

interviewed in 2019–20 said, "The Instructional Leadership Team and I [are] much more intentional in terms of looking at academic data and looking at our behavior data and really trying to connect the dots ... and make sure ... groups are working in unison [toward] our one school improvement goal."

Satisfaction with other supports. Figure 14 reports the percentage of survey respondents who reported that they were "satisfied" or "very satisfied" with (a) the degree to which PtL helped them reflect on their work habits and (b) the information that PtL provided them about how to lead schools during the pandemic. The percentage of principals reporting satisfaction with the degree that PtL helped them reflect on their work habits decreased from 2020 (97%) to 2021 (83%). However, although relatively fewer principals reported satisfaction with the information that PtL provided them about how to lead schools during the pandemic, this percentage increased from 2020 (66%) to 2021 (77%).

Figure 14. Percentage of Principals Who Reported Being "Very Satisfied" or "Satisfied" With Other PtL Supports



Note. Data are based on the Annual Principal Survey, 2020 and 2021.

One principal who was interviewed in 2018–19 discussed how they hoped their approaches to school improvement will change as a result of engaging in PtL:

> I want to be able to educate my teachers on this [project] and really emphasize the importance of this work, and help them realize how effective it is and really get them to be able to not put any blame on the students ... to start looking at some of our instructional practices and some of our own instructional shortcomings and be open and willing to improve some of those instructional strategies within our classroom.

Two principals who were interviewed in 2020–21 discussed the additional funding provided by the PtL program as a benefit. One of these principals stated that it's "not just the professional

development, but the financial support that goes along with it, that's helped make implementing some of these things possible."

Impact of PtL

AIR evaluated the impact of PtL program participation on three sets of outcomes: (a) principal leadership effectiveness as captured by the Illinois 5Essentials Survey, (b) schoolwide student achievement, and (c) principal retention. Our primary contrast compares spring 2021 outcomes in PtL schools with spring 2021 outcomes of schools that were similar to PtL schools in the baseline year (2017–18), before PtL implementation began in the participating schools.

AIR used quasi-experimental designs to compare the outcomes of schools participating in the PtL program with those of nonparticipating (i.e., comparison) schools with similar characteristics. Specifically, we used a difference-in-differences (DID) design to evaluate the impact of PtL participation on leadership effectiveness, we used a comparative interrupted time series (CITS) design to evaluate the impact of PtL participation on student achievement, and we used a matched-comparison group design to evaluate the impact of PtL on principal retention. See Appendix C for a description of AIR's analytic models.

The findings from AIR's independent analysis of PtL impact can be summarized as follows:

- Accounting for baseline scores, school level, and student characteristics, schools that participated in PtL had lower 5Essentials Effective Leaders scores, on average, than comparison schools in spring of the final year of the intervention. The difference is equivalent to an effect size of −0.064 and is not statistically significant ($p < .05$).
- After controlling for other factors included in the statistical model, we estimate that students in intervention schools scored 0.065 standard deviation (SD) higher in ELA and 0.104 SD higher in math in spring of the final year of the intervention. These differences, which are equivalent to percentile rank increases of 2.6 in ELA and 4.1 in math, are not statistically significant ($p < .05$), although the p value for the estimated positive impact on student math achievement is 0.06.
- After controlling for baseline school and principal characteristics, we estimate that principal retention was 12.7 percentage points higher in intervention schools than in comparison schools, an effect size of 0.33 SD. However, this difference is not statistically significant ($p < .05$).

From March 2020 through the end of the intervention, some PtL principals may have become somewhat disengaged from the full PtL PD and coaching due to competing professional and

personal priorities, which may have attenuated the impact of PtL. It is likely that the change in priorities necessitated by the pandemic attenuated PtL's impact.

Leadership Effectiveness

Leadership Effectiveness Outcome Measures

To measure principal leadership practices, AIR used the Illinois 5Essentials Survey, a validated school condition and culture survey administered annually and statewide to teachers and students (Hart et al., 2020), as a measure of general leadership effectiveness in schools. The five "measures" have demonstrated reliability, each with a Cronbach's alpha of 0.8 or higher (Klugman et al., 2015). Illinois public schools are required to participate in the survey each year, although districts may elect to participate in approved alternate school climate surveys.

Specifically, we used the Effective Leaders scale from the survey. This scale includes four measures: Instructional Leadership, Program Coherence, Teacher Influence, and Teacher-Principal Trust. Instructional Leadership reflects the degree to which teachers see the school leadership team as instructional leaders. Program Coherence focuses on the extent that school programs are coordinated and consistent with their goals for student learning. Teacher Influence focuses on the extent that teachers have influence in a range of school policies and practices. Teacher-Principal Trust focuses on the extent that teachers and principals share mutual trust and respect. Effective Leaders scores range from 0 to 100.

The analysis is based on data from the 2016–17, 2017–18, 2018–19, and 2019–21 administrations of the survey. Because, prior to 2018–19, schools conducted the Illinois 5Essentials Survey only every other year (rather than annually), the AIR team does not have 2017–18 Effective Leaders scores as our baseline measure for some schools. In the cases of schools without 2017–18 Effective Leaders scores, we used the scores from 2016–17 as the baseline measure. However, because of low teacher rates of response to the Illinois 5Essentials Survey, some schools in the analysis lack baseline Effective Leaders scores from either year. Thus, AIR excluded schools with no Effective Leaders scores from either 2016–17 or 2017–18 from the baseline equivalence and impact analyses.

Numbers of Schools in the Analysis Sample

Because 5Essentials Effective Leaders outcomes are not available for schools where less than 50% of teachers completed the survey, we were not able to include the full set of PtL schools in our analysis. Our analysis sample for measuring program impact on schoolwide leadership effectiveness includes 104 schools with outcome data for both the baseline and summative outcome years: 26 (of 37) intervention schools and 78 comparison schools (Table 11). Schools with missing data for the baseline or summative outcome years were excluded from the analysis of baseline equivalence and program impact. The total number of schools included in

the 2018–19 outcome analysis is fewer than 104 due to lower response rates in 2018–19, and 5Essentials scores were not available for 2019–20 due to the COVID-19 pandemic. If the impact of PtL in these 26 schools is not representative of the impact of all 37 schools participating in the intervention, our estimates of the impact of PtL on leadership effectiveness may be biased.

Table 11. Number of Schools in the Analysis of Program Impact on Leadership Effectiveness, by Year

Group	Baseline	2018–19	2020–21
Intervention	26	25	26
Comparison	78	75	78
Total	104	100	104

Representativeness of Teachers Within Schools

To receive the rating *Meets WWC Standards With Reservations*, cluster-level assignment studies must demonstrate that individuals in the analytic sample in the baseline and outcome time periods are representative of clusters (WWC, 2022). For our analysis of program impact on leadership effectiveness, we demonstrate that teachers who responded to the Illinois 5Essentials Survey during the baseline (2016–17 or 2017–18) and outcome (2020–21) years are representative of all teachers in their schools at that time. We first obtained teacher response rates in each school in the analytic sample each year and then calculated average response rates in intervention and comparison schools, as well as average response rates across both groups of schools, each year. Results are presented in Table 12.

Table 12. 5Essentials Teacher Response Rates at Baseline and Outcome

Group	Baseline (2016–17 or 2017–18)		Outcome (2020–21)	
	Average response rate	Average attrition	Average response rate	Average attrition
Intervention	80.6%	19.4%	78.6%	21.4%
Comparison	86.2%	13.8%	78.3%	21.7%
Overall	84.8%	15.2%	78.4%	21.6%

Note. Rates are averages across schools.

At baseline (2016–17 or 2017–18), overall attrition was 15.2%, and differential attrition was 5.6 percentage points. In 2020–21, overall attrition was 21.6%, and differential attrition was less than 1 percentage point. These overall and differential attrition rates are within the acceptable range under both optimistic and cautious assumptions (WWC, 2022).

Assessment of Baseline Equivalence

AIR used propensity score matching to identify a suitable comparison group for the impact analysis. Each intervention school was matched to three similar comparison schools, and comparison schools were identified separately for elementary, middle, and high schools. Schools were matched based on the following characteristics: rural status; percentage of students in the school who were ELs, had a disability, were eligible for the National School Lunch Program, and were non-White in 2017–18 (the baseline year); student enrollment in 2017–18; and baseline Effective Leaders scores.

Estimating program impact in a quasi-experimental design study such as the PtL project requires that baseline equivalence between the intervention and comparison groups must be established in the analytic sample (WWC, 2022). Because schools in the analysis sample have baseline measures from different years, AIR used linear regression to calculate the baseline equivalence of 5Essentials scores. The model used fixed effects to compare intervention and comparison schools with the same school level (elementary, middle, or high) and year of baseline outcome (2016–17 or 2017–18). Accounting for school level and year of baseline measure, Effective Leaders scores were 0.84 points lower at baseline in intervention schools than in comparison schools, on average (Table 13). The absolute value of the standardized mean difference (SMD) is 0.049, which is lower than the 0.05 threshold set by WWC to demonstrate baseline equivalence between intervention and comparison groups (WWC, 2022).

Table 13. Baseline Equivalence of 5Essentials Effective Leaders Scores

5Essentials measure	Comparison schools			Intervention schools			Difference	SMD
	N	Mean	SD	*N*	Adjusted mean	SD		
Effective Leaders	78	43.090	17.049	26	43.929	17.371	0.839	0.049

Note. *N* is the number of schools. Mean is the mean outcome across schools. Adjusted mean is the regression-adjusted mean outcome across schools. SD is the standard deviation of outcomes across schools. Difference is the difference between the regression-adjusted mean outcome in intervention schools and the mean outcome in comparison schools. SMD is the standardized mean difference in outcomes at baseline. See Appendix C for additional details.

Program Impact

The AIR study team employed a DID model to examine the impact of the PtL program on leadership effectiveness. The analysis accounted for the following school characteristics: school level (elementary, middle, or high), enrollment, the percentage of Black or African American students, the percentage of Hispanic or Latino students, the percentage of Asian American students, the percentage of White students, the percentage of students eligible for the National School Lunch Program, the percentage of ELs, and the percentage of students with a disability.

To account for correlation of school-level outcomes over time, the analysis accounted for clustering of observations within schools by including school-level random intercept terms in the analytic model (Appendix C).

To meet WWC standards, DID analyses must demonstrate that the correlation between baseline and outcome measures is 0.60 or higher. The correlation between baseline and 2021 Effective Leaders scores is 0.49.

Estimates of program impact on 5Essentials Effective Leaders scores are presented in Table 14. Accounting for baseline scores, school level, and characteristics of students attending the school, schools that participated in PtL had Effective Leaders scores that were 0.87 points lower, on average, than comparison schools. The difference is not statistically significant. Standard error estimates from Table 41 in Appendix C indicate that our analysis of PtL's impact on leadership effectiveness had sufficient power to detect effect sizes of 0.654 or greater.

Table 14. Program Impact on 5Essentials Effective Leaders Scores

5Essentials measure	Comparison schools			Intervention schools			Effect size	*p* Value
	N	Mean	SD	*N*	Adjusted mean	SD		
Effective Leaders	78	48.564	12.866	26	47.692	15.885	-0.064	.773

Note. *N* is the number of schools. Mean is the mean outcome across schools. Adjusted mean is the regression-adjusted mean outcome across schools. SD is the standard deviation of outcomes across schools. *P* value is the estimated probability that the observed regression-adjusted difference between intervention and comparison schools would be as large (in absolute value) as observed if the intervention had no impact on the outcome. See Appendix C for additional details.

Student Achievement

Student Achievement Outcome Measures

The assessments used to measure the ELA and math achievement of students in Grades 3 through 8 and 11 differ across student grades and school years as follows:

- Grades 3 through 8, 2014–15 to 2017–18: Partnership for Assessment of Readiness for College and Careers ELA and math student-level test scores
- Grades 3 through 8, 2018–19 and 2020–21: IAR ELA and math student-level test scores
- Grade 11, 2016–17 to 2018–19 and 2020–21: SAT® evidence-based reading (ELA) and math

Since the 2016–17 school year, all Illinois public school students in Grade 11 have been required to complete the SAT®. The SAT® serves as the state assessment for state and federal

accountability. The SAT® is provided at no cost to public school districts in Illinois and is offered to students during the school day.

As in most places, students in Illinois schools were not tested in spring 2020 because of the COVID-19 pandemic. Therefore, the AIR study team could not evaluate the impact of the PtL program on student achievement in the 2019–20 school year.

Numbers of Schools and Students in the Analysis Sample

The numbers of intervention and comparison schools included in the analysis each year are presented in Table 15. The analysis sample includes 36 intervention and 108 comparison schools in 2016–17, 2017–18, 2018–19, and 2020–21. Because the Illinois State Board of Education (ISBE) did not require Grade 11 to take the SAT® prior to 2016–17, 11 intervention and 33 comparison high schools were excluded from the 2014–15 and 2015–16 analyses.

Table 15. Number of Schools in the Analysis of Program Impact on English Language Arts and Math Achievement, by Year

Group	2014–15	2015–16	2016–17	2017–18	2018–19	2020–21
Intervention	25	25	36	36	36	36
Comparison	75	75	108	108	108	108
Total	100	100	144	144	144	144

Note. Because the Illinois State Board of Education did not require all Grade 11 students to take the SAT® prior to 2016–17, high schools were excluded from the 2014–15 and 2015–16 analyses.

The numbers of students included in the analysis each year are presented in Table 16 and Table 17. For the analysis of program impact on ELA achievement, the number of students in intervention schools included in the analysis ranges from 7,378 in 2015–16 to 9,994 in 2016–17, and the number of students in comparison schools included in the analysis ranges from 25,230 in 2015–16 to 32,716 in 2016–17. For the analysis of program impact on math achievement, the number of students in intervention schools included in the analysis ranges from 7,028 in 2015–16 to 9,802 in 2018–19, and the number of students in comparison schools included in the analysis ranges from 25,247 in 2015–16 to 32,765 in 2016–17.

Table 16. Number of Students in the Analysis of Program Impact on English Language Arts Achievement, by Year

Group	2014–15	2015–16	2016–17	2017–18	2018–19	2020–21
Intervention schools	7,487	7,378	9,994	9,696	9,826	7,790
Comparison schools	25,535	25,230	32,716	32,350	32,501	27,010
Total	33,022	32,608	42,710	42,046	42,327	34,800

Note. Because the Illinois State Board of Education did not require all Grade 11 students to take the SAT® prior to 2016–17, high school students were excluded from the 2014–15 and 2015–16 analyses.

Table 17. Number of Students in the Analysis of Program Impact on Math Achievement, by Year

Group	2014–15	2015–16	2016–17	2017–18	2018–19	2020–21
Intervention schools	7,539	7,028	9,676	9,391	9,802	7,737
Comparison schools	25,559	25,247	32,765	32,354	32,486	26,774
Total	33,098	32,275	42,441	41,745	42,288	34,511

Note. Because the Illinois State Board of Education did not require all Grade 11 students to take the SAT® prior to 2016–17, high school students were excluded from the 2014–15 and 2015–16 analyses.

Representativeness of Students Within Schools

The number of Grades 3 through 8 and 11 students tested was noticeably lower in 2020–21 than in 2016–17, 2017–18, or 2018–19. To demonstrate that students included in the analytic sample in 2020–21 were representative of all Grades 3 through 8 and 11 students in intervention and comparison schools, AIR calculated overall and differential rates of "attrition"—the percentage of students in Grades 3 through 8 and 11 who were enrolled in intervention and comparison schools but were not tested—in both ELA and math in 2020–21. The results are presented in Table 18. The overall attrition rate was 16% in ELA and 17% in math, and the differential attrition rates were 3% in both subjects. These rates meet WWC standards for representativeness at both the optimistic and cautious boundaries (WWC, 2022).

Table 18. Student English Language Arts and Math Assessment Participation Rates, 2020–21

Group	English language arts			Math		
	Enrolled	Tested	Attrition	Enrolled	Tested	Attrition
Intervention schools	9,576	7,790	19%	9,576	7,737	19%
Comparison schools	31,999	27,010	16%	31,997	26,774	16%
Overall	41,575	34,800	16%	41,573	34,511	17%

Note. Because the Illinois State Board of Education did not require all Grade 11 students to take the SAT® prior to 2016–17, high school students were excluded from the 2014–15 and 2015–16 analyses.

Assessment of Baseline Equivalence

AIR used propensity score matching to identify a suitable comparison group for the impact analysis. Each intervention school was matched to three similar comparison schools, and comparison schools were identified separately for elementary, middle, and high schools. Schools were matched based on the following characteristics: percentage of students in the school who were ELs, had a disability, were eligible for the National School Lunch Program, and were non-White in 2017–18 (the baseline year); number of students tested in 2017–18; and schoolwide ELA and math proficiency rates in spring 2017 and 2018. Elementary and middle schools also were matched on schoolwide ELA and math proficiency rates in spring 2015 and 2016.

Because our analysis of achievement impact is based on student-level data, but entire schools were assigned to the treatment or comparison condition, AIR used multilevel mixed-effects regression models that account for clustering of students within schools to measure baseline equivalence in ELA and math achievement between the intervention and comparison groups (Appendix C).

SMDs in baseline achievement between intervention and comparison schools were very small: after accounting for school level and clustering of students within schools, students in intervention schools scored 0.02 SD lower in ELA and 0.01 SD lower in math than students in comparison schools at baseline (Table 19). These values are lower than the 0.05 threshold set by WWC to demonstrate baseline equivalence between intervention and comparison groups without requiring adjustment for baseline scores in the impact analysis (WWC, 2022). In other words, intervention schools and comparison schools were nearly identical at baseline on achievement in both ELA and math.

Table 19. Baseline Equivalence of Student Achievement Outcomes

Achievement measure	Students in comparison schools			Students in intervention schools			SMD
	N	Mean	SD	*N*	Adjusted mean	SD	
English language arts	32,350	0.005	1.005	9,696	–0.012	0.984	0.017
Math	32,354	0.010	1.002	9,391	–0.001	0.992	0.011

Note. Achievement measure is an individual student's 2017–18 English language arts or math score, standardized within each grade and subject. *N* is the number of students. Mean is the mean outcome across students. SD is the standard deviation of outcomes across students. Adjusted mean is the regression-adjusted mean outcome across students. SMD is the standardized mean difference in outcomes at baseline. See Appendix C for additional details.

Characteristics of students included in the student achievement impact analysis for the baseline year (2017–18) are reported in Table 20. Although WWC does not require intervention and comparison samples to be equivalent on background characteristics, demonstration of similarity on these types of characteristics provides reassurance that the comparison schools demographically resembled the intervention schools and constituted an appropriate comparison group. SMDs are less than 0.25 in absolute value across all characteristics reported in Table 20 for both ELA and math.

Table 20. Characteristics of Students Included in the Student Achievement Impact Analysis at Baseline (2017–18)

Student characteristic	English language arts			Math		
	Comparison mean	Intervention mean	SMD	Comparison mean	Intervention mean	SMD
English learner	6.0%	4.5%	–0.184	6.2%	4.8%	–0.161
Eligible for the National School Lunch Program	39.9%	38.3%	–0.041	39.9%	39.4%	–0.014
Students with a disability	12.6%	12.2%	–0.025	12.6%	12.6%	–0.003
Non-White	36.2%	36.1%	–0.004	36.3%	36.2%	–0.002
Grade 3	8.2%	9.0%	0.062	8.2%	9.2%	0.081
Grade 4	9.6%	9.3%	–0.021	9.6%	9.6%	0.001
Grade 5	9.8%	9.4%	–0.022	9.8%	9.7%	–0.002
Grade 6	16.4%	15.0%	–0.060	16.4%	15.5%	–0.038
Grade 7	16.3%	16.1%	–0.005	16.3%	15.8%	–0.021
Grade 8	16.6%	15.6%	–0.042	16.5%	13.7%	–0.130
Grade 11	23.3%	25.5%	0.072	23.3%	26.3%	0.099

Note. Reported means are unadjusted means across students in intervention and comparison schools during the baseline year (2017–18). Numbers of students included in the English language arts and math student achievement impact analyses in 2017–18 are reported in Table 16 and Table 17, respectively. SMD is the standardized mean difference as measured by Cox's index.

Program Impact

AIR used CITS models to estimate program impact on student achievement in ELA and math. Details of the statistical models are presented in Appendix C. The models use statistical controls to account for average differences between intervention and comparison schools in student achievement levels and trends during the baseline period (2014–15 through 2017–18).

The models account for individual student characteristics, including EL status, eligibility for the National School Lunch Program, disability status, grade, race, and Hispanic ethnicity. The models also account for school level (elementary, middle, or high) and urbanicity as well as characteristics of the school that change over time, including enrollment and percentages of students who are ELs; are eligible for the National School Lunch Program; have a disability; and are Hispanic, African American, Asian American, or White.

In addition, because students are the unit of observation in our analysis, but assignment to the intervention and comparison groups was at the school level, our models account for the clustering of students within schools (Abadie et al., 2017). Specifically, we used multilevel mixed-effects regression models that account for clustering of students within schools. Further details of the statistical models are presented in Appendix C.

The models compare deviations from expected trends (based on the baseline years) in achievement in intervention schools in the outcome year (especially 2020–21) with deviations in expected trends in comparison schools in the same year. As with the other outcomes, we consider regression-adjusted differences in student achievement between intervention and comparison schools in 2020–21 to be the confirmatory measure of program impact on student achievement.

Estimates of the impact of PtL on student achievement through the third year of the intervention (2020–21) are presented in Table 21. After controlling for the factors included in the statistical model, we estimate that students in intervention schools scored 0.065 SD higher in ELA and 0.104 SD higher in math. These differences are equivalent to percentile rank increases of 2.6 in ELA and 4.1 in math. We are unable to reject the null hypothesis of no program impact on student achievement with a p value less than .05, although the p value for the estimated positive impact on student math achievement is .06. Standard error estimates from Table 44 and Table 45 indicate that our analysis of PtL's impact on student achievement had sufficient power to detect effect sizes as small as 0.171 in ELA and 0.162 in math.

Table 21. Program Impact on Student Achievement in 2020–21

Achievement measure	Students in comparison schools			Students in intervention schools			Effect size	*p* Value
	N	Mean	SD	*N*	Adjusted mean	SD		
English language arts	27,010	0.004	1.008	7,790	0.069	0.973	0.065	.288
Math	26,774	–0.005	0.997	7,737	0.099	1.010	0.104	.063

Note. *N* is the number of students. Mean is the mean outcome across students. SD is the standard deviation of outcomes across students. Adjusted mean is the regression-adjusted mean outcome across students. *P* value is the estimated probability that the observed regression-adjusted difference between students in intervention and comparison schools would be as large (in absolute value) as observed if the intervention had no impact on the outcome. See Appendix C for additional details.

Principal Retention

Principal Retention Outcome Measure

To analyze program impact on principal retention, AIR obtained individual-level personnel data from ISBE. The measure of principal retention is a binary indicator equaling 1 if the school's principal in spring 2018 continued to serve as the school's principal through spring 2021 and 0 otherwise. The sample used to analyze the impact of PtL on principal retention includes all 37 intervention and 111 matched-comparison schools.

Selection of Comparison Schools

AIR used propensity score matching to identify a matched set of comparison schools for analysis of program impact on principal retention. Because the intervention commenced in late summer 2018, we consider spring 2018 to be the appropriate baseline time period for this outcome. Schools were matched exactly on school level (elementary, middle, or high), and estimates of schools' propensity to participate in the intervention were based on seven principal and school characteristics: principal retention in the school between spring 2017 and spring 2018, whether the school's principal in spring 2018 was a novice principal,[4] whether the school's principal in spring 2018 was non-White, schoolwide ELA proficiency rates in spring 2018, schoolwide math proficiency rates in spring 2018, the percentage of students at the school who were eligible for the National School Lunch Program in spring 2018, and schoolwide enrollment in spring 2018.

[4] We define a principal's first, second, and third years as a school principal as the period in which the principal is a novice principal.

Assessment of Baseline Equivalence

Baseline equivalence for principal retention and principal novice status (i.e., binary measures) was calculated using Cox's index, and baseline equivalence for ELA and math proficiency and the percentage of students at the school eligible for the National School Lunch Program was calculated using Hedges' *g*.

PtL intervention and comparison schools were similar at baseline (Table 22). Retention rates, the percentage of novice principals, schoolwide percentages of students eligible for the National School Lunch Program, and schoolwide ELA proficiency rates were slightly higher in intervention schools at baseline, on average, whereas schoolwide math proficiency rates were slightly lower. None of the SDs were greater than 0.11 in absolute value. Because the SMD of the retention rate—the outcome variable—was greater than 0.05 (but still less than 0.25), our impact analyses adjust for the measure at baseline.

Table 22. Baseline Equivalence for Principal Retention Analysis

School or principal characteristic	Comparison schools			Intervention schools			Difference	SMD
	N	Mean	SD	*N*	Mean	SD		
Retention rate	111	84.7%	0.362	37	86.5%	0.347	0.018	0.088
Novice principal	111	28.8%	0.455	37	32.4%	0.475	0.036	0.102
Schoolwide percentage of students eligible for the National School Lunch Program	111	39.6%	0.167	37	40.4%	0.171	0.008	0.047
Schoolwide English language arts proficiency rate	111	36.4%	0.136	37	36.8%	0.11	0.004	0.031
Schoolwide math proficiency rate	111	33.5%	0.150	37	32.9%	0.138	–0.006	–0.041

Note. *N* is the number of schools. Mean is the mean outcome across schools. SD is the standard deviation across schools. SMD is the standardized mean difference in outcomes at baseline. SMD is calculated using Cox's index for the retention rate and percentage of principals who are novice principals, and Hedges' *g* is used for all other baseline measures. Retention is principal retention in the same school between spring 2017 and spring 2018. See Appendix C for additional details.

Program Impact

We used a matched-comparison design to estimate program impact on principal retention. The model accounts for school level (elementary, middle, or high), principal characteristics at baseline (novice status, non-White race/ethnicity, and degree/certificate attainment), and school characteristics at baseline (principal retention between spring 2017 and spring 2018; enrollment; schoolwide ELA and math proficiency rates; and percentages of students who are White, are ELs, are eligible for the National School Lunch Program, and have a disability).

The estimate of the impact of PtL on principal retention during the third year of the intervention (2020–21) is presented in Table 23. After controlling for other factors included in the statistical model, we estimate that principal retention was 12.7 percentage points higher in intervention schools than in comparison schools, which is equivalent to an effect size of 0.33. Despite the relatively large effect size estimate, we are unable to reject the null hypothesis of no program impact on principal retention with a *p* value less than .05.

Table 23. Program Impact on Principal Retention

Outcome	Comparison schools			Intervention schools			Difference	Effect size	*p* Value
	N	Mean	SD	*N*	Adjusted mean	SD			
Retention	111	56.8%	0.498	37	69.5%	0.475	0.127	0.332	.155

Note. Retention is principal retention in the same school between spring 2018 and spring 2021. *N* is the number of schools. Mean is the percentage of schools in which the principal was retained. SD is the standard deviation of outcomes across schools. Adjusted mean is the regression-adjusted percentage of schools in which the principal was retained. Effect size is calculated using Cox's index. *P* value is the estimated probability that the observed regression-adjusted difference between intervention and comparison schools would be as large (in absolute value) as observed if the intervention had no impact on the outcome. See Appendix C for additional details.

Conclusion

PtL is an innovative, research-informed 3-year principal PD program funded by ED to address specific leadership learning needs and inform the field about what works in principal PD. Funded by a 5-year EIR grant, the PtL program has been designed to be rigorous, replicable, and sustainable. AIR, the independent evaluator of PtL, has completed an implementation and impact study of the program. AIR's evaluation was guided by four RQs:

- **RQ 1:** To what degree was PtL implemented with fidelity across participating PtL schools?
- **RQ 2:** To what extent did school leadership quality and school culture improve in schools that participated in PtL in comparison to similar schools that did not participate in PtL?

- **RQ 3:** To what extent did PtL participation have an impact on student learning in English language arts and math in comparison to similar schools that did not participate in PtL?
- **RQ4:** To what extent did PtL participation have an impact on principal retention in comparison to similar schools that did not participate in PtL?

For the most part, PtL did not meet fidelity of implementation thresholds, although some indicators were implemented with fidelity, and principals were satisfied with the components of PtL. AIR's analyses indicate that PtL did not have a significant positive impact on principal leadership effectiveness, student achievement, or principal retention. However, students in intervention schools scored 0.104 SD higher in math in spring of the final year of the intervention, and the *p* value for the estimated positive impact on student math achievement was .06.

A number of factors could explain the lack of program impact on outcomes of interest. At this point, we cannot determine whether the program would have achieved its desired impact if it had been implemented with higher levels of fidelity. However, it reasonable to assume that the COVID-19 pandemic influenced program impact. In March 2020, the middle year of the PtL program, all Illinois schools were closed because of the COVID-19 pandemic. For all Illinois public schools, in-person teaching and learning abruptly ceased. School administrators shifted gears and priorities to ensure that students and staff had access to food, technology, and resources at home as well as to transition to remote/virtual learning. From March 2020 through the end of the intervention, some PtL principals may have become somewhat disengaged from the full PtL PD and coaching due to competing professional and personal priorities, which may have attenuated the impact of PtL.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* (NBER Working Paper 24003). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w24003/w24003.pdf

Bartenen, B. (2019). *Identifying principal improvement*. (EdWorkingPaper 19-136). Annenberg Institute at Brown University. https://doi.org/10.26300/kcrz-tm19

Clifford, M., Behrstock-Sherratt, E., & Fetters, J. (2012). *The ripple effect: A synthesis of research on principal influence to inform performance evaluation design.* American Institutes for Research. https://eric.ed.gov/?id=ED530748

Cosner, S., & Jones, M. F. (2016). Leading school-wide improvement in low-performing schools facing conditions of accountability: Key actions and considerations. *Journal of Educational Administration, 54*(1), 41–57. https://eric.ed.gov/?id=EJ1087678

Cosner, S., Tozer, S., & Zavitkovsky, P. (2016). Enacting a cycle of inquiry capstone research project in doctoral-level leadership preparation. In V. A. Storey & K. A. Hesbol (Eds.), *Contemporary approaches to dissertation development and research methods* (pp. 162–183). IGI Global.

George W. Bush Institute. (2016). *Principal talent management according to the evidence: A review of the literature.* George W. Bush Institute, Education Reform Initiative. https://eric.ed.gov/?id=ED587185

Glaser, B. (1965). The constant comparative method of qualitative analysis. *Social Problems, 12*(4), 436–445. https://doi.org/10.2307/798843

Grissom, J. A., Anna Egalite, A. J., & Lindsay, C. A. (2021). *How principals affect students and schools: A systematic synthesis of two decades of research.* The Wallace Foundation. https://eric.ed.gov/?id=ED611009

Hart, H., Young, C., Chen, A., Zou, A., & Allensworth, E. M. (2020). *Supporting school improvement: Early findings from reexamination of the 5Essentials Survey*. University of Chicago Consortium on School Research. https://eric.ed.gov/?id=ED608120

Herman, R., Gates, S. M., Arifkhanovea, A., Barrett, M., Bega, A., Chavez-Herrerias, E. R., Han, E., Harris, M., Migacheva, K., Ross, R., Leschitz, J. T., & Wrabel, S. (2017). *School leadership interventions under the Every Student Succeeds Act: Evidence review.* RAND Corporation. https://eric.ed.gov/?id=ED581652

Klugman, J., Gordon, M. F., Sebring, P. B., & Sporte, S. E. (2015). *A first look at the 5Essentials in Illinois schools.* University of Chicago Consortium on Chicago School Research. https://eric.ed.gov/?id=ED577587

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *How leadership influences student learning*. The Wallace Foundation. https://eric.ed.gov/?id=ED485932

Steele, J. L., Steiner, E. D., & Hamilton, L. S. (2021). Priming the leadership pipeline: School performance and climate under an urban school leadership residency program. *Educational Administration Quarterly, 57*(2), 221–256. https://eric.ed.gov/?id=EJ1291815

What Works Clearinghouse. (2022). *What Works Clearinghouse: Procedures and standards handbook, version 5.0.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://eric.ed.gov/?id=ED621928

Appendix A. Fidelity of Implementation Tables

Appendix A provides additional data on the fidelity of implementation of each program component. Each table reports the number of schools that achieved high, moderate, and low fidelity by school year across all indicators and program components. Table 24, Table 25, Table 26, and Table 35 correspond to figures in the Fidelity of PtL Implementation section of this report. All other tables report the fidelity of implementation of each indicator by school level, rural status, and school year.

Table 24. Number of Principals Implementing Cycles of Inquiry (COI) Components With High, Moderate, and Low Implementation Fidelity, by Year

Indicator	Year	High	Moderate	Low	Total
Principals complete six initial training sessions.	2018–19	22	10	5	37
Principals complete six initial training sessions.	2019–20	23	10	4	37
Principals complete six initial training sessions.	2020–21	24	9	4	37
Principals participate in 60 minutes of COI coaching per month.	2018–19	1	3	33	37
Principals participate in 60 minutes of COI coaching per month.	2019–20	15	17	5	37
Principals participate in 60 minutes of COI coaching per month.	2020–21	9	13	15	37

Table 25. Number of Principals Implementing Time Utilization Analysis (TUA) Components With High, Moderate, and Low Implementation Fidelity, by Year

Indicator	Year	High	Moderate	Low	Total
Principals complete 1-hour, one-on-one, in-person TUA training.	2018–19	37	0	0	37
Principals complete 1-hour, one-on-one, in-person TUA training.	2019–20	37	0	0	37
Principals complete 1-hour, one-on-one, in-person TUA training.	2020–21	NA	NA	NA	NA
Principals participate in the TUA process.	2018–19	NA	NA	NA	NA
Principals participate in the TUA process.	2019–20	0	4	33	37
Principals participate in the TUA process.	2020–21	NA	NA	NA	NA

Indicator	Year	High	Moderate	Low	Total
Principals participate in 45 minutes per month of TUA coaching while school is in session.	2018–19	0	0	37	37
Principals participate in 45 minutes per month of TUA coaching while school is in session.	2019–20	0	0	37	37
Principals participate in 45 minutes per month of TUA coaching while school is in session.	2020–21	0	0	37	37

Note. NA is not applicable.

Table 26. Number of Principals Implementing Leadership Framework Components With High, Moderate, and Low Implementation Fidelity, by Year

Indicator	Year	High	Moderate	Low	Total
Principals identify an instructional leadership team (ILT).	2018–19	NA	NA	NA	NA
Principals identify an ILT.	2019–20	36	0	1	37
Principals identify an ILT.	2020–21	31	0	6	37
Principals meet with the ILT for 60 minutes per month while school is in session.	2018–19	NA	NA	NA	37
Principals meet with the ILT for 60 minutes per month while school is in session.	2019–20	13	14	10	37
Principals meet with the ILT for 60 minutes per month while school is in session.	2020–21	15	2	20	37
ILT members meet with teacher teams every other week while school is in session.	2018–19	NA	NA	NA	37
ILT members meet with teacher teams every other week while school is in session.	2019–20	24	9	4	37
ILT members meet with teacher teams every other week while school is in session.	2020–21	24	3	10	37

Note. NA is not applicable.

Table 27. Number of Principals Completing Six Initial Training Sessions With High, Moderate, and Low Implementation Fidelity, by School Level, Rural Status, and Year

School level and rural status	Year	High	Moderate	Low	Total
Elementary	2018–19	9	5	2	16
Elementary	2019–20	9	5	2	16
Elementary	2020–21	10	4	2	16
Middle	2018–19	5	4	1	10
Middle	2019–20	6	4	0	10
Middle	2020–21	6	4	0	10
High	2018–19	8	1	2	11
High	2019–20	8	1	2	11
High	2020–21	8	1	2	11
Rural	2018–19	10	5	3	18
Rural	2019–20	10	5	3	18
Rural	2020–21	10	5	3	18
Nonrural	2018–19	12	5	2	19
Nonrural	2019–20	13	5	1	19
Nonrural	2020–21	14	4	1	19

Table 28. Number of Principals Implementing the Cycles of Inquiry (COI) Indicator "Principals participate in 60 minutes of COI training per month" With High, Moderate, and Low Fidelity, by School Level, Rural Status, and Year

School level and rural status	Year	High	Moderate	Low	Total
Elementary	2018–19	1	0	15	16
Elementary	2019–20	6	8	2	16
Elementary	2020–21	3	4	9	16
Middle	2018–19	0	0	10	10
Middle	2019–20	6	2	2	10
Middle	2020–21	3	7	0	10
High	2018–19	0	3	8	11
High	2019–20	3	7	1	11
High	2020–21	3	2	6	11
Rural	2018–19	0	0	18	18
Rural	2019–20	4	10	4	18
Rural	2020–21	0	8	10	18
Nonrural	2018–19	1	3	15	19
Nonrural	2019–20	11	7	1	19
Nonrural	2020–21	9	5	5	19

Table 29. Number of Principals Who Completed 1-Hour, One-on-One, In-Person Time Utilization Analysis Training With High, Moderate, and Low Fidelity, by School Level, Rural Status, and Year

School level and rural status	Year	High	Moderate	Low	Total
Elementary	2018–19	16	0	0	16
Elementary	2019–20	16	0	0	16
Elementary	2020–21	NA	NA	NA	NA
Middle	2018–19	10	0	0	10
Middle	2019–20	10	0	0	10
Middle	2020–21	NA	NA	NA	NA
High	2018–19	11	0	0	11
High	2019–20	11	0	0	11
High	2020–21	NA	NA	NA	NA
Rural	2018–19	18	0	0	18
Rural	2019–20	18	0	0	18
Rural	2020–21	NA	NA	NA	NA
Nonrural	2018–19	19	0	0	19
Nonrural	2019–20	19	0	0	19
Nonrural	2020–21	NA	NA	NA	NA

Note. NA is not applicable.

Table 30. Number of Principals Who Participated in the Time Utilization Analysis Process With High, Moderate, and Low Fidelity, by School Level, Rural Status, and Year

School level and rural status	Year	High	Moderate	Low	Total
Elementary	2018–19	NA	NA	NA	NA
Elementary	2019–20	0	3	13	16
Elementary	2020–21	NA	NA	NA	NA
Middle	2018–19	NA	NA	NA	NA
Middle	2019–20	0	0	10	10
Middle	2020–21	NA	NA	NA	NA
High	2018–19	NA	NA	NA	NA
High	2019–20	0	1	10	11
High	2020–21	NA	NA	NA	NA
Rural	2018–19	NA	NA	NA	NA
Rural	2019–20	0	1	17	18
Rural	2020–21	NA	NA	NA	NA
Nonrural	2018–19	NA	NA	NA	NA
Nonrural	2019–20	0	3	16	19
Nonrural	2020–21	NA	NA	NA	NA

Note. NA is not applicable.

Table 31. Number of Principals Implementing the Time Utilization Analysis (TUA) Indicator "Principals participate in 45 minutes per month of TUA coaching while school is in session," by School Level, Rural Status, and Year

School level and rural status	Year	High	Moderate	Low	Total
Elementary	2018–19	0	0	16	16
Elementary	2019–20	0	0	16	16
Elementary	2020–21	0	0	16	16
Middle	2018–19	0	0	10	10
Middle	2019–20	0	0	10	10
Middle	2020–21	0	0	10	10
High	2018–19	0	0	11	11
High	2019–20	0	0	11	11
High	2020–21	0	0	11	11
Rural	2018–19	0	0	18	18
Rural	2019–20	0	0	18	18
Rural	2020–21	0	0	18	18
Nonrural	2018–19	0	0	19	19
Nonrural	2019–20	0	0	19	19
Nonrural	2020–21	0	0	19	19

Table 32. Number of Principals Implementing the Leadership Framework Indicator "Principals identify an instructional leadership team" With High, Moderate, and Low Fidelity, by School Level, Rural Status, and Year

School level and rural status	Year	High	Moderate	Low	Total
Elementary	2018–19	NA	NA	NA	NA
Elementary	2019–20	15	0	1	16
Elementary	2020–21	11	0	5	16
Middle	2018–19	NA	NA	NA	NA
Middle	2019–20	10	0	0	10
Middle	2020–21	10	0	0	10
High	2018–19	NA	NA	NA	NA
High	2019–20	11	0	0	11
High	2020–21	10	0	1	11
Rural	2018–19	NA	NA	NA	NA
Rural	2019–20	17	0	1	18
Rural	2020–21	13	0	5	18
Nonrural	2018–19	NA	NA	NA	NA
Nonrural	2019–20	19	0	0	19
Nonrural	2020–21	18	0	1	19

Note. NA is not applicable.

Table 33. Number of Principals Implementing the Leadership Framework Indicator "Principals meet with the instructional leadership team for 60 minutes per month while school is in session" With High, Moderate, and Low Fidelity, by School Level, Rural Status, and Year

School level and rural status	Year	High	Moderate	Low	Total
Elementary	2018–19	NA	NA	NA	NA
Elementary	2019–20	6	3	7	16
Elementary	2020–21	6	0	10	16
Middle	2018–19	NA	NA	NA	NA
Middle	2019–20	2	5	3	10
Middle	2020–21	2	1	7	10
High	2018–19	NA	NA	NA	NA
High	2019–20	5	6	0	11
High	2020–21	7	1	3	11
Rural	2018–19	NA	NA	NA	NA
Rural	2019–20	7	5	6	18
Rural	2020–21	6	2	10	18
Nonrural	2018–19	NA	NA	NA	NA
Nonrural	2019–20	6	9	4	19
Nonrural	2020–21	9	0	10	19

Note. NA is not applicable.

Table 34. Number of Principals Implementing the Leadership Framework Indicator "Instructional leadership team members meet with teacher teams every other week while school is in session" With High, Moderate, and Low Fidelity, by School Level, Rural Status, and Year

School level and rural status	Year	High	Moderate	Low	Total
Elementary	2018–19	NA	NA	NA	NA
Elementary	2019–20	8	5	3	16
Elementary	2020–21	8	2	6	16
Middle	2018–19	NA	NA	NA	NA
Middle	2019–20	10	0	0	10
Middle	2020–21	10	0	0	10
High	2018–19	NA	NA	NA	NA
High	2019–20	6	4	1	11
High	2020–21	6	1	4	11
Rural	2018–19	NA	NA	NA	NA
Rural	2019–20	9	6	3	18
Rural	2020–21	9	0	9	18
Nonrural	2018–19	NA	NA	NA	NA
Nonrural	2019–20	15	3	1	19
Nonrural	2020–21	15	3	1	19

Note. NA is not applicable.

Table 35. Number of Schools Meeting the Minimum Threshold for Adequate Fidelity of Implementation, by Program Component and Year

Program component	Year	Met minimum threshold	Did not meet minimum threshold	Total
Cycles of Inquiry	2018–19	22	15	37
Cycles of Inquiry	2019–20	32	5	37
Cycles of Inquiry	2020–21	28	9	37
Leadership Framework	2018–19	NA	NA	NA
Leadership Framework	2019–20	34	3	37
Leadership Framework	2020–21	26	11	37
Time Utilization Analysis	2018–19	0	37	37
Time Utilization Analysis	2019–20	0	37	37
Time Utilization Analysis	2020–21	37	0	37

Note. NA is not applicable.

Appendix B. Additional Results: Annual Principal Survey

Appendix B provides additional information from the principals who responded to the survey each year and the number and percentage of principals who selected each survey response category. Table 36 reports the percentage of principals who responded to the survey, by school level, locale, and Regional Office of Education. Table 37 through Table 39 report the percentage of principals who selected each level of satisfaction.

Table 36. Distributions of School Principals Responding to the Survey, by School Level, Locale, and Regional Office of Education

	Sample		2018–19		2019–20		2020–21	
	N	%	*N*	%	*N*	%	*N*	%
All schools included in the analysis	37	100%	32	86%	29	78%	30	81%
Locale								
Nonrural	19	51%	17	53%	12	41%	14	47%
Rural	18	49%	15	47%	17	59%	16	53%
School Level								
Elementary	16	43%	15	47%	13	45%	14	47%
Middle	10	27%	9	28%	5	17%	7	23%
High	11	30%	8	25%	11	38%	9	30%
Regional Office of Education								
01 Adams/Pike	2	5%	2	6%	2	7%	2	7%
17 Dewitt/Livingston/McLean	20	54%	18	56%	16	55%	16	53%
19 DuPage	11	30%	10	31%	7	24%	8	27%
28 Bureau/Henry/Stark	4	11%	2	6%	4	14%	4	13%

Table 37. Principal Satisfaction With Cycles of Inquiry (COI) Aspects of PtL

Item	Year	Total schools	Not satisfied		Slightly satisfied		Satisfied		Very satisfied		Missing	
			N	%	*N*	%	*N*	%	*N*	%	*N*	%
Amount of COI training[a]	2018–19	32	0	0%	3	9%	17	53%	11	34%	1	3%
	2019–20	29	0	0%	3	10%	12	41%	14	48%	0	0%
	2020–21	30	0	0%	5	17%	14	47%	11	37%	0	0%
Degree that the program supports my instructional leadership team	2019–20	29	0	0%	0	0%	14	48%	15	52%	0	0%
	2020–21	30	0	0%	1	3%	16	53%	12	40%	1	3%
Degree that the program improves school team's use of COI	2019–20	29	0	0%	1	3%	17	59%	11	38%	0	0%
	2020–21	30	0	0%	2	7%	16	53%	11	37%	1	3%

Note. Data are based on the Annual Principal Survey, 2019, 2020, and 2021.

[a] In 2020–21, the survey text was "Amount of training for COI."

Table 38. Principal Satisfaction With the Coaching Aspect of the Intervention

Item	Year	Total schools	Not satisfied		Slightly satisfied		Satisfied		Very satisfied		Missing	
			N	%	*N*	%	*N*	%	*N*	%	*N*	%
Coaching quality	2018–19	32	0	0%	0	0%	11	34%	20	63%	1	3%
	2019–20	29	0	0%	0	0%	6	21%	23	79%	0	0%
	2020–21	30	0	0%	0	0%	4	13%	26	87%	0	0%
Number of hours of coaching[a]	2018–19	32	0	0%	2	6%	17	53%	12	38%	1	3%
	2019–20	29	0	0%	1	3%	6	21%	22	76%	0	0%
	2020–21	30	0	0%	0	0%	8	27%	22	73%	0	0%
Coach responsiveness	2018–19	32	0	0%	2	6%	8	25%	20	63%	2	6%
	2019–20	29	0	0%	0	0%	4	14%	25	86%	0	0%
	2020–21	30	0	0%	0	0%	3	10%	27	90%	0	0%
Practicality of coaching for use in my school	2018–19	32	0	0%	3	9%	11	34%	17	53%	1	3%
	2019–20	29	0	0%	2	7%	9	31%	18	62%	0	0%
	2020–21	30	0	0%	1	3%	10	33%	19	63%	0	0%
Degree that coaching addresses problems of practice/student learning that I care most about[b]	2018–19	32	0	0%	1	3%	13	41%	16	50%	2	6%
	2019–20	29	0	0%	0	0%	9	31%	20	69%	0	0%
	2020–21	30	0	0%	2	7%	10	33%	17	57%	1	3%

Note. Data are based on the Annual Principal Survey, 2019, 2020, and 2021.

[a] In 2020–21, the survey text was "Total coaching hours."

[b] In 2018–19 and 2019–20, the survey text was "Degree that coaching addresses problems of practice that I care most about," and in 2020–21, the survey text was "Degree that coaching addresses problems of student learning that I care most about."

Table 39. Principal Satisfaction With Other PtL Supports

Item	Year	Total schools	Not satisfied		Slightly satisfied		Satisfied		Very satisfied		Missing	
			N	%	*N*	%	*N*	%	*N*	%	*N*	%
Degree that the program helps me reflect on work habits	2019–20	29	0	0%	1	3%	14	48%	14	48%	0	0%
	2020–21	30	0	0%	4	13%	12	40%	13	43%	1	3%
Information and advice about how to lead schools through the COVID-19 pandemic	2019–20	29	2	7%	8	28%	14	48%	5	17%	0	0%
	2020–21	30	2	7%	4	13%	15	50%	8	27%	1	3%

Note. Data are based on the Annual Principal Survey, 2019, 2020, and 2021.

Appendix C. Baseline Equivalence and Program Impact

Effective Leaders Analysis

Baseline Equivalence

Prior to 2018–19, schools conducted the Illinois 5Essentials Survey only every other year (rather than annually). As a result, the American Institutes for Research® (AIR®) study team does not have 2017–18 Effective Leaders scores for some schools participating in the intervention. In those cases, we used Effective Leaders scores from 2016–17 as the baseline measure. However, because of low teacher response to the Illinois 5Essentials Survey, some intervention schools lack baseline 5Essentials Effective Leaders scores from either year. Schools with missing 5Essentials scores for the baseline or summative outcome years were excluded from the analysis of baseline equivalence and program impact.

Because the percentages of elementary, middle, and high schools with baseline Effective Leaders scores differ between the intervention and comparison groups, and because the percentages of schools in the intervention and comparison groups with baseline scores from 2016–17 (instead of 2017–18) differ, AIR used a regression approach to measure baseline equivalence in Effective Leaders scores between the intervention and comparison groups.

To test baseline equivalence for our analysis of program impact on 5Essentials Effective Leaders scores, the AIR study team estimated a regression model, as shown in the following equation, to examine the difference between treatment and comparison schools at the baseline year:

$$y_s = \beta_0 + \beta_1 Treat_s + \boldsymbol{z_s'}\boldsymbol{\beta_2} + \varepsilon_s$$

The variables in the model are as follows:

- y_s represents the baseline score for school *s*.
- $Treat_s$ is a binary variable indicating whether school *s* is a treatment school.
- $\boldsymbol{z_s'}$ is a vector of indicator variables for the interaction of school level and year of baseline scores as follows:
  - The school is an elementary school with baseline 5Essentials scores from 2016–17.
  - The school is a middle school with baseline 5Essentials scores from 2017–18.
  - The school is a middle school with baseline 5Essentials scores from 2016–17.
  - The school is a high school with baseline 5Essentials scores from 2017–18.
  - The school is a high school with baseline 5Essentials scores from 2016–17.

  - Elementary schools with baseline 5Essentials scores from 2017–18 make up the reference group.
- ε_s is the error term.

$\hat{\beta}_1$is the estimate of baseline difference between the treatment and comparison schools. This estimate was divided by the pooled standard deviation (SD) of baseline 5Essentials scores to obtain the standardized mean difference between the intervention and comparison schools at baseline.

Coefficient and standard error estimates from the statistical models measuring baseline equivalence of 5Essentials Effective Leaders scores between Partners to Lead (PtL) and comparison schools are presented in Table 40.

Table 40. Effective Leaders Baseline Equivalence Model Coefficient and Standard Error Estimates

Model covariate PtL	Coefficient estimate	Standard error	*p* Value
Treatment school	0.839	3.897	.830
Elementary school with baseline 5Essentials scores from 2016–17	3.058	5.315	.566
Middle school with baseline 5Essentials scores from 2017–18	–1.845	5.699	.747
Middle school with baseline 5Essentials scores from 2016–17	2.776	5.629	.623
High school with baseline 5Essentials scores from 2017–18	–6.556	6.142	.288
High school with baseline 5Essentials scores from 2016–17	–7.313	7.020	.300
Intercept	44.690	4.443	.000

Impact Analysis

AIR implemented a difference-in-differences model, as shown in the following equation, to estimate the impact of the intervention on 5Essentials Effective Leaders scores after the first (2018–19) and third (2020–21) years of program implementation:

$$y_{st} = \beta_0 + \beta_1 Treat_s + \beta_2 Post1_t + \beta_3 Treat_s * Post1_t + \beta_4 Post3_t + \beta_5 Treat_s * Post3_t + \boldsymbol{x}_{st}'\boldsymbol{\beta_8} + \gamma_s + \varepsilon_{st}$$

The variables in the model are as follows:

- y_{st} represents the outcome measure for school *s* in year *t*.
- $Treat_s$ is a binary variable indicating whether school *s* is ever a treatment school.
- $Post1$ equals 1 if the year is 2018–19.

- $Post3$ equals 1 if the year is 2020–21.
- $\boldsymbol{x}_{st}^{'}$ is a vector of school-level characteristics as follows:
  - The school is an elementary school with baseline 5Essentials scores from 2017–18.
  - The school is a middle school with baseline 5Essentials scores from 2016–17.
  - The school is a middle school with baseline 5Essentials scores from 2017–18.
  - The school is a high school with baseline 5Essentials scores from 2016–17.
  - The school is a high school with baseline 5Essentials scores from 2017–18.
  - Elementary schools with baseline 5Essentials scores from 2016–17 make up the reference group.
  - Total student enrollment.
  - The percentage of students who are Black.
  - The percentage of students who are White.
  - The percentage of students who are Hispanic.
  - The percentage of students who are English learners (ELs).
  - The percentage of students with a disability.
  - The percentage of students who are eligible for the National School Lunch Program.
- γ_s is a school random effect.
- ε_{st} is the error term.

To account for the correlation of outcomes within schools over time, we cluster the standard errors within schools. Schools with missing 5Essentials scores for the baseline or summative outcome years were excluded from the analysis of baseline equivalence and program impact. We used dummy imputation to account for missing covariates (WWC, 2022). $\hat{\beta}_5$ is the estimate of the impact of the intervention on 5Essentials Effective Leaders scores after the third year (2020–21) of the intervention.

Coefficient and standard error estimates from the statistical models measuring the impact of PtL on 5Essentials Effective Leaders scores are presented in Table 41.

Table 41. Effective Leaders Impact Model Coefficient and Standard Error Estimates

Model covariate PtL	Coefficient estimate	Standard error	*P* Value
Treatment school	1.807	3.451	.601
Year is 2019	–0.979	4.013	.807
Treatment school and year is 2019	0.057	3.232	.986
Year is 2021	1.211	3.886	.755
Treatment school and year is 2021	–0.872	3.191	.785
Elementary school with baseline 5Essentials scores from 2017–18	–8.931	4.007	.023
Middle school with baseline 5Essentials scores from 2016–17	1.189	4.127	.773
Middle school with baseline 5Essentials scores from 2017–18	–2.404	3.996	.547
High school with baseline 5Essentials scores from 2016–17	–7.479	5.556	.178
High school with baseline 5Essentials scores from 2017–18	–2.287	5.265	.664
Enrollment	–0.004	0.003	.151
The percentage of students who are White	–0.102	0.145	.479
The percentage of students who are White is missing	11.160	11.890	.348
The percentage of students who are Black	–0.173	0.177	.328
The percentage of students who are Black is missing	–0.782	2.305	.734
The percentage of students who are Hispanic	–0.065	0.148	.662
The percentage of students who are Hispanic is missing	0.521	2.789	.852
The percentage of students who are English learners	0.084	0.179	.640
The percentage of students who are English learners is missing	2.604	2.568	.311
The percentage of students with disabilities	0.241	0.205	.239
The percentage of students who are eligible for the National School Lunch Program	–0.177	0.076	.020
The percentage of students who are eligible for the National School Lunch Program is missing	–23.180	12.990	.074
Intercept	62.840	13.190	.000

Student Achievement Analysis

Baseline Equivalence

To test baseline equivalence for our analysis of student achievement, the AIR study team estimated the following model based on students in treatment and comparison groups during the 2017–18 (baseline) school year:

$$y_{is} = \beta_0 + \beta_1 Treat_s + \beta_2 middle_s + \beta_3 high_s + \gamma_s + \varepsilon_{is}$$

The model was estimated separately for English language arts (ELA) and math. The variables in the model are as follows:

- y_{is} represents the standardized test score[5] of student *i* in school *s*.
- $Treat_s$ is a binary variable indicating whether the school is a treatment school.
- $middle_s$ and $high_s$ are binary variables indicating whether the school is a middle school or a high school.
- γ_s is a school random effect.
- ε_{is} is the error term.

$\hat{\beta}_1$ is the estimate of the baseline difference between the treatment and comparison schools.

Coefficient and standard error estimates from the statistical models measuring baseline equivalence of student ELA and math scores between PtL and comparison schools are presented in Table 42 and Table 43.

Table 42. Student English Language Arts (ELA) Achievement Baseline Equivalence Model Coefficient and Standard Error Estimates

Model covariate PtL ELA	Coefficient estimate	Standard error	*P* Value
Treatment school	–0.017	0.062	.784
The school is a middle school	–0.065	0.064	.307
The school is a high school	0.019	0.066	.772
Intercept	–0.017	0.043	.693

[5] The standardized test score has a mean of 0 and SD of 1 within each grade, subject, and year.

Table 43. Student Math Achievement Baseline Equivalence Model Coefficient and Standard Error Estimates

Model covariate PtL math	Coefficient estimate	Standard error	*p* Value
Treatment school	–0.011	0.070	.873
The school is a middle school	–0.076	0.071	.286
The school is a high school	0.012	0.075	.870
Intercept	–0.020	0.048	.688

Impact Analysis

To estimate the impact of the intervention on student learning outcomes after the first (2018–19) and third (2020–21) years of program implementation, AIR estimated the following comparative interrupted time series model:

$$y_{ist} = \beta_0 + \beta_1 Treat_s + \beta_2 Time_t + \beta_3 Time_t * Treat_s + \beta_4 Post1_t + \beta_5 Treat_s * Post1_t + \beta_6 Post3_t + \beta_7 Treat_s * Post3_t + \boldsymbol{x1}'_{\boldsymbol{ist}}\boldsymbol{\beta_8} + \boldsymbol{x2}'_{\boldsymbol{st}}\boldsymbol{\beta_9} + \gamma_s + \varepsilon_{ist}$$

Although the model was estimated separately for ELA and math, we present one model here, and we exclude "subject" subscripts. The variables in the model are as follows:

- y_{ist} represents the standardized test score[6] of student *i* in school *s* and year *t*.
- $Treat_s$ is a binary variable indicating whether school *s* is ever a treatment school.
- $Time_t$ is the year.
- $Post1_t$ equals 1 if the year is 2018–19.
- $Post3_t$ equals 1 if the year is 2020–21.
- $\boldsymbol{x1}'_{\boldsymbol{ist}}$ is a vector of student characteristics as follows:
  - The student is Black.
  - The student is Hispanic.
  - The student is Asian.
  - The student is Native Hawaiian or other Pacific Islander, Native American, or two or more races.
  - The student has a disability.
  - The student is an EL.
  - The student is eligible for the National School Lunch Program.

[6] The standardized test score has a mean of 0 and SD of 1 within each grade, subject, and year.

  - The student's grade level.
- $\boldsymbol{x2}'_{\boldsymbol{st}}$ is a vector of school-level characteristics as follows:
  - The school is a middle school.
  - The school is a high school.
  - Total student enrollment.
  - The percentage of students who are Black.
  - The percentage of students who are Hispanic.
  - The percentage of students who are Asian.
  - The percentage of students who are Native Hawaiian or other Pacific Islander, Native American, or two or more races.
  - The percentage of students who have a disability.
  - The percentage of students who are ELs.
  - The percentage of students who are eligible for the National School Lunch Program.
  - The school is a rural school.
- γ_s is a school random effect to account for the clustering of students within schools (Abadie et al., 2017).
- ε_{ist} is the error term.

$\hat{\beta}_7$ is an estimate of the impact of the intervention on student achievement after the third year (2020–21) of the intervention.

We limited the analytic sample to high schools with student outcome data each year from 2016–17 to 2020–21 and elementary and middle schools with outcomes each year from 2014–15 to 2020–21. Because the Illinois State Board of Education (ISBE) did not require all Grade 11 students to take the SAT® prior to 2016–17, high school outcomes for 2014–15 and 2015–16 were excluded from the analysis. Students with missing outcome data were excluded from the analysis. All students with available outcome data also had nonmissing data for all model covariates.

Coefficient and standard error estimates from the statistical models measuring the impact of PtL on student ELA and math achievement are presented in Table 44 and Table 45, respectively.

Table 44. Student English Language Arts (ELA) Achievement Impact Model Coefficient and Standard Error Estimates

Model covariate PtL ELA	Coefficient estimate	Standard error	*p* Value
Treatment school	14.520	25.500	.569
Year	0.013	0.006	.052
Treatment × year	–0.007	0.013	.568
Year is 2019	0.000	0.020	.985
Year is 2021	–0.026	0.031	.393
Treatment school and year is 2019	0.043	0.041	.289
Treatment school and year is 2021	0.065	0.061	.288
The student is Black	–0.421	0.007	.000
The student is Hispanic	–0.155	0.005	.000
The student is Asian	0.333	0.009	.000
The student is Native Hawaiian or other Pacific Islander, Native American, or two or more races	–0.071	0.009	.000
The student is an English learner	–0.678	0.008	.000
This student is eligible for the National School Lunch Program	–0.344	0.004	.000
The student has a disability	–0.998	0.005	.000
The student is in Grade 4	–0.042	0.008	.000
The student is in Grade 5	–0.063	0.008	.000
The student is in Grade 6	–0.085	0.012	.000
The student is in Grade 7	–0.100	0.013	.000
The student is in Grade 8	–0.113	0.013	.000
The student is in Grade 11	–0.088	0.061	.148
The school is a middle school	–0.169	0.070	.016
The school is a high school	–0.133	0.047	.005
Number of students tested in the school	0.000	0.000	.000
The percentage of students who are Black	–0.245	0.160	.125
The percentage of students who are Hispanic	–0.008	0.108	.938
The percentage of students who are Asian	1.055	0.280	.000
The percentage of students who are Native Hawaiian or other Pacific Islander, Native American, or two or more races	–0.203	0.322	.528
The percentage of students who are English learners	–0.110	0.185	.553

Model covariate PtL ELA	Coefficient estimate	Standard error	*p* Value
The percentage of students who are eligible for the National School Lunch Program	–0.148	0.066	.026
The percentage of students with disabilities	–0.066	0.180	.714
Intercept	–25.040	13.100	.056

Table 45. Student Math Achievement Impact Model Coefficient and Standard Error Estimates

Model covariate PtL Math	Coefficient estimate	Standard error	*p* Value
Treatment school	23.290	23.170	.315
Year	0.011	0.006	.074
Treatment × year	–0.012	0.012	.314
Year is 2019	–0.015	0.018	.406
Year is 2021	–0.032	0.028	.254
Treatment school and year is 2019	0.049	0.037	.184
Treatment school and year is 2021	0.104	0.056	.063
The student is Black	–0.490	0.007	.000
The student is Hispanic	–0.209	0.005	.000
The student is Asian	0.415	0.009	.000
The student is Native Hawaiian or other Pacific Islander, Native American, or two or more races	–0.123	0.009	.000
The student is an English learner	–0.520	0.008	.000
This student is eligible for the National School Lunch Program	–0.358	0.004	.000
The student has a disability	–0.904	0.005	.000
The student is in Grade 4	–0.030	0.008	.000
The student is in Grade 5	–0.049	0.008	.000
The student is in Grade 6	–0.051	0.012	.000
The student is in Grade 7	–0.059	0.013	.000
The student is in Grade 8	–0.069	0.013	.000
The student is in Grade 11	–0.183	0.062	.003
The school is a middle school	–0.065	0.074	.378
The school is a high school	–0.154	0.053	.004
Number of students tested in the school	0.000	0.000	.000

Model covariate PtL Math	Coefficient estimate	Standard error	*p* Value
The percentage of students who are Black	–0.281	0.173	.104
The percentage of students who are Hispanic	0.036	0.113	.750
The percentage of students who are Asian	1.003	0.289	.001
The percentage of students who are Native Hawaiian or other Pacific Islander, Native American, or two or more races	–0.407	0.307	.186
The percentage of students who are English learners	–0.171	0.178	.337
The percentage of students who are eligible for the National School Lunch Program	–0.173	0.066	.008
The percentage of students with disabilities	0.174	0.169	.303
Intercept	–20.850	11.920	.080

To prepare the data for student achievement impact analysis, the AIR study team filtered the data as follows:

- Retained records for students in Grades 3 through 8 and 11 only
- Retained Partnership for Assessment of Readiness for College and Careers student achievement records with the following test codes: ELA03–ELA08 and MAT03–MAT08
- Retained SAT® scores for students in Grade 11 only
- Dropped student achievement records with invalid test scores
- Dropped records with invalid reasons for not testing (not tested reason code desc ! = "NULL")
- Dropped student achievement records in which the student's enrolled grade did not match the student's tested grade
- If a student had more than one record, dropped the duplicate record for "on or before May 1 home school" = 0 and kept the duplicate record for "on or before May 1 home school" = 1
- If students still had records for more than one school, including schools that are not in the study, dropped all remaining records
- Dropped all records for a given test for students who have more than one record for the same test

Principal Retention Analysis

Impact Analysis

AIR used a matched-comparison group design to estimate the impact of PtL on principal retention from spring 2018, prior to the start of the intervention, through spring 2021. A principal was "retained" if they remained at the same school from spring 2018 to spring 2021, and was not retained otherwise. Because principal retention is a binary outcome (e.g., 1 if the school's principal in spring 2018 continued to serve as the principal through spring 2021, and 0 otherwise), we used a probit regression model for the analysis. The probit model used to estimate program impact on retention can be represented as follows:

$$P(y_s = 1) = P(\beta_0 + \beta_1 Treat_s + \boldsymbol{x_s'\beta_2} + \varepsilon_s > 0) = P(\beta_0 + \beta_1 Treat_s + \boldsymbol{x_s'\beta_2} > -\varepsilon_s)$$
$$= G(\beta_0 + \beta_1 Treat_s + \boldsymbol{x_s'\beta_2})$$

The functions and variables in this model are as follows:

- $P(y_s = 1)$ is the probability that the school's principal in spring 2018 served as principal of the school through spring 2021.
- $Treat_s$ is a binary variable indicating that school *s* is in the intervention group.
- $\boldsymbol{x_s'}$ is a vector of variables describing the school and its principal in spring 2018, prior to the start of the intervention, as follows:
  - Characteristics of the school's principal in 2018:
    - The principal has 2 or fewer completed years of experience as a principal.
    - The principal is not White.
    - The principal has a doctoral degree.
    - The principal has a certificate of advanced study or specialist degree.
    - (Principals with master's degrees are the omitted category.)
  - Characteristics of the school in 2018:
    - The school's principal in spring 2018 also was the school's principal in spring 2017, as a measure of baseline retention.
    - The school is a middle school.
    - The school is a high school.
    - The number of students enrolled in the school.
    - The percentage of students who are non-Hispanic White.
    - The percentage of students who have disabilities.

    - The percentage of students who are ELs.
    - The percentage of students who are eligible for the National School Lunch Program.
    - The percentage of students who are proficient in ELA.
    - The percentage of students who are proficient in math.
- ε_s is the error term.
- $G(\cdot)$ is the cumulative distribution function for $-\varepsilon$, which is normally distributed.

Analyses were based on personnel data provided by ISBE, which did not include any missing data elements. Coefficient and standard error estimates from the statistical model estimating the impact of PtL on principal retention are presented in Table 46.

Table 46. Principal Retention Impact Model Coefficient and Standard Error Estimates

Model covariate	Coefficient estimate	Standard error	*P* Value
Treatment school	0.398	0.280	.155
The school's principal in spring 2018 also was the school's principal in spring 2017	–0.261	0.374	.485
The spring 2018 principal is a novice principal	0.247	0.283	.383
The spring 2018 principal is not White	–0.482	0.656	.463
The spring 2018 principal has a doctoral degree	–1.125	0.374	.003
The spring 2018 principal has a certificate of advanced study or specialist degree	–0.910	0.305	.003
The school is a middle school	–0.409	0.352	.245
The school is a high school	–0.061	0.349	.861
Number of students enrolled in the school in spring 2018	0.000	0.000	.998
The percentage of students who are White in spring 2018	0.708	0.960	.461
The percentage of students with disabilities in spring 2018	–5.464	3.321	.100
The percentage of students who are English learners in spring 2018	1.770	2.201	.421
The percentage of students who are eligible for the National School Lunch Program in spring 2018	–1.453	1.372	.290
Schoolwide English language arts proficiency rate in spring 2018	0.955	1.822	.600
Schoolwide math proficiency rate in spring 2018	–1.031	1.649	.532
Intercept	1.522	1.675	.363

About the American Institutes for Research®

Established in 1946, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of education, health, and the workforce. AIR's work is driven by its mission to generate and use rigorous evidence that contributes to a better, more equitable world. With headquarters in Arlington, Virginia, AIR has offices across the U.S. and abroad. For more information, visit AIR.ORG.

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

21048_04/23