U.S. Department of Education
October 2019

The Effects of a Principal Professional Development Program Focused on Instructional Leadership: Appendices

Mariesa Herrmann
Melissa Clark
Susanne James-Burdumy
Christina Tuttle
Tim Kautz
Virginia Knechtel
Dallas Dotter
Claire Smither Wulsin
John Deke
Mathematica

ies INSTITUTE OF EDUCATION SCIENCES

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE

CONTENTS

APPENDIX A: THE STUDY'S PRINCIPAL PROFESSIONAL DEVELOPMENT PROGRAM 1

A. Focus and structure of the professional development program 2

B. Characteristics of professional development program staff 7

C. Implementation support for the study's principal professional development program 8

D. Costs of the study's principal professional development program 8

APPENDIX B: STUDY DESIGN, DATA COLLECTION, AND ANALYTIC METHODS 9

A. Study design 10

1. Sample selection 10

2. Random assignment 12

B. Data collection 14

C. Analytic methods 16

1. Constructing outcome measures 16

2. Estimating effects 22

3. Estimating the relationship between characteristics of the professional development program, its effects on principals' practices, and its effects on student achievement 25

APPENDIX C: SUPPLEMENTAL TABLES 26

A. Additional details on findings in the report 27

1. Effects on student achievement and school and teacher outcomes 27

2. Effects on principals' practices 31

3. Implementation of the study's principal professional development program and the overall professional development that principals received from the study and other sources 39

4. Effects on student achievement by district and blocks of schools 42

5. Relationship between characteristics of the professional development program and its effects on principal practices and student achievement 47

B. Additional findings that are not in the report 52

1. Effects on principals' practices 52

2. Implementation of the study's principal professional development program and the overall professional development that principals received 55

C. Supplemental information for systematic reviews 60

D. Realized minimum detectable effects 66

APPENDIX D: INTERPRETING STUDY FINDINGS: ALTERNATIVES TO *P*-VALUES AND STATISTICAL SIGNIFICANCE 68
A. BASIE (BAyeSian Interpretation of Estimates) 69
B. Interpretation of study findings using the BASIE approach 70
1. Selecting and synthesizing prior evidence 70
2. Findings 71
3. Sensitivity analysis 72
ENDNOTES 76
REFERENCES 78
ACKNOWLEDGEMENTS 80
DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST 81

TABLES

A.1. Components of the professional development program 2

A.2. Percentage of time spent on three school leadership areas, by program component and overall 3

A.3. Planned content of summer institute, group training sessions, and professional learning community sessions 3

A.4. Characteristics of staff delivering the program 8

B.1. Comparison of study districts and public school districts nationally 11

B.2. Comparison of study schools and public elementary schools nationally 12

B.3. Comparison of baseline characteristics of treatment and control students and schools 13

B.4. Comparison of baseline characteristics of principals and teachers in treatment and control schools 13

B.5. Data sources 14

B.6. Response rates for data sources used to estimate effects 15

B.7. Overall model fit 17

B.8. Items in final scales and estimated reliabilities 17

B.9. Covariates included in the models used to estimate effects 23

B.10. Analysis samples 23

B.11. Subgroups examined 24

B.12. Treatment group characteristics examined 25

C.1. Effects on student achievement (supplement to report Figure 2) 28

C.2. Effects on student achievement, by principal experience level (supplement to report text) 28

C.3. Effects on student achievement, by average baseline achievement of study schools (supplement to report text) 29

C.4. Effects on school climate (supplement to report Figure 3) 30

C.5. Effects on principal and teacher retention (supplement to report Figure 4) 31

C.6. Effects on principals' time use (supplement to report Figure 5 and report text) 32

C.7. Effects on duration of teacher observations (supplement to report text) 33

C.8. Effects on principals' instructional leadership practices (supplement to report Figure 6, 1, and report text) 34

C.9. Effects on teacher-reported principals' instructional leadership practices, by principal experience level and school average baseline achievement (supplement to report text) 36

C.10. Effects on principals' human capital management and organizational leadership practices (supplement to report text) 37

C.11. Amount of professional development and other supports that principals received (supplement to report Figure 9) 40

C.12. Content of supports that principals received (supplement to report text) 41

C.13. Correlations between the professional development program's effects on principals' instructional leadership practices and its effects on student achievement (supplement to report text) 48

C.14. Correlations between the professional development program's effects on principals' human capital management practices and its effects on student achievement (supplement to report text) 49

C.15. Correlations between the professional development program's effects on principals' organizational leadership practices and its effects on student achievement (supplement to report text) 49

C.16. Correlations between characteristics of the professional development program and its effects on principals' leadership practices (supplement to report text) 50

C.17. Effects on principals' leadership practices, by characteristics of the professional development program 51

C.18. Effects on principals' time use (lower and upper bounds) 53

C.19. Effects on teacher-reported principals' human capital management and organizational leadership practices, by principal experience level and school average baseline achievement 54

C.20. Principals' perceptions of the summer institute, group training sessions, and professional learning community sessions, Year 1 56

C.21. Principals' perceptions of coaching 57

C.22. Usefulness of professional development and other supports that principals received 58

C.23. Principals' perceptions of supports received 59

C.24. Attrition of study schools, by outcome 61

C.25. Missing outcome data, by outcome and year 62

C.26. Effects of professional development program, by outcome and year 63

C.27. Missing student achievement outcome data, alternate student sample 64

C.28. Effects of professional development program on student achievement, alternate student sample 65

C.29. Realized minimum detectable effects for key outcomes 66

D.1. Prior evidence from the What Works Clearinghouse 71

D.2. Effects on student achievement and the probability that the professional development program had effects of a certain size 72

D.3. Effects on frequency of instructional support and feedback from principal and the probability that the professional development program had effects of a certain size 72

D.4. Sensitivity to the prior evidence for effects on student achievement 73

D.5. Sensitivity to the prior evidence for the effects on frequency of instructional support and feedback from principal 74

D.6. Difference between traditional effect estimates and Bayesian shrunken estimates 75

FIGURES

A.1. Percent of coaching time spent on meetings and practical applications of coaching 6
A.2. The four phases of the inquiry cycle 7
B.1. Results from district recruitment effort 10
C.1. Effects on students' English language arts achievement in Year 1, by district and random assignment block 42
C.2. Effects on students' math achievement in Year 1, by district and random assignment block 43
C.3. Effects on students' English language arts achievement in Year 2, by district and random assignment block 44
C.4. Effects on students' math achievement in Year 2, by district and random assignment block 45
C.5. Effects on students' English language arts achievement in Year 3, by district and random assignment block 46
C.6. Effects on students' math achievement in Year 3, by district and random assignment block 47

These appendices provide supplementary materials for NCEE report 2020-0002, “The Effects of a Principal Professional Development Program Focused on Instructional Leadership.”

APPENDIX A

THE STUDY'S PRINCIPAL PROFESSIONAL DEVELOPMENT PROGRAM

The Center for Educational Leadership (CEL) at the University of Washington provided the study's principal professional development program. CEL's approach is based on the theory that, if principals have a clear understanding of what high-quality instruction looks like and know how they can support it in their schools' classrooms, instruction will become more effective and student achievement will improve. Two complementary instructional frameworks guide CEL's approach:

1. CEL's 5 Dimensions of Teaching and Learning (5D)TM framework defines high-quality instruction on multiple dimensions within five areas: "purpose; student engagement; curriculum and pedagogy; assessment for student learning; and classroom environment and culture."
2. CEL's 4 Dimensions of Instructional Leadership (4D)TM framework describes ways leaders can work with teachers to improve instruction. The four dimensions include: "vision, mission, and learning-focused culture; improvement of instructional practice; allocation of resources; and management of systems and processes."

The 5D instructional framework describes how teachers can provide high-quality instruction and includes targeted questions to guide principals' observations of and feedback to teachers. CEL used this framework to help principals understand what they should be looking for when observing a classroom. (When study districts had their own instructional framework for evaluating teachers, CEL aligned the 5D with the districts' instructional framework.) Similarly, the 4D instructional leadership framework describes what principals should do to support high-quality instruction. CEL used this framework's guiding questions to identify specific areas in which individual principals might benefit from additional support.

Exhibit 1. Focus of the study's principal professional development program

Instructional leadership (primary focus)

- Conducting classroom observations
- Providing feedback to teachers

Human capital management

- Analyzing data to tailor professional development to teachers' needs

Organizational leadership

- Developing and communicating a plan to improve school culture

A. Focus and structure of the professional development program

The professional development program included four components: (1) a summer institute, (2) group trainings, (3) professional learning communities (PLCs), and (4) individualized coaching. Across all four components, CEL planned to deliver 188 hours of professional development over two years (Table A.1).

Table A.1. Components of the professional development program

		Number of hours planned		
Component	Format	Year 1	Year 2	Total
Summer institute	In-person meetings	28	[a]	28
Group trainings during the school year	In-person meetings	54	—	54
Professional learning community sessions	Virtual meetings	6	[a]	6
Individualized coaching	In-person and virtual meetings	50	50	100
Total		**138**	**50**	**188**

Source: Professional development program documentation.

[a] Principals new to the study in Year 2 were offered a two-day summer institute and four professional learning community sessions. Principals who participated in Year 1 did not participate in these supplemental components in Year 2.

— Not offered.

Although the program covered three main areas of principals' leadership practices (Exhibit 1), it primarily emphasized instructional leadership. In total, across the four study components, 70 percent of the study's professional development time focused on instructional leadership (Table A.2). This ranged from a low of 54 percent of the time spent in

professional learning community sessions to a high of 77 percent of the time spent in group training sessions. Human capital management and organizational leadership received much less attention (16 and 14 percent overall, respectively).

Table A.2. Percentage of time spent on three school leadership areas, by program component and overall

Leadership area	Summer institute	Group training sessions	Professional learning community sessions	Coaching		Overall
				Year 1	Year 2	
Instructional leadership	71.3	77.0	53.7	61.0	69.5	70.4
Human capital management	20.0	12.8	22.4	13.5	16.7	15.6
Organizational leadership	8.7	10.2	23.9	25.5	13.8	14.0

Sources: Observation forms and coaching logs completed for each coaching session during Years 1 and 2. Observation forms collected for 8 summer institute sessions, 56 of 64 formal group training sessions, and 32 professional learning community sessions during Year 1. Coaching logs completed for 50 principals during Year 1 and 49 principals during Year 2.

Note: Analysis gives each principal equal weight. The overall average weights each program component by total hours.

Table A.3 summarizes the primary topics and key activities of three of the program components; the topics and activities in the last component (coaching) varied according to the needs of each principal.

Table A.3. Planned content of summer institute, group training sessions, and professional learning community sessions

Session	Principal leadership area(s)	Primary topics	Key activities
Summer institute (in person)[a]			
Day 1	Instructional leadership	• Introduction to instructional leadership (four dimensions of leadership)	• Discuss in small groups • Review case studies • Conduct self-assessment
Day 2	Instructional leadership Organizational leadership	• Establishing a learning environment focused on student outcomes • Introduction to the inquiry cycle	• Review case studies • Document notes and observations in a journal
Day 3	Instructional leadership	• Understanding an instructional framework (five dimensions of teaching and learning) • Collecting classroom observation data	• Practice observations using video
Day 4	Human capital management	• Understanding and planning for professional development needs within the school	• Conduct self-assessment • Document notes and observations in a journal
Group training sessions (in person)[b]			
1	Instructional leadership Organizational leadership	• Introducing three foundational skills of effective instructional leaders: • Developing a theory for how a change in their practices might affect teachers and students • Finding time for instructional leadership activities • Conducting classroom observations	• Review case study • Review school data • Draft theory showing how a change in principals' practices might affect teachers and students • Draft annual, monthly, and weekly calendars • Prepare for classroom observations

Session	Principal leadership area(s)	Primary topics	Key activities
2	Instructional leadership	• Conducting classroom observations	• Document information from observations • Share findings
3	Instructional leadership	• Conducting classroom observations	• Document information from observations • Share findings
4+5	Instructional leadership	• Providing targeted feedback to teachers	• View video examples • Plan conversations
6	Instructional leadership	• Conducting classroom observations focused on specific issues	• Document information from observations • Share findings
7	Instructional leadership	• Conducting classroom observations focused on specific issues	• Document information from observations • Share findings
8	Human capital management	• Using data to determine professional development needs • Planning for professional development needs within the school	• Review research on teacher learning • Review school data • Create a professional development plan
Professional learning community sessions (virtual)[c]			
1	Organizational leadership	• Building a strong culture in your school • Conducting professional learning communities in the school	• Review case study • Discuss with group
2	Instructional leadership	• Observing and analyzing instruction	• Document information from observations • Discuss with group
3	Instructional leadership	• Providing feedback to teachers	• Plan feedback using example formats • Discuss with group
4	Human capital management	• Providing professional development and support to teachers and staff	• Discuss with group

Source: Authors' compilation based on Center for Educational Leadership curriculum materials and training agendas from Years 1 and 2.

[a]In Year 2, only principals new to the study participated in the summer institute, which lasted two days. The Year 2 summer institute covered the same topics as the Year 1 institute but with a condensed presentation.

[b]New principals in Year 2 did not receive any formal group training.

[c]New principals in Year 2 were offered four professional learning community sessions in that year.

Each of the four study components is described in more detail below.

Summer institute. CEL held the summer institute in person over four days before the first study school year. (In the summer before the second study school year, CEL hosted a condensed, two-day summer institute for the seven principals who were new to the study schools that year.) The summer institute allowed CEL to efficiently present the introductory content to the 50 treatment group principals from multiple districts at one time, while simultaneously building relationships between the coaches and principals who would work together throughout the study. CEL instructors used group work, case studies, reviews of research, and video to introduce principals to the 4D and 5D instructional frameworks and the key components of the professional development. The summer institute also showed principals how to document their classroom observations to help them analyze the quality of instruction in their schools. Specifically, principals were asked to use an instructional framework (either the 5D or their district's framework) to guide their

observations; document them using nonjudgmental, fact-based, comprehensive descriptions of what teachers and students did in the classroom; and use the observations and associated documentation to identify areas in which each teacher could improve instruction.

Group trainings. The group trainings included eight day-long, in-person training sessions in each district, held throughout the first study school year. No group trainings were offered in the second year of the study. The CEL coach assigned to each district (and sometimes, another CEL staff person who specialized in the training content) led the group trainings.

Six of the eight group trainings provided principals with hands-on experience observing teachers and giving them feedback.

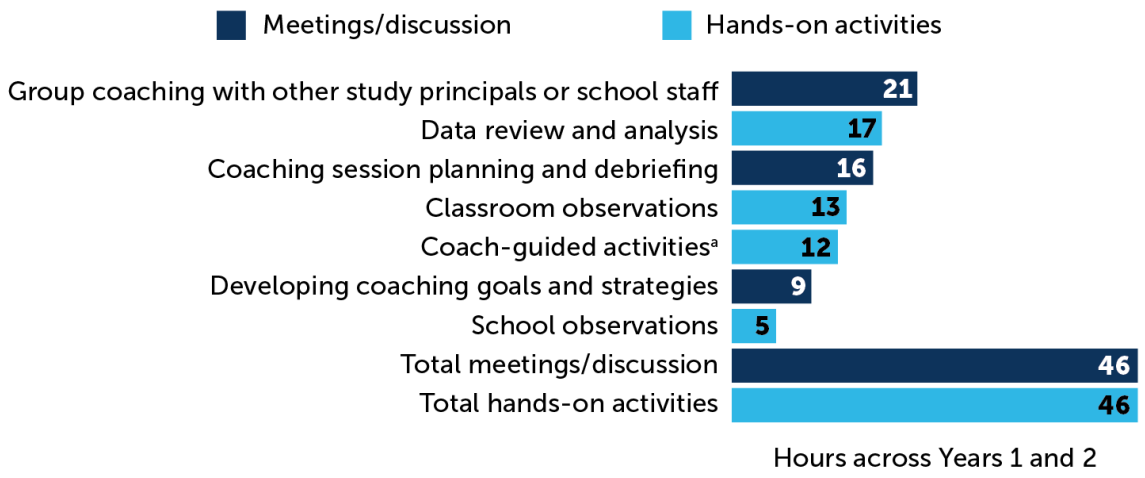
- In four of the sessions, all principals participating in the program in each district (between 5 and 10 principals) conducted observations together and then had the opportunity to put the strategies discussed above into practice. To start the day, the principals and coach would meet to discuss the focus of the observations and the coach would review the guidelines for the observations (for example, that the group would enter a classroom together, watch for five minutes, walk around the classroom, and take separate notes for each class). The coach and principals would then conduct four or five 15- to 20-minute observations of different classrooms. They would then debrief for about an hour. During this time, they would individually code their notes from the observations and share what they observed and recorded with the group, with the coach pressing for specific evidence. Finally, the group would identify strengths and weaknesses across the classrooms they observed, develop theories to explain those patterns, and develop strategies to address the weaknesses.
- Another two sessions included a "feedback institute" that focused on providing support and feedback to teachers. During these sessions, CEL used videos and roleplay to help principals practice planning and conducting conversations with teachers. Rather than providing direct guidance on how teachers should change their instruction, principals were encouraged to ask teachers questions to help them reflect on their instructional approach and how they might change it to improve student performance.

PLCs. The PLCs included four 90-minute virtual sessions held throughout the first study school year. The sessions, which were led by CEL staff, enabled principals to discuss specific issues with other coaches and principals outside their district and obtain new perspectives on specific challenges. Each session focused on one of the same four topics covered in the group trainings—conducting classroom observations, providing feedback to teachers, planning professional development and support for teachers, and building a strong school culture. In the second study school year, CEL facilitated four PLCs for principals who were new to the study.

Individualized coaching. CEL provided individualized coaching to principals across both years of the study. One coach was assigned to work with all of the study principals in a given district. That coach led 8 to 10 one-on-one (or small-group) half-day, in-person sessions, supplemented with periodic check-ins via phone or email. CEL described its coaching strategy as using a "strengths-based approach," focusing on what principals already do well and encouraging them to build on what is working, rather than emphasizing areas needing improvement. Because most of the individualized coaching was in-person, it provided time for face-to-face discussion to build trust between coaches and principals, as well as opportunities for coaches to demonstrate or model skills and provide feedback on principals' practices in their schools and classrooms.

Coaching was evenly split across meetings and practical applications. About half the time principals spent with their coaches (46 hours) included meetings and discussion; the other half (46 hours) involved hands-on activities, such as conducting observations, reviewing data, and modeling or role-playing with the coaches (Figure A.1).

Figure A.1. Percent of coaching time spent on meetings and practical applications of coaching



Source: Coaching logs (50 principals in Year 1 and 49 principals in Year 2).

Note: The light blue bars for hands-on activities sum to 47 due to rounding.

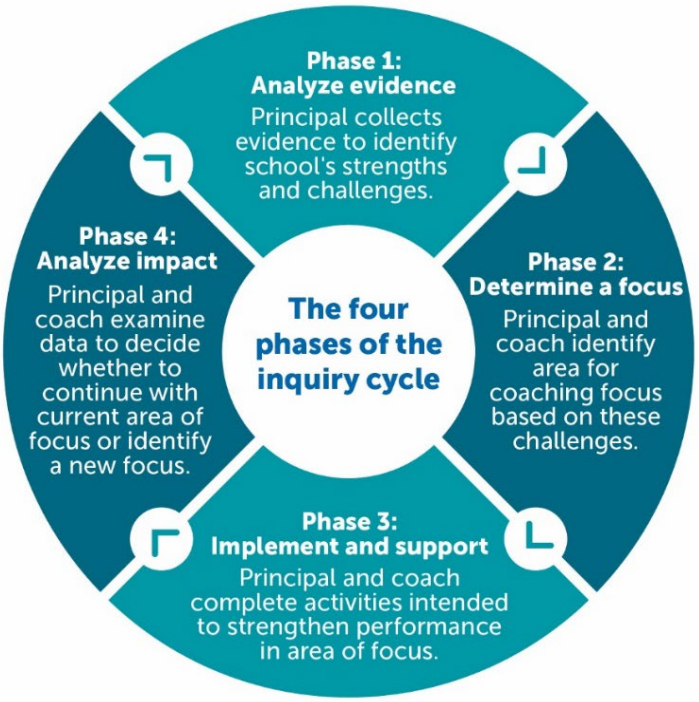
Figure reads: Across both study years, principals spent 21 hours, on average, participating in group coaching with other study principals or school staff.

[a] Coach-guided activities include modeling, role-playing with coaches, practicing challenging conversations, and using tools (such as documents to help principals gather data and assess teacher skills).

When working with principals, CEL coaches used an inquiry cycle—a sequence of four steps or phases to identify and address challenges related to student learning, teaching, and leadership in each school (Figure A.2). In the first phase, principals collected data on their schools—for example, from classroom observations or student test scores—that they analyzed with the assistance of their coach. In the second phase, the coach and principal used the analysis from the first phase to determine an "area of focus" or specific problem to address during the coaching. In the third phase, the principal and coach worked together to develop and implement a plan with specific strategies to address the problem they identified in the prior phase. Finally, in the fourth phase, the coach worked with the principal to analyze the impact of the plan and adjust the plan accordingly. This could have entailed minor adjustments to the area of focus for the principal's next cycle or changing it entirely.

For example, if a principal and coach reviewed formative assessment data and determined that 3rd grade English language arts scores in a school were low (Phase 1—analyze evidence), they would discuss different strategies that the principal might take—with guidance from their coach—to better support reading instruction in the school (Phase 2—determine a focus). Based on these discussions, the principal and coach might agree on a plan to begin more frequent observations of 3rd grade classrooms during reading instruction and provide more frequent feedback to 3rd grade teachers about their English language arts instruction (Phase 3—implement and support). If a common set of instructional issues across teachers was identified through those observations and the coach and principal determined that the issues need additional attention (Phase 4—analyze impact), the next cycle might include a plan for the principal to work with the 3rd grade team during common planning time to address the issues or to arrange professional development for teachers on the specific issue identified.

Figure A.2. The four phases of the inquiry cycle



Source: Center for Educational Leadership coaching materials.

B. Characteristics of professional development program staff

To select staff (coaches, summer institute instructors, and group training instructors) to deliver the professional development program, CEL drew primarily on its extensive network of independent consultants who had already been trained in CEL processes and protocols. CEL specified eight qualifications for study staff:

1. Master's degree in teaching and learning, educational leadership, or equivalent education and experience
2. Extensive knowledge of and experience in K–12 education, including leadership responsibilities—preferably as a district-level leader
3. A history of successfully teaching adults and understanding principles of adult learning
4. Extensive knowledge of high-quality instruction, including how to identify it and how to support teachers' instructional practice
5. Extensive knowledge of and experience in developing and supporting principals as instructional leaders
6. Current knowledge of teachers' and principals' evaluation models and processes
7. Expertise in presenting to small and large groups, leading trainings, and adjusting the content being presented based on groups' needs
8. Experience working for CEL and using CEL materials
9. Consistent with CEL's expectations, staff who delivered the program had substantial relevant experience (Table A.4).

Table A.4. Characteristics of staff delivering the program

Characteristic of staff[a]	Mean	Minimum	Maximum
Years of teaching experience	8.7	0	20
Years of experience as a principal	5.6	0	17
Years of experience in a district-level leadership position[b]	6.0	0	24
Years of experience with Center for Educational Leadership	2.7	0	15
Percentage holding an advanced degree in teaching and learning or educational leadership	87.0	n.a.	n.a.
Number of Center for Educational Leadership staff	15	15	15

Source: Center for Educational Leadership staff résumés.

[a] Staff include district coaches, summer institute facilitators, content experts providing group training, and professional learning community facilitators.

[b] Includes experience as assistant or associate superintendent, superintendent, vice president of educational services, director of teaching and learning, director of assessment, executive director, director of personnel, or director of schools.

n.a. = not applicable.

C. Implementation support for the study's principal professional development program

To increase the chances that the professional development program would run smoothly in the participating districts, the study's technical assistance team communicated frequently with district and CEL staff to carefully monitor all program activities. The team:

1. Reviewed the credentials of CEL staff to confirm they were consistent with the qualifications established for the study
2. Reviewed materials used in the summer institute, group trainings, and PLC sessions for completeness and quality of presentation
3. Monitored the summer institute, group trainings, PLC sessions, and coaching to ensure the activities were delivered as intended
4. Met regularly with CEL and district staff to review data on principals' participation (see Appendix C), the content of the program, and plans for upcoming activities, and to gather feedback from districts

D. Costs of the study's principal professional development program

The cost of the professional development program as implemented for the study was approximately $65,000 per principal over two years. These per-principal costs are atypically high, due to circumstances related to the evaluation. For example, the four-day summer institute was held in a central location, which required facility and travel costs for all summer institute instructors and participating principals for the length of the institute. During the two school years of program implementation, CEL coaches and group training instructors also traveled to each participating district six or more times to administer the in-person professional development components. Because only five principals participated in the program in most districts, those fixed costs (per principal) are higher than they would have been if they had been spread across more principals. In addition, CEL coaches participated in activities required by the study, including completing a detailed coaching log after each session with participating principals and attending regular meetings with coaches across all of the study districts.

APPENDIX B

STUDY DESIGN, DATA COLLECTION, AND ANALYTIC METHODS

This appendix discusses the design, data collection, and analytic methods for the study.

A. Study design

The study team recruited districts and schools to participate in the study and randomly assigned schools to receive the study's principal professional development program or not. This section describes how the study team selected the districts and schools for the study.

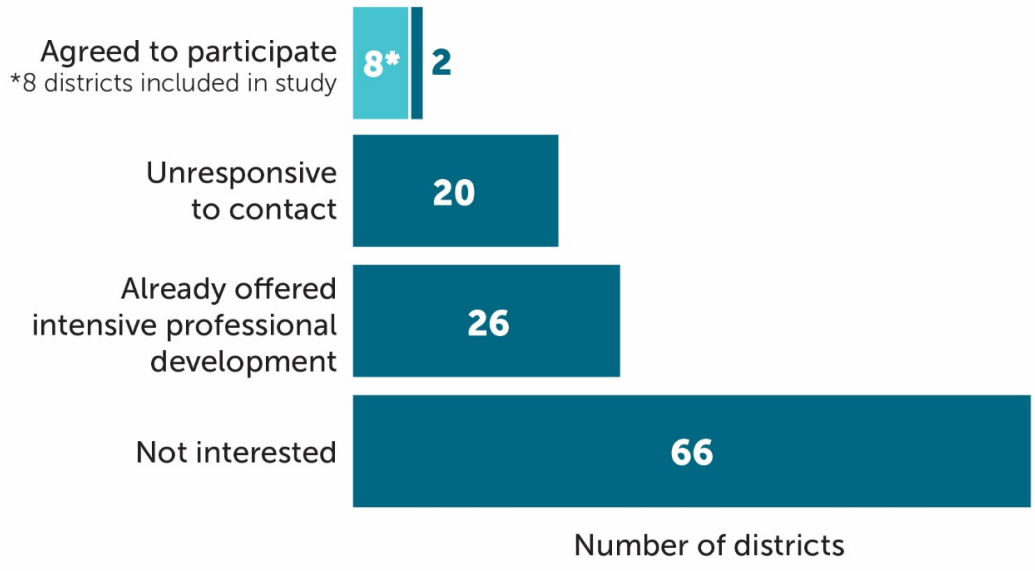
1. Sample selection

The study focused on districts that did not already offer intensive professional development to their principals. This ensured that there would be a meaningful contrast between principals who participated in the program and those who did not. Focusing on districts with less intensive professional development also made it more feasible for principals to participate the professional development from both the study and the districts. Within these districts, the study focused on principals in high-poverty elementary schools. High-poverty schools may have the greatest need for interventions to promote effective leadership,[i] as they often have less experienced principals and more instructional challenges than other schools.[ii] In addition, because the effects of principal professional development could differ for elementary and secondary schools, we focused on one level—elementary schools—to help ensure that the study's sample size would be sufficient to detect effects on student test scores at that level.

To identify eligible districts and schools for the study, the study team used the Common Core of Data to identify districts with at least 20 high-poverty elementary schools. The team defined high-poverty schools as those with at least 40 percent of students eligible for free or reduced-price lunch in the most recent school year for which data were available (2011–2012). The study team then excluded districts with existing intensive principal professional development or other major leadership initiatives, including the National Institute for School Leadership (NISL), the Wallace Foundation's Principal Pipeline and Principal Supervisor Initiatives, the Bill & Melinda Gates Foundation's Intensive Partnership Program, and the University of Virginia School Turnaround Program.

Applying these criteria resulted in a list of 122 districts that the study team contacted during the recruitment effort. Through this contact, if the study team learned that a district was already implementing intensive principal professional development, they excluded it from participation. Ultimately, 10 districts agreed to participate in the study, and the study included 8 of the 10. (The study excluded the other 2 because it had already met its recruitment targets.) Figure B.1 presents the results of the recruitment effort for the 122 eligible districts.

Figure B.1. Results from district recruitment effort



Across the eight participating districts, a total of 100 schools participated in the study. To select the study schools in each district, the study team generated a list of elementary schools in which at least 40 percent of students were eligible for free or reduced-price lunch. Districts then identified the 10 schools that (1) were not undergoing particular stress, (2) had a principal who was not expected to retire or leave the school during the study period, and (3) were not receiving School Improvement Grants or other interventions that could conflict with study activities.[iii] Districts asked the selected principals to sign a commitment form indicating that they would participate in (a) the program if they were assigned to receive it and (b) all data collection activities (regardless of whether they were assigned to participate in the program).

The study team selected a sample of 100 schools to yield a minimum detectable effect of 0.10 standard deviations on student test scores. The study achieved minimum detectable effects that ranged from 0.06 to 0.13 standard deviations across years and subjects. Appendix C and Table C.29 provide more detailed information on the minimum detectable effects that the study achieved.

Given the study's focus on districts with at least 20 high-poverty elementary schools, study districts and schools differed from typical districts and schools nationwide. Table B.1 shows that study districts differed from school districts nationwide in multiple ways. For example, study districts were larger, more concentrated in the South, and higher-poverty than public school districts nationwide. Table B.2 shows that study schools differed from public elementary schools in multiple ways. For example, study schools were higher-poverty, larger, and less likely to be magnet or charter schools than public elementary schools nationwide.

Table B.1. Comparison of study districts and public school districts nationally

Characteristic (percentages unless otherwise noted)	Mean		Difference	*p*-value
	Study districts	All public school districts		
Student characteristics				
Student racial and ethnic distribution				
Black, non-Hispanic	32	8	24*	0.025
Hispanic	39	15	24*	0.013
White, non-Hispanic	24	70	-46*	0.000
Other	5	8	-3*	0.000
Students eligible for free or reduced-price lunch	67	48	19*	0.000
English language learners	13	5	8*	0.003
Students with individualized education plan	11	15	-4*	0.002
Size				
Number of schools (average)	73	7	66*	0.000
Number of students (average)	47,422	3,475	43,947*	0.000
District location				
Urban	38	8	29	0.110
Suburban	63	23	40*	0.030
Town	0	18	-18*	0.000
Rural	0	51	-51*	0.000
Geographic region				
Northeast	0	21	-21*	0.000
Midwest	0	36	-36*	0.000
South				
South Atlantic	38	5	32	0.077
East South Central	0	4	-4*	0.000
West South Central	63	14	48*	0.008
West	0	20	-20*	0.000
Number of districts	8	13,547–14,040		

Source: Common Core of Data (2014–2015 school year).

Note: Table excludes districts that contained only charter schools. The difference between study districts and all public school districts may not equal the difference shown in the table due to rounding.

* Difference is statistically significant at the .05 level, two-tailed test.

Table B.2. Comparison of study schools and public elementary schools nationally

Characteristic (percentages unless otherwise noted)	Mean		Difference	*p*-value
	All study schools	All public elementary schools		
Student racial and ethnic distribution				
Black, non-Hispanic	39	15	24*	0.000
Hispanic	37	24	13*	0.000
White, non-Hispanic	19	52	-33*	0.000
Other	5	9	-5*	0.000
Students eligible for free or reduced-price lunch	75	55	20*	0.000
Number of students (average)	596	458	138*	0.000
Student-to-teacher ratio (average)	15	17	-1*	0.000
Schoolwide Title I status[a]	99	81	18*	0.000
Magnet school	0	4	-4*	0.000
Charter school	0	6	-6*	0.000
Number of schools	100	40,244–53,000		

Source: Common Core of Data (2014–2015 school year).

Note: The difference between study schools and all public elementary schools may not equal the difference shown in the table due to rounding.

[a] Schoolwide Title I status refers to schools with student populations that are at least 40 percent low income and are Title I-eligible. This means that the schools are classified by state and federal regulations as high poverty and eligible for additional financial assistance.

* Difference is statistically significant at the .05 level, two-tailed test.

2. Random assignment

The study team randomly assigned schools to a treatment group whose principals were offered the professional development program or to a control group whose principals were not. The primary goal of random assignment was to create treatment and control groups that were similar before the start of the program. That way, any differences in outcomes between the two groups could be reliably attributed to the effects of the program. To help ensure that the treatment and control schools would be similar in terms of key baseline characteristics, in each district, the study team grouped schools into pairs, or random assignment blocks, based on the following characteristics: grade span, average English language arts and math student achievement from the 2013–2014 school year, school size, and percentage of students eligible for free or reduced-price lunch. Within each pair, the team randomly assigned one school to the treatment group and one school to the control group.

The resulting treatment and control groups had similar baseline characteristics. Table B.3 shows that students and schools in the treatment and control groups had similar student achievement and student demographic characteristics at baseline. Similarly, Table B.4 shows that teachers and principals in treatment and control schools had similar education levels, experience, and demographic characteristics at baseline.

Table B.3. Comparison of baseline characteristics of treatment and control students and schools

Baseline student characteristics (percentages unless otherwise noted)	Mean		Difference	*p*-value
	Treatment schools	Control schools		
Student outcomes (percentiles in state)				
English language arts achievement	42	41	1	0.632
Math achievement	41	41	0	0.921
Student characteristics				
Female	50	49	1	0.363
Student racial and ethnic distribution				
Asian	3	3	0	0.888
Black	39	42	-3	0.659
Hispanic	38	37	1	0.834
White	35	31	4	0.497
Other	3	2	0	0.922
Students eligible for free or reduced-price lunch	76	76	0	0.999
English language learners	18	14	4	0.233
Students with individualized education plan	11	10	0	0.785
School characteristics				
Number of students (average)	573	599	-26	0.484
Student-to-teacher ratio (average)	15	15	0	0.995
Schoolwide Title I status[a]	100	100	0	n.a.
Number of students	5,816–11,725	5,842–12,198		
Number of schools	50	50		

Sources: Student outcomes and characteristics come from administrative student records (2014–2015 school year). School characteristics come from Common Core of Data (2014–2015 school year).

Note: The difference between treatment and control schools may not equal the difference shown in the table due to rounding. None of the differences is statistically significant at the .05 level, two-tailed test.

[a] Schoolwide Title I status refers to schools with student populations that are at least 40 percent low income and are Title I-eligible. This means that state and federal regulations classify the schools as high poverty and eligible for additional financial assistance.

n.a. = not applicable.

Table B.4. Comparison of baseline characteristics of principals and teachers in treatment and control schools

Baseline characteristic (percentages unless otherwise noted)	Mean		Difference	*p*-value
	Treatment schools	Control schools		
Principal characteristics				
Female	74	83	-9	0.355
Principal racial and ethnic distribution				
Asian	0	0	0	n.a.
Black	29	26	4	0.734
Hispanic	<6	<11	-5	0.398
White	68	69	-1	0.935
Other	0	0	0	n.a.
Principal education level				
Doctorate or professional degree	<9	<12	-3	0.646
Master's degree or education specialist	91	91	0	1.000
Bachelor's degree	<6	<3	3	0.317

Baseline characteristic (percentages unless otherwise noted)	Mean		Difference	*p*-value
	Treatment schools	Control schools		
Other or no degree	0	0	0	n.a.
Years of service as an administrator[a]	5	5	-1	0.485
Teacher characteristics				
Female	89	90	0	0.734
Teacher racial and ethnic distribution				
Asian	1	1	0	0.550
Black	16	17	-1	0.819
Hispanic	19	21	-2	0.660
White	72	73	-1	0.918
Other	2	3	0	0.603
Teacher education level				
Doctorate or professional degree	0	0	0	0.146
Master's degree or education specialist	29	31	-2	0.485
Bachelor's degree	70	68	2	0.445
Other or no degree	1	0	0	0.680
Years of service as a teacher[b]	11	10	1	0.430
Number of principals	34–50	35–50		
Number of teachers	902–1,494	970–1,512		
Number of schools	34–50	35–50		

Source: Administrative educator records (2014–2015 school year).

Note: The difference between treatment and control schools may not equal the difference shown in the table due to rounding. None of the differences is statistically significant at the .05 level, two-tailed test.

[a] Years of service include all years as an administrator before and including the 2014–2015 school year.

[b] Years of service include all years taught before and including the 2014–2015 school year.

< or > indicates that we have withheld the exact percentage to protect respondent confidentiality in accordance with National Center for Education Statistics statistical standards, but that the percentage is less than or greater than the number following the < or > symbol.[iv]

n.a. = not applicable.

B. Data collection

The study team collected data from several sources to assess the effects of the study's principal professional development program and describe how it was implemented. Table B.5 lists these data sources. Table B.6 lists the response rates for the data sources used to measure the effects of the program.

Table B.5. Data sources

Data source	Data obtained	Timing of data collected	Respondent
Data to measure effects			
Principal time-use log	Time principal spent on various leadership practices	Four weeks during each study school year (2015–2016 and 2016–2017)	Principals (treatment and control)
Principal survey[a]	Principal leadership practices, professional development participation, and background characteristics	Spring 2016 and spring 2017	Principals (treatment and control)
Teacher survey	Teacher perception of principal leadership practices and school climate	Spring 2016 and spring 2017	Teachers (treatment and control)

Data source	Data obtained	Timing of data collected	Respondent
Principal and teacher records	Grade and school assignment data, retention, and background characteristics for teachers and principals	Fall 2016 and fall 2017	Principals and teachers (treatment and control)
Student records	Student achievement, behavioral outcomes, and background characteristics	Fall 2016, fall 2017, and fall 2019	Students (treatment and control)
Data to measure implementation			
Participation forms	Summer institute, group training, and professional learning community session attendance and dosage	Each summer institute, group training, and professional learning community session[b]	Study observers
Observation forms	Content and fidelity of implementation for summer institute, group training, and professional learning community sessions	Each summer institute, group training, and professional learning community session[b]	Study observers
End-of-session forms	Quality and usefulness of summer institute, group training during the school year, and professional learning community sessions	Each summer institute, group training, and professional learning community session	Principals (treatment only)
Coaching logs	Content and dosage of coaching for each principal	Each coaching session	Coaches
Staff résumés	Educational background and professional experience for each coach or instructor	Before summer institute	Coach or instructor

Note: The 2015–2016 school year was Year 1 and 2016–2017 was Year 2.

[a] Data from the principal survey are used to describe the implementation and effects of the professional development program.

[b] In Year 1, each participant was offered four days of the summer institute, eight days of group training during the school year, and four professional learning community sessions. The study team completed observation forms for all of these sessions, except for two group training sessions in which the team observed only half of the districts. In Year 2, new participants were offered two days of the summer institute and four professional learning community sessions. The study team completed observation forms for all of these sessions in Year 2.

Table B.6. Response rates for data sources used to estimate effects

	Response rates (percentages)					
Data collected	Overall	Treatment group	Control group	Overall	Treatment group	Control group
	Year 1			Year 2		
Principal time-use log						
Week 1	100	100	100	99	98	100
Week 2	100	100	100	100	100	100
Week 3	99	100	98	99	98	100
Week 4	100	100	100	96	98	94
Principal survey	95	94	96	96	96	96
Teacher survey	91	90	91	89	88	90
Principal and teacher records						
Principal retention	100	100	100	100	100	100
Teacher retention[a]	100	100	100	100	100	100
Student records[b]						
English language arts	90	91	89	92	92	91
Math	92	93	91	93	95	92

[a] Teacher retention response rates were calculated for the seven districts that provided data on the number of teachers in the district in each year.

[b] Five of the eight districts provided student records data for Year 3. Among these districts, Year 3 response rates for English language arts scores were 94 percent overall (93 percent for the treatment group and 95 percent for the control group), and Year 3 response rates for math scores were 98 percent overall (98 percent for the treatment group and 98 percent for the control group).

C. Analytic methods

This section describes the approach for examining the effects of the study's principal professional development program. First, we describe how we constructed outcome measures for the study. Second, we provide details on the study's analytic methods, including the methods used to estimate effects of the program on these outcomes and the methods used to estimate the relationship between the characteristics of the program, its effects on principals' practices, and its effects on student achievement.

1. Constructing outcome measures

The study examined measures of principals' practices and school climate, principals' time use, principal and teacher retention, and students' test scores. This section discusses the methods used to construct these measures and standardize test scores across districts.

Measures of principals' practices and school climate. To measure these outcomes, we combined items from the surveys into scales meant to capture a common underlying construct (such as school climate). Compared with analyzing all items separately, combining outcomes into scales can (1) reduce the measurement error associated with a given construct, leading to a more precise estimate of effects; and (2) reduce the number of outcomes in the analysis, limiting the possibility of finding many significant effects simply due to chance. To create the scales, we averaged the survey items in each group.

We conducted a confirmatory factor analysis to ensure that the scales adequately captured the underlying constructs from the teacher and principal surveys. Although many of the items in the teacher and principal surveys came from validated surveys, most of the scales are new in that they combine items from multiple existing surveys, along with some newly developed items. We used a confirmatory—rather than an exploratory— approach to validate the new scales because the study's conceptual model provided a theoretical basis for the groups of items. We used the Year 1 data to validate the scales and then used the same scales for the Year 2 analysis.

We conducted the confirmatory analysis in two iterative stages:

1. **Estimated a confirmatory factor model using data from the Year 1 surveys.** First, we grouped survey items based on the construct they aimed to measure. Second, based on these groupings, we estimated a structural equation model separately for the teacher and principal surveys. We used the weighted least squares with mean and variance adjustment (robust) estimator, which studies have shown to be both robust and feasible for models with categorical measures and relatively high numbers of factors, such as those in this study.[v] We modeled categorical, Likert-type variables using an ordered probit model to allow for the possibility that responses are nonlinear in the underlying factor (for example, reporting disagree versus strongly agree might mean something different from reporting agree versus disagree). Third, we assessed the model fit, reliability, validity, and number of factors within each group of items.
2. **Adjusted groupings based on the results of the initial analysis.** After making some small adjustments to the groupings based on the results of the initial factor analysis, we re-estimated the model with the revised groupings.

Overall, the final models fit the data well. (Table B.7 provides overall fit statistics, and Table B.8 shows reliabilities and the final set of items included in each scale.) The model for the teacher survey met standard criteria for overall fit, reliability, and validity. The overall fit was acceptable, based on the standard criteria of a root mean square error of approximation that was less than 0.05,[vi] and comparative fit index and Tucker Lewis index that were greater than 0.90.[vii] All of the scales were reliable, based on estimated reliabilities that exceeded the standard cutoff of 0.70.[viii] Only one of 66 correlations between factors exceeded 0.85, suggesting that the final groupings exhibit strong discriminant validity.[ix] In particular, the

correlation between principals' competence in providing instructional supports (teacher report) and school climate was 0.90. Because these scales capture distinct concepts, we did not consider combining them into a single scale. Horn's parallel analysis also provided no evidence that any groups of items captured more than one latent factor.[x]

The model for the principal survey also met criteria for overall fit and discriminant validity. In addition, the reliabilities exceeded 0.70 for all but two of the scales. Because the reliabilities for these two scales—frequency with which principal arranged professional development for teachers and frequency of principal's communication about school improvement—were close to the threshold of 0.70 (0.65 and 0.64), we retained them for the analysis but also confirmed that effects on the individual items followed the same patterns as the overall scales.

Table B.7. Overall model fit

				Root mean square error of approximation (RMSEA)			Comparative fit	
Survey	Sample size	Number of items	Number of constructs	Estimate	90 percent confidence interval	Probability RMSEA < .05	CFI	TLI
Principal	95	44	8	0.039	0.025–0.050	0.95	0.94	0.93
Teacher	1,136	109	12	0.031	0.030–0.031	1.00	0.97	0.97

CFI = comparative fit index; RMSEA = root mean square error of approximation; TLI = Tucker Lewis index.

Table B.8. Items in final scales and estimated reliabilities

Construct and associated items	Number of items	Reliability (omega)
Principal survey		
Principal's competence in providing instructional supports (principals' report) Extent to which principals agreed with the following (strongly disagreed, disagreed, agreed, or strongly agreed): • I know what effective teaching looks like • I know what teaching practices to look for when I'm conducting classroom observations • I feel comfortable having difficult conversations with teachers in my school • I feel comfortable suggesting specific teaching actions to teachers, based on student achievement data, teacher effectiveness data, or classroom observation data • I feel competent helping teachers identify their areas of instructional practice that need improvement • I feel competent helping teachers recognize their accomplishments or identify their areas of strength • I know how to ask teachers questions soliciting their own reflection on teaching practices • I know how to give teachers feedback on their instruction that provides them with actionable steps for improvement • I know where to find resources to support teacher instructional practice outside of my areas of expertise	9	0.94

Construct and associated items	Number of items	Reliability (omega)
Principal's teacher observation skills (principals' report) Extent to which principals did the following when visiting classrooms to observe instruction (not at all, to a small extent, to a moderate extent, or to a great extent): • Focused observations on specific areas or issues unique to the teacher's needs • Recorded descriptions of specific things that the teacher and students did or said during a classroom observation • Analyzed data collected during classroom observations to identify trends in instructional practice • Found classroom observations useful for their leadership practice	4	0.89
Frequency with which principal arranged professional development for teachers (principals' report) How often principals completed the following activities (never, yearly, quarterly, monthly, or weekly), with units converted to number of times per year in the final scale: • Helping a teacher locate formal professional development opportunities to support his or her goals • Arranging an informal learning opportunity to support a teacher's growth • Connecting a teacher to a content expert • Connecting a teacher to a network of teachers formed specifically for the professional development of teachers	4	0.65
Coherence of school improvement plan (principals' report) Extent to which principals agreed with the following (strongly disagreed, disagreed, agreed, or strongly agreed): • The administration collaborated with teachers collaboratively to shape the plans in the school • Plans for improvement in the school included indicators to measure progress toward goals • Plans for improvement in the school aligned with evidence from teacher performance evaluations, observations of classroom teaching, or student performance data • Plans for improvement in the school were consistent with teachers' goals for individual growth • Plans for improvement in the school clearly outlined steps that teachers should take to improve their teaching	5	0.89
Frequency of principal's communication about school improvement (principals' report) • How often principals did the following activities (never, yearly, quarterly, monthly, or weekly), with units converted to number of times per year in the final scale: • Communicating goals for improving instructional quality in school to teachers or other school staff • Updating staff on progress toward the school vision or goals for improvement • Incorporating a clear vision for the school into regular communications • Delegating these actions surrounding school culture and vision to another member of the staff	4	0.64

Construct and associated items	Number of items	Reliability (omega)
School climate (principals' report) Extent to which principals report the following are a problem for their school (not at all, to a small extent, to a moderate extent, or to a great extent): • Student absenteeism • Widespread disorder in classrooms • Racial tensions among students • Bullying or harassment among students • Physical conflicts among students • Students' acts of disrespect for teachers • Conflicts between students and teachers or verbal abuse of teachers • Students' possession of weapons	8	0.84
Teacher survey		
Principal's competence in providing instructional supports (teachers' report) Extent to which teachers agreed with the following (strongly disagreed, disagreed, agreed, or strongly agreed): • My principal knows what effective teaching looks like • My principal made teachers feel comfortable to try new things in the classroom • My principal communicated clear standards for student learning • My principal was transparent about performance expectations for teachers • My principal expected teachers to continually learn and grow • My principal encouraged teachers to implement what they learned in professional development • My principal informed teachers about resources they could use to improve their instruction • My principal encouraged teachers to share ideas and work together to improve their teaching • My principal praised or encouraged teachers for their efforts to improve their teaching • My principal worked directly with teachers to help them improve their instruction • My principal knew what was going on in classrooms • My principal changed instructional assignments to match teachers' expertise with students' needs	12	0.97
Coherence of school improvement plan (teachers' report) Extent to which teachers agreed with the following about plans for school improvement (strongly disagreed, disagreed, agreed, or strongly agreed): • Teachers collaborated with the administration to shape plans • Plans for school improvement included indicators to measure progress toward goals • Plans for school improvement were consistent with teachers' own goals for their individual growth • Plans for school improvement clearly outlined specific steps that teachers could take to improve their teaching	4	0.84

Construct and associated items	Number of items	Reliability (omega)
Frequency of principal's communication about school improvement (teachers' report) How often teachers reported their principals had done the following activities (never, yearly, quarterly, monthly, weekly), with units converted to number of times per year in the final scale: • Discussed his or her goals for improving our school's instructional quality with teachers • Communicated progress toward goals for improving our school's instructional quality to teachers • Communicated a clear vision for our school's instructional quality through his or her regular communications	3	0.86
School climate (teachers' report) Extent to which teachers agreed with the following (strongly disagreed, disagreed, agreed, or strongly agreed): • There is a great deal of cooperative effort among the staff members in my school • The school administration's behavior toward the staff is supportive and encouraging • Teachers at my school trust each other • In my school, teachers are encouraged to experiment in their classrooms • In my school, teachers are expected to continually learn and seek new ideas for teaching • Academic achievement is recognized and acknowledged in my school • The level of student misbehavior in this school (such as noise, horseplay, or fighting in the halls, cafeteria, or student lounge) interferes with my teaching • Students in my school respect others who get good grades • Teachers at my school encourage students to keep trying even when the work is challenging • Teachers at my school set high expectations for academic work	10	0.91
Frequency instructional support and feedback from principal (teachers' report) How often teachers reported having received each support (never, yearly, quarterly, monthly, or weekly) with units converted to number of times per year in the final scale: • Principal observed classroom instruction • Principal gave specific feedback on the quality of teaching as part of a state- or district-mandated evaluation • Principal gave specific feedback on the quality of teaching not as part of a state- or district-mandated evaluation • Principal worked to develop specific instructional practice goals • Principal examined data to determine whether instructional practice goals were met • Principal discussed grade- or school-level student achievement data, teacher effectiveness data, or classroom observation data • Principal made data (student achievement data, teacher effectiveness data, or classroom observation data) or reports available • Principal suggested specific teaching actions, based on student achievement data, teacher effectiveness data, or classroom observation data • Principal reviewed teaching plans to ensure that they aligned with curriculum standards • Principal shared instructional materials or curricula to support instructional goals • Principal helped establish classroom systems or routines to improve student engagement or support high expectations for students	11	0.88

Construct and associated items	Number of items	Reliability (omega)
Usefulness of all types of support from principal (teachers' report) Extent to which teachers reported that the following supports were useful (not very useful, somewhat useful, moderately useful, or very useful): • Principal observed classroom instruction • Principal gave specific feedback on the quality of teaching as part of a state- or district-mandated evaluation • Principal gave specific feedback on the quality of teaching not as part of a state- or district-mandated evaluation • Principal worked to develop specific instructional practice goals • Principal examined data to determine whether instructional practice goals were met • Principal discussed grade- or school-level student achievement data, teacher effectiveness data, or classroom observation data • Principal made data (student achievement data, teacher effectiveness data, or classroom observation data) or reports available • Principal suggested specific teaching actions, based on student achievement data, teacher effectiveness data, or classroom observation data • Principal reviewed teaching plans to ensure that they aligned with curriculum standards • Principal shared instructional materials or curricula to support instructional goals • Principal helped establish classroom systems or routines to improve student engagement or support high expectations for students	11	0.98
Usefulness of feedback received from principal (teachers' report) Extent to which teachers reported that feedback they received (not at all, to a small extent, to a moderate extent, or to a great extent): • Addressed the pressing issues in their classroom(s) • Included questions soliciting their own reflection on teaching practices • Involved them talking more than the evaluator • Used evidence as a starting point for reflection • Provided them with actionable steps for improvement • Focused on improving aspects of their teaching practice that are realistic for them to change • Identified trends in their instructional practice, based on analysis of evidence, such as changes over time or patterns across different populations of students • Related to feedback they received earlier in the year on the same issue or area for improvement • Provided descriptions of specific things that their students and they did or said during a classroom observation • Included recognition of their accomplishments or helped them identify areas of strength • Helped them identify their areas of instructional practice that need improvement	11	0.96

Principals' time use. To measure how principals spent their time, we calculated the amount of time principals reported spending on each of 10 leadership activities listed in the study's time use logs. Principals completed 20 daily logs a year throughout both study school years (over five consecutive days, during four weeks selected by the study team throughout each year). In each daily log, principals reported how they spent their time during each hour-long period of the school day. For each activity reported in that hour-long period, they indicated the time they spent on the activity, within ranges (1 to 14 minutes, 15 to 29 minutes, 30 to 44 minutes, and 45 minutes to an hour). At the beginning and end of the school day, they also had the option of reporting "more than one hour."

Our main analysis uses the midpoint of each range to determine the number of minutes principals spent on each activity, following the approach used by Camburn et al.[xi] If the principal selected "more than one hour" for these estimates, we assumed a range of 60 to 75 minutes and took the midpoint of this range. To assess the sensitivity of the findings to using the midpoint for each range, we conducted sensitivity tests that used (1) the minimum for each range and (2) the maximum for each range (see Appendix C, Table C.18).

Principal and teacher retention. To measure effects on principal and teacher retention, we used administrative data on principals' and teachers' school assignments to determine whether principals and teachers remained in their schools. For principal retention, we calculated whether the principal of the school before random assignment remained the principal of that school at the end of Year 1 (one-year retention), Year 2 (two-year retention), and in the year after the program ended (three-year retention). Likewise, for teachers, we calculated one-, two-, and three-year retention rates for teachers who taught in the school before random assignment.

Student test scores. To measure student achievement, we used students' test scores on state assessments in math and English language arts, standardized across the different states in the study. To standardize, we converted each of the test scores to *z*-scores by subtracting the statewide mean and dividing by the statewide standard deviation for that year, grade, and subject. After estimating effects on the standardized scores, we converted the estimates into test score percentiles to make them easier to interpret.

2. Estimating effects

In this section, we describe our approach to estimating the effects of the professional development program on school, principal, teacher, and student outcomes.

Main estimation model. To estimate the effects of the program, we used the following model:

$$Y_{ij} = \alpha + \beta T_j + \delta P_{ij} + \gamma Z_j + \varepsilon_{ij}, \quad (1)$$

where Y_{ij} is the outcome of interest for individual (principal, teacher, or student) i in school j; α is an intercept term; T_j is an indicator equal to one if the school was assigned to the treatment group and zero otherwise; P_{ij} is a vector of baseline school- and student-level characteristics (only the student outcome models include this vector), Z_j is a vector of fixed effects corresponding to the study's random assignment blocks; δ and γ are coefficient vectors; and ε_{ij} is a random error term. The coefficient β represents the average effect of the program.

We estimated Equation (1) separately for each year of implementation of the program. In addition, we also estimated the effects of the program on student achievement in the year after implementation was complete (Year 3). In Years 2 and 3, the estimates reflect the cumulative effect of offering schools the program for two years, as opposed to the effect of individual principals participating in the program for two years; 84 percent of principals in treatment schools participated in the program for the full two years.

We estimated the model using ordinary least squares. The estimated standard errors account for the clustering of educator and student outcomes at the school level. We also calculated the unadjusted mean outcomes for the control group and mean outcomes for the treatment group (the unadjusted mean outcomes for the control group plus the average effect of the program).

Covariates. All models controlled for random assignment block fixed effects to improve the precision of the estimated effects. In addition, the models for student outcomes controlled for baseline school- and student-level covariates. The models for principal and teacher outcomes did not control for school- or educator-level covariates, because the smaller sample sizes for these analyses limited the number of covariates we could include. Table B.9 summarizes the covariates included in each type of model.

Table B.9. Covariates included in the models used to estimate effects

Model type	Student-level covariates	School-level covariates	Random assignment block fixed effects
Principal and teacher outcomes	No	No	Yes
Student outcomes	• Standardized English language arts and math test scores from the baseline year • Gender • Race and ethnicity • Free or reduced-price lunch eligibility • English learner status • Special education status • State- and grade-level fixed effects	• School size (number of students) • School's student-to-teacher ratio	Yes

Weights. We weighted educator and student outcomes so each school contributed equally to the estimated effects. Thus, each school received the same weight in the analysis, regardless of the number of teachers or students in the school.

Treatment of missing data. The analysis included only individuals who had nonmissing values of the outcome variables. Simulations have suggested that, for randomized controlled trials, this approach may have led to only a small amount of bias (0.05 standard deviations or less) when outcome data are missing at random among individuals with the same covariate values.[xii]

The analysis included students with missing covariate values. We replaced the missing covariate values with a placeholder (zero) and created an indicator for the covariate having a missing value, which we included in the model. Simulations by Puma et al.[xiii] have shown that this approach to handling missing covariate data is likely to keep estimation bias at less than 0.05 standard deviations.

Samples. Table B.10 describes the samples for the analyses of effects. For the student analyses, we defined the sample as students who were enrolled in the study schools at the beginning of Year 1. This definition helped to ensure that the estimates reflected the effects of the program on individual students. We examined outcomes for these students in all three years, including for students who had moved to other schools in the district by Years 2 or 3. (These estimates exclude any effects due to student mobility if the program affected the types of students who were enrolled in each year, as these effects are likely to be of less interest to policymakers.)

Table B.10. Analysis samples

Model and outcomes	Samples
Principal survey outcomes and time use	Current principals in Years 1 and 2
Teacher survey outcomes	Current teachers in Years 1 and 2
Principal retention	Principals of study schools at baseline
Teacher retention	Teachers in study schools at baseline, Year 1, and Year 2
Student achievement	Students enrolled in study schools at the beginning of Year 1 (followed into Years 2 and 3)

Estimation of effects for subgroups. We estimated the effects of the program for two subgroups of schools—principal experience and baseline achievement level—that could be defined for both treatment and control schools based on their characteristics at the start of the study. Therefore, we can calculate rigorous estimates of the effects of the program separately for these subgroups. Table B.11 defines the subgroup characteristics that we analyzed and describes how we grouped schools.

Table B.11. Subgroups examined

Subgroup	Definitions of subgroups
Principal experience level	
More experienced	Baseline principal had at least three years of experience as a principal
Less experienced	Baseline principal had fewer than three years of experience as a principal
Baseline achievement level	
High	Average baseline achievement in math and English language arts above the 75th percentile for study schools
Medium	Average baseline achievement in math and English language arts between 25th and 75th percentile for study schools
Low	Average baseline achievement in math and English language arts below 25th percentile for study schools

In each set of subgroup analyses, we organized schools into two or three subgroups, such as high, medium, and low baseline achievement. To estimate effects on subgroups, we modified Equation (1) as follows:

(2) $Y_{ij} = \alpha + \beta_1 T_j + \gamma_2 Group2_j + \gamma_3 Group3_j + \beta_2 (T_j \times Group2_j) + \beta_3 (T_j \times Group3_j) + \delta P_{ij} + \gamma Z_j + \varepsilon_{ij}$,

where $Group2_j$ and $Group3_j$ represent two of the three subgroups, and $Group1_j$ is the omitted category. In this model, the effects of the program on subgroups 1, 2, and 3 are β_1, $(\beta_1 + \beta_2)$, and $(\beta_1 + \beta_3)$. When schools are organized into two subgroups, such as high and low, Equation (2) would be the same, except it would not include indicators and interaction terms involving $Group3_j$.

Estimation of effects based on treatment group characteristics. Three other school characteristics—the extent to which coaching focused on instructional leadership, completion of coaching-assigned activities, and coach experience level—were based on the characteristics of the program (Table B.12). Thus, they were available only for treatment schools and could be measured only *after* the start of the study. These analyses are therefore less rigorous than the subgroup analyses described above and provide only suggestive evidence. We estimated these models using an equation similar to Equation (2). We defined the characteristic for each pair of schools based on the value of the characteristic for the treatment school in the pair.

Table B.12. Treatment group characteristics examined

Characteristic	Definitions of characteristics
Focus of coaching on instructional leadership	
Greater	At least two-thirds of coaching sessions (the sample median) had a primary focus on instructional leadership
Less	Fewer than two-thirds of coaching sessions (the sample median) had a primary focus on instructional leadership
Completion of coaching-assigned activities	
High	Principals completed more than the median number of coach-assigned activities
Low	Principals completed fewer than the median number of coach-assigned activities
Coach experience level	
Less experienced	Coach had at least three years of experience working for the Center for Educational Leadership
More experienced	Coach had fewer than three years of experience working for the Center for Educational Leadership

3. Estimating the relationship between characteristics of the professional development program, its effects on principals' practices, and its effects on student achievement

In this section, we describe our approach to estimating the relationship between characteristics of the professional development program, its effects on principals' practices, and its effects on student achievement.

For this analysis, we first estimated the effects of the program on student achievement and principal practices in each treatment school. To do so, we estimated the effect of the program in the random assignment block to which the treatment school belonged. We used a modified version of Equation (1) for student achievement and principal practice outcomes, in which the treatment indicator was replaced by a vector of interaction terms between the treatment indicator and indicators for each of the 50 random assignment blocks:

$$(3) \quad Y_{ij} = \alpha + \sum_{b=1}^{50} \beta_b (T_j \times B_j^b) + \delta P_{ij} + \gamma Z_j + \varepsilon_{ij},$$

where B_j^b is an indicator for random assignment block *b*, β_b represents the effect of the program in block *b*, and all other variables are the same as those in Equation (1). As in Equation (1), the principal practice outcome models exclude P_{ij}, the vector of baseline school- and student-level characteristics.

We then estimated a series of bivariate correlations to examine the relationship between characteristics of the program and its effects on principals' practices and student achievement.

- To learn how the program's effects on principals' practices were related to its effects on student achievement, we estimated a series of bivariate correlations between the program's block-specific effects on principal practices and student achievement.
- To learn about how different characteristics of the program were related to its effects on principal practices, we estimated bivariate correlations examining the relationship between characteristics of the program in each treatment school and its effects on principal practices in that treatment school's random assignment block. The effects on principal practices included in this analysis were those that were statistically significantly correlated with effects on student achievement.

APPENDIX C

SUPPLEMENTAL TABLES

This appendix supplements the findings presented in the report. It includes additional details on findings presented in the report, additional findings that are not in the report, and supplemental information for systematic reviews.

A. Additional details on findings in the report

This section includes additional information on findings in the report on (1) the effects of the study's principal professional development program on student achievement and school and teacher outcomes, (2) the effects of the program on principals' practices, (3) implementation of the program and the professional development that principals received from the study and other sources, (4) the effects of the program on student achievement by district and blocks of schools, and (5) the relationship between characteristics of the program and its effects on principals' practices and student achievement.

1. Effects on student achievement and school and teacher outcomes

In this section, we present key supporting analyses on the effects of the professional development program on student achievement and school and teacher outcomes. Figure 2 in the report shows that the program had no effect on average English language arts or math achievement in Years 1, 2, or 3. Table C.1 presents the estimated effects of the program on student achievement and corresponding *p*-values. Table C.2 shows that the program did not affect student achievement for subgroups of students based on principals' years of experience. Table C.3 shows that the program had different effects for some subgroups based on schools' baseline achievement levels.

Table C.1. Effects on student achievement (supplement to report Figure 2)

Student achievement (percentile in state)	Year 1				Year 2				Year 3			
	Treat-ment	Control	Effect	*p*-value	Treat-ment	Control	Effect	*p*-value	Treat-ment	Control	Effect	*p*-value
English language arts	41	40	0	0.706	41	41	0	0.964	42	40	2	0.299
Math	39	40	-1	0.073	40	39	1	0.599	40	39	1	0.503
Number of students	11,423–11,725	11,876–12,198			9,724–9,989	9,724–9,803			4,907–5,149	5,064–5,251		
Number of schools	50	50			50	50			30	30		

Source: Administrative student records for the 2015–2016, 2016–2017, and 2017–2018 school years.

Note: None of the effects is statistically significant at the .05 level, two-tailed test.

Table C.2. Effects on student achievement, by principal experience level (supplement to report text)

Student achievement (percentile in state)	Year 1				Year 2				Year 3			
	Treat-ment	Control	Effect	*p*-value	Treat-ment	Control	Effect	*p*-value	Treat-ment	Control	Effect	*p*-value
English language arts												
Less experienced principals	40	39	1	0.428	42	41	1	0.589	44	42	3	0.348
More experienced principals	41	41	0	0.877	41	41	0	0.643	42	42	1	0.703
Math												
Less experienced principals	39	40	-1	0.623	41	41	0	0.959	45	46	-1	0.877
More experienced principals	39	41	-2	0.071	41	40	1	0.462	42	40	2	0.361
Number of students	11,423–11,725	11,876–12,198			9,724–9,989	9,724–9,803			4,907–5,149	5,064–5,251		
Number of schools	50	50			50	50			30	30		

Source: Administrative student records for the 2015–2016, 2016–2017, and 2017–2018 school years.

Note: Principal experience is defined based on the experience level of the principal at baseline. Less experienced principals are those in their first three years as a principal at baseline. More experienced principals are those with three or more years of experience as a principal at baseline. None of the effects is statistically significant at the .05 level, two-tailed test.

Table C.3. Effects on student achievement, by average baseline achievement of study schools (supplement to report text)

Student achievement (percentile in state)	Year 1				Year 2				Year 3			
	Treat-ment	Control	Effect	*p*-value	Treat-ment	Control	Effect	*p*-value	Treat-ment	Control	Effect	*p*-value
English language arts												
High baseline achievement	42	42	0	0.800	45	44	1	0.725	53	50	3	0.092
Medium baseline achievement	35	37	-1	0.068	35	37	-2*	0.042	40	41	-1	0.568
Low baseline achievement	33	30	3	0.056	33	31	2	0.211	34	30	4*	0.050
Math												
High baseline achievement	39	41	-2	0.198	44	44	0	0.881	50	46	4	0.084
Medium baseline achievement	32	34	-2*	0.034	34	35	-1	0.385	39	38	0	0.853
Low baseline achievement	31	31	0	0.779	32	30	2	0.366	31	31	0	0.887
Number of students	11,423–11,725	11,876–12,198			9,724–9,989	9,724–9,803			4,907–5,149	5,064–5,251		
Number of schools	50	50			50	50			30	30		

Source: Administrative student records for the 2015–2016, 2016–2017, and 2017–2018 school years.

Note: Low baseline achievement is defined as schools with average student achievement in the bottom 25 percent of the sample, medium baseline achievement is defined as schools with average student achievement in the middle 50 percent, and high baseline achievement is defined as schools with average student achievement in the top 25 percent.

*Effect is statistically significant at the .05 level, two-tailed test.

Figure 3 in the report shows that the program did not affect principals' or teachers' perceptions of school climate. Table C.4 presents the estimated effects on school climate and *p*-values.

Table C.4. Effects on school climate (supplement to report Figure 3)

Outcome (1- to 4-point scale)	Year 1				Year 2			
	Treatment	Control	Effect	*p*-value	Treatment	Control	Effect	*p*-value
School climate (principals' report)[a]	3.1	3.2	-0.1	0.314	3.0	3.1	-0.1	0.440
School climate (teachers' report)[b]	3.0	3.1	-0.1	0.072	3.0	3.0	0.0	0.541
Number of principals	45	45			46	46		
Number of teachers	543	584			527	568		

Sources: Principal and teacher surveys, spring 2016 and spring 2017.

Note: None of the effects is statistically significant at the .05 level, two-tailed test.

[a] School climate, as reported by principals, includes the extent to which principals reported the school having problems with student absenteeism, widespread disorder in classrooms, and conflicts between students and teachers. The scale indicates whether each issue is a problem to a (1) great extent, (2) moderate extent, (3) small extent, or (4) not at all.

[b] School climate, as reported by teachers, includes the extent to which teachers reported cooperative effort among staff members in the school, the school administration being supportive and engaging, and not having problems with student misbehavior interfering with their teaching. The scale indicates whether teachers (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

Figure 4 in the report shows that the program did not affect overall principal or teacher retention over a three-year period for staff who worked in study schools before the study began, other than a small negative effect on teacher retention in Year 1. Table C.5 presents the estimated effects on principal and teacher retention and *p*-values.

Table C.5. Effects on principal and teacher retention (supplement to report Figure 4)

Retention rate among educators who worked in schools at baseline (percentages)	Treatment	Control	Effect	*p*-value
Principals				
Over one year (baseline to Year 1)	84	86	-2	0.742
Over two years (baseline to Year 2)	74	64	10	0.165
Over three years (baseline to Year 3)	56	54	2	0.821
Teachers				
Over one year (baseline to Year 1)	79	83	-3*	0.008
Over two years (baseline to Year 2)	59	59	0	0.967
Over three years (baseline to Year 3)	55	52	3	0.054
Number of principals	50	50		
Number of teachers	1,465–1,604	1,515–1,662		
Number of schools	40	40		

Sources: Implementation data and administrative educator records for the 2014–2015 through 2017–2018 school years.

Note: Seven of the eight participating districts provided administrative educator records for teachers.

*Effect is statistically significant at the .05 level, two-tailed test.

2. Effects on principals' practices

In this section, we present additional details on report findings on the effects of the professional development program on principals' practices. Figure 5 in the report shows that the program did not affect the amount of time that principals spent on instructional leadership in Year 2. Table C.6 presents the estimated effects and *p*-values for time use in Year 2 and shows that patterns of time use in Year 1 are similar to those shown for Year 2.

Figure 6 and Table 1 in the report show that the program had some negative effects on principals' instructional leadership practices, including the frequency of instructional support and feedback they provided to teachers and their competence in providing instructional supports. Table C.7 shows that the typical classroom observation lasted about 30 minutes. Table C.8 presents the estimated effects and *p*-values related to report Figure 6 and Table 1, along with findings on principals' perceptions of the quality of instructional leadership that are referenced only in the report text. Table C.9 shows that the program had some negative effects on principal's instructional leadership practices for subgroups of principals and schools based on principals' years of experience and schools' baseline level of achievement.

As discussed in the report, although the program primarily emphasized instructional leadership, it also covered human capital management and organizational leadership practices to help support learning and instruction, such as arranging professional development for teachers and developing a plan for school improvement. Table C.10 shows that the program had few effects on these human capital management and organizational leadership practices.

Table C.6. Effects on principals' time use (supplement to report Figure 5 and report text)

	Year 1				Year 2			
	Mean outcome				Mean outcome			
Number of hours per week spent on:	Treatment	Control	Effect	*p*-value	Treatment	Control	Effect	*p*-value
Organizational leadership								
Student affairs	7	7	0	0.642	7	8	0	0.571
Administration	6	5	0	0.502	5	5	0	0.437
Other								
School improvement efforts	2	2	0	0.865	3	3	0	0.381
Community and parent outreach	2	2	0	0.818	2	2	0	0.523
Other work-related activities	1	1	0	0.744	1	1	0	0.611
Instructional leadership								
Evaluation	5	5	0	0.674	5	4	0	0.490
Teacher feedback	4	4	0	0.515	4	4	0	0.785
Curriculum	3	3	0	0.809	2	3	-1*	0.031
Human capital management								
Recruiting teachers	2	2	1	0.061	2	2	-1	0.158
Personnel policies	2	2	0	0.220	1	1	0	0.924
Own professional growth	**5**	**4**	**1***	**0.004**	**4**	**3**	**1***	**0.025**
Nonwork activities	**2**	**1**	**1**	**0.173**	**4**	**4**	**0**	**0.611**
Total hours	41	38	4*	0.024	41	41	0	0.912
Number of principals	50	50			50	50		

Source: Principal time use logs, 2015–2016 and 2016–2017 school years.

Note: Total hours calculated across all 20 rounds of principal log. For each 15-minute window throughout the day, principals indicated whether they spent time on each activity. Instead of filling in the precise number of minutes spent on each activity during each hour-long period of the school day, principals reported their time use in ranges (1 to 14 minutes, 15 to 29 minutes, 30 to 44 minutes, and 45 to 60 minutes). The estimates assume the number of hours the principal spent is the average of the upper and lower bounds for each time range. The difference between the treatment and control estimates may not equal the effect shown in the table due to rounding.

*Effect is statistically significant at the .05 level, two-tailed test.

Table C.7. Effects on duration of teacher observations (supplement to report text)

	Year 1				Year 2			
	Mean outcome				Mean outcome			
Duration (in minutes)	Treatment	Control	Effect	*p*- value	Treatment	Control	Effect	*p*-value
Principals' reports of duration of a typical observation	34	37	-4	0.311	26	33	-7	0.066
Teachers' reports of duration of a typical observation	23	23	1	0.395	25	22	3	0.179
Number of principals	38	38			41	41		
Number of teachers	504	548			496	534		

Source: Principal survey and teacher survey, spring 2016 and spring 2017.

Note: The difference between the treatment and control estimates may not equal the effect shown in the table due to rounding.

*Effect is statistically significant at the .05 level, two-tailed test.

Table C.8. Effects on principals' instructional leadership practices (supplement to report Figure 6, Table 1, and report text)

Respondent		Year 1 Mean outcome Treatment	Year 1 Mean outcome Control	Year 1 Effect	Year 1 *p*-value	Year 2 Mean outcome Treatment	Year 2 Mean outcome Control	Year 2 Effect	Year 2 *p*-value
Frequency of instructional support (number of times per year)									
Classroom observations	Principals	19	20	-1	0.818	22	19	3	0.571
Instructional support and feedback from principal[a]	Teachers	9	11	-2*	0.001	9	11	-2*	0.012
Quality of instructional support (1- to 4-point scale unless otherwise noted)									
Principal's teacher observation skills[b]	Principals	3.3	3.3	-0.1	0.631	3.3	3.3	0.0	0.702
Principal's competence in providing instructional supports[c]	Principals	3.5	3.5	0.0	0.698	3.5	3.5	0.0	1.000
Principal's competence in providing instructional supports[d]	Teachers	3.1	3.2	-0.1*	0.037	3.1	3.2	0.0	0.657
Usefulness of feedback received from principal[e]	Teachers	2.8	2.9	-0.1	0.279	2.9	2.9	0.0	0.806
Usefulness of all types of instructional support from principal[f]	Teachers	3.1	3.2	0.0	0.411	3.1	3.2	0.0	0.705
Interactions with principal about instruction were useful (percentages)	Teachers	64	67	-3	0.437	60	66	-6*	0.041
Instructional feedback between principal and someone else were consistent (percentages)	Teachers	73	78	-5*	0.030	76	76	0	0.975
Number of principals		39–45	39–45			45–46	45–46		
Number of teachers		366–547	425–586			367–532	413–571		

Sources: Principal and teacher surveys, spring 2016 and spring 2017.

Note: The difference between the treatment and control estimates may not equal the effect shown in the table due to rounding. See Appendix B, Table B.8 for information on the items included in each scale below.

[a] "Instructional support and feedback from principal" includes classroom observations, feedback on teaching, developing specific instructional practice goals, using data to determine progress and suggest specific teaching actions, and other instructional supports.

[b] "Principals' teacher observation skills" includes whether principals report that they focus their observations on specific areas or issues unique to the teachers' needs, record descriptions of specific things the teacher and students did or said during a classroom observation, and analyze data collected during classroom observations to identify trends in instructional practice. The scale indicates whether principals reported doing each item (1) not at all, (2) to a small extent, (3) to a moderate extent, or (4) to a great extent.

[c] "Principals' competence in providing instructional support (principals' reports)" includes whether principals feel that they know what effective teaching looks like; feel competent helping teachers identify their areas of instructional practice that need improvement; know how to give teachers' feedback on their instruction that provides them with actionable steps for improvement; and know where to find resources to support teacher instructional practice outside of their areas of expertise. The scale indicates whether principals (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

[d] "Principals' competence in providing instructional support (teachers' reports)" includes whether teachers feel that principals know what effective teaching looks like; directly work with teachers to improve instruction; communicate clear standards for student learning and expectations for teacher performance; and encourage teachers to use what they learn from professional development, resources on teaching, and each other to improve instruction. The scale indicates whether teachers (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

[e] "Usefulness of feedback received from principal" includes whether teachers feel that the feedback addressed pressing issues in their classroom, provided them with actionable steps for improvement, and helped them identify areas of instructional practice in which they need improvement. The scale indicates whether teachers felt the feedback met certain criteria (1) not at all, (2) to a small extent, (3) to a moderate extent, or (4) to a great extent.

[f] "Usefulness of all types of instructional support from principal" includes whether teachers feel that the following types of instructional support were useful: classroom observations, feedback on teaching, developing specific instructional practice goals, using data to determine progress and suggest specific teaching actions, and other instructional supports. The scale indicates whether teachers felt the instructional support was (1) not very useful, (2) somewhat useful, (3) moderately useful, or (4) very useful.

*Effect is statistically significant at the .05 level, two-tailed test.

Table C.9. Effects on teacher-reported principals' instructional leadership practices, by principal experience level and school average baseline achievement (supplement to report text)

Outcome (units)	Year	Overall effect	Principal experience level		Baseline achievement		
			More experienced	Less experienced	High	Medium	Low
Frequency of instructional support (number of times per year)							
Instructional support and feedback from principal[a]	Year 1	-2*	-1	-4*	-2*	-3*	1
	Year 2	-2*	0	-6*	-1	-2	-1
Quality of instructional support (1- to 4-point scale unless otherwise noted)							
Principal's competence in providing instructional supports[b]	Year 1	-0.1*	-0.1	-0.1	0.0	-0.2	0.0
	Year 2	0.0	0.0	-0.2	0.0	-0.1	0.0
Usefulness of feedback received from principal[c]	Year 1	-0.1	-0.1	-0.1	-0.2	-0.1	0.1
	Year 2	0.0	0.0	-0.1	-0.2	0.0	0.1
Usefulness of all types of instructional support from principal[d]	Year 1	0.0	0.0	-0.1	-0.2	0.0	-0.1
	Year 2	0.0	0.1	-0.3*	0.0	-0.1	0.0
Interactions with principal about instruction were useful (percentages)	Year 1	-3	-2	-3	-1	-9	8
	Year 2	-6*	-1	-17*	-6	-5	-10
Instructional feedback between principal and someone else was consistent (percentages)	Year 1	-5*	-4	-9*	-6	-6	-2
	Year 2	0	3	-5	-1	0	-1
Number of teachers		745–1,135	545–842	200–293	189–304	353–539	203–297

Source: Teacher surveys, spring 2016 and spring 2017.

Note: Low baseline achievement is defined as schools with average student achievement in the bottom 25 percent of the sample, medium baseline achievement is defined as schools with average student achievement in the middle 50 percent, and high baseline achievement is defined as schools with average student achievement in the top 25 percent. Less experienced principals are those in their first three years as a principal. More experienced principals are those with three or more years of experience as a principal. See Appendix B, Table B.8 for information on the items included in each scale below.

[a] "Instructional support and feedback from principal" includes classroom observations, feedback on teaching, developing specific instructional practice goals, using data to determine progress and suggest specific teaching actions, and other instructional supports.

[b] "Principals' competence in providing instructional support" includes whether teachers feel that principals know what effective teaching looks like; directly work with teachers to improve instruction; communicate clear standards for student learning and expectations for teacher performance; and encourage teachers to use what they learn from professional development, resources on teaching, and each other to improve instruction. The scale indicates whether teachers (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

[c] "Usefulness of feedback received from principal" includes whether teachers feel that the feedback addressed pressing issues in their classroom, provided them with actionable steps for improvement, and helped them identify areas of instructional practice in which they need improvement. The scale indicates whether teachers felt the feedback met certain criteria (1) not at all, (2) to a small extent, (3) to a moderate extent, or (4) to a great extent.

[d] "Usefulness of all types of instructional support from principal" includes whether teachers feel that the following types of instructional support were useful: classroom observations, feedback on teaching, developing specific instructional practice goals, using data to determine progress and suggest specific teaching actions, and other instructional supports. The scale indicates whether teachers felt the instructional support was (1) not very useful, (2) somewhat useful, (3) moderately useful, or (4) very useful.

*Effect is statistically significant at the .05 level, two-tailed test.

Table C.10. Effects on principals' human capital management and organizational leadership practices (supplement to report text)

Respondent		Year 1 Mean outcome Treatment	Year 1 Mean outcome Control	Year 1 Effect	Year 1 *p*-value	Year 2 Mean outcome Treatment	Year 2 Mean outcome Control	Year 2 Effect	Year 2 *p*-value
Human capital management (number of times per year unless otherwise noted)									
Frequency with which principal arranged professional development for teachers[a]	Principals	15	14	2	0.480	15	15	0	0.979
Principal uses data (teacher evaluations, teacher observations, and student performance data) to determine content of teacher professional development (percentage)	Principals	>78	>93	-16*	0.018	91	87	4	0.533
Hours of formal professional development teachers received	Teachers	45	46	-1	0.857	50	42	8*	0.013
Organizational leadership (1- to 4-point scale unless otherwise noted)									
Coherence of school improvement plan[b]	Principals	3.3	3.2	0.1	0.525	3.4	3.3	0.1	0.182
Coherence of school improvement plan[c]	Teachers	3.0	3.0	-0.1	0.091	3.0	3.0	0.0	0.571
Frequency of communication about school improvement[d] (number of times per year)	Principals	16	15	1	0.808	20	16	4	0.090
Frequency of communication about school improvement[e] (number of times per year)	Teachers	12	16	-4*	0.000	12	15	-3*	0.009
Number of principals		45	45			46	46		
Number of teachers		491–541	538–585			489–528	533–570		

Sources: Principal and teacher surveys, spring 2016 and spring 2017.

Note: The difference between the treatment and control estimates may not equal the effect shown in the table due to rounding. See Appendix B, Table B.8 for information on the items included in each scale below.

[a] "Frequency with which principal arranged professional development for teachers" includes helping a teacher locate formal professional development to support his or her goals, arranging an informal learning opportunity to support a teacher's growth, connecting a teacher to a content expert, and connecting a teacher to a network of teachers formed specifically for the professional development of teachers.

[b] "Coherence of school improvement plan (principals' reports)" includes whether principals felt that the administration collaborated with teachers to shape the plans in the school. Plans for improvement in the school included indicators to measure progress toward goals, aligned with evidence on teachers' or students' performance; were consistent with teachers' goals for individual growth; and clearly outlined steps that teachers should take to improve their teaching. The scale indicates whether principals (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

[c] "Coherence of school improvement plan (teachers' reports)" includes whether teachers felt that the administration collaborated with teachers to shape the plans in the school. Plans for improvement in the school included indicators to measure progress toward goals, aligned with evidence on teachers' or students' performance; were consistent with teachers' goals for individual growth; and clearly outlined steps that teachers should take to improve their teaching. The scale indicates whether teachers (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

[d] "Frequency of principal's communication around school improvement (principals' reports)" includes the principal communicating goals for improving instructional quality in school to teachers or other school staff, updating staff on progress toward the school vision or goals for improvement, incorporating a clear vision for the school into regular communications, and delegating these actions surrounding school culture and vision to another member of the staff.

[e] "Frequency of principal's communication around school improvement (teachers' reports)" includes the principal discussing his or her goals for improving the school's instructional quality with teachers, communicating progress toward goals for improving the school's instructional quality to teachers, and communicating a clear vision for the school's instructional quality through his or her regular communications.

*Effect is statistically significant at the .05 level, two-tailed test.

< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics[xiv] statistical standards, but that the percentage is less than or greater than the number following the < or > symbol.[xv]

3. Implementation of the study's principal professional development program and the overall professional development that principals received from the study and other sources

This section presents additional details on findings from the report on implementation of the study's principal professional development program and the overall professional development that principals received. Figure 9 in the report shows that the program increased the average number of hours of professional development that principals received. Table C.11 presents the estimated effects on overall hours of professional development and *p*-values. Table C.12 shows that the program increased the percentage of principals who received professional development on content related to instructional feedback, classroom observations, and communication around school improvement goals.

Table C.11. Amount of professional development and other supports that principals received (supplement to report Figure 9)

	Year 1				Year 2			
	Mean outcome				Mean outcome			
Mean hours received	Treatment	Control	Effect	*p*-value	Treatment	Control	Effect	*p*-value
Formal degree program or university courses (online or in person)	19	7	12	0.180	12	4	8	0.101
Formal group learning sessions, such as workshops, conferences, or seminars	107	70	37*	0.035	67	68	-1	0.963
One-on-one development opportunities, such as leadership mentoring or coaching	31	11	20*	0.000	58	19	39*	0.043
Other development opportunities, such as participating in professional development for teachers or a professional learning community	62	32	30*	0.039	40	41	-1	0.897
Total hours across all types of development	220	120	100*	0.002	178	132	45	0.275
Number of principals	26–45	26–45			29–46	29–46		

Source: Principal survey, spring 2016 and spring 2017.

Note: For Year 1, questions refer to professional development and supports received since September 1, 2015. For Year 2, questions refer to professional development and supports received since September 1, 2016. We counted principals who did not receive the support at all as receiving zero hours.

*Effect is statistically significant at the .05 level, two-tailed test.

< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics[xvi] statistical standards, but that the percentage is less than or greater than the number following the < or > symbol.[xvii]

Table C.12. Content of supports that principals received (supplement to report text)

Percentage of principals reporting receiving support in each of the following content areas:	Year 1				Year 2			
	Mean percentage				Mean percentage			
	Treatment	Control	Effect	*p*-value	Treatment	Control	Effect	*p*-value
Managing school staff (such as hiring and promoting staff, assigning teachers to grades and students, or designing professional development for staff)	73	64	9	0.323	64	64	0	1.000
Observing classroom instruction	>93	>71	22*	0.003	91	71	20*	0.011
Providing feedback to teachers on their instruction	100	69	31*	0.000	91	71	20*	0.011
Instructional practices or the curriculum being taught in their school	>93	>93	2	0.323	>93	>82	11	0.058
Setting and communicating school improvement goals or progress toward school improvement	>93	>73	20*	0.005	84	78	7	0.411
Community and parent outreach; student affairs; or school operations, finances, and administration	60	67	-7	0.519	67	60	7	0.519
Number of principals	45	45			45	45		

Source: Principal survey, spring 2016 and spring 2017.

Note: For Year 1, questions refer to professional development and supports received since September 1, 2015. For Year 2, questions refer to professional development and supports received since September 1, 2016.

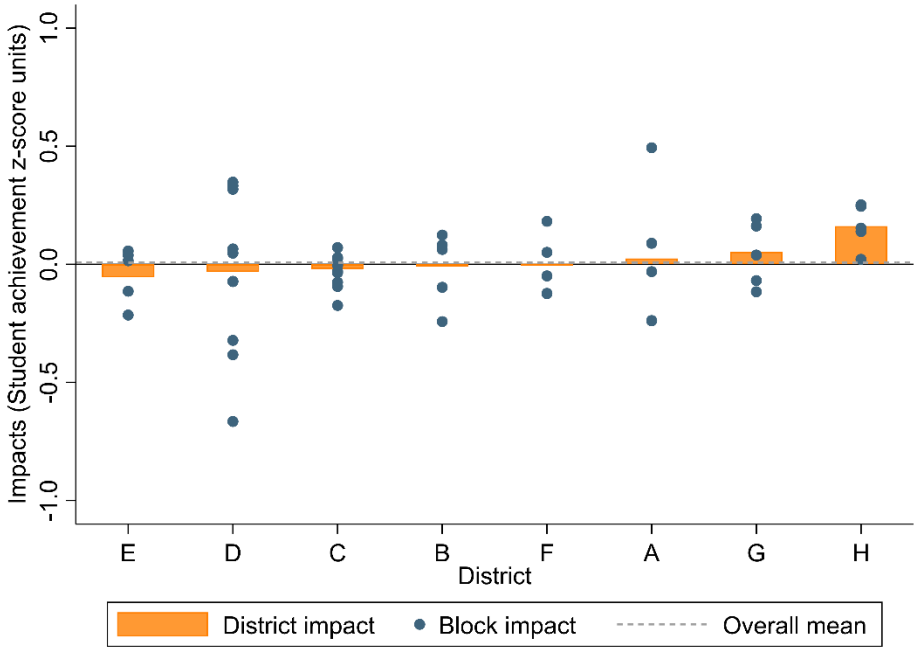
*Effect is statistically significant at the .05 level, two-tailed test.

< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics[xviii] statistical standards, but that the percentage is less than or greater than the number following the < or > symbol.[xix]

4. Effects on student achievement by district and blocks of schools

In this section, we present findings on the effects of the professional development program on student achievement by district and random assignment block (Figures C.1–C.6). Estimated effects did not vary to a statistically significant degree across districts. However, estimated effects did vary across blocks of schools by a statistically significant degree and a substantial magnitude in each subject and year. For example, in English Language Arts in Year 1, the effects of the program ranged from -0.67 to 0.49, and without regard to statistical significance, effects were positive (greater than 0.05) in 19 blocks, negative (less than -0.05) in 18 blocks, and close to zero (within 0.05 of zero) in 13 blocks.

Figure C.1. Effects on students' English language arts achievement in Year 1, by district and random assignment block

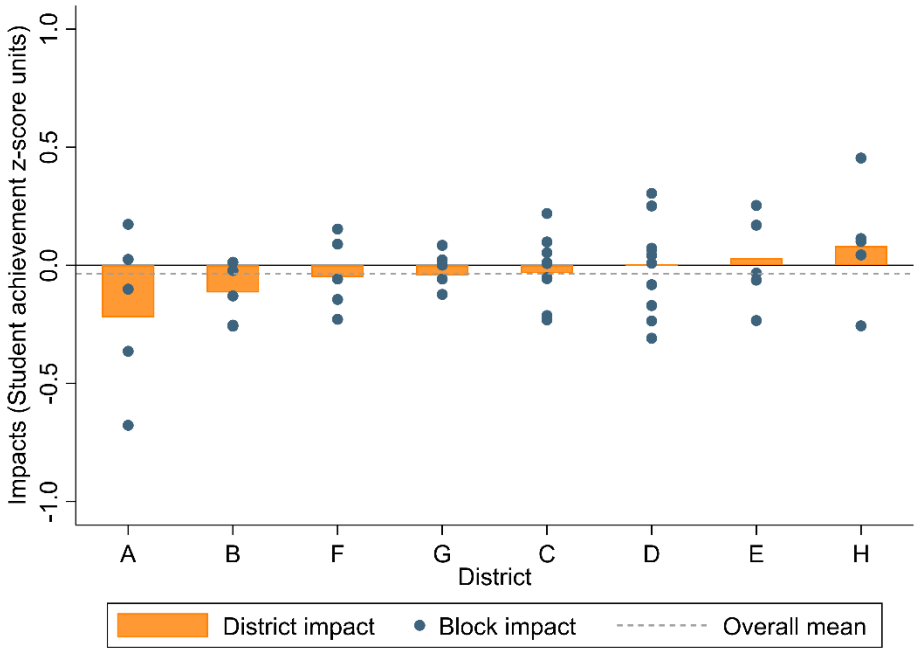


Source: Student administrative data (23,299 students).

Note: A Monte Carlo permutation test of the null hypothesis that effects do not vary across districts has a *p*-value greater than .05. For the mixed model used to estimate block-level heterogeneity in effects, a likelihood ratio test of the null hypothesis that effects do not vary across blocks has a *p*-value less than .05.

Figure reads: In District E, the professional development program lowered average student English language arts achievement by 0.06 student z-score units after the first year of implementation.

Figure C.2. Effects on students' math achievement in Year 1, by district and random assignment block

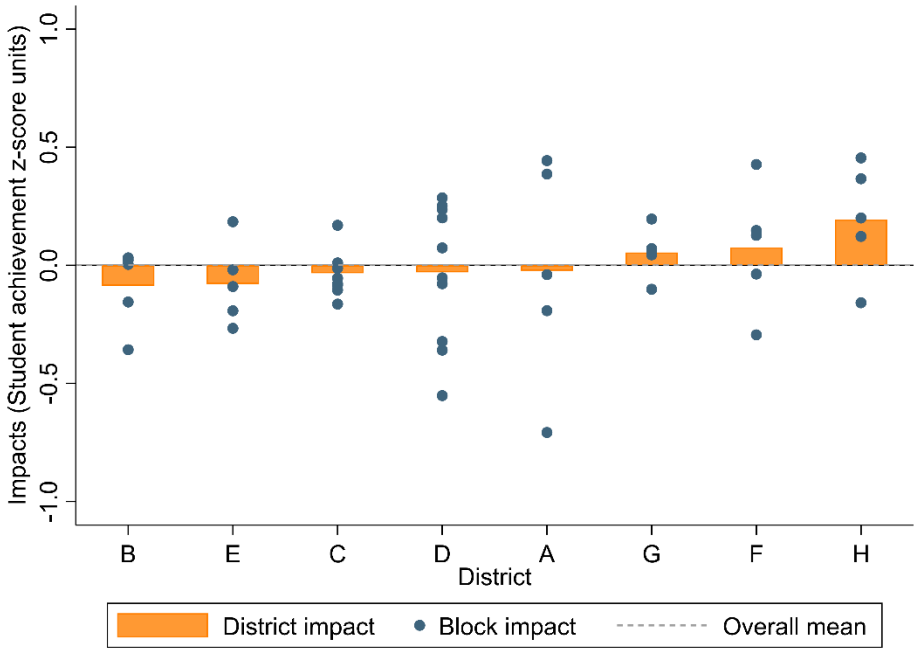


Source: Student administrative data (23,923 students).

Note: A Monte Carlo permutation test of the null hypothesis that effects do not vary across districts has a *p*-value greater than .05. For the mixed model used to estimate block-level heterogeneity in effects, a likelihood ratio test of the null hypothesis that effects do not vary across blocks has a *p*-value less than .05.

Figure reads: In District A, the professional development program lowered average student math achievement by 0.22 student z-score units after the first year of implementation.

Figure C.3. Effects on students' English language arts achievement in Year 2, by district and random assignment block

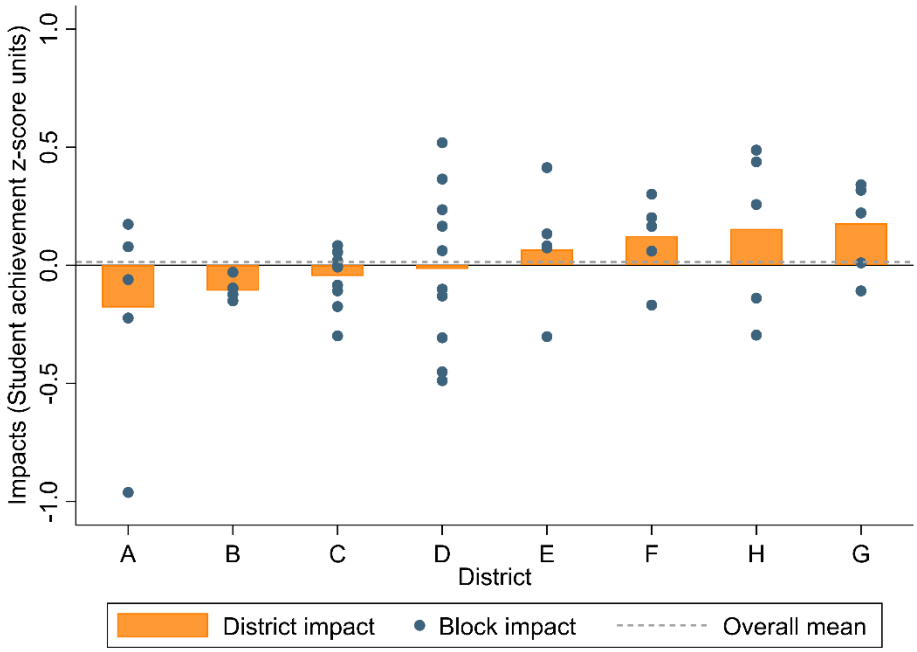


Source: Student administrative data (19,448 students).

Note: A Monte Carlo permutation test of the null hypothesis that effects do not vary across districts has a *p*-value greater than .05. For the mixed model used to estimate block-level heterogeneity in effects, a likelihood ratio test of the null hypothesis that effects do not vary across blocks has a *p*-value less than .05.

Figure reads: In District B, the professional development program lowered average student English language arts achievement by 0.09 student z-score units after the second year of implementation.

Figure C.4. Effects on students' math achievement in Year 2, by district and random assignment block

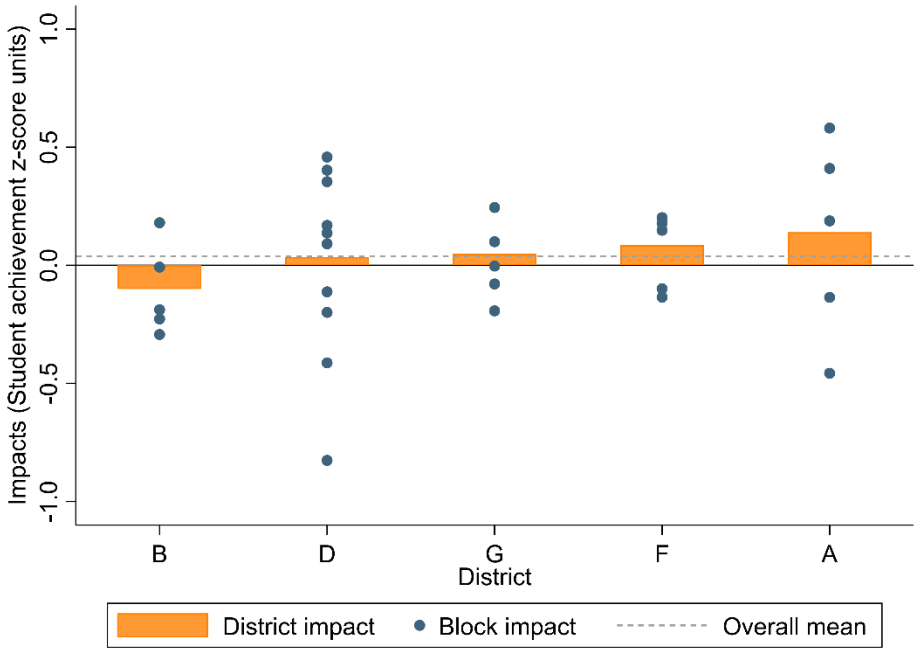


Source: Student administrative data (19,792 students).

Note: A Monte Carlo permutation test of the null hypothesis that effects do not vary across districts has a *p*-value greater than .05. For the mixed model used to estimate block-level heterogeneity in effects, a likelihood ratio test of the null hypothesis that effects do not vary across blocks has a *p*-value less than .05.

Figure reads: In District A, the professional development program lowered average student math achievement by 0.18 student z-score units after the second year of implementation.

Figure C.5. Effects on students' English language arts achievement in Year 3, by district and random assignment block

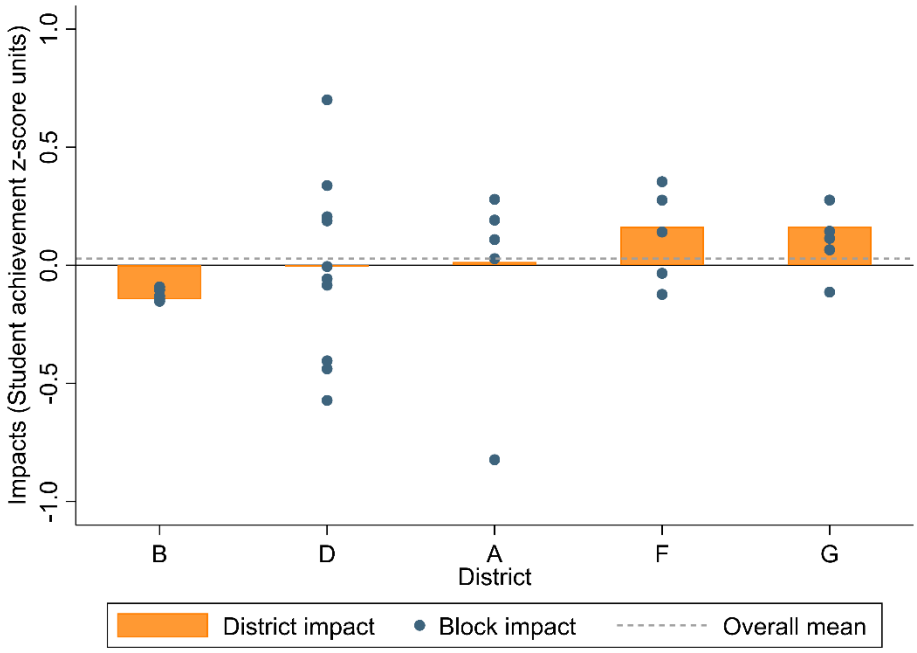


Source: Student administrative data (9,956 students).

Note: A Monte Carlo permutation test of the null hypothesis that effects do not vary across districts has a *p*-value greater than .05. For the mixed model used to estimate block-level heterogeneity in effects, a likelihood ratio test of the null hypothesis that effects do not vary across blocks has a *p*-value less than .05.

Figure reads: In District B, the professional development program lowered average student English language arts achievement by 0.10 student z-score units a year after implementation was complete.

Figure C.6. Effects on students' math achievement in Year 3, by district and random assignment block



Source: Student administrative data (10,390 students).

Note: A Monte Carlo permutation test of the null hypothesis that effects do not vary across districts has a *p*-value greater than .05. For the mixed model used to estimate block-level heterogeneity in effects, a likelihood ratio test of the null hypothesis that effects do not vary across blocks has a p-value less than .05.

Figure reads: In District B, the professional development program lowered average student math achievement by 0.14 student z-score units a year after implementation was complete.

5. Relationship between characteristics of the professional development program and its effects on principal practices and student achievement

This section presents findings on the relationship between the program's effects on all the principal practices examined in the study and its effects on student achievement, as well as on the characteristics of the professional development program and its effects on principal practices.

We also examined the relationship between effects on principals' practices and effects on student achievement. Table C.13 shows that effects on 7 of the 8 instructional leadership practices examined were positively associated with effects on student achievement in Year 2, and 3 of these associations were statistically significant. Table C.14 shows that none of the effects on the three human capital management practices examined was significantly associated with the program's effects on student achievement in either year. Table C.15 shows that effects on two of the four organizational leadership practices examined (teachers' and principals' reports of the coherence of the school improvement plan) were positively associated with the effects on student achievement in Year 2.

To examine how characteristics of the program related to principal's practices, we examined the relationship between several characteristics of the program and the effects on the four teacher-reported principal practices that were significantly associated with effects on student achievement. Table C.16 examines the correlations between effects on these four practices and two characteristics of the program: the percentage of coaching time spent on hands-on activities and the percentage of coaching time spent developing and implementing plans to address specific problems. It shows that the percentage of coaching time spent developing and implementing plans to address specific problems was not significantly associated with any of these four practices in Year 2. Table C.17 examines how effects on these four practices differ across different groups of principals and schools, defined by key characteristics of the coaching (coach experience level, focus of coaching on instructional leadership, and principals' completion of coach-assigned activities). It shows that

the program had few positive effects even among principals with more experienced coaches, principals whose coaching had a greater focus on instructional leadership, or those who were most engaged in the program (as measured by completion of coach-assigned activities).

Table C.13. Correlations between the professional development program's effects on principals' instructional leadership practices and its effects on student achievement (supplement to report text)

	Math		English language arts	
Instructional leadership practices	Year 1 Correlation (*p*-value)	Year 2 Correlation (*p*-value)	Year 1 Correlation (*p*-value)	Year 2 Correlation (*p*-value)
Teachers' reports				
Frequency of instructional support and feedback from principal[a]	0.05 (0.707)	0.34* (0.016)	-0.05 (0.718)	0.28* (0.047)
Principals' competence in providing instructional support[b]	0.04 (0.762)	0.30* (0.033)	-0.04 (0.778)	0.30* (0.033)
Interactions with principal about instruction were useful (percentage)	0.15 (0.287)	0.28* (0.049)	0.09 (0.539)	0.26 (0.065)
Instructional feedback from principal and others was consistent (percentage)	-0.02 (0.869)	0.09 (0.540)	-0.13 (0.354)	0.14 (0.329)
Usefulness of feedback received from principal[c]	0.05 (0.722)	0.22 (0.129)	-0.08 (0.568)	0.24 (0.096)
Usefulness of all types of instructional support from principal[d]	0.13 (0.364)	0.17 (0.232)	-0.02 (0.910)	0.14 (0.330)
Principals' reports				
Number of teacher observations conducted	-0.03 (0.849)	-0.06 (0.706)	0.12 (0.456)	-0.09 (0.539)
Principal's teacher observation skills[e]	-0.13 (0.414)	0.09 (0.552)	0.06 (0.712)	0.22 (0.138)
Number of random assignment blocks	39–50	45–50	39–50	45–50

Source: Teacher survey, principal survey, and administrative student records for 2015–2016 and 2016–2017 school years.

[a] "Frequency of instructional support and feedback from principal" includes classroom observations, feedback on teaching, developing specific instructional practice goals, using data to determine progress and suggest specific teaching actions, and other instructional supports.

[b] "Principals' competence in providing instructional support" includes whether teachers feel that principals know what effective teaching looks like; work directly with teachers to improve instruction; communicate clear standards for student learning and expectations for teacher performance; and encourage teachers to use what they learn from professional development, resources on teaching, and each other to improve instruction. The scale indicates whether teachers (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

[c] "Usefulness of feedback received from principal" includes whether teachers feel that the feedback addressed pressing issues in their classroom, provided them with actionable steps for improvement, and helped them identify areas of instructional practice in which they need improvement. The scale indicates whether teachers felt the feedback met certain criteria (1) not at all, (2) to a small extent, (3) to a moderate extent, or (4) to a great extent.

[d] "Usefulness of all types of instructional support from principal" includes whether teachers feel that the following types of instructional support were useful: classroom observations, feedback on teaching, developing specific instructional practice goals, using data to determine progress and suggest specific teaching actions, and other instructional supports. The scale indicates whether teachers felt the instructional support was (1) not very useful, (2) somewhat useful, (3) moderately useful, or (4) very useful.

[e] "Principals' teacher observation skills" includes whether principals report that they focus their observations on specific areas or issues unique to the teachers' needs, record descriptions of specific things the teacher and students did or said during a classroom observation, and analyze data collected during classroom observations to identify trends in instructional practice. The scale indicates whether principals reported doing each item (1) not at all, (2) to a small extent, (3) to a moderate extent, or (4) to a great extent.

*Correlation between block-level effects is statistically significant at the .05 level, two-tailed test.

Table C.14. Correlations between the professional development program's effects on principals' human capital management practices and its effects on student achievement (supplement to report text)

Human capital management practices	Math		English language arts	
	Year 1 Correlation (*p*-value)	Year 2 Correlation (*p*-value)	Year 1 Correlation (*p*-value)	Year 2 Correlation (*p*-value)
Teachers' reports				
Hours of formal professional development teacher received	-0.05 (0.715)	0.06 (0.680)	0.08 (0.596)	0.09 (0.549)
Principals' reports				
Frequency with which principal arranged professional development for teachers[a]	-0.29 (0.055)	0.07 (0.652)	-0.18 (0.245)	-0.04 (0.803)
Principal uses data to determine content of teacher professional development (percentage)	-0.12 (0.437)	0.12 (0.426)	0.13 (0.378)	0.12 (0.426)
Number of random assignment blocks	45–50	46–50	45–50	46–50

Source: Teacher survey, principal survey, and administrative student records for 2015–2016 and 2016–2017 school years.

[a] "Frequency with which principal arranged professional development for teachers" includes helping a teacher locate formal professional development to support his or her goals, arranging an informal learning opportunity to support a teacher's growth, connecting a teacher to a content expert, and connecting a teacher to a network of teachers formed specifically for the professional development of teachers.

*Correlation between block-level effects is statistically significant at the .05 level, two-tailed test.

Table C.15. Correlations between the professional development program's effects on principals' organizational leadership practices and its effects on student achievement (supplement to report text)

Organizational leadership practices	Math		English language arts	
	Year 1 Correlation (*p*-value)	Year 2 Correlation (*p*-value)	Year 1 Correlation (*p*-value)	Year 2 Correlation (*p*-value)
Teachers' reports				
Coherence of school improvement plan[a]	0.01 (0.971)	0.41* (0.003)	-0.09 (0.540)	0.40* (0.004)
Frequency of communication about school improvement[b]	-0.07 (0.607)	0.08 (0.600)	-0.09 (0.555)	0.13 (0.386)
Principals' reports				
Coherence of school improvement plan[c]	0.00 (0.976)	0.22 (0.144)	0.14 (0.350)	0.34* (0.021)
Frequency of communication about school improvement[d]	-0.17 (0.252)	-0.14 (0.339)	-0.07 (0.656)	-0.19 (0.212)
Number of random assignment blocks	45–50	46–50	45–50	46–50

Source: Teacher survey, principal survey, and administrative student records for 2015–2016 and 2016–2017 school years.

[a] "Coherence of school improvement plan" (teachers' reports) includes whether teachers felt that the administration collaborated with teachers to shape the plans in the school. Plans for improvement in the school included indicators to measure progress toward goals, aligned with evidence on teachers' or students' performance; were consistent with teachers' goals for individual growth; and clearly outlined steps that teachers should take to improve their teaching. The scale indicates whether teachers (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

[b] "Frequency of communication about school improvement" (teachers' reports) includes the principal discussing his or her goals for improving the school's instructional quality with teachers, communicating progress toward goals for improving the school's instructional quality to teachers, and communicating a clear vision for the school's instructional quality through his or her regular communications.

[c] "Coherence of school improvement plan" (principals' reports) includes whether principals felt that the administration collaborated with teachers to shape the plans in the school. Plans for improvement in the school included indicators to measure progress toward goals, aligned with evidence on teachers' or students' performance; were consistent with teachers' goals for individual growth; and clearly outlined steps that teachers should take

to improve their teaching. The scale indicates whether principals (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

[d] "Frequency of communication about school improvement" (principals' reports) includes the principal communicating goals for improving instructional quality in school to teachers or other school staff, updating staff on progress toward the school vision or goals for improvement, incorporating a clear vision for the school into regular communications, and delegating these actions surrounding school culture and vision to another member of the staff.

*Correlation between block-level effects is statistically significant at the .05 level, two-tailed test.

Table C.16. Correlations between characteristics of the professional development program and its effects on principals' leadership practices (supplement to report text)

	Percentage of coaching time spent on hands-on activities		Percentage of coaching time spent developing and implementing plans to address specific problems	
Principals' practices (teachers' reports)	Year 1 Correlation (*p*-value)	Year 2 Correlation (*p*-value)	Year 1 Correlation (*p*-value)	Year 2 Correlation (*p*-value)
Frequency of instructional support and feedback from principal[a]	-0.07 (0.63)	-0.18 (0.21)	0.11 (0.44)	0.24 (0.10)
Principals' competence in providing instructional support[b]	-0.09 (0.54)	-0.02 (0.92)	0.04 (0.76)	0.10 (0.49)
Interactions with principal about instruction were useful (percentage)	-0.09 (0.53)	0.02 (0.90)	-0.17 (0.25)	0.19 (0.20)
Coherence of school improvement plan[c]	-0.17 (0.24)	-0.06 (0.67)	0.10 (0.51)	0.18 (0.21)
Number of random assignment blocks	50	49	50	49

Sources: Coaching logs completed for each coaching session and teacher survey for 2015–2016 and 2016–2017 school years.

[a] "Frequency of instructional support and feedback from principal" includes classroom observations, feedback on teaching, developing specific instructional practice goals, using data to determine progress and suggest specific teaching actions, and other instructional supports.

[b] "Principals' competence in providing instructional support" includes whether teachers feel that principals know what effective teaching looks like; work directly with teachers to improve instruction; communicate clear standards for student learning and expectations for teacher performance; and encourage teachers to use what they learn from professional development, resources on teaching, and each other to improve instruction. The scale indicates whether teachers (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

[c] "Coherence of school improvement plan" includes whether teachers felt that the administration collaborated with teachers to shape the plans in the school. Plans for improvement in the school included indicators to measure progress toward goals, aligned with evidence on teachers' or students' performance; were consistent with teachers' goals for individual growth; and clearly outlined steps that teachers should take to improve their teaching. The scale indicates whether teachers (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

*Correlation between block-level effects is statistically significant at the .05 level, two-tailed test.

Table C.17. Effects on principals' leadership practices, by characteristics of the professional development program

Outcome (1- to 4-point scale unless otherwise noted)	Year	Overall effect	Coach experience		Focus of coaching on instructional leadership		Completion of coach-assigned activities	
			More experienced	Less experienced	Greater	Less	High	Low
Frequency of instructional support and feedback from principal[a] (number of times per year)	Year 1	-2*	0	-3*	-3*	-1	-1	-3*
	Year 2	-2*	0	-3*	1	-4*	-1	-3*
Principals' competence in providing instructional support[b]	Year 1	-0.1*	0.0	-0.2*	-0.1	-0.1*	-0.1	-0.1
	Year 2	0.0	0.0	-0.1	0.1	-0.2*	0.0	-0.1
Interactions with principal about instruction were useful (percentages)	Year 1	-3	-1	-3	-3	-3	-3	-3
	Year 2	-6*	-2	-9*	5	-15*	-4	-9*
Coherence of school improvement plan[c]	Year 1	-0.1	0.1*	-0.2*	-0.1	-0.1	0.0	-0.1*
	Year 2	0.0	0.0	-0.1	0.1	-0.2*	0.0	-0.1
Number of teachers		745–1135	316–510	429–628	348–513	379–601	380–595	365–540

Source: Teacher surveys, spring 2016 and spring 2017.

Note: Greater focus on instructional leadership means that principals spent 67 percent or more of coaching time on instructional leadership activities in the corresponding year. Lesser focus on instructional leadership means that principals spent less than 67 percent of coaching time on instructional leadership activities. High completion of coach-assigned activities means that principals completed more than the median number of coach-assigned activities in the corresponding year. Low completion of coach-assigned activities means the principal completed fewer than the median number of coach-assigned activities. More experienced coaches had at least three years of experience working for the Center for Educational Leadership. Less experienced coaches had fewer than three years of experience working for the Center for Educational Leadership. See Appendix B, Table B.8 for information on the items included in each scale below.

[a] "Frequency of instructional support and feedback from principal" includes classroom observations, feedback on teaching, developing specific instructional practice goals, using data to determine progress and suggest specific teaching actions, and other instructional supports.

[b] "Principals' competence in providing instructional support" includes whether teachers feel that principals know what effective teaching looks like; directly work with teachers to improve instruction; communicate clear standards for student learning and expectations for teacher performance; and encourage teachers to use what they learn from professional development, resources on teaching, and each other to improve instruction. The scale indicates whether teachers (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their principal.

[c] "Coherence of school improvement plan" includes whether teachers felt that the administration collaborated with teachers to shape the plans in the school. Plans for improvement in the school included indicators to measure progress toward goals, aligned with evidence on teachers' or students' performance; were consistent with teachers' goals for individual growth; and clearly outlined steps that teachers should take to improve their teaching. The scale indicates whether teachers (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

* Effect is statistically significant at the .05 level, two-tailed test.

B. Additional findings that are not in the report

This section includes additional findings that are not in the report. These findings are related to the (1) effects of the program on principals' practices and (2) implementation of the program and the overall professional development that principals received.

1. Effects on principals' practices

In this section, we present additional findings on the effects of the professional development program on principals' practices. To examine the effect of the program on principals' time use, we coded the number of minutes that principals indicated spending on each activity. Our main analysis in the report used the midpoint of the ranges that principals selected to indicate how much time they spent on an activity (1 to 14 minutes, 15 to 29 minutes, 30 to 44 minutes, and 45 minutes to an hour). To assess the sensitivity of the findings to using the midpoint for each range, we conducted sensitivity tests that used (1) the minimum for each range (a lower bound) and (2) the maximum for each range (an upper bound). Table C.18 shows that the findings for time use were not sensitive to using these alternate methods to code principals' time use.

The report showed that the program had few effects on the human capital management and organizational leadership practices that it covered, on average. Table C.19 shows that it also had few effects on these practices for subgroups of principals based on principals' years of experience or schools' baseline level of achievement.

Table C.18. Effects on principals' time use (lower and upper bounds)

	Lower bound				Upper bound			
	Year 1		Year 2		Year 1		Year 2	
Number of hours per week spent on:	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
Organizational leadership								
Student affairs	6	6	6	7	8	9	9	9
Administration	5	4	4	4	7	6	6	5
Other								
School improvement efforts	2	2	2	2	3	3	3	3
Community and parent outreach	2	2	2	2	3	3	3	3
Other work-related activities	1	1	1	1	1	1	1	1
Instructional leadership								
Evaluation	4	4	4	4	6	5	6	5
Teacher feedback	3	3	4	4	5	4	5	5
Curriculum	2	2	2	2	3	3	3*	3
Human capital management								
Recruiting teachers	2	1	2	2	3	2	2	3
Personnel policies	1	1	1	1	2	2	2	2
Own professional growth	**4***	**3**	**3**	**2**	**6***	**4**	**4***	**3**
Nonwork activities	**2**	**1**	**3**	**3**	**2**	**2**	**5**	**5**
Total hours	34*	31	34	34	49*	45	49	48
Number of principals	50	50	50	50	50	50	50	50

Source: Principal time use logs, 2015–2016 and 2016–2017 school years.

Note: Total hours calculated across all 20 rounds of principal log. For each 15-minute window throughout the day, principals indicated whether they spent time on each activity. Instead of filling in the precise number of minutes spent on each activity during each hour-long period of the school day, principals reported their time use in ranges (1 to 14 minutes, 15 to 29 minutes, 30 to 44 minutes, and 45 to 60 minutes). The lower-bound estimates assume the minimum possible time for each range, and the upper-bound estimates assume the maximum possible time for each range.

*Effect is statistically significant at the .05 level, two-tailed test.

Table C.19. Effects on teacher-reported principals' human capital management and organizational leadership practices, by principal experience level and school average baseline achievement

Outcome (units)	Year	Overall effect	Principal experience level		Baseline achievement		
			More experienced	Less experienced	High	Medium	Low
Human capital management (hours)							
Number of hours of formal professional development teachers received	Year 1	-1	0	-1	-8	1	7
	Year 2	8*	6	16*	13	4	9
Organizational leadership (1- to 4-point scale unless otherwise noted)							
Coherence of school improvement plan[a]	Year 1	-0.1	-0.1	-0.1	-0.1	-0.1	0.1
	Year 2	0.0	0.0	-0.2	0.0	-0.1	0.0
Frequency of principal's communication about school improvement[b] (number of times per year)	Year 1	-4*	-5*	-3	-3	-6*	-2
	Year 2	-3*	-2	-4*	-1	-5	-1
Number of teachers		745–1,135	545–842	200–293	189–304	353–539	203–297

Source: Teacher surveys, spring 2016 and spring 2017.

Note: Low baseline achievement is defined as schools with average student achievement in the bottom 25 percent of the sample, medium baseline achievement is defined as schools with average student achievement in the middle 50 percent, and high baseline achievement is defined as schools with average student achievement in the top 25 percent. Less experienced principals are those in their first three years as a principal. More experienced principals are those with three or more years of experience as a principal. See Appendix B, Table B.8 for information on the items included in each scale below.

[a] "Coherence of school improvement plan" includes whether teachers felt that the administration collaborated with teachers to shape the plans in the school; and plans for improvement in the school included indicators to measure progress toward goals, aligned with evidence on teachers' or students' performance, were consistent with teachers' goals for individual growth, and clearly outlined steps that teachers should take to improve their teaching. The scale indicates whether principals (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

[b] "Frequency of principal's communication around school improvement" includes the principal discussing his or her goals for improving the school's instructional quality with teachers, communicating progress toward goals for improving the school's instructional quality to teachers, and communicating a clear vision for the school's instructional quality through his or her regular communications.

*Effect is statistically significant at the .05 level, two-tailed test.

2. Implementation of the study's principal professional development program and the overall professional development that principals received

This section presents additional findings on implementation of the study's principal professional development program and the overall professional development that principals received. Tables C.20 and C.21 show that principals had positive views of all four components of the program, including the group trainings, summer institute, professional learning communities, and coaching. Table C.22 shows that the program improved principals' perceptions of the usefulness of the professional development they received in Year 1 but not in Year 2. Table C.23 shows that the program improved principals' perceptions of the professional development they received across several dimensions, including degree of alignment with their improvement goals and opportunities to improve aspects of their work.

Table C.20. Principals' perceptions of the summer institute, group training sessions, and professional learning community sessions, Year 1

Aspects of training	Percentage of principals who somewhat or strongly agreed			
	Summer institute[a]	Group training sessions[b]	Professional learning community sessions[c]	Overall experience in these three components[d]
Usefulness of content: the professional development …				
Was relevant to principal's professional growth needs	100.0	99.7	98.9	n.a.
Included new ideas, strategies, or information	100.0	98.5	95.0	n.a.
Included content that helped principal better manage time and resources	n.a.	n.a.	n.a.	83
Was useful	n.a.	n.a.	n.a.	>94
Was a good use of the principal's time	100.0	96.7	92.2	n.a.
Built on earlier sessions	n.a.	99.4	98.3	n.a.
Was well organized	100.0	98.8	97.8	
Actionable steps: the professional development …				
Gave principal a clear understanding of the immediate, specific actions he or she could take in response to the training	100.0	98.5	97.2	>94
Suggested specific actions that would improve principal's practice	100.0	99.1	98.3	n.a.
Suggested specific actions principal was likely to implement in his or her practice	100.0	99.1	97.8	100
Helped principal identify ways to measure progress toward school improvement goals	n.a.	n.a.	n.a.	89
Helped principal make more accurate assessments of teachers' performance	n.a.	n.a.	n.a.	>94
Number of principals	49	337	179	47

Sources: Session evaluation forms administered during 2015–2016 school year and principal survey, spring 2016.

Note: Responses were comparable for the subset of seven new principals who received group training in Year 2. Responses for Year 2 are not reported due to the small number of those principals.

[a] Principals completed a single evaluation form at the close of the four-day summer institute. Thus, the maximum number of responses is 50.

[b] Principals completed a separate evaluation form at the end of each group training session, except for Group Training Sessions 4 and 5, which took place on consecutive days; principals completed a single session evaluation form at the end of the combined session. The form asked principals to evaluate the training session they had just completed. Thus, the maximum number of responses is 350 (50 principals multiplied by 7 sessions).

[c] Principals completed a separate evaluation form at the end of each professional learning community session. Thus, the maximum number of responses is 200 (50 principals multiplied by 4 sessions).

[d] Principals received these questions once at the end of the school year. The form asked them to reflect on their overall experiences in all the trainings that were delivered to them in groups throughout the year. Thus, the maximum number of responses is 50.

n.a. = not applicable; question was not asked on this survey.

< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics[xx] statistical standards, but that the percentage is less than or greater than the number following the < or > symbol.[xxi]

Table C.21. Principals' perceptions of coaching

Aspects of coaching	Percentage of principals who somewhat or strongly agreed	
	Year 1	Year 2
Usefulness of content: the coaching sessions …		
Helped principal better manage his or her time and resources	83	81
Were useful	94	>94
Actionable steps: the coaching sessions …		
Gave principal a clear understanding of the immediate, specific actions he or she could take in response to the coaching	>94	100
Suggested specific actions the principal is likely to implement	94	>94
Helped principal identify ways to measure progress toward school improvement goals	89	94
Helped principal make more accurate assessments of teachers' performance	100	>94
Organization: the coaching session …		
Was well organized	100	94
Number of principals	47	47

Source: Principal survey, spring 2016 and spring 2017.

Note: For these items on the principal survey, principals reflected on their overall coaching experience at the end of the school year.

< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics[xxii] statistical standards, but that the percentage is less than or greater than the number following the < or > symbol.[xxiii]

Table C.22. Usefulness of professional development and other supports that principals received

Percentage of principals reporting this support was moderately or very useful	Year 1				Year 2			
	Mean outcome				Mean outcome			
	Treatment	Control	Effect	*p*-value	Treatment	Control	Effect	*p*-value
Formal group learning sessions, such as workshops, conferences, or seminars	93	77	16*	0.018	83	83	0	1.000
One-on-one development opportunities, such as leadership mentoring or coaching	>88	>77	12	0.265	90	76	14	0.212
Other development opportunities, such as participating in a professional development for teachers or a professional learning community	>93	>73	20*	0.019	81	81	0	1.000
Total hours across all types of development	220	120	100*	0.002	178	132	45	0.275
Number of principals	26–45	26–45			29–46	29–46		

Source: Principal survey, spring 2016 and spring 2017.

Note: For Year 1, questions refer to professional development and supports received since September 1, 2015. For Year 2, questions refer to professional development and supports received since September 1, 2016. The survey asked principals who did not receive each type of support to skip questions regarding usefulness of that support. Because only four principals reported participating in a formal degree program or university courses, we did not examine the reported usefulness of these supports.

* Effect is statistically significant at the .05 level, two-tailed test.

< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics[xxiv] statistical standards, but that the percentage is less than or greater than the number following the < or > symbol.[xxv]

Table C.23. Principals' perceptions of supports received

Percentage of principals reporting the supports they received during the given school year had the following characteristics to a great or moderate extent:	Year 1				Year 2			
	Mean percentage				Mean percentage			
	Treatment	Control	Effect	*p*-value	Treatment	Control	Effect	*p*-value
Degree of alignment								
Were aligned with their own improvement goals	>93	>73	20*	0.011	>93	>70	24*	0.002
Were aligned with schoolwide goals or plans for improvement	93	89	4	0.486	100	89	11*	0.024
Were aligned with evidence from their principal evaluations or leadership framework	80	69	11	0.280	93	62	31*	0.001
Suggesting specific actions								
Sent a consistent message about specific actions they can take to facilitate school improvement	93	69	24*	0.006	>93	>61	33*	0.000
Engaged them in setting specific goals to improve their school leadership	91	60	31*	0.002	89	63	26*	0.004
Gave them specific information about steps they can take to improve teaching in their schools	>93	>80	13	0.057	93	76	17*	0.019
Improvement and innovation in principals' practice								
Provided them opportunities to improve aspects of their work	91	71	20*	0.027	93	78	16*	0.033
Helped them pay closer attention to particular things they were doing in their work	>93	>69	24*	0.003	89	70	20*	0.018
Led them to try new things in their practice or work	>93	>62	31*	0.001	91	63	28*	0.000
Improving teachers' practices								
Helped them think about supports teachers need to improve their practice	>93	>84	9	0.103	100	76	24*	0.000
Taught them to gather and analyze evidence about the instructional quality in their school	>93	>67	27*	0.001	89	67	22*	0.003
Gave them specific information about providing effective feedback to teachers	93	67	27*	0.004	>93	>61	33*	0.000
Number of principals	45	45			45–46	45–46		

Source: Principal survey, spring 2016 and spring 2017.

Note: For Year 1, questions refer to professional development and supports received since September 1, 2015. For Year 2, questions refer to professional development and supports received since September 1, 2016.

* Effect is statistically significant at the .05 level, two-tailed test.

< or > indicates that we have withheld the exact percentage to protect respondents' confidentiality in accordance with National Center for Education Statistics[xxvi] statistical standards, but that the percentage is less than or greater than the number following the < or > symbol.[xxvii]

C. Supplemental information for systematic reviews

Systematic reviews of evidence on the effects of educational interventions often require specific types of information to evaluate the quality of a study. This section reports additional information that a systematic review might need to assess the quality of the findings for two key samples. First, it presents information for the main samples used to analyze principal and teacher retention and student achievement: principals and teachers who worked in study schools at baseline (the end of the school year before random assignment was conducted) and students who were enrolled in study schools at the beginning of Year 1. The study's principal professional development program likely would not have affected the types of students enrolled in study schools at the beginning of Year 1. However, to demonstrate that findings are similar for students enrolled in study schools before random assignment, this section also presents information for a second sample of students—those enrolled in study schools at baseline.

Main samples (principals and teachers who worked in study schools at baseline and students enrolled in study schools at the beginning of Year 1).

This section provides statistics for the main samples used to analyze principal and teacher retention and student achievement outcomes:

Attrition and missing outcome data. Table C.24 shows that there was no school attrition for principal retention, teacher retention, and student achievement outcomes, in districts that provided these data. (The teacher retention analysis omitted one district [with 20 schools] that could not provide data on teacher retention, and the student achievement analysis for Year 3 omitted three districts [with 40 schools] that could not provide these data.) Table C.25 shows that individual principals, teachers, and students also had low rates of missing outcome data in the districts that provided these data. The principal and teacher retention analyses include all principals and teachers who worked in study schools at baseline, and the student achievement analyses include about 90 percent or more of the students enrolled in study schools at the beginning of Year 1, for all three years of student achievement data.

Summary statistics and effects of the professional development program. Table C.26 presents the means, standard deviations, and estimates of the effects of the program. Because the student achievement models used *z*-scores, the statistics for student achievement outcomes are expressed in *z*-score units. (In the report, the study team converted the results to percentiles for easier interpretation.)

Alternate student sample (students enrolled in study schools at baseline). This section provides corresponding statistics for the sample of students enrolled in study schools at baseline:

Missing outcome data. Table C.27 shows that individual students in this sample had low rates of missing outcome data, with the student achievement analyses including about 90 percent or more of the students enrolled in study schools at baseline.

Summary statistics and effects of the program. Table C.28 presents the means, standard deviations, and estimates of the effects of the program for this alternate sample. Findings resemble those for students enrolled in study schools at the beginning of Year 1.

Table C.24. Attrition of study schools, by outcome

Outcome	Number of schools randomly assigned			Number of schools that remained in analysis sample			Attrition (percentages)			
	Overall	Treatment	Control	Overall	Treatment	Control	Overall	Treatment	Control	Differential
Principal retention										
Year 1	100	50	50	100	50	50	0	0	0	0
Year 2	100	50	50	100	50	50	0	0	0	0
Teacher retention[a]										
Year 1	80	40	40	80	40	40	0	0	0	0
Year 2	80	40	40	80	40	40	0	0	0	0
Student achievement										
Year 1	100	50	50	100	50	50	0	0	0	0
Year 2	100	50	50	100	50	50	0	0	0	0
Year 3[b]	60	30	30	60	30	30	0	0	0	0

Sources: Administrative educator records for the 2014–2015 through 2017–2018 school years and administrative student records for the 2014–2015 through 2016–2017 school years.

[a] One of the eight districts (with 20 schools) could not provide data on teacher retention and are excluded from the teacher retention tabulations.

[b] Three of the eight districts (with 40 schools) could not provide data on student achievement in Year 3 and are excluded from these tabulations.

Table C.25. Missing outcome data, by outcome and year

Outcome and year	Number in original sample			Number that remained in analysis sample			Missing outcome data (percentages)			
	Overall	Treatment	Control	Overall	Treatment	Control	Overall	Treatment	Control	Differential
Principal retention										
Baseline to Year 1	100	50	50	100	50	50	0	0	0	0
Baseline to Year 2	100	50	50	100	50	50	0	0	0	0
Baseline to Year 3	100	50	50	100	50	50	0	0	0	0
Teacher retention										
Baseline to Year 1	3,012	1,496	1,516	3,012	1,496	1,516	0	0	0	0
Baseline to Year 2	3,012	1,496	1,516	3,012	1,496	1,516	0	0	0	0
Baseline to Year 3	3,012	1,496	1,516	3,012	1,496	1,516	0	0	0	0
Student achievement										
English language arts, Year 1	26,089	12,643	13,446	23,299	11,423	11,876	11	10	12	-2
Math, Year 1	26,089	12,643	13,446	23,923	11,725	12,198	8	7	9	-2
English language arts, Year 2	21,409	10,679	10,730	19,448	9,724	9,724	9	9	9	0
Math, Year 2	21,409	10,679	10,730	19,792	9,989	9,803	8	6	9	-2
English language arts, Year 3	10,622	5,271	5,351	9,971	4,907	5,064	6	7	5	2
Math, Year 3	10,622	5,271	5,351	10,400	5,149	5,251	2	2	2	0

Sources: Administrative educator records for the 2014–2015 through 2016–2017 school years and administrative student records for the 2014–2015 through 2017–2018 school years.

Note: Students in the Year 1 sample are those enrolled in grades 3–5 in study schools at the beginning of Year 1. Students in the Year 2 sample are those enrolled in grades 2–4 in study schools at the beginning of Year 1. Students in the Year 3 sample are those enrolled in grades 1–3 in study schools at the beginning of Year 1. The difference between the percentages of the treatment and control group that are missing outcomes may not equal the differential rate shown in the table due to rounding.

Table C.26. Effects of professional development program, by outcome and year

Outcome and year	Treatment group		Control group		Effect	p-value	Number of individuals	Number of schools
	Adjusted mean	Unadjusted standard deviation	Adjusted mean	Unadjusted standard deviation				
Principal retention (percentages)								
Baseline to Year 1	84	37	86	35	-2	0.742	100	100
Baseline to Year 2	74	44	64	48	10	0.165	100	100
Baseline to Year 3	56	50	54	50	2	0.821	100	100
Teacher retention (percentages)								
Baseline to Year 1	79	41	83	38	-3*	0.008	3,012	80
Baseline to Year 2	59	49	59	49	0	0.967	3,012	80
Baseline to Year 3	55	50	52	50	3	0.054	3,012	80
Student achievement (*z*-score units)								
English language arts, Year 1	-0.24	0.97	-0.25	0.95	0.01	0.706	23,299	100
Math, Year 1	-0.29	0.96	-0.26	0.93	-0.04	0.073	23,923	100
English language arts, Year 2	-0.24	0.97	-0.24	0.94	0.00	0.964	19,448	100
Math, Year 2	-0.26	0.97	-0.28	0.94	0.01	0.599	19,792	100
English language arts, Year 3	-0.21	1.00	-0.25	0.98	0.04	0.299	9,971	60
Math, Year 3	-0.26	1.00	-0.29	0.97	0.03	0.503	10,400	60

Sources: Administrative educator records for the 2014–2015 through 2016–2017 school years and administrative student records for the 2014–2015 through 2017–2018 school years.

Note: Students in the Year 1 sample are those enrolled in grades 3–5 in study schools at the beginning of Year 1. Students in the Year 2 sample are those enrolled in grades 2–4 in study schools at the beginning of Year 1. Students in the Year 3 sample are those enrolled in grades 1–3 in study schools at the beginning of Year 1. Means were adjusted using the regression model described in Appendix B. Unadjusted standard deviations were the standard deviations across principals for principal retention outcomes, across teachers for teacher retention outcomes, and across students for student achievement outcomes. The difference between the treatment and control adjusted means may not equal the effect shown in the table due to rounding.

* Effect is statistically significant at the .05 level, two-tailed test.

Table C.27. Missing student achievement outcome data, alternate student sample

	Number in original sample			Final number that remained in analysis sample			Missing outcome data (percentages)			
Student achievement	Overall	Treatment	Control	Overall	Treatment	Control	Overall	Treatment	Control	Differential
English language arts, Year 1	21,871	10,827	11,044	19,622	9,800	9,822	10	9	11	-2
Math, Year 1	21,871	10,827	11,044	20,105	10,047	10,058	8	7	9	-2
English language arts, Year 2	19,095	9,613	9,482	17,357	8,753	8,604	9	9	9	0
Math, Year 2	19,095	9,613	9,482	17,647	9,002	8,645	8	6	9	-2
English language arts, Year 3	9,647	4,840	4,807	9,072	4,508	4,564	6	7	5	2
Math, Year 3	9,647	4,840	4,807	9,448	4,732	4,716	2	2	2	0

Source: Administrative student records for the 2014–2015 and 2016–2017 school years.

Note: Students in the Year 1 sample are those enrolled in grades 2–4 in study schools at baseline. Students in the Year 2 sample are those enrolled in grades 1–3 in study schools at baseline. Students in the Year 3 sample are those enrolled in grades kindergarten–2 in study schools at baseline. The difference between the percentages of the treatment and control group that are missing outcomes may not equal the differential rate shown in the table due to rounding.

Table C.28. Effects of professional development program on student achievement, alternate student sample

Student achievement (*z*-score units)	Treatment group		Control group		Effect	p-value	Number of students	Number of schools
	Adjusted mean	Unadjusted standard deviation	Adjusted mean	Unadjusted standard deviation				
English language arts, Year 1	-0.22	0.96	-0.22	0.94	0.00	0.946	19,622	100
Math, Year 1	-0.26	0.96	-0.22	0.93	-0.04*	0.047	20,105	100
English language arts, Year 2	-0.23	0.97	-0.23	0.94	0.00	0.965	17,357	100
Math, Year 2	-0.25	0.97	-0.26	0.95	0.01	0.809	17,647	100
English language arts, Year 3	-0.22	1.00	-0.25	0.99	0.03	0.389	9,072	60
Math, Year 3	-0.26	0.99	-0.30	0.98	0.03	0.429	9,448	60

Source: Administrative student records for the 2014–2015 and 2016–2017 school years.

Note: Students in the Year 1 sample are those enrolled in grades 2–4 in study schools at baseline. Students in the Year 2 sample are those enrolled in grades 1–3 in study schools at baseline. Students in the Year 3 sample are those enrolled in grades kindergarten–2 in study schools at baseline. Means were adjusted using the regression model described in Appendix B. Unadjusted standard deviations were the standard deviations across students. The difference between the treatment and control adjusted means may not equal the effect shown in the table due to rounding. None of the effects is statistically significant at the .05 level, two-tailed test.

D. Realized minimum detectable effects

To summarize the level of precision in this study, Table C.29 shows, for each key outcome, the realized values of the minimum detectable effects based on the study's actual data and approach. The minimum detectable effect is the smallest true effect for which the study had an 80 percent probability of obtaining an estimate that was statistically significant at the 5 percent level. We report the minimum detectable effect both in terms of the units reported in the main analysis and in terms of standard deviations of the outcome within the sample, for comparability with other studies.

Table C.29. Realized minimum detectable effects for key outcomes

	Minimum detectable effect		
	Units reported in the analysis		Within-sample standard deviation units
Outcome	Units	Estimate	Estimate
Student achievement			
English language arts, Year 1	Percentile in state	2.0	0.06
Math, Year 1	Percentile in state	2.2	0.06
English language arts, Year 2	Percentile in state	2.5	0.07
Math, Year 2	Percentile in state	2.9	0.08
English language arts, Year 3	Percentile in state	4.1	0.11
Math, Year 3	Percentile in state	4.6	0.13
School climate			
School climate (principals' report), Year 1[a]	1- to 4-point scale	0.23	0.59
School climate (teachers' report), Year 1[b]	1- to 4-point scale	0.09	0.18
School climate (principals' report), Year 2[a]	1- to 4-point scale	0.29	0.56
School climate (teachers' report), Year 2[b]	1- to 4-point scale	0.13	0.24
Retention			
Principal retention over one year (baseline to Year 1)	Percentage	17.0	0.48
Teacher retention over one year (baseline to Year 1)	Percentage	3.6	0.10
Principal retention over two years (baseline to Year 2)	Percentage	20.0	0.44
Teacher retention over two years (baseline to Year 2)	Percentage	4.5	0.10
Principal retention over three years (baseline to Year 3)	Percentage	24.7	0.50
Teacher retention over three years (baseline to Year 3)	Percentage	4.4	0.09
Principals' time use (Year 2)[c]			
Organizational leadership			
Student affairs	Hours per week	2.3	0.53
Administration	Hours per week	1.5	0.54
Other	Hours per week	1.8	0.56
Instructional leadership			
Evaluation	Hours per week	1.5	0.51
Teacher feedback	Hours per week	1.3	0.50
Curriculum	Hours per week	1.0	0.55
Human capital management			
Recruiting teachers	Hours per week	1.1	0.57

Outcome	Minimum detectable effect		
	Units reported in the analysis		Within-sample standard deviation units
	Units	Estimate	Estimate
Personnel policies	Hours per week	0.6	0.42
Own professional growth	Hours per week	1.1	0.52
Nonwork activities	Hours per week	2.3	0.59
Principals' instructional leadership practices			
Classroom observations (principals' report), Year 1	Number of times per year	15.2	0.64
Instructional support and feedback from principal (teachers' report), Year 1[d]	Number of times per year	1.6	0.19
Classroom observations (principals' report), Year 2	Number of times per year	14.8	0.62
Instructional support and feedback from principal (teachers' report), Year 2[d]	Number of times per year	1.9	0.23

Sources: Administrative student records for the 2015–2016, 2016–2017, and 2016–2018 school years, principal and teacher surveys spring 2016 and spring 2017, implementation data for the 2014–2015 through 2016–2017 school years, administrative educator records for the 2014–2015 through 2017–2018 school years, principal time use logs for the 2016–2017 school year.

Note: For each outcome, we calculated the minimum detectable effect as 2.8 multiplied by the standard error of the associated effect estimate. Seven of the eight participating districts provided administrative educator records for teachers.

[a] School climate, as reported by principals, includes the extent to which principals reported the school having problems with student absenteeism, widespread disorder in classrooms, and conflicts between students and teachers. The scale indicates whether each issue is a problem to a (1) great extent, (2) moderate extent, (3) small extent, or (4) not at all.

[b] School climate, as reported by teachers, includes the extent to which teachers reported cooperative effort among staff members in the school, the school administration being supportive and engaging, and not having problems with student misbehavior interfering with their teaching. The scale indicates whether teachers (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree with statements about their school.

[c] Total hours calculated across all 20 rounds of principal log. For each 15-minute window throughout the day, principals indicated whether they spent time on each activity. Instead of filling in the precise number of minutes spent on each activity during each hour-long period of the school day, principals reported their time use in ranges (1 to 14 minutes, 15 to 29 minutes, 30 to 44 minutes, and 45 to 60 minutes). The estimates assume the number of hours the principal spent is the average of the upper and lower bounds for each time range.

[d] "Instructional support and feedback from principal" includes classroom observations, feedback on teaching, developing specific instructional practice goals, using data to determine progress and suggest specific teaching actions, and other instructional supports.

APPENDIX D

INTERPRETING STUDY FINDINGS: ALTERNATIVES TO *P*-VALUES AND STATISTICAL SIGNIFICANCE

The main body of this report presents estimates of the effects of the study's principal professional development program, along with discussion of whether the effects are statistically significant. Although this is a common approach to presenting findings of educational evaluations, readers often think statistical significance (p-value < 0.05) means that there is at least a 95 percent chance that an intervention had an effect. However, that conclusion is incorrect and can lead to serious misinterpretation of study findings. Similarly, a lack of statistical significance does not necessarily mean that there is a low probability an intervention had an effect. The consequences of p-value misinterpretation can be so severe that several researchers have urged the field to abandon the use of p-values and statistical significance.[i]

In this appendix, we present an alternative approach to the p-value, known as BASIE (BAyeSian Interpretation of Estimates).[ii] We then apply this approach to estimate the probability that the professional development program had an effect on key study outcomes. The information using this alternative approach suggests that the program was unlikely to have a meaningful positive effect on the outcomes examined. While in this case the BASIE and p-value approaches lead to a similar overall conclusion, the BASIE approach provides information on the likelihood and size of effects that is richer than the p-value approach and less likely to be misinterpreted.

A. BASIE (BAyeSian Interpretation of Estimates)

The BASIE approach directly estimates the probability that the true effect of an intervention is of a certain size. It does so by applying Bayesian methods, drawing on both the effect directly estimated from the study's data and prior evidence about how common it is for interventions to have effects. However, the BASIE approach differs from how researchers often apply Bayesian methods in two key ways: First, a common concern with Bayesian methods is that they can be subjective. Instead of drawing on prior *evidence*, they sometimes rely on prior *beliefs* about the effects of an intervention.[iii] The BASIE framework avoids this concern by drawing only on prior evidence from similar evaluations, rather than on the researcher's beliefs about the intervention's effects. Second, under the standard Bayesian approach, reseachers often only report the Bayesian shrunken estimate (which is a weighted average of the traditional effect estimate and prior evidence). In contrast, the BASIE approach encourages researchers to report both the traditional effect estimate (based only on study data) and the Bayesian shrunken estimate.[iv]

Under the BASIE approach, to estimate the probability that the true effect of an intervention is of a certain size, a researcher needs to know (1) the effect estimate and standard error for the intervention that was evaluated, and (2) how common it is for *generally similar* interventions to have effects. The prior evidence tells us how common it is to achieve effects of different sizes, such as how common it is to achieve positive effects or effects greater than 0.20 standard deviations. Effect estimates from a particular study that are similar to the prior evidence are judged to be more credible; effect estimates that are very different are deemed less credible.

Given its importance, prior evidence must be selected thoughtfully and analyzed appropriately. BASIE applies five guidelines for selecting and analyzing prior evidence:

1. Use prior evidence, not prior belief. The controversy surrounding Bayesian methods stems largely from concerns about basing the prior on personal beliefs. Under BASIE, the prior is based only on prior evidence.
2. Select prior evidence that meets systematic standards for quality. The What Works Clearinghouse (WWC) provides an excellent source of systematically vetted prior evidence on the effectiveness of educational interventions.
3. Statistically adjust evidence for variation in precision and possible bias due to the file drawer problem. When calculating the mean and standard deviation of effect estimates from prior evidence, researchers should give greater weight to more precise estimates. Giving greater weight to more precisely estimated effects is standard practice in meta-analysis.[v] In addition, researchers should adjust for any observed correlation between effect estimates and the standard error of those estimates. Such a correlation could suggest that the researcher conducted several different versions of the effect estimation but chose to present only those results that were most favorable (sometimes referred to as the *file drawer problem*, suggesting that the researcher might choose to leave the less favorable results buried in a file drawer).

4. **Ideally, the evaluated intervention should be a member of the same "population" as the prior evidence.** For example, the WWC focuses on education interventions, so studies that meet WWC standards provide a natural set of prior evidence for education evaluations seeking to apply the BASIE approach. Researchers could draw on all evidence that meets WWC standards (regardless of the specific nature of the interventions) for prior evidence. Alternatively, if an evaluated intervention is a member of an identifiable subset of evidence in the WWC (for example, reading interventions), that subset could be used as the prior evidence.
5. **Examine and report sensitivity of findings to the selection of prior evidence.** As with many other methodological choices made in impact analysis (for example, which covariates to include in a regression, the approach to accounting for clustering, how to weight the data), researchers can arrive at different conclusions regarding the most appropriate set of prior evidence to use. As with those other methodological choices, researchers should prespecify and conduct analyses to assess the sensitivity of findings to the selection of prior evidence.

B. Interpretation of study findings using the BASIE approach

The study team applied the BASIE approach to learn about the probability that the study's principal professional development program had an effect on three key study outcomes: (1) students' English language arts achievement, (2) student's math achievement, and (3) frequency of instructional support and feedback teachers received from principals. These probabilities provide an alternative to the *p*-value that helps readers to interpret the study's findings more richly and accurately.

1. Selecting and synthesizing prior evidence

The study team used four sets of prior evidence from the WWC to construct the probability estimates:

1. Effects on language/literacy outcomes (used to interpret the effect of the program on English language arts test scores)
2. Effects on math achievement (used to interpret the effect of the program on math test scores)
3. All effects that meet evidence standards in the WWC (used to interpret effects on frequency of instructional support and feedback because the WWC does not include a clearly identifiable set of prior findings directly relevant to this outcome)
4. All effects from professional development interventions that meet evidence standards in the WWC (used for sensitivity tests)

All four sets of prior evidence included only findings from the WWC that met evidence standards, with or without reservations.

The team then synthesized the prior evidence and adjusted for variation in the precision of prior estimates and potential bias due to the file drawer problem. To do so, we used a Bayesian meta-analysis.[vi]

- Bayesian meta-analysis implicitly gives greater weight to estimates that are more precisely estimated.[vii]
- To adjust for potential bias due to the file drawer problem, we ran a Bayesian meta-regression (a Bayesian meta-analysis with regression adjustment) of effect size estimates on the standard error of those estimates.[viii] The constant term from that regression is the expected value of intervention effects, and the coefficient on the standard error indicates how many findings are left in the file drawer, on average. The correlation between effect size and standard error that we observed in the WWC database corresponds to approximately three unreported estimates in the file drawer for every reported estimate in the WWC.

Table D.1 shows the means and standard deviations of effect sizes from each of the four sets of prior evidence discussed above. We report means and standard deviations of the prior evidence both before and after the two adjustments described above. In presenting findings below using the BASIE approach, we use only the means and standard deviations

that include both adjustments. In addition, we also report a prior distribution in Table D.1 based on all findings from WWC that met standards, but with a mean centered at zero rather than the actual mean of these findings. This distribution allows us to apply a more conservative correction for potential bias arising from the file drawer problem.

Table D.1. Prior evidence from the What Works Clearinghouse

Number of interventions	Number of estimated effects	Average effect size			Standard deviation of effect sizes		
		Unadjusted	Precision adjusted	File drawer adjustment	Unadjusted	Precision adjusted	File drawer adjustment
All findings for language/literacy outcomes that met standards							
185	1,320	0.28	0.20	**0.02**	0.43	0.24	**0.20**
All findings for mathematics outcomes that met standards							
129	417	0.20	0.16	**0.04**	0.41	0.22	**0.19**
All findings from WWC that met standards							
306	2,367	0.22	0.20	**0.04**	0.42	0.26	**0.23**
All findings on effects of professional development interventions from WWC that met standards							
19	93	0.09	0.07	**-0.01**	0.21	0.13	**0.12**
All findings from WWC that met standards, with mean centered at zero							
--	--		0			0.23	

2. Findings

Table D.2 shows the estimated effects of the professional development program on students' English language arts and math scores, along with four probability statements for each effect. The estimated effects are based purely on study data; they are unaffected by the prior evidence.[ix] However, the probability statements do depend on the prior evidence reported in Table D.1. The probability statements for effects on reading test scores are based on all findings that met standards from the WWC for language/literacy outcomes. The statements for effects on math test scores are based on all findings that met standards from the WWC for mathematics outcomes.

In all three years, the program's effect on students' English language arts test scores is more likely positive than negative, but also unlikely to be greater than 2 percentile points in all years. This suggests that it is unlikely that the program had meaningful effects on students' English language arts achievement in any of the three years. This finding is consistent with the interpretation based on the *p*-value that the program had no statistically significant effects on students' English language arts achievement but provides additional information. For example, there is a 15 percent probability that the effect on English language arts scores in Year 1 is greater than 1 percentile point, but a 1 percent chance that it is greater than 2 percentile points.

In contrast, it is very likely that the program's effect on students' math test scores in the first year was negative (a 96 percent probability). The estimated effect on Year 1 math scores is –1 percentile point, with a *p*-value of 0.073—an effect deemed "not statistically significant" in the main body of the report because the *p*-value was greater than 0.05. This result shows how focusing simply on whether the effect estimate is statistically significant (*p*-value less than 0.05) might mislead readers about the probability that the intervention had a negative effect on math achievement. In the second year, the effect on math test scores is much less likely to be negative (a 29 percent probability).

Table D.2. Effects on student achievement and the probability that the professional development program had effects of a certain size

Outcome	Estimated effect	*p*-value	Probability that the true effect is: Less than 0	Greater than 1 p.p.	Greater than 2 p.p.	Greater than 4 p.p.
Year 1						
English language arts	0 p.p.	0.706	35%	15%	1%	< 1%
Math	-1 p.p.	0.073	96%	< 1%	< 1%	< 1%
Year 2						
English language arts	0 p.p.	0.964	48%	15%	2%	< 1%
Math	1 p.p.	0.599	29%	33%	8%	< 1%
Year 3						
English language arts	2 p.p.	0.299	15%	63%	36%	4%
Math	1 p.p.	0.503	23%	53%	29%	4%

Source: Administrative student records for the 2015–2016, 2016–2017, and 2017–2018 school years.

Note: None of the effects is statistically significant at the .05 level, two-tailed test. The probability that the effect was below or above the specified levels is calculated using the estimated effect, estimated standard error, and prior evidence from the What Works Clearinghouse that met standards. For the effect on English language arts scores, we restricted the prior evidence to effects on language/literacy outcomes. For the effect on math test scores, we restricted the prior evidence to effects on math test scores. More precisely estimated prior evidence receives a larger weight, and the prior evidence is adjusted for potential bias due to the file drawer problem.

p.p. = percentile point.

Table D.3 shows the effects of the program on the frequency of instructional support and feedback principals provided to teachers. In this case, there is a very high probability (99 percent or greater in both years) that the effects were less than zero. This finding is consistent with the interpretation based on the *p*-value that, in both years, the estimated effects of the program were negative and statistically significant (as shown in Table C.10), but also provides additional information about the likely size of the effect. For example, more likely than not, the true effect lies between -1 and -2 (the probability that the effect is between -1 and -2 is 57 percent in year 1 [93 percent - 36 percent] and in year 2 [78 percent - 21 percent]).

Table D.3. Effects on frequency of instructional support and feedback from principal and the probability that the professional development program had effects of a certain size

Year	Estimated effect (number of times per year)	*p*-value	Probability that the true effect is: Less than 0	Less than –1	Less than –2
Year 1	–2*	0.001	> 99%	93%	36%
Year 2	–2*	0.012	99%	78%	21%

Sources: Principal and teacher surveys, spring 2016 and spring 2017.

Note: The probability that the effect was below the specified levels is calculated using the estimated effect, estimated standard error, and all prior evidence from the What Works Clearinghouse that met standards. More precisely estimated prior evidence receives a larger weight, and the prior evidence is adjusted for potential bias due to the file drawer problem. Frequency of instructional support and feedback from principal includes classroom observations, feedback on teaching, developing specific instructional practice goals, using data to determine progress and suggest specific teaching actions, and other instructional supports.

* Effect is statistically significant at the .05 level, two-tailed test.

3. Sensitivity analysis

We examined the sensitivity of findings to different sets of prior evidence listed in Table D.1. The probability statements in Table D.4 (effects on math and reading test scores) are not very sensitive to which prior evidence is used. Across the four sets of prior evidence along with the distribution centered at zero, the alternative probability estimates generally differ from the main estimates by no more than five percentage points. This shows that the choice of which prior evidence to include in the analysis (all effects that meet evidence standards in the WWC, or a more limited subset of effects from

similar interventions) does not have major effects on the interpretation of this study's main findings. Some probability statements in Table D.5 (effects on frequency of instructional support and feedback) are more sensitive to which prior evidence is used. The probability that the true effect is less than zero differ by no more than one percentage point across different sets of prior evidence. However, the probability that the true effect is less than -1 or -2 times per year is more sensitive to the prior evidence (up to a 20-percentage point difference).

Finally, we also examined the sensitivity of the effect estimates themselves to the incorporation of prior evidence. The effect estimates reported throughout the report and in this appendix are based only on study data. In Table D.6 we also show Bayesian effect estimates that do take into account prior evidence. The Bayesian estimates represent the effect estimate that is most likely correct, given both the study data and the prior evidence. These Bayesian estimates are sometimes called *shrunken* estimates because they *shrink* the traditional effect estimates towards the center of the prior evidence distribution. Except for the year three impact on English language arts scores, the shrunken estimates are identical to the effect estimates calculated from only the study data.

Table D.4. Sensitivity to the prior evidence for effects on student achievement

Description of prior evidence from the WWC that met standards	Prior distribution		Probability that the true effect is:			
	Mean	s.d.	Less than 0	Greater than 1 p.p.	Greater than 2 p.p.	Greater than 4 p.p.
English language arts, Year 1						
All findings	0.04	0.23	35%	15%	1%	< 1%
Language/literacy[a]	*0.02*	*0.20*	*35%*	*15%*	*1%*	*< 1%*
Math	0.04	0.19	35%	15%	1%	< 1%
Professional development interventions	-0.01	0.12	36%	14%	1%	< 1%
All findings, centered at 0	0	0.23	35%	15%	1%	< 1%
Math, Year 1						
All findings	0.04	0.23	96%	< 1%	< 1%	< 1%
Language/literacy	0.02	0.20	96%	< 1%	< 1%	< 1%
Math[a]	*0.04*	*0.19*	*96%*	*< 1%*	*< 1%*	*< 1%*
Professional development interventions	-0.01	0.12	96%	< 1%	< 1%	< 1%
All findings, centered at 0	0	0.23	96%	< 1%	< 1%	< 1%
English language arts, Year 2						
All findings	0.04	0.23	47%	15%	2%	< 1%
Language/literacy[a]	*0.02*	*0.20*	*48%*	*15%*	*2%*	*< 1%*
Math	0.04	0.19	47%	15%	2%	< 1%
Professional development interventions	-0.01	0.12	49%	14%	1%	< 1%
All findings, centered at 0	0	0.23	48%	14%	2%	< 1%
Math, Year 2						
All findings	0.04	0.23	29%	33%	8%	< 1%
Language/literacy	0.02	0.20	30%	33%	8%	< 1%
Math[a]	*0.04*	*0.19*	*29%*	*33%*	*8%*	*< 1%*
Professional development interventions	-0.01	0.12	31%	31%	7%	< 1%
All findings, centered at 0	0	0.23	30%	33%	8%	< 1%
English language arts, Year 3						
All findings	0.04	0.23	14%	64%	36%	4%
Language/literacy[a]	*0.02*	*0.20*	*15%*	*63%*	*35%*	*4%*
Math	0.04	0.19	14%	64%	36%	4%

Description of prior evidence from the WWC that met standards	Prior distribution		Probability that the true effect is:			
	Mean	s.d.	Less than 0	Greater than 1 p.p.	Greater than 2 p.p.	Greater than 4 p.p.
Professional development interventions	-0.01	0.12	16%	60%	31%	3%
All findings, centered at 0	0	0.23	15%	63%	35%	4%
Math, Year 3						
All findings	0.04	0.23	24%	53%	29%	4%
Language/literacy	0.02	0.20	25%	52%	29%	4%
Math[a]	*0.04*	*0.19*	*24%*	*53%*	*29%*	*4%*
Professional development interventions	-0.01	0.12	27%	48%	25%	2%
All findings, centered at 0	0	0.23	25%	52%	28%	4%

Source: Administrative student records for the 2015–2016, 2016–2017, and 2017–2018 school years.

Note: None of the effects is statistically significant at the .05 level, two-tailed test. The probability that the true effect was above or below a certain threshold is calculated using the estimated effect, estimated standard error, and prior evidence from the What Works Clearinghouse that met standards.

[a]The italicized rows are based on the same prior evidence used in the main estimates reported in Table D.2. The nonitalicized rows show the sensitivity of the findings to alternative sets of prior evidence.

s.d. = standard deviation.

p.p. = percentile point.

Table D.5. Sensitivity to the prior evidence for the effects on frequency of instructional support and feedback from principal

Description of prior evidence	Prior distribution		Probability that the true effect on the number of times instructional support and feedback are provided per year is:		
	Mean	s.d.	Less than 0	Less than –1	Less than –2
Year 1					
All findings[a]	0.04	0.23	> 99%	93%	36%
Language/literacy	0.02	0.20	> 99%	93%	33%
Math	0.04	0.19	> 99%	92%	30%
Professional development interventions	-0.01	0.12	> 99%	86%	17%
All findings, centered at 0	0	0.23	> 99%	94%	37%
Year 2					
All findings[a]	0.04	0.23	99%	78%	21%
Language/literacy	0.02	0.20	99%	76%	19%
Math	0.04	0.19	99%	74%	16%
Professional development interventions	-0.01	0.12	98%	64%	8%
All findings, centered at 0	0	0.23	99%	80%	22%

Sources: Principal and teacher surveys, spring 2016 and spring 2017.

Note: The probability that the true effect was above or below a certain threshold is calculated using the estimated effect, estimated standard error, and prior evidence from the What Works Clearinghouse that met standards. Frequency of instructional support and feedback from principal includes classroom observations, feedback on teaching, developing specific instructional practice goals, using data to determine progress and suggest specific teaching actions, and other instructional supports.

* Effect is statistically significant at the .05 level, two-tailed test.

[a]The italicized rows are based on the same prior evidence used in the main estimates reported in Table D.3. The nonitalicized rows show the sensitivity of the findings to alternative sets of prior evidence.

s.d. = standard deviation.

Table D.6. Difference between traditional effect estimates and Bayesian shrunken estimates

Outcome	Traditional estimate	Bayesian shrunken estimates
Effects on English language arts achievement (percentile points)		
Year 1	0	0
Year 2	0	0
Year 3	2	1
Effects on math achievement (percentile points)		
Year 1	-1	-1
Year 2	1	1
Year 3	1	1
Effects on frequency of instructional support and feedback (number of times per year)		
Year 1	-2	-2
Year 2	-2	-2

Sources: Administrative student records for the 2015–2016, 2016–2017, and 2017–2018 school years and Principal and teacher surveys, spring 2016 and spring 2017.

Note: The traditional estimate is based only on study data. The Bayesian estimate *shrinks* the traditional estimate towards the center of the prior distribution. The same prior evidence is used for the findings reported in this table as was used in Tables D.2 and D.3.

ENDNOTES

[I.] Corcoran et al. 2012

[II.] Branch et al. 2009; Grissom and Loeb 2009; Clotfelter et al. 2006

[III.] Two of the eight participating districts included 20 schools in the study (rather than 10).

[IV.] U.S. Department of Education 2002

[V.] Brown 2015; Muthén and Muthén 2010

[VI.] Browne and Cudeck 1992

[VII.] Bentler 1990

[VIII.] Bland and Altman 1997

[IX.] Brown 2015

[X.] Horn 1965

[XI.] Camburn et al. (2010)

[XII.] Puma et al. 2009

[XIII.] Puma et al. 2009

[XIV.] National Center for Education Statistics 2002

[XV.] U.S. Department of Education 2002

[XVI.] National Center for Education Statistics 2002

[XVII.] U.S. Department of Education 2002

[XVIII.] National Center for Education Statistics 2002

[XIX.] U.S. Department of Education 2002

[XX.] National Center for Education Statistics 2002

[XXI.] U.S. Department of Education 2002

[XXII.] National Center for Education Statistics 2002

[XXIII.] U.S. Department of Education 2002

[XXIV.] National Center for Education Statistics 2002

[XXV.] U.S. Department of Education 2002

[XXVI.] Wasserstein and Lazar 2016; Greenland et al. 2016; American Statistician 2018; Amrhein et al. 2019

[XXVII.] Deke and Finucane 2019

[XXVIII.] de Finetti 1974; Kaplan 2019

[xxix] None of the components of BASIE is methodologically new—it draws on guidance from many sources (Gigerenzer and Hoffrage 1995; Gelman and Weakliem 2009; Gelman 2001, 2012, 2015a, 2015b, 2016; Gelman and Shalizi 2013). See Deke and Fincane (2019) for more information about the BASIE approach.

[xxx] Cooper et al. 2009

[xxxi] Cooper et al. 2009; Gelman et al. 2013

[xxxii] The WWC database does not include standard errors. However, it does include the effect (in effect size units) and a p-value, which the study team used to estimate the standard error. This calculation is not exact because many p-values in the database are approximations calculated by the WWC.

[xxxiii] This adjustment is motivated by the idea that for any given study, the effect estimate observed in the literature is the largest of all effect estimates calculated by the author (with the rest unseen in a file drawer). In other words, it is a maximum order statistic, which is well-approximated by a linear function of the standard error (Royston 1981).

[xxxiv] We conducted the analyses in effect size units. For consistency with the findings presented in the main body of the report, we converted the estimated effects on student test scores into percentile point equivalents using the standard normal distribution. We converted the estimated effects on frequency of instructional support and feedback from principal to number of times per year by multiplying by the standard deviation of that variable.

REFERENCES

American Statistician. "Statistical Inference in the 21st Century: A World Beyond p < 0.05." [Special issue]. *American Statistician,* vol. 73, suppl. 1, 2018.

Amrhein, V., S. Greenland, and B. McShane. "Scientists Rise Up Against Statistical Significance." *Nature,* vol. 567, no. 7748, 2019, pp. 305–307.

Bentler, Peter M. "Fit Indexes, Lagrange Multipliers, Constraint Changes and Incomplete Data in Structural Models." *Multivariate Behavioral Research,* vol. 25, no. 2, 1990, pp. 163–172.

Bland, J. Martin, and Douglas G. Altman. "Cronbach's Alpha." *BMJ (Clinical Research Ed.),* vol. 314, no. 7080, 1997, p. 572.

Branch, Gregory, Eric Hanushek, and Steven Rivkin. "Estimating Principal Effectiveness." Working paper 32. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, 2009.

Brown, Timothy A. *Confirmatory Factor Analysis for Applied Research.* New York: Guilford Publications, 2015.

Browne, Michael W., and Robert Cudeck. "Alternative Ways of Assessing Model Fit." *Sociological Methods Research,* vol. 21, no. 2, 1992, pp. 230–258.

Camburn, Eric, James P. Spillane, and James Sebastian. "Assessing the Utility of a Daily Log for Measuring Principal Leadership Practice." *Educational Administration Quarterly,* vol. 46, no. 5, 2010, pp. 707–737.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources,* vol. 41, no. 4, fall 2006, pp. 778–820.

Cooper, H.M., L.V. Hedges, and J. Valentine (eds.). *The Handbook of Research Synthesis and Meta-Analysis* (second edition). New York: Russell Sage Foundation, 2009.

Corcoran, Sean P., Amy Ellen Schwartz, and Meryle Weinstein. "Training Your Own: The Impact of New York City's Aspiring Principals Program on Student Achievement." *Educational Evaluation and Policy Analysis,* vol. 34, no. 2, 2012, pp. 232–253.

de Finetti, B. *Theory of Probability: A Critical Introductory Treatment.* New York: Wiley, 1974.

Deke, J., and M. Finucane. "Moving Beyond Statistical Significance: The BASIE (BAyeSian Interpretation of Estimates) Framework for Interpreting Findings from Impact Evaluations." OPRE Report 2019-35. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, 2019.

Gelman, A. "Ethics and Statistics: Ethics and the Statistical Use of Prior Information." *CHANCE,* vol. 25, no.4, 2012, pp. 52–54.

Gelman, A. "Induction and Deduction in Bayesian Data Analysis." *Rationality, Markets and Morals,* vol. 2*,* 2001, pp. 67–78.

Gelman, A. "Prior *information*, not prior *belief*." Statistical Modeling, Causal Inference, and Social Science blog, 2015a. Available at http://andrewgelman.com/2015/07/15/prior-information-not-prior-belief/. Accessed June 7, 2018.

Gelman, A. "The General Problem I Have with Noninformatively-Derived Bayesian Probabilities Is That They Tend to Be Too Strong." Statistical Modeling, Causal Inference, and Social Science blog, 2015b. Available at http://andrewgelman.com/2015/05/01/general-problem-noninformatively-derived-bayesian-probabilities-tend-strong/. Accessed June 7, 2018.

Gelman, A. "What Is the 'True Prior Distribution'? A Hard-Nosed Answer." Statistical Modeling, Causal Inference, and Social Science blog, 2016. Available at http://andrewgelman.com/2016/04/23/what-is-the-true-prior-distribution-a-hard-nosed-answer/. Accessed June 7, 2018.

Gelman, A., J.B. Carlin, G.S. Stern, D.B. Dunson, A. Vehtari, & D.B. Rubin. *Bayesian Data Analysis* (third edition.). Boca Raton, FL: CRC Press, 2013.

Gelman, A., and C.R. Shalizi. "Philosophy and the Practice of Bayesian Statistics." *British Journal of Mathematical and Statistical Psychology,* vol. 66, no. 1, 2013, pp. 8–38. doi:10.1111/j.2044-8317.2011.02037x

Gelman, A., and D. Weakliem. "Of Beauty, Sex and Power." *American Scientist,* vol. 97, no. 4, 2009, pp. 310–316.

Gigerenzer, G., and U. Hoffrage. "How to Improve Bayesian Reasoning Without Instruction: Frequency Formats." *Psychological Review,* vol. 102, no. 4, 1995, pp. 684–704.

Greenland, S., S.J. Senn, K.J. Rothman, J.B. Carlin, C. Poole, S.N. Goodman, and D.G. Altman. "Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations." *European Journal of Epidemiology*, vol. 31, no. 4, 2016, pp. 337–350.

Grissom, Jason A., and Susanna Loeb. "Triangulating Principal Effectiveness: How Perspectives of Parents, Teachers, and Assistant Principals Identify the Central Importance of Managerial Skills." CALDER working paper 35. Washington, DC: Urban Institute, 2009.

Horn, John L. "A Rationale and Test for the Number of Factors in Factor Analysis." *Psychometrika,* vol. 30, no. 2, 1965, pp. 179–185.

Kaplan, D. "Bayesian Inference for Social Policy Research." OPRE Report 2019-36. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, 2019.

Muthén, Linda K., and Bengt O. Muthén. *Mplus: Statistical Analysis with Latent Variables: User's Guide* (sixth edition). Los Angeles, CA: Muthén & Muthén, 1998–2010.

Puma, Michael, Robert Olsen, Stephen Bell, and Cristofer Price. "What to Do When Data Are Missing in Group Randomized Controlled Trials." NCEE 2009-0049. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, October 2009.

Royston, J.P. "Expected Normal Order Statistics (Exact and Approximate)." Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 31, no. 2, 1982, pp. 161–165.

U.S. Department of Education, National Center for Education Statistics. "NCES Statistical Standards." NCES 2003-601. Washington, DC: U.S. Department of Education, National Center for Education Statistics, 2002. Available at https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2003601. Accessed November 8, 2018.

Wasserstein, Ronald L., and Nicole A. Lazar. "The ASA's Statement on *p*-Values: Context, Process, and Purpose." *American Statistician,* vol. 70, no. 2, 2016, pp. 129–133.

ACKNOWLEDGEMENTS

This study would not have been possible without the contributions of many people. We are grateful for the cooperation of the many coaches, principals, teachers, district leaders, and staff who helped implement the study's principal professional development program, assisted with data collection, and provided important information that shaped the study. A team of leaders and expert consultants from the Center for Educational Leadership at the University of Washington was primarily responsible for developing the study's professional development program and providing the professional development and technical assistance for the study. They included Steve Fink, June Rimmer, Lisa Rooney, and Donna Anderson-Davis. A team of staff from the American Institutes for Research supported implementation of the study's professional development program. They included Ben Kalina, Tammie Knights, Nicol Christie, Marlene Darwin, Debbie Davidson-Gibbs, Mariann Lemke, Elaine Liebesman, and Roshni Menon.

Several people helped collect and analyze data for the study. Kristina Rall led the administration of principal and teacher surveys, with support from Stacie Feldman, Jaimie Grazi, and Maya Reid. A large team of staff from Mathematica, American Institutes for Research, and Social Policy Research Associates helped recruit schools into the study. They included Tim Silva, Nancy Duda, Chris Jones, Jacob Hartog, Linda Mendenko, Chandra Larsen, Sukey Leshnick, Heather Lewis-Charp, Brandon Nicholson, and Castle Sinicrope. A team of staff from Mathematica and Social Policy Research Associates helped collect and analyze administrative and implementation data for the study. They included Seth Brown, Megan Christianson, Mika Clark, Mayte Cruz, Amy Defnet, Morgan Levine, Matt Jacobus, Jessie Mazeika, Eduardo Ortiz, Cassie Pickens-Jewell, Dory Seidel, Charles Tilley, Tara Wildszewski, and Clare Wolfendale.

Many people contributed to the study design, interpretation of the study's findings, and production of this report. The study received useful advice from our technical working group, which included Eric Camburn, Roger Goddard, Jason Grissom, Jason Huff, Carolyn Kelley, Jim Kemple, Susanna Loeb, John Nunnery, and Jeff Smith. Matt Clifford at the American Institutes for Research and Ellen Goldring at Vanderbilt University helped shape and guide the evaluation and provided expert contributions and advice at various stages. Mark Dynarski at Pemberton Research provided important input on the study design, analysis plans, and report. Allen Schirm provided valuable input on the statistical interpretation of the study findings. Maura Butler edited the report, Sheena Flowers and Brigitte Tran created graphics, and Shantal James prepared the report for publication.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

The research team for this evaluation included staff from Mathematica and its subcontractors and consultants, American Institutes for Research, Pemberton Research, Social Policy Research Associates, Ellen Goldring of Vanderbilt University, and Allen Schirm. None of the research team members has financial interests that could be affected by findings from this evaluation. None of the 9 members of the two technical working groups, convened by the research team over the course of the study to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

U.S. Department of Education

Betsy DeVos
Secretary

Institute of Education Sciences

Mark Schneider
Director

National Center for Education Evaluation and Regional Assistance

Matthew Soldner
Commissioner

October 2019

This report was prepared for the Institute of Education Sciences under Contract ED-IES-14-C0028. The project officer is Elizabeth Warner in the National Center for Education Evaluation and Regional Assistance.

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

This report is available on the Institute of Education Sciences website at http://ies.ed.gov/ncee.

Upon request, this report is available in alternate formats, such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.