



BENJAMIN K. MASTER, HEATHER L. SCHWARTZ, FATIH UNLU, JONATHAN SCHWEIG,  
LOUIS T. MARIANO, ELAINE LIN WANG

# Effects of the Executive Development Program and Aligned Coaching for School Principals in Three U.S. States

---

Investing in Innovation Study Final Report

For more information on this publication, visit [www.rand.org/t/RRA259-1](http://www.rand.org/t/RRA259-1)

**Library of Congress Cataloging-in-Publication Data** is available for this publication.

ISBN: 978-1-9774-0512-8

Published by the RAND Corporation, Santa Monica, Calif.

© Copyright 2020 RAND Corporation

**RAND**® is a registered trademark.

#### Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit [www.rand.org/pubs/permissions](http://www.rand.org/pubs/permissions).

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

#### Support RAND

Make a tax-deductible charitable contribution at  
[www.rand.org/giving/contribute](http://www.rand.org/giving/contribute)

[www.rand.org](http://www.rand.org)

## Preface

---

With funding from the U.S. Department of Education’s Investing in Innovation (i3) grant program, the National Institute for School Leadership (NISL) provided a professional development program and aligned coaching to current principals in three states. NISL’s Executive Development Program (EDP) is a year-long professional development program that has served thousands of principals in 23 states since 2014. With funds from the i3 grant, NISL provided the EDP professional learning to new principals and coupled it with intensive leadership coaching from trained NISL coaches.

This final report analyzes the effects of the EDP and paired coaching in this i3-funded intervention, with results that are three years following random assignment of schools to treatment. The report is primarily intended for an academic audience of researchers (with a focus on methods) and, secondarily, for policymakers.

RAND is the independent research partner conducting the evaluation of the EDP and coaching study. This report is the second in a series of publications by the RAND research team about the EDP and paired coaching. We published the first report, titled *Putting Professional Learning to Work: What Principals Do with Their Executive Development Learning* (Wang et al.), in 2019. We will publish a third publication in fall 2020 summarizing the outcomes of a second ongoing federally funded study of the EDP and paired coaching.

This study was undertaken by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through postsecondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking. Criterion Education sponsored the report, with funding from NISL through the i3 grant from the U.S. Department of Education (under grant number U411B140042). More information about RAND can be found at [www.rand.org](http://www.rand.org). Questions about this report should be directed to Ben Master at [bmaster@rand.org](mailto:bmaster@rand.org), and questions about RAND Education and Labor should be directed to [educationandlabor@rand.org](mailto:educationandlabor@rand.org).

# Contents

---

Preface.....	iii
Figure .....	v
Tables.....	vi
Summary.....	viii
Acknowledgments.....	xi
Abbreviations.....	xii
1. Introduction.....	1
Description of the Investing in Innovation Study.....	1
Preview of Key Findings .....	5
About This Report .....	6
Organization of This Report.....	7
2. Description of the Principal Professional Learning and Prior Research About Its	
Effectiveness .....	8
The EDP and Coaching .....	8
Theory of Action .....	9
Prior Research About the Effects of Principal Professional Learning Programs .....	11
3. Data, Methods, and Analytic Samples.....	19
Data.....	19
Survey Administration.....	23
Analytic Methods .....	24
Missing Data.....	34
Adjustments for Multiple Comparisons .....	35
Analytic Samples.....	36
4. Results.....	42
Results About Implementation .....	42
Impact Results: Student Outcomes.....	51
Impact Results: Leadership Practices and School Practices.....	55
5. Discussion and Conclusion.....	59
Summary of Findings .....	59
Limitations.....	60
Discussion.....	61
Appendix. Technical Appendix .....	65
References.....	80

## Figure

---

Figure 2.1. Logic Model for the EDP and Coaching .....	10
--	----

## Tables

---

Table 1.1. Treatment and Control Conditions .....	1
Table 1.2. Final Sample of Schools and Students in the Research Study.....	4
Table 3.1. Study Topics by Data Source and Where to Find the Analysis.....	19
Table 3.2. Types of Data Collected and Timepoints for Collection .....	21
Table 3.3. Summary of Statewide Achievement Exams, by Subject, State, and Grade Level.....	22
Table 3.4. Student and School Characteristics in Experimental Analyses Analytic Sample, Overall and by State and Cohort.....	37
Table 3.5. Student and School Characteristics in Experimental Analyses Analytic Sample, Overall and by Treatment and Control Groups.....	39
Table 4.1. Proportion of Principals Who Fully Participated in the EDP and Coaching.....	42
Table 4.2. Participation Rate in the EDP, Overall and by State .....	43
Table 4.3. Reasons for Partial Participation or Nonparticipation in the EDP.....	43
Table 4.4. Participation in Coaching, Overall and by State.....	44
Table 4.5. Reasons for Partial Participation or Nonparticipation in Coaching.....	44
Table 4.6. Perceptions of the EDP, the Action Learning Project, and Coaching .....	46
Table 4.7. How Principals Applied EDP Concepts .....	47
Table 4.8. Principals’ Responses to the Question “What Top Three Lessons or Ideas Have You Used the Most from NISL Coaching, if Any?” .....	48
Table 4.9. Principals’ Responses to the Question “What Was the Most Beneficial Activity You Did with Your Coach?” .....	49
Table 4.10. Estimated Impact of Offering the EDP and Coaching on Student Achievement .....	51
Table 4.11. Estimated Range of Impacts of Participating in the EDP and Coaching on Student Achievement and Attendance in Schools .....	53
Table 4.12. Estimated Impact of Full Participation in the EDP and Coaching on Student Achievement and Attendance .....	54
Table 4.13. Impact of Offering and of Full Participation in the EDP and Coaching on Principal Leadership Practices .....	56
Table 4.14. Impact of Offering and of Full Participation in the EDP and Coaching on School Policy, Culture, and Practices .....	57
Table A.1. Principal Reports the School Has a Strategic Plan .....	65
Table A.2. Teachers Report That Principal Is an Effective Leader .....	66
Table A.3. Principal Indicates School Has Differentiated Roles for Teachers.....	66
Table A.4. Teachers Indicate School Has Differentiated Roles for Teachers .....	67
Table A.5. Teachers Indicate They Have Input into Staffing, Curriculum, or Professional Development.....	67

Table A.6. Principal Ensures Curriculum Is Aligned and Evidence-Based .....	67
Table A.7. Principal Indicates High Priority for Use of Student Assessment Data.....	68
Table A.8. Principal Indicates School Offers High-Quality Professional Development for Teachers .....	68
Table A.9. Teachers Indicate That Administrators Observe Teachers’ Classrooms .....	68
Table A.10. Teachers Indicate That Teachers Observe Other Teachers’ Classrooms.....	69
Table A.11. Teachers Say They Have Received Actionable Feedback.....	69
Table A.12. Principal Reports There Is a Positive School Culture.....	69
Table A.13. Principal and Teachers Feel School Is Unsafe and Disorderly.....	70
Table A.14. Teachers Endorse Their School .....	70
Table A.15. Principal’s Sense of Educators’ Efficacy.....	70
Table A.16. Teachers’ Sense of Self-Efficacy.....	71
Table A.17. Principal Believes Teachers Are Committed to High-Quality Academics.....	71
Table A.18. Teachers Report the School Has High-Quality Academics .....	72
Table A.19. Teachers Focus on Students’ Higher Order Skills.....	73
Table A.20. Principal Indicates Instruction Is Personalized to Students .....	73
Table A.21. Principal Prioritizes NISL’s Concepts of Learning .....	73
Table A.22. Principal Indicates Teachers Collaborate About Matters of Instruction and Professional Development .....	74
Table A.23. Teachers Indicate They Collaborate About Instruction .....	74
Table A.24. Principal Has Autonomy to Deploy School Resources .....	75
Table A.25. Principal Indicates They Have Curricular Autonomy .....	75
Table A.26. Standardized Treatment Versus Control Group Differences on Baseline Test Scores, by Sample Used for Analyses of Full Participation in EDP and Coaching.....	76
Table A.27. Percentage of Initially Randomized Students Without Achievement Outcomes in Year Three, by Subject and Assignment Group .....	76
Table A.28. Mean Responses Overall and by State for Selected Survey Questions About Participating Principals’ Experiences with the EDP.....	77
Table A.29. Impact of Offering the EDP and Coaching on Student Achievement on State- Specific Exams.....	78
Table A.30. Impact of Full Participation in the EDP and Coaching on Student Achievement on State-Specific Exams .....	78
Table A.31. Impact of the EDP and Coaching on Student Grade Progression.....	79
Table A.32. State-by-State Estimates of the Effects of EDP and Coaching on Selected Survey Factors.....	79

## Summary

---

Substantial research has shown that effective school leadership is associated with improved instructional practices and greater student learning in schools. However, we know much less about the extent to which professional development programs for school principals can improve leadership practices and student and school outcomes at scale and, if so, what conditions are needed for such programs to have an impact.

This study evaluated the effects of a large-scale implementation of the National Institute for School Leadership's (NISL's) Executive Development Program (EDP) and paired coaching for K–12 school principals. The EDP is a widely used principal professional development program that previously has been shown to have a positive influence on student achievement outcomes (Nunnery, Yen, and Ross, 2011; Nunnery et al., 2011). The EDP is an in-person program for school principals that is typically delivered in 24 full-day sessions, two per month, over 12 months. The EDP provides principals with supports to build their skills, knowledge, and resources to set the direction for teachers, improve their instruction, and create an effective school environment rooted in professional learning. The EDP also includes diagnostic tools principals can use for their school planning. By the third unit of the 12-unit EDP, principals develop a hands-on Action Learning Project (ALP) about their school that the leaders refine and implement over the rest of the EDP program and beyond.

In this study, NISL-certified coaches offered at least 60 hours of one-on-one coaching to principals. Most of these coaching hours were face-to-face meetings at the principal's school, and the balance of coaching was done remotely via phone or email. Typically, the coaching focused on use of the EDP content and tools and on the implementation of the principal's ALP.

The implementation of the EDP and coaching spanned three states, 332 schools, and 118 school districts. The professional development was offered to a pair-randomized sample of 166 novice principals in middle schools (i.e., schools that included grades six through eight) in participating districts. These 166 principals made up the treatment group. The remaining 166 schools were offered the EDP (but not coaching) three years later. This set of principals made up the control group. The professional development and the research study were sponsored by a grant from the U.S. Department of Education's Investing in Innovation (i3) program.

Our study examined the implementation of the EDP and coaching professional development, the perceptions of participants, and the impacts of the professional development. We considered both the impacts of the offer of and the impacts of full participation in the EDP and coaching on student academic outcomes and on school practices, as measured by principal and teacher surveys.

## Implementation of the EDP and Coaching

The rates at which the principals who were offered the EDP and coaching fully participated were low. We define *full participation* as completion of at least ten of the 12 EDP units, which equates to attending 20 or more of the 24-day, in-person professional development program *and* receipt of 60 or more hours of NISL coaching. By this definition, only 58 (35 percent) of the 166 principals offered the treatment participated fully. Reasons for nonparticipation included principal mobility, districts opting out of the study, and principals opting out. Rates of full participation varied by state, ranging from a low of 15 percent of eligible principals fully participating in one state to a high of 49 percent of eligible principals participating in another state. Low participation rates dilute the measured effects of the intervention in our experimental analysis.

Although participation rates were low, principals who did participate in the EDP and/or coaching rated their experience very highly. Eight out of ten principals who participated in the EDP and took our survey reported that the program helped them improve their school and that they would recommend the EDP to a friend who was a principal. A similar proportion enthusiastically endorsed NISL coaching as well. Participating principals reported embracing core concepts from the EDP curriculum and said they took concrete leadership actions in response to the program. The most frequently cited actions related to distributing leadership within their buildings, adapting practices using research about how people learn, and incorporating the concept of “all means all,” which refers to ensuring that all students have access to high-quality learning opportunities, in school leadership decisions.

## Impacts of the EDP and Coaching

We did not find significant effects from either (1) offering the EDP and coaching or (2) principals’ actual participation in the EDP and coaching on student achievement in English language arts or mathematics, on student attendance rates, or on student grade progression rates within the first three years following the start of the program. Although our estimated effects varied by state, none of the effects were large enough to be statistically significant.

We did find effects from participation in the EDP and coaching in two areas of leadership practice as reported by principals on surveys conducted more than two years after the start of the intervention. (RAND’s surveys of teachers and of principals examined those aspects of principals’ leadership practices and their schools’ instructional policies and culture that the EDP targets.) The first effect on leadership practices was related to the school having a strategic plan, and the second was the personalization of instruction for students. In both cases, we found large positive effects in which principals fully participating in the EDP and coaching reported agreement that was substantially higher than principals from the control group reported. We did not, however, find evidence of significant effects from similar surveys of teachers.

## Conclusions

As the low take-up rates might suggest, the EDP and coaching did not sufficiently affect school practices to generate detectable effects on student achievement within three school years, in spite of having some apparent effects on school practice. There were, however, some positive signs in terms of principals' positive perceptions of the professional development and coaching and in terms of effects on important school practices.

We also found that the take-up rates, the reported experiences of participants, and our estimates of impact varied somewhat across state contexts. According to this and other suggestive evidence of state-by-state differences, we hypothesize that local buy-in and capacity to fully participate in the intensive professional development program most likely influenced the degree to which the intervention was successful. Overall, our research points to the importance of the notion that local buy-in and integration with other district and state systems may be critical to the capacity of a principal professional development program to drive meaningful changes in student and school outcomes at scale.

## Acknowledgments

---

The authors thank the NISL staff in the national office and the NISL staff in each of the three participating states who worked with the RAND team to provide and help interpret administrative data essential to this report. We would especially like to thank Claire Hollywood at NISL and Laura Groth, formerly at NISL, for helping us to better understand the implementation and context surrounding the NISL i3 intervention.

We also wish to thank the Johns Hopkins University evaluation team as the original evaluator of this i3 grant and, specifically, Steven Ross, Betsy Wolf, and Gavin Latham for their assistance in transferring data and records that they had collected for the study. It was their team who conducted the original randomization and conceived of the overall study design. They graciously gave us their time to explain their randomization process and data collection procedures and to share the data that they had collected.

We wish to thank Brian Phillips and Tiffany Tsai at RAND, who played major roles in data preparation and survey administration. This report benefited substantively from quality assurance feedback from Susan Straus, RAND quality assurance manager, and our two reviewers, Matthew Baird and Matthew Clifford. Any flaws that remain are solely the authors' responsibility.

## Abbreviations

---

ALP	Action Learning Project
BLPD	Balanced Leadership Professional Development
CCSS	Common Core State Standards
EDP	Executive Development Program
EFA	exploratory factor analysis
ELA	English language arts
GBM	generalized boosted methods
i3	Investing in Innovation
ITT	intent-to-treat
LEA	local education agency
MICE	multiple imputation through chained equations
NISL	National Institute for School Leadership
SEA	state education agency
SEED	Supporting Effective Educator Development
TOT	treatment-on-the-treated
RCT	randomized controlled trial
RQ	research question
WWC	What Works Clearinghouse

# 1. Introduction

---

In this chapter, we first describe the i3 study, then list the research questions, preview the main findings, and set out the organization of the report.

## Description of the Investing in Innovation Study

The U.S. Department of Education awarded an Investing in Innovation (i3) grant to the National Institute for School Leadership (NISL) in fiscal year 2014 to provide intensive professional development and coaching to current principals in three states.<sup>1</sup> The key i3 grant activities were to

- provide the 12-month Executive Development Program (EDP) to principals at more than 300 schools across three states
- provide intensive leadership coaching for up to 30 months that was aligned with the EDP curriculum.

This evaluation of NISL’s efforts in the i3 states involved a randomized controlled trial (RCT) conducted in participating school districts. Table 1.1 summarizes the treatment and control group conditions, which we describe next.

**Table 1.1. Treatment and Control Conditions**

Sample	Eligibility Criteria	Offered Services During First Three School Years of Four-Year i3 Study	Offered Services in the Fourth Year
Treatment group	<ul style="list-style-type: none"> <li>• Novice principal, defined as 0–5 years of principal experience at any school as of the time of randomization.</li> <li>• Principal of a middle school, defined as a school containing grades 6–8.</li> <li>• Had not taken any part of the EDP prior to the i3 study.</li> </ul>	<ul style="list-style-type: none"> <li>• The EDP, which is a 24-day, in-person professional development program. There are 12 units in the EDP curriculum. NISL-certified facilitators led regionally held EDP sessions that occurred on weekdays. The EDP also provided participants with online materials, such as diagnostic tools and case studies, for principals’ use in their schools.</li> <li>• Sixty or more hours of one-on-one coaching. A NISL-certified coach most often met with the principal at their school for face-to-face coaching, but also provided coaching remotely. Coaching hours were spread over as many as 30 months.</li> </ul>	<ul style="list-style-type: none"> <li>• None</li> </ul>
Control group	<ul style="list-style-type: none"> <li>• Same as treatment group</li> </ul>	<ul style="list-style-type: none"> <li>• None</li> </ul>	<ul style="list-style-type: none"> <li>• The 12-unit, 24-day EDP</li> </ul>

---

<sup>1</sup> We mask the names of the three states throughout this report. One state is in the west, one is in the south, and one is in the southeast of the United States.

The RCT was designed to test the effects of providing both the EDP program and intensive coaching to a cohort of middle school principals with zero to five years of experience at any school and who had no prior exposure to the EDP.<sup>2</sup> Schools were randomly chosen to have their principals offered the EDP professional learning and coaching immediately (the treatment group) or to delay the offer of the EDP only (no coaching) for three years. The i3 grant paid for the EDP course for the participants in the study (both the treatment and the control group) and the NISL coaching (which was offered to just the treatment group).

There were two reasons for the focus on novice principals of middle schools:

1. NISL hypothesized that the EDP and coaching intervention might be more influential for principals who were new to the role.
2. The middle school grade span aligned with the i3 grant study period. All schools in this study were middle schools that included students in grades six through eight. With a three-year treatment period, the study examines the effects of the EDP and coaching for novice principals on students who, at the outset of the treatment period, were sixth graders. The end of the treatment period coincided with those same students completing eighth grade.

As we explain in the following paragraphs, there were two cohorts of treatment and control groups in this study. With one minor initial difference,<sup>3</sup> the offered treatment was the same in both cohorts: the EDP course and NISL coaching. In both cohorts, participating districts agreed that the researchers would randomly assign their middle school principals to either the treatment or the control group in the summer prior to the start of the EDP. In both cohorts, the 12-month EDP started in September—for Cohort 1, September 2015, and for Cohort 2, September 2016—and the NISL coaching started in January—for Cohort 1, January 2016, and for Cohort 2, January 2017—and lasted up to 30 months. Coaches in both cohorts were to prioritize in-person, every-other-month visits to principals' schools to deliver coaching, although coaching could also occur by phone and by web-meetings at the principal's preference. NISL-certified EDP facilitators led the EDP courses in all three states. As is usual, these sessions were held in group settings on weekdays, for which a principal needed to travel to a regional center to attend. For

---

<sup>2</sup> Districts across the country sponsor NISL to provide the EDP for its principals. Therefore, it was possible that principals in a participating district in the i3 study had prior exposure to the EDP. Therefore, the eligibility criteria for principals randomized to immediate or delayed EDP excluded principals who had attended some or all of the EDP in the past. In the three states included in this study, prior participation in the EDP was relatively rare.

<sup>3</sup> NISL's original expectation for i3 coaching of Cohort 1 principals was to provide 70 cumulative hours of coaching over a period of 17 months, which equates to 1.5 school years. The cumulative hours would be a combination of in-person, phone, email, and webcalls. The in-person visits were to occur at a frequency of once every other month and to last three to six hours each visit, with supplemental remote coaching. In fall 2016, NISL decided to extend the period of coaching (but not increase cumulative hours) to 30 months rather than 17. However, six-hour visits proved to be too much for participants because the visit consumed a full school day. Therefore, NISL decided in mid-2017 to lower the expected cumulative hours of coaching to 60 total hours, which was the sum of hours resulting from shortened school visits to three to four hours that occurred every other month. After changing the coaching expectations to 60 hours for Cohort 1 over 30 months and receiving positive feedback from coaches, NISL designed Cohort 2 to follow the same model of 60 cumulative hours delivered over a total period of 30 months.

both cohorts, we define *fully compliant* principals as those who received the full 60 hours of NISL coaching and who attended ten or more units of the 12-unit EDP. The principals randomly selected to be in the control group in both cohorts were offered the EDP program three years after the treatment group principals.

In two of the three states in this study, a *national NISL coach*, meaning a coach who does not necessarily live in the state, provided the coaching directly to the participating principals in two of the three states. In the third state, all but one of the coaches were NISL-certified, retired principals from that state. Whether national or local, all coaches assisted school leaders with creating and refining their ALPs and provided any other support needed to implement strategic changes aligned with the EDP.

For Cohort 1, NISL state coordinators in three states recruited local education agencies (LEAs) starting in spring 2015 to participate in the study. Each state coordinator solicited the participation of all superintendents of districts in the state that had traditional public middle schools. The state coordinators each lived in the state where they worked and had prior experience working as a principal, superintendent, or state education agency (SEA) employee in that state and thus knew the LEA context and many of the superintendents. Interested LEAs signed a memorandum of understanding with NISL for novice middle school principals to participate in the i3 grant-funded program, with the understanding that some of the middle school principals they listed as potential participants would receive professional development on a delayed schedule as part of the control group. The first cohort consisted of 109 original LEAs across three states, with a combined total of 271 middle school principals from those states. Just before the start of school year 2015–2016, the original research team randomized the first cohort of study schools in each of three states.<sup>4</sup>

In LEAs for which two or more study schools had observably similar student demographics and test scores, researchers identified these two schools as pairs and assigned one school at random from each pair into the treatment group.<sup>5</sup> In cases (66 out of 161 pairs) for which only one school in a district was in the study sample, when there was an odd school still unpaired in a district, or when no sufficiently similar school could be found within a district for a school, the researchers matched two observably similar study schools from the same state and randomly assigned one of the schools from each pair into the treatment group.

---

<sup>4</sup> The Johns Hopkins University research team members were the original evaluators of the i3 grant. In winter 2017, NISL selected RAND to replace the original research team.

<sup>5</sup> Specifically, the original research team identified schools that were closely matched on multiple school characteristics according to publicly available data, including school average test scores for up to four prior years, upward versus downward trends in test scores over those prior years, the percentage of students eligible for free lunch, and the percentage of students classified as English language learners. Ideal pair matches were identified according to being a sufficiently close match for as many of these separate characteristics as possible, with priority given to matching on average test score in cases of ties. The same criteria were used for identifying within-district and within-state pairs.

To proactively address the nonparticipation of some Cohort 1 districts and schools that emerged in the first year of the study (i.e., 2015–2016), NISL reallocated available i3 funds to recruit additional districts in all three states to establish a second cohort of principals for the study. There was sufficient interest from additional LEAs in one of the three states to establish a second cohort. This second cohort consisted of 61 schools, of which the 31 treatment schools started the EDP and then coaching one school year after the first cohort. By spring 2016, the NISL state coordinator in that state recruited an additional nine LEAs, which made up the second cohort in the study. That same research team then randomized the second cohort just prior to school year 2016–2017 using the same methods as they did for the first cohort. This second cohort started treatment one school year after the first cohort.

Looking at the two cohorts combined, NISL recruited a total of 118 school districts from the three states, from which 332 schools were randomized in pairs into treatment and control groups. However, a combination of school closures, schools that were alternative schools and lacked achievement outcome measures, and schools that had been initially misclassified as spanning grades six through eight reduced the final analytic sample from 332 to 323 schools with student achievement outcome data in the third year of the study.

Table 1.2 provides the final number of schools and students analyzed in this report. Of the 323 schools included in our analysis, 161 were randomly assigned to treatment, and the rest were assigned to control. In this report, we focus on outcomes for all students who were in sixth grade in study schools within the first six weeks of the school year following random assignment and for whom outcome data were available in the third school year following random assignment. This is a total analytic sample of 63,337 students (across 323 schools) for whom outcome data are available in all years. We provide additional descriptive characteristics of the schools and students included in our analyses in Chapter 4.

**Table 1.2. Final Sample of Schools and Students in the Research Study**

<b>Sample</b>	<b>Control Schools</b>	<b>Treatment Schools</b>	<b>Control Students</b>	<b>Treatment Students</b>
Cohort 1				
State A	51	51	9,284	10,280
State B	42	42	11,176	11,143
State C	39	37	3,806	4,077
<i>Total in Cohort 1</i>	<i>132</i>	<i>130</i>	<i>24,266</i>	<i>25,500</i>
Cohort 2 (State B only)	30	31	6,755	6,816
<b>Total of Cohorts 1 and 2</b>	<b>162</b>	<b>161</b>	<b>31,021</b>	<b>32,316</b>

NOTE: Treatment group principals from Cohort 1 started the EDP and coaching in the 2015–2016 school year, and treatment group principals in Cohort 2 started the EDP and coaching in the 2016–2017 school year. Student enrollment is from the initially randomized cohort of students with available testing information from 2017–2018 for Cohort 1 and from 2018–2019 for Cohort 2.

In this study, we estimated the impact of the EDP and coaching intervention on student attendance, grade progression, and achievement, as measured on spring state standardized tests in English language arts (ELA) and mathematics about three full school years after treatment group principals first began the yearlong EDP. For Cohort 1, this means we examined spring 2018 test scores; for Cohort 2, we examined spring 2019 test scores. In addition to student outcomes, we analyzed the extent to which principals participated in the EDP and coaching, their perceptions of both components, and principal and staff survey-reported perceptions of school practices aligned to the EDP. Our research questions were as follows.

### *Research Questions About Implementation*

1. What proportion of principals who were offered the EDP and coaching participated?
2. Among those who participated, what were their perceptions of the EDP and of coaching?
3. What form did coaching take, and what were the main topics of coaching?

### *Research Questions About Impacts of the EDP and Coaching*

4. What was the effect of offering the EDP and coaching on student achievement, attendance, and grade promotion rates after three years?<sup>6</sup>
5. What was the effect of actually participating in the EDP and coaching on student achievement, attendance, and grade promotion rates after three years?
6. What was the effect of offering and of participation in the EDP and coaching on principals' leadership practices and on school policy, culture, and practices?

Questions 1, 4, and 6 are relevant for LEAs or SEAs regarding the likely take-up rates and effect of offering (not mandating) the EDP and coaching. Questions 2, 3, 5, and 6 are relevant for developers of principal professional learning programs, as well as for principals, LEA, and SEAs who wish to know principals' perceptions and whether actual participation in the EDP would be likely to benefit students and school practices.

## Preview of Key Findings

This section provides a brief overview of our key findings. We present more-complete findings in Chapter 4.

Overall, across the three-state sample, we did not find that either offering the EDP and coaching to novice, middle-school principals or the principals' actual participation in the EDP

---

<sup>6</sup> Following the requirement for i3-funded studies to prespecify research questions as *confirmatory* or *exploratory*, the research team prespecified research questions 4 and 5 as the *confirmatory contrasts*, meaning the core research questions designed to estimate the causal impact of the EDP and coaching on student achievement. These two research questions, which we examined across our pooled sample of schools spanning all three states, are also the analyses for which we had the most statistical power to detect any true effects. We classified analysis of state-specific effects and analysis related to research question 6 as *exploratory*, meaning that they offer suggestive, but not definitive, evidence about how and whether the EDP and coaching influence intermediate outcomes, such as school culture and instructional practices, or final outcomes, such as student achievement.

and coaching influenced student achievement, student attendance rates, or student grade progression rates within the first three school years following the start of the program. We found no significant effects in either mathematics or ELA achievement. Although there were some differences in the estimated effects of offering the program in each of the individual states, none of the state-specific estimates were statistically significant, nor were any other subgroup differences that we explored.

Principals who opted to participate in the EDP and coaching reported very positive experiences of both. However, the take-up of the intervention was fairly low, with only 57 percent of principals fully participating in the EDP and just 35 percent fully participating in both the EDP and coaching. The two main reasons for less than full participation were when principals opted out of all or part of the program and when principals transitioned out of their initial study school.

We did find effects of full participation in the EDP and coaching on two survey-reported school practices that the EDP teaches principals to do. The first was the principal describing the school as having a strategic plan, and the second was the personalization of instruction for students. In both cases, we found large positive effects in which fully participating principals in the EDP and coaching reported agreement that was substantially higher than principals from the control group. We did not, however, find evidence of significant effects from teacher-reported school practice measures. We hypothesize that changes in the principals' practice may not yet have affected the teachers' perceptions of school practice to a degree that was detectable in our sample.

Overall, the results of this study indicate that, in line with the low take-up rates we observed, the EDP and coaching did not sufficiently affect school practices in enough schools to generate detectable effects on student achievement within three school years of the start of the program, in spite of having some effects on school practice. We found suggestive evidence that state and district buy-in and coach-principal matching were important factors that influenced principals' participation in the EDP and coaching.

## About This Report

This report is the second in a series of publications by the RAND research team about the EDP and paired coaching. Our first report, published in 2019, included nine in-depth case studies about the school reforms that principals enacted in their buildings in the years following the EDP (Wang et al., 2019). Four of these case studies were from i3 treatment group schools, and the other five were from a separate federally funded study of the EDP. We will publish the results of that separate study in fall 2020.

## Organization of This Report

In Chapter 2, we provide background for the study by first explaining the EDP and coaching, what prior research shows about this kind of principal professional learning and support, and the design of the research study. In Chapter 3, we set out our data sources, analytic methods, and analytic samples. In Chapter 4, we provide the results for the implementation and outcome research questions. In Chapter 5, we interpret the results in a discussion of the findings and their implications for educators.

## 2. Description of the Principal Professional Learning and Prior Research About Its Effectiveness

---

This chapter provides an overview of the EDP, paired coaching, and prior research about both types of principal development.

### The EDP and Coaching

The EDP is an in-person professional learning program for current school principals that is typically delivered over 24 full-day sessions. NISL first offered the EDP in 2014. As of 2019, more than 15,000 principals across 23 states have participated in the EDP. A NISL-certified facilitator delivers these sessions using a written curriculum that NISL developed. The sessions typically occur on two workdays per month (not weekends) over 12 months. These sessions might be held regionally within a state or, if there is a sufficient number of principals from a single district in the EDP, within a single district. NISL offers some make-up dates for school leaders who miss the original scheduled days.

Grounded in research about how students and adults learn, the EDP provides principals with supports to build their skills, knowledge, and resources to set the direction for teachers, improve their instruction, and create an effective school environment rooted in professional learning. The two-day units cover the topics listed in Box 2.1.

#### **Box 2.1**

##### **Topics Covered in the EDP**

###### **Course 1: World-Class Schooling: Vision and Goals**

- Unit 1: The Educational Challenge
- Unit 2: The Principal as Strategic Thinker
- Unit 3: Elements of Standards-Aligned Instructional Systems

###### **Course 2: Focus on Teaching and Learning**

- Unit 4: Foundations of Effective Learning
- Unit 5: Leadership in the Instructional Core: English Language Arts and History
- Unit 6: Leadership in the Instructional Core: Science and Mathematics
- Unit 7: Coaching for High-Quality Teaching

###### **Course 3: Sustaining Transformation Through Capacity and Commitment**

- Unit 8: Promoting the Learning Organization
- Unit 9: Teams for Instructional Leadership
- Unit 10: Ethical Leadership for Equity
- Unit 11: Driving and Sustaining Transformation
- Unit 12: Final Case Simulation and Presentations

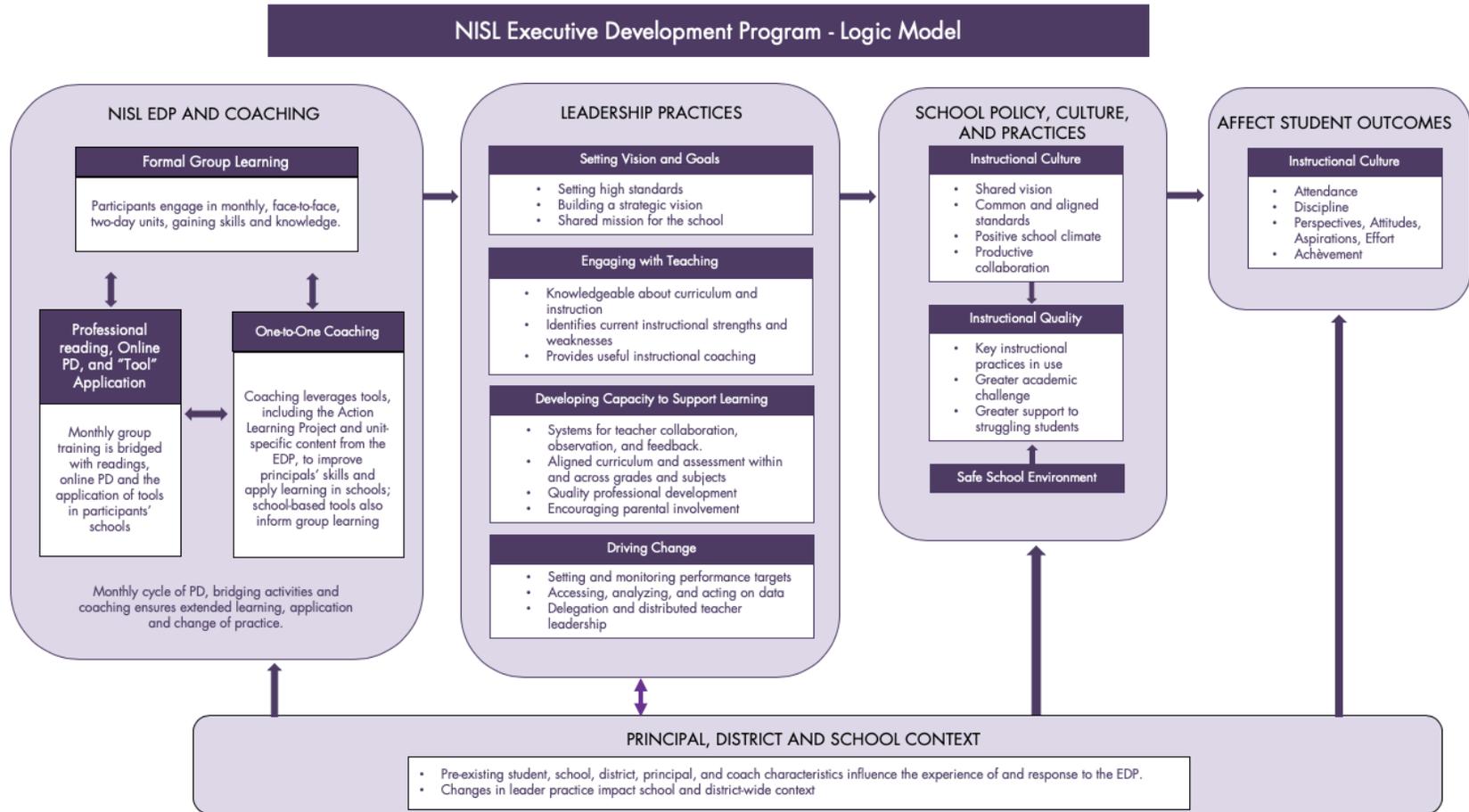
At the end of the third EDP unit, school leaders develop a first draft of their Action Learning Project (ALP) for their school that outlines one or more goals and the strategies and action steps that need to be completed to reach those goals. School leaders refine the ALP throughout their professional learning and use it after they complete the EDP to guide their continued work on the topic the principal specified in their ALP. The following are examples of actual ALP topics that we profile in a companion publication that includes nine case studies (Wang et al., 2019): wider use of formative assessment practices, school culture and climate, small-group reading instruction, alignment of ELA curriculum across and within grade levels, and teacher Professional Learning Communities.

In addition to the EDP, NISL directly employs coaches (who are often retired principals) and offers a five-day school leadership coaching program. Through this five-day program, NISL supports principal supervisors, coaches, or current principals in developing the capacities to be effective coaches of EDP-certified principals in their district. NISL views having a foundation in the EDP as essential to benefit from the coaching, which explicitly and exclusively focuses on applying strategies and concepts from the EDP. The role of the coach is to serve as a thought partner to the principal in the use of EDP content and tools and in the implementation of the principal's ALP. The coaching is to be topically focused on areas NISL has identified in the EDP as promoting student learning. The coach might conduct walk-throughs with a principal, review the principal's ALP, refer principals to EDP tools, review data with the principal, or model for the principal certain actions, such as coaching teachers.

## Theory of Action

For the i3 grant, NISL refined its logic model for the EDP and coaching effects. This logic model is shown in Figure 2.1 and summarized below. As we describe in Chapter 3, we designed our data collection to cover the main domains shown in Figure 2.1.

**Figure 2.1. Logic Model for the EDP and Coaching**



NOTE: PD = professional development.

As the left-most box in Figure 2.1 shows, NISL’s theory of action includes three forms of inputs that collectively provide a principal’s professional development: (1) formal group learning, which refers to the 12-unit EDP, (2) one-on-one coaching from a NISL-certified coach, which helps the principal apply EDP concepts to their ALP, and (3) readings offered by the coach or in the EDP curriculum for the principal to consume during their ongoing development. All EDP participants were also provided with a user account to access online materials that accompany the EDP curriculum. Among the materials are diagnostic assessments that the principal completes during the EDP and that the coach walks through with principals during the coaching sessions. The online materials also include case studies and research articles referenced in the EDP curriculum.

Moving from left to right in Figure 2.1, the three-part professional learning is intended to influence four aspects of a principal’s leadership: (1) their ability to set a compelling vision and goals to reach that vision, (2) their increased capacity as an instructional leader, (3) their creation or improvement of systems such as professional learning communities and alignment of curricular programs across and within grade levels, and (4) their strategic skills to drive change in the building, such as setting performance targets, using data to guide their decisions, and creating leadership roles to distribute leadership beyond just the principal-assistant nucleus.

NISL theorizes that the combination of these leadership changes will improve the day-to-day core functions of the school, as shown in the third box of Figure 2.1. NISL theorizes that the changes in leadership should make instruction throughout the school more coherent and of higher quality and that the culture among teachers and students should also improve.

Finally, the improvements in the culture and instructional quality referenced in the third box should result in higher student achievement, higher academic and behavioral expectations for students, improved student attendance, and decreased student disciplinary actions, as shown in the far-right box in Figure 2.1.

The horizontal box at the bottom of Figure 2.1 shows contextual factors, such as LEA school and principal accountability systems and the principal’s disposition and other traits, that NISL hypothesizes will moderate the theorized effects of the EDP and coaching.

## Prior Research About the Effects of Principal Professional Learning Programs

To situate our research in the existing body of work on professional learning programs for principals, we present a brief literature review in this section. Because of the dual focus of NISL’s work on formal group learning for principals (i.e., the EDP) and one-on-one coaching for principals, we divided our review of prior research into these two topics.

To identify studies for our review of formal group learning for principals, we conducted online searches using the search terms “professional learning” or “professional development” combined with the term “principal.” We limited the review to the most rigorous studies—i.e.,

ones with experimental or quasi-experimental designs—of the effects of formal principal professional learning programs on various outcomes in U.S. schools serving students in any grade level from prekindergarten through 12. We further limited our search to studies published in the past decade (2010–2020). We included only programs designed to further the learning of current principals and excluded initial preparation programs. In addition, we consulted two recent reviews that used Every Student Succeeds Act evidence guidelines or What Works Clearinghouse (WWC) standards (i.e., George W. Bush Institute, 2016, and Herman et al., 2017).

For our review of the effect of coaching on outcomes of interest, we initially conducted online searches for published literature meeting the same criteria as above (i.e., experimental or quasi-experimental studies within the past decade), but using the search term “coaching” with “principal.” We focused specifically on coaching and did not include more-general literature on the mentoring of principals or on principal supervisors, who may or may not act as coaches to principals. Also, we included only studies that centered on coaching as the intervention, excluding formal professional learning programs with a coaching component. Moreover, we were not interested in parsing different models of coaching, such as blended coaching or cognitive coaching. As with the review of professional learning programs, we included only studies involving current principals. Using those criteria, however, the searches yielded very few studies, leading us to relax our requirements (e.g., to accept studies from 2000 to 2009 and studies with less rigorous designs). In our review below, we note when we reference these types of studies.

### *Impact of Principal Development Programs on Various Outcomes*

We found limited research evidence from experimental or quasi-experimental studies of professional development programs for current principals. We first present findings related to effects on student academic achievement, followed by effects on nonacademic school outcomes and, finally, effects on principal practices. In summary, evidence with respect to all three types of outcomes is mixed.

What evidence there is about principal development programs indicates mixed outcomes on *student academic achievement*. Camburn et al.’s (2007) review identified only two rigorously designed studies. We identified four additional qualifying studies published since then, three of which are included in more recent evidence reviews of principal professional development programs (see George W. Bush Institute, 2016, and Herman et al., 2017). We focus our review on these four most recent studies. Two of the four found statistically significant impacts of the program on student achievement, and the other two did not. The two with positive academic effects were studies of the effects of NISL’s EDP in schools where the intervention was participated in with fidelity (Nunnery et al., 2011; Nunnery, Yen, and Ross, 2011). We summarize each of the four studies as follows:

- In the first (Nunnery, Yen, and Ross, 2011), participating schools had math (but not ELA) achievement gains over a three-year period (2006–2009) that were statistically significantly higher than matched schools. In this Pennsylvania study, the state department of education commissioned NISL’s EDP. Importantly, for our discussion of i3 results in the last chapter of this report, there are some important similarities and differences between the i3 grant intervention and study and the Pennsylvania study, as well as the Massachusetts study described in the next bullet.

The similarities are that the Massachusetts and Pennsylvania studies focused on providing NISL’s EDP to novice principals and that the research study periods in all three studies (this i3 study, the Pennsylvania study, and the Massachusetts study) examined student achievement in the three school years after the principal first began the EDP.

There are five substantive differences between this i3 study and the two state-specific studies. The first distinction is that the i3 study included NISL coaching, whereas the Pennsylvania and Massachusetts studies did not. The second distinction is that the Pennsylvania and Massachusetts SEAs, rather than a federal grant, paid for NISL’s services. (In the case of the Massachusetts study, the combination of the state and districts paid.) Third, the Pennsylvania and Massachusetts districts sought volunteer principals to participate in the EDP as part of a districtwide initiative, rather than a randomized controlled trial in which the superintendent assigned only a portion of principals (middle school only) to attend. Fourth, the EDP was 27 days, rather than 24 days, with 13 units rather than 12 and typically spread over 15–18 months rather than 12 months. However, the EDP curriculum covered many of the same topics as the version of the EDP delivered in this i3 study. The fifth difference is that NISL in those years engaged the sponsoring district in a professional development step that preceded the EDP, which was not offered in the i3 study. During this step, NISL typically began engagements with clients via a *Leadership Team cohort* composed of key district leaders and principals, who were selected as facilitator candidates. Leadership Team cohort members went through the EDP as participants but were also provided facilitator binders so they had access to everything certified facilitators had. In this way, some central district staff were fully exposed to the content of the EDP.

- In the second Massachusetts study (Nunnery et al., 2011), from 2007 to 2010, the novice principals in 38 schools participating in NISL’s EDP saw higher achievement gains in both math and ELA than in comparison schools. More specifically, no effect was detected in the first two years, but participating schools by the third year experienced higher positive growth compared with control schools.
- The third is a three-year study of McREL International’s Balanced Leadership Professional Development (BLPD) program, which found no impact on academic achievement (Jacob et al., 2015).<sup>7</sup> In this study, 126 rural public schools in Michigan that included grades three through five (among other grades they may serve) were recruited. Half were randomized into the treatment condition, and the balance were randomized into

---

<sup>7</sup> This same study appears in the George W. Bush Institute, 2016, review under the following reference (p. 60):

Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2014). Exploring the causal impact of the McREL Balanced Leadership Program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis*, 52, 187–220.

a “business as usual” control group. Participants in the program received ten two-day sessions over three academic years (2008–2011). The program content centered on 21 individual leadership responsibilities identified as having a positive relationship with student achievement. No impact of the BLPD program was detected, however, on students’ achievement scores for either reading or mathematics state assessments.

- In the fourth, Herrmann et al., 2019, studied a principal professional development program provided by the Center for Educational Leadership at the University of Washington. The program consisted of four components: a summer institute, group trainings, professional learning communities, and individualized coaching, with an emphasis on instructional leadership. In all, the program consisted of 188 planned hours of professional development over two years. Results of the study of 100 schools, randomized equally into treatment and control conditions, indicated that the program had no effect on average mathematics or ELA achievement in years one, two, or three.

Two of these four studies investigated the impact of principal development programs on *nonacademic school-related outcomes*, and the results were again mixed. One study did not find an effect on principals’ or teachers’ perceptions of school climate (Herrmann et al., 2019), but Jacob et al., 2015, did. In that study, principals in the BLPD program reported a better climate and better collaboration among staff, though teachers’ ratings did not reflect differences (Jacob et al., 2015). Similarly, Herrmann et al., 2019, found no effect in principal or teacher retention, whereas the BLPD program appeared to have had an effect on staff stability (i.e., reduced principal and teacher turnover; Jacob et al., 2015). The WWC review characterized the BLPD program’s effects on school leader retention as “substantively important positive effects” but not “statistically significant” (WWC, 2020b). Specifically, the mean principal turnover after three years was 24 percent at BLPD schools compared with 40 percent at comparison schools.

Although principals participating in leadership development programs felt more efficacious (Jacob et al., 2015; Miller et al., 2016), effects of the programs on *principals’ practices* appear to be limited. According to principal self-reports in one study, principals receiving BLPD had statistically significantly larger pre- to post-gain scores than comparison principals on a majority of constructs on the list of leadership responsibilities principals are expected to perform (Miller et al., 2016). According to teacher perceptions in another evaluation of BLPD, however, principals participating in the program were not deemed more effective leaders (Jacob et al., 2015). The program also did not appear to affect the amount of time principals spent on instructional leadership (Herrmann et al., 2019); in fact, it had some negative effects on their instructional leadership practices, such as providing less frequent and less competent instructional supports and feedback to teachers (Herrmann et al., 2019).

### *Impact of Coaching of Principals on Various Outcomes*

We organize this review of the impact of coaching by the type of outcomes, starting with student achievement, followed by principals’ leadership practices, followed by a focus on instruction. In summary, we did not find convincing evidence that principal coaching boosts

student academic achievement. We did find evidence that coaching influences principals' leadership behaviors and focus on instruction. We conclude this coaching review with research that identifies particular aspects of the coaching relationship as more effective.

The dearth of research examining the direct link between coaching of principals and *student achievement* has been well noted (e.g., Barnett and O'Mahony, 2008; Grissom and Harrington, 2010; Hobson, 2003). We identified no experimental or quasi-experimental study investigating this question. However, studies with less rigorous designs have. For example, Grissom and Harrington (2010) found that, among different modes of administrator professional development (such as university courses, formal principal networks, and individual coaching), only formal mentoring and coaching was positively associated (and statistically significant at  $p < 0.10$ ) with student outcomes. The authors cautioned, however, about a possible selection effect based on the backgrounds of principals who participated in coaching opportunities. Two other studies examining the effect of a coaching program on student achievement using the schools' academic performance index found positive results. The schools of the majority of participating principals experienced growth that exceeded the average targeted growth of the county (Bossi, 2008; Warren and Kelsen, 2013). Finally, in a national survey, Wise and Cavazos, 2017, asked respondents whether, in their opinion, student achievement has grown as a result of leadership coaching. Nearly 72 percent of principals receiving coaching believed so. These principals rated the competencies of their coaches higher than principals who did not perceive achievement growth.

Although also extremely limited, empirical research on the effects of coaching on the *leadership practices of principals* suggests that coaching is linked to a positive effect on principals' behaviors. Again, the vast majority of these studies are not experimental or quasi-experimental in design. We identified four of the most quantitatively rigorous studies in the following bullets. We note that recent qualitative implementation studies of various coaching programs also suggest that coaching enhances principals' abilities to lead school improvement efforts (e.g., Klar et al., 2019; Lackritz et al., 2019).

- Goff et al., 2014, conducted a multiyear, randomized experiment with 52 elementary and middle school principals in an urban district to explore the differential impact of feedback and coaching on principals' leadership behavior and perspectives for two groups of principals. Group 1 was composed of principals who received feedback from teachers alone. Group 2 was composed of principals who received not only feedback from their teachers but also support from a performance-based coach. Principals in the second group participated in an average of eight sessions of coaching. The researchers found that the Group 2 condition—teacher feedback and coaching together—improved principals' ability to support teachers' leadership development, as reflected in actions such as sharing feedback results with teachers and discussing their leadership and goals with teachers. This "Principal Leadership Development" construct was measured by seven items on a survey. The items, anchored in national leadership standards and the Vanderbilt Assessment of Leadership in Education, were designed to operationalize behaviors that leaders might engage in as a result of receiving feedback and/or coaching

related to instructional leadership (Goldring et al., 2009; Murphy et al., 2011). Coaches, all of whom had previously been principals and were trained to be a coach, were particularly instrumental in facilitating principals' self-reflection. They helped principals interpret the feedback they received from teachers into action, clarify and prioritize issues that needed to be addressed in their schools, and initiate change, among other benefits. This study did not find, however, evidence of effects of coaching on principals' efforts to support individual teachers' instructional development. The authors conclude that although more research is needed to understand the dosage of coaching required to produce desired effects, their study suggests that coaching is a viable and potentially cost-effective method for principals to further develop leadership competencies.

- Wise and Hammack, 2011, conducted a survey study that involved 65 principals with a coach who rated a set of 20 competencies about their coach, such as "The coach establishes a specific, results-oriented coaching plan" or "The coach knows when to push me and under what conditions." The 65 principals also rated the extent to which the coaching they received "affected the presence and/or implementation" of a set of nine best practices, such as school leaders emphasizing continuous improvement, holding teachers accountable for supporting students, and teachers providing differentiated instruction. The authors calculated the mean rating across the nine competencies and correlated that rating with each of the 20 competencies. All correlations were moderate and statistically significant, ranging from 0.396 to 0.729, suggesting that coaching and the enactment of best practices are likely related.
- Similarly, Warren and Kelsen, 2013, had 18 principals in urban underperforming schools rate their ability to carry out each of nine leadership responsibilities related to "building a purposeful community" prior to working with a coach and after working with a coach and asked whether the coach supported the development of the leadership behavior. Results showed that all nine leadership responsibilities showed significant growth and that principals recognized the efforts of their coach in their progress.
- Grissom and Harrington, 2010, used secondary data to investigate the relationship between the mode of administrator professional development (i.e., university courses, formal principal networks, coaching) and teachers' perceptions of principal effectiveness. Principals who participated in a formal mentoring or coaching program were rated significantly higher by teachers than principals who did not. According to the authors, the magnitude of the effect suggests "that participating in mentorship [or coaching] is worth about 7 percent of a standard deviation on the principal effectiveness scale, a moderate relationship" (Grissom and Harrington, 2010, p. 601). Again, caution about possible selection bias in this study is warranted.

Research suggests a positive relationship between coaching and principals' *focus on instruction*. Principals who are coached appear to spend more time working with teachers on their instruction (Strong, Barrett, and Bloom, 2003). For example, research from the New Teacher Center at the University of California Santa Cruz suggests that beginning administrators participating in just-in-time individualized leadership coaching for three to six hours a month from a certified School Leadership Coach are "more proactive and focused on systemic instructional issues than those who do not receive such support" (Bloom, Danilovich, and Fogel, 2005, p. 31). And findings about the Metropolitan Independent School District Principal

Coaching Initiative, which provides coaching for novice and experienced principals, suggest that coached principals became more learner-centered (Lee, 2010).

To date, few experimental studies have been conducted to investigate what particular *characteristics of principal coaching* make it effective for principals and their practice. Therefore, questions about the dosage and frequency of coaching, mode of coaching, coaches' background, and more remain unanswered.

There are, however, some useful insights from other studies about professional development more generally, of which coaching is one type. Desimone and Pak, 2017, distilled five research-based features of high-quality professional development:

- content focus (activities are focused on subject-matter content and how students learn that content)
- active learning (opportunities for participants to observe, receive feedback, and engage actively)
- coherence (content, goals, and activities of professional learning are consistent with the school goals, teacher knowledge and beliefs, the needs of students, and contextual reforms and policies)
- sustained duration (professional development is ongoing and includes at least 20 hours of contact time)
- collective participation (participants from the same grade, subject, or school participate together as an interactive learning community).

The authors applied these features as concepts for the instructional coaching of teachers.

In a study related to Goff et al., 2014, Huff, Preston, and Goldring, 2013, examined quality of coaching. Profiling two principal-coach pairs, they noted that the dosage between the two pairs was similar; however, one principal demonstrated “far deeper changes in her perceptions and practices” compared with the other (Huff, Preston, and Goldring, 2013, p. 518). The difference seemed attributable to a higher quality of coaching, characterized by more-detailed (rather than superficial) discussions. In lower-quality coaching sessions, coaches asked simple questions, largely agreed with what the principal shared, and did little to push for full understanding of an issue. In more-effective sessions, coaches made more-concerted efforts to reframe a principal's thinking or change a principal's perceptions and assumptions. According to the authors, hallmarks of higher-quality coaching include the following key strategies:

- posing targeted questions that require principals to review specific conditions in their schools
- engaging principals in role plays and scenarios requiring principals to explain how they would respond
- revisiting issues and action plans over multiple sessions
- establishing routines in which the coach reviews the principal's progress on an ongoing effort and fields questions or concerns about new issues.

Other qualitative studies have identified what principals regard as characteristics of effective coaching. In a study of a university-based coaching program designed to help new principal

preparation program graduates in their first leadership position, 17 new principal leaders rated the following two aspects of coaching positively:

- having personalized professional support from an experienced administrator
- knowing that the discussions and feedback they had with their coach were focused on their particular context and situation and not simply generic in nature (Silver et al., 2009).

Responses to open-ended questions on a 2010 survey study of 94 principals receiving leadership coaching comported with Silver et al.'s findings and added the following three characteristics: (1) Principals voiced that trust was an essential element in a coaching relationship, (2) they valued coaches who pushed their thinking with probing questions, and (3) they valued coaches who are experts with a deep knowledge base (Wise, 2010).

Finally, researchers synthesizing lessons learned from implementation studies suggest that successful coaching programs engage in rigorous screening, selection, and training of coaches; match coach and principal with leadership experience in similar types of schools; and devote time to building a relationship (Barnett and O'Mahony, 2008; Hobson, 2003). Furthermore, the coach and principal should maintain regular and sustained contact (though no specific time is specified) and not be inhibited by geographic distance.

### 3. Data, Methods, and Analytic Samples

---

In this chapter, we explain the sources of data we analyze and the methods by which we analyze them. We also provide descriptive statistics about the principals, schools, teachers, and students in this study.

#### Data

RAND collected a series of primary and secondary data to examine the implementation and effects of the EDP and coaching for middle school novice principals. We used NISL’s logic model for the EDP and coaching (shown in Figure 2.1) to determine which aspects of implementation and which outcomes to examine in this study. Table 3.1 lists the specific topics we analyzed in the study, by data source. Because two of these topics are covered in the first report, we include a third column in Table 3.1 indicating where to find the results—either in this report or in our companion case study and interviews report. In the methods section that follows, we describe how we developed the survey factors shown in this table.

**Table 3.1. Study Topics by Data Source and Where to Find the Analysis**

<b>Topic</b>	<b>Data Source</b>	<b>Where to Find the Analysis</b>
Participation in the EDP and coaching		
EDP session attendance	NISL attendance records	Chapter 4
Number of coaching sessions	NISL coach logs	Chapter 4
Principals’ perceptions of the EDP and of coaching	Principal survey	Chapter 4
Coaches’ perceptions of coaching	Coach survey	Chapter 4
Leadership practices		
Strategic plan for the school	Principal survey	Chapter 4
Effective leader	Teacher survey	Chapter 4
School has differentiated roles for teachers	Teacher survey, principal survey	Chapter 4
Teachers have input on staffing, curriculum, or professional development	Teacher survey	Chapter 4
Curriculum is aligned and evidence-based	Principal survey	Chapter 4
High priority on use of student assessment data	Principal survey	Chapter 4
School offers high-quality professional development for teachers	Principal survey	Chapter 4
Administrators observe teachers’ classrooms	Teacher survey	Chapter 4
Teachers observe other teachers’ classrooms	Teacher survey	Chapter 4
Teachers receive actionable feedback	Teacher survey	Chapter 4

<b>Topic</b>	<b>Data Source</b>	<b>Where to Find the Analysis</b>
School policy, culture, and practices		
Putting the principal's ALP into practice	Case study interviews	Wang et al., 2019
Positive school culture	Principal survey	Chapter 4
School is unsafe and disorderly	Teacher survey, principal survey	Chapter 4
Teachers endorse the school	Teacher survey	Chapter 4
Educators' self-efficacy	Teacher survey, principal survey	Chapter 4
High-performing teachers are assigned to struggling students	Principal survey	Chapter 4
High-quality academics	Teacher survey, principal survey	Chapter 4
Students' higher-order thinking skills	Principal survey	Chapter 4
Principal focuses on NISL concepts of learning	Principal survey	Chapter 4
Personalized instruction for students	Principal survey	Chapter 4
Teachers collaborate about matters of instruction and professional development	Teacher survey, principal survey	Chapter 4
School assigns coaches to low-performing teachers	Principal survey	Chapter 4
Student outcomes		
Mathematics and ELA outcomes	State administrative data	Chapter 4
Student attendance	State administrative data	Chapter 4
Grade progression	State administrative data	Chapter 4
Contextual factors that moderate implementation		
District priorities and alignment with the EDP and coaching	Case study interviews	Wang et al., 2019
Principal, coach, and district characteristics that moderated EDP or coaching efficacy	NISL state coordinators	Chapter 5
Principal has curricular autonomy	Principal survey	Chapter 4
Principal has autonomy to deploy school resources	Principal survey	Chapter 4

NOTE: Unless otherwise noted, all survey data described in this report are from the final, not the first, survey administered to the respondent group—e.g., the final principal survey.

In Table 3.2, we list the timing of data collection and the analytic sample size for each data source. Following the table, we describe how we collected these data.

**Table 3.2. Types of Data Collected and Timepoints for Collection**

Type of Data Collected	Analytic Sample Size (and Response Rate, if Relevant)	When the Collection Occurred
Surveys		
First coach survey, Cohort 1	27 (84%)	Mar 2017 to May 2017
First principal survey, Cohort 1	118 (57%)	Oct 2017 to Dec 2017
First principal survey, Cohort 2	27 (55%)	Mar 2018 to May 2018
Second, final coach survey, both Cohorts 1 and 2	15 (50%)	May 2018 to Sep 2018
Second, final principal survey, both Cohorts 1 and 2	159 (65%)	Jan 2019 to Apr 2019
First and only teacher survey, both Cohorts 1 and 2	1,645 (62%)	Jan 2019 to Apr 2019
Interviews		
Case study interviews of principals, coaches, district leaders, and teachers	9 case studies in total, 4 of which are from i3 study	Spring 2017 to fall 2018
Phone interviews of i3 principals	17 (43%)	Dec 2018 to Feb 2019
NISL state coordinators and national NISL staff	4 (100%)	Feb 2020
Documents		
NISL study roster and records documenting the status of schools, principals, and districts in the study	332	Spring 2015 to Jun 2019
EDP attendance records	166 (i.e., each treatment school)	Cohort 1: Sep 2015 to Aug 2016 Cohort 2: Nov 2016 to Oct 2017
NISL coach logs <sup>a</sup>	2,664 logs entered in electronic logging system	Cohort 1: Jan 2016 to Jun 2018 Cohort 2: Jan 2017 to Jun 2019
Case study school implementation documents	9 schools	Spring 2017 to fall 2018
Student outcomes		
Academic achievement data (see Table 3.3 for details)	63,337 students with outcome data in 2018–2019	School year 2014–2015 through 2018–2019

NOTE: The original research study team conducted additional surveys of some of the i3 participating study schools during the period of fall 2015 to spring 2016. These surveys were of some schools' principals, teachers, and students and were conducted at timepoints that, in most cases, postdated the start of implementation. We do not include these data, given that findings from shortly after the treatment began were less helpful than subsequent surveys conducted after the EDP and coaching had time to be fully implemented. The survey response rate is the number of completed surveys divided by the number of eligible, invited individuals. Some schools and districts opted not to participate in surveys.

<sup>a</sup> NISL required its coaches to log each communication, although enforcement became more routine once NISL transitioned to an electronic log system. During the study period, NISL coaches transitioned from paper to electronic coaching logs. Effective January 1, 2017, coaches used the NISL portal only to fill out a digital log after each meeting (whether phone, web, email, or face-to-face) to document the length of the session, summarize the topics covered, and summarize next steps. In the first year of the i3 study, coaches filled out paper logs until August 2016, at which point some of the coaches transitioned to electronic logs. The final coach log file analyzed for this study includes data from both the electronic and transcribed hard copy coach logs.

## *Achievement Outcome Measures and Testing Context by State*

Our primary outcomes of interest were students’ ELA and mathematics statewide test scores on tests administered in grades six through eight. Schools included in this study were classified as middle schools because they all included grades six through eight, though some also included other grade levels. Although some students in participating schools attended grade levels lower than six or higher than eight, we limited our focus to grades six through eight. More specifically, our analyses focused on a single cohort of students who were in grade six as of the first school year following randomization and were expected to be in grade eight as of the third school year following randomization. Across states, achievement data were available for all students and schools in our analytic sample from at least school year 2014–2015 through school year 2018–2019.

Table 3.3 provides a summary of the statewide exams included in this study. We hypothesized that the EDP and coaching might affect any tested subject area because the interventions did not have a particular subject-area focus and instead emphasized overall school improvement. Therefore, we also examined (in a separate exploratory analysis) specific test outcomes that single states administered, such as science, algebra, and geometry.

All three states in this study adopted either the Common Core State Standards (CCSS) or grade-level standards that are highly aligned to CCSS. And all the exams considered in this report were designed to be aligned with the state’s standards. In States A and B, the statewide testing regime was consistent between the 2014–2015 and 2018–2019 school years, whereas the testing regime in State C was consistent between school years 2015–2016 and 2018–2019, but a different series of tests was used in school year 2014–2015 (the year prior to random assignment).

**Table 3.3. Summary of Statewide Achievement Exams, by Subject, State, and Grade Level**

<b>State</b>	<b>Grade Levels Analyzed</b>	<b>Subject-Area Grouping</b>
State A		
Mathematics	6–8	Mathematics
ELA	6–8	ELA
State B		
Mathematics	6–8	Mathematics
ELA	6–8	ELA
Algebra 1	7 and 8	Mathematics
Geometry	7 and 8	Mathematics
State C		
Mathematics	6–8	Mathematics
ELA	6–8	ELA
Science	8	Science

SOURCES: State data files of student achievement.

## *Attendance and Grade Progression Outcome Measures*

In addition to analyzing student achievement outcomes, we examined whether the intervention affected student average daily attendance rates and the rate of on-time grade progression. We defined *on-time grade progression* as the proportion of the study students who were initially in grade six in the first study year and who were enrolled three years later in grade eight. Our analysis used available administrative records provided by the states.

## Survey Administration

In this section, we describe how we administered the principal, teacher, and coach surveys that we analyze in this report. In addition to inviting all participating principals from both the treatment and control groups to take our survey, we invited up to 17 teachers per school (again, both treatment and control schools) to take our teacher survey.<sup>8</sup> We randomly sampled teachers to invite from teacher rosters we collected from participating states and districts or directly from school leaders. Both the teacher and the principal survey took approximately 30 minutes to complete. Because we used the same procedures for the principal and the teacher survey and fielded them at the same time, we combine the administrative details for these two surveys in the next two paragraphs.

To field the principal and teacher surveys, we first sent an email to each superintendent about a month prior to the survey re-explaining the study and previewing the upcoming survey timeline. Then we mailed a postcard to each principal, timed to arrive a week ahead of the survey launch and telling them to look for surveys that would arrive in hard copy and via email at their school. Over the course of several months, we sent each survey participant up to five emails, as well as three sets of hard-copy letters mailed to their schools inviting them to take the survey electronically or on paper. The large majority of the participants took the survey electronically rather than on paper—e.g., 85 percent of principals who took our final survey in winter 2019 took it electronically, as did 90 percent of teachers. Regardless of the mode in which they completed the survey, we pooled the answers from the identical online and hard-copy surveys for our analysis.<sup>9</sup>

In our first hard-copy mailing to each principal and each teacher, we included a \$10 pre-incentive gift card, and we offered a \$20 post-completion gift card as an incentive to respond. Prior to mailing the third and final hard copy of the survey to nonrespondents, we also placed phone calls using a call script to principals at each school where more than half of the teachers we invited from that school had not yet responded.

---

<sup>8</sup> A few districts declined to allow principals or teachers to participate in surveys.

<sup>9</sup> The rate at which respondents in the treatment and control groups responded via mailed versus electronic surveys was virtually identical.

We administered the coach survey electronically, with no accompanying hard-copy version. It took approximately 15–30 minutes to complete, depending on the number of principals a given coach worked with. With each wave of the coach survey, we first sent NISL an email to forward to their i3 coach employees to explain the purpose, benefits, and confidentiality of the upcoming survey. We then emailed each coach directly with a link to the coach survey, sending up to five emails in total to nonrespondents. To coach respondents, we emailed a \$20 gift code to thank them for their time.

## Analytic Methods

### *Measures of EDP Participation*

To assess the proportion of originally assigned principals to the treatment and control group who attended the EDP, we drew on two data sources: NISL’s EDP attendance tracker and NISL’s study roster documenting interactions with all principals and districts in the study. The first file allowed us to determine which of the 12 EDP units each i3 study participant attended. NISL’s roster details information about attrition from the study or crossovers from random assignment and allowed us to document reasons that individuals did not participate in the EDP and in coaching.<sup>10</sup>

### *Measures of Coaching and Analysis of Coaching*

We used the electronic logs that the 34 NISL i3 coaches created to determine the proportion of the principals assigned to the treatment group who ultimately participated in one or more sessions of coaching, the number of coaching sessions each individual had, and the topical focus of those sessions.

The i3 treatment group principals’ winter 2019 survey responses were our data source for their perceptions of the EDP, their ALP, and coaching. We also thematically coded their open-ended written answers to the following two questions: “What is the single concept from EDP that you have used the most in your school” ( $N = 41$  responses) and “In a sentence or two, write one specific example of how you have applied that concept in your school” ( $N = 40$  responses).<sup>11</sup>

---

<sup>10</sup> In addition to treatment-assigned schools that did not participate in the EDP and/or coaching, there were five control-assigned schools that received exposure to the EDP through the grant. Four were schools where an EDP-certified principal from a treatment school transitioned into a control-assigned school during the study period, and one was a school that purchased the EDP instead of waiting to receive it.

<sup>11</sup> As described in the technical appendix for Wang et al., 2019, to analyze responses to open-ended interview questions, we coded the interview transcripts in Dedoose. Two of the researchers who collected the data reviewed interview notes in the Excel spreadsheet and select interview transcripts and performed open coding of emergent themes related to the value of the EDP, the value of coaching, the most-beneficial coaching activities, school improvement activities, and suggestions for improving coaching and the EDP. The researchers clustered these emergent themes into coding schemes. The two researchers met with the larger research team to discuss the emergent themes and draft codes, refine rules around their use, and reach consensus about how the codes should be

We supplemented these with data from phone interviews of 17 principals who participated in both the EDP and coaching. For these 30-minute interviews, we used a semistructured interview protocol that is included in the appendix of Wang et al., 2019, p. 4:

The interview protocol asked principals to rate their satisfaction with the EDP, coaching, and the ALP; to explain those ratings; to reflect on top lessons learned from the EDP and coaching; to recall the main areas for school improvement that they worked on during the EDP; and to explain how, if at all, the EDP and coaching influenced their work.

A team of three researchers conducted the interviews, which were audio recorded and later transcribed. The researchers also made notes in an Excel spreadsheet during the interviews to track principals' responses to close-ended items (e.g., years of experience as a principal, years participating in the EDP, level of agreement with statements about the EDP and coaching) to facilitate analysis.

### *Measures of School Leadership, Culture, and School Instructional Practices*

We drew our measures of school implementation from the winter 2019 surveys of principals and teachers from both the treatment and control groups.

To create the survey, we first reviewed several candidate surveys, such as The 5 Essentials, The Teaching, Empowering, Leading and Learning survey, Schools and Staffing Survey, and the TALIS survey, to include validated items in our surveys of principals and teachers. However, we found they did not directly capture the concepts covered in the EDP (see NISL's logic model in Figure 2.1), so we wrote the vast majority of the items on the principal and teacher survey and included some isolated items from prior surveys and one full survey scale: the unsafe and disorderly school scale (Organisation for Economic Co-Operation and Development, 2010). All survey instruments are available upon request. The principal and teacher surveys covered the following topics:

- leadership practices
- school culture
- school safety
- standards, curriculum, instruction, and assessment
- teacher collaboration, professional development, and leadership opportunities
- principals' views of the EDP and coaching (for principals assigned to the EDP or coaching only)
- respondent's background characteristics.

To distill the i3 teacher and principal survey data, we applied the survey scales that we had developed several months prior from identically worded surveys of principals and teachers for a second RCT study (funded by a Supporting Effective Educator Development [SEED] grant) that RAND researchers are conducting for NISL in regard to the EDP and coaching. We did this

---

applied. During coding, the team of three researchers had ongoing conversations about how text should be coded, modifying the coding scheme and discussing questions until consensus was reached.

because the SEED study had larger samples of principals ( $N = 367$ ), teachers ( $N = 5,161$ ), and schools ( $N = 654$ ) than the i3 study and because we had intentionally fielded identical surveys in both studies to allow for subsequent pooled analyses.<sup>12</sup>

When analyzing the SEED survey results, we examined the dimensionality and internal structure of the survey results and we performed exploratory factor analyses (EFAs). The primary objective of these analyses was rank reduction (Alwin, 1973; Bollen and Lennox, 1991; Cronbach, 1976): Specifically, we used factor analysis as our primary approach to justify the creation of summary scales, reducing the number of variables that would be included in subsequent analyses. To the extent possible, we sought to form scales that included items that were related in ways that were consistent with theory and interpretation. We describe these analyses below. Each of the survey scales is listed in Tables A.1 through A.25 in the appendix.

We describe here the EFAs we conducted with the SEED grant surveys to develop the survey scales, which we also used with the i3 surveys. Both the SEED and i3 principal surveys contained 82 items, of which we included 80 items in our analyses. The SEED and the i3 teacher survey contained 91 items, of which we included 83 items in our analyses. We did not include items in our EFA for which respondents could select all that apply (e.g., “Which of the following resources does your school provide for students who are not succeeding academically. Mark all that apply.” or “What are the main criteria used to assign teachers to leadership roles in your school?”), and we did not include yes/no items (e.g., “Are low-performing teachers assigned a coach to strengthen their practice?”). For both the principal and teacher surveys, the items were all Likert-type items and were generally scored on four-point scales, though the scales had a variety of anchors (e.g.,  $1 = \text{None of the time}$ ,  $4 = \text{All of the time}$ ;  $1 = \text{Not at all}$ ,  $4 = \text{To a great extent}$ ).<sup>13</sup>

Our approach to EFA was iterative and driven by both empirical and theoretical considerations. We began by inspecting eigenvalues and scree plots and by conducting parallel analyses to determine a range of factor extractions. We used maximum likelihood factor analysis and an oblique rotation (geomin) method so that the extracted factors were allowed to correlate freely. We then examined the parameter estimates from these rotated solutions to determine the number of factors to retain. Ultimately, decisions about which solution to retain were based on

---

<sup>12</sup> Although the number of principal respondents from the i3 study is adequate according to some guidelines in the literature (see Comrey and Lee, 1992, and Tabachnik and Fidell, 2001), larger samples reduce sampling error and improve the quality of inferences.

<sup>13</sup> Conventional factor analytic techniques were developed using scales with common response anchors and uniform scale steps. Our models do not accommodate these differences in item format. However, all the items were constructed using a similar interpretation: Higher agreement or more prevalence is consistent with higher-quality practices. Because EFA is based on correlations among items, and the maximum likelihood estimator treats all items as continuous, we do not believe that differences in the number of scale steps is problematic here (see Muthén and Kaplan, 1985), particularly given that the primary purpose of the factor analysis is rank reduction. Models that can accommodate these differences in item format rely on more-complicated statistical machinery and, accordingly, larger sample sizes to provide appropriate inferences.

criteria recommended by Pedhazur and Schmelkin, 1991: Every item had a high loading on only one factor (i.e., we attempted to minimize cross loadings), and each factor had high loadings for only some items. We also attempted to ensure that the retained factors were conceptually clear and substantively interpretable. All factor analyses were conducted in Mplus.

The teacher survey data were slightly more complex in structure than the principal survey data, because the data were hierarchical in structure: Teachers were clustered within schools.<sup>14</sup> Because of this structure, survey responses could be used in two distinct ways, and there could potentially be two levels of analysis that were of substantive interest. First, the teacher responses could be used to analyze the individual perceptions of teachers. Second, the individual teacher responses could be aggregated to yield measures of the shared perception of the school environment (Lüdtke et al., 2009).

To accommodate the hierarchical structure of the teacher survey data, we used multilevel EFA techniques that allow for the simultaneous extraction of factors both within schools (using a pooled within-school correlation matrix) and between schools (using a school-level correlation matrix). Specifically, we used an approach outlined by Hox, Moerbeek, and van de Schoot, 2010: First, we estimated intraclass correlations to determine the proportion of item variance that is between schools. Then, we examined eigenvalues and scree plots for both correlation matrices separately to determine a number of factors to extract at each level. Then, we used a partially saturated approach to extraction (Hox, Moerbeek, and van de Schoot, 2010; Ryu and West, 2009), in which an unrestricted model (all correlations were estimated directly) was fit at one level and factors were extracted at the other level. The unrestricted model fit the data perfectly, and therefore the partially saturated approach allowed for level-specific evaluation of factor models.

In all, we identified 15 factors from the principal survey and 11 factors from the teacher survey. In Tables A.1 through A.25 in the appendix, we list the full wording and response scales of each of the items in the survey scales we developed.

### *Intent-to-Treat Methodology to Examine Effects of Offering the EDP and Coaching*

To answer our first confirmatory research question (research question [RQ] 4), we used an intent-to-treat (ITT) analysis to identify the effects of offering the EDP and coaching to principals randomly assigned to the treatment condition. The ITT analysis was conducted using two-level hierarchical linear models that nest students within schools. Students were included in our analyses if they were enrolled in grade six in a study school at any time within the first six weeks of the start of the first school year after random assignment occurred. We analyzed each

---

<sup>14</sup> In fact, this hierarchical structure complicates the issue of addressing differences in item format even further. For one, all the variables are continuous at the school level (the unit of analysis). Hierarchical multivariate models that use categorical inputs are largely unexplored in the research literature and demand large numbers of both level 1 (individual) and level 2 (cluster) units. Given these constraints and attending to the fact that the primary purpose of these analyses is data reduction, we have proceeded with a conventional multilevel EFA framework.

state separately, because each state administered different student achievement tests, available control covariates differed by state, and study participants' compliance rates also differed by state.

These models were estimated using the *xtmixed* procedure in Stata. Our overarching approach was to compare outcomes for schools assigned to the treatment and control conditions, while controlling for pretreatment observable student and school characteristics. These covariates were design factors in the selection of school pair blocks in the randomization. Therefore, the models did not explicitly control for school pairs to preserve degrees of freedom, which is an appropriate strategy in block randomized designs such as this one (Raudenbush, Martinez, and Spybrook, 2007) and is consistent with WWC standards (WWC, 2020a).

The general form of the model for research question 4 can be expressed as follows:

**Level 1 (Students):**

$$Y_{ij} = \beta_{0j} + \sum_{n=1}^N \beta_{nj} X_{nij} + \varepsilon_{ij}$$

**Level 2 (Schools):**

$$\beta_{0j} = \gamma_{00} + \gamma_{01} T_j + \sum_{m=1}^M \gamma_{0(1+m)} Z_{mj} \sum_{m=1}^M \gamma_{0(1+m)} Z_{mj} + \omega_{0j}$$

$$\beta_{nj} = \gamma_{n0} \quad n = 1, \dots, N$$

where

- $Y_{ij}$  = state test score of student  $i$  in school  $j$  (standardized separately for each school year, state, subject, and grade level combination).
- $X_{nij}$  =  $n$ th student-level covariate, including grade, baseline-year fifth-grade test scores in both reading and mathematics, demographics (gender, race, and ethnicity), socioeconomic status, disability status, and English language learner status.
- $T_j$  = treatment indicator for school  $j$  (set to 1 if the school principal was randomly assigned to receive the EDP and aligned coaching and set to 0 otherwise).
- $Z_{mj}$  =  $m$ th school or principal-level covariate, including school-level averages of demographics socioeconomic status, disability status, English language learner status, and school averages of prior year test scores. We also include school enrollment, average prior-year daily attendance, and suspension rates.
- $\varepsilon_{ij}$  = student-level residual assumed to be normally distributed with mean of 0 and variance of  $\sigma_{\varepsilon}^2$ .
- $\omega_{0j}$  = school-level residual assumed to be independent of the student-level residual and normally distributed with mean of 0 and variance of  $\sigma_{\omega}^2$ .

The parameter of interest that captures the ITT effect is given by  $\gamma_{01}$ . As explained in more detail below, we estimated separate models for each subject of interest in each state and cohort

combination (as applicable).<sup>15</sup> In addition, because not all students in the initially randomized cohort progressed at the same pace across grade levels, we pooled scores across all grade levels tested in that subject in each state. To facilitate this, the model included controls for grade-level fixed effects, and test scores were standardized for each school year, state, subject, and grade-level combination.<sup>16</sup>

For all exams listed, we standardized the student test scores within the school year and test instrument across our available sample of test-takers. In State C, the available sample included all test-takers statewide, whereas we standardized scores in States A and B across the study-specific students for whom we received data. In State B, most students took the generic statewide mathematics exams in grades seven and eight, but some students instead took the more advanced subject-specific algebra and geometry exams instead; for these exams, we have data on the levels (1 to 5) that students achieved.<sup>17</sup> We used data from all three mathematics exams in State B to arrive at a single, standardized math exam score for each student in that state. Specifically, we standardized outcomes for each exam separately and then populated an overall mathematics test score for each student using the standardized score for whichever exam they took, favoring the more advanced exams in very rare cases in which a student took more than one exam. Versions of tests that were designed for students with special needs were not included as outcomes in our analysis, and students who take only these exams are missing outcome data and are counted as attritors in our analysis.

We also conducted a series of exploratory analyses. First, for all research questions, we examined state-by-state effects. Second, we examined heterogeneity in the effect estimates by school and student characteristics, including school and student baseline average achievement and attendance levels, demographics, socioeconomic indicators, and school size. To do this, we used a machine-learning technique developed by Athey and Imbens (Athey and Imbens, 2015; Athey and Imbens, 2016) and implemented by Davis and Heller (Davis and Heller, 2017; Davis and Heller, forthcoming). This approach used many regression trees that partitioned the data by the student- and school-level covariates to build a flexible model of the outcome as a function of the covariates. Each tree was built using a random subset of the data to maximize heterogeneity in the estimated effects across nodes. The trees made up a causal forest, which was used with another random subset of the data (which was not used in the tree generation process) to predict

---

<sup>15</sup> In a subsequent subsection, we describe how we pool these state- and cohort-specific estimates to create the sample average effect for a given subject. We assessed the robustness of the results by fitting an alternative model that pooled data across all states and cohorts. This alternative approach produced similar results to those presented here.

<sup>16</sup> In preliminary analyses, we also conducted a specification check in which students who did not reach grade eight in the third year following random assignment were instead excluded from the analysis and treated as attritors. Results were virtually the same as those presented in the report. These results are available upon request.

<sup>17</sup> We examined and found no effect of treatment on the rates at which students took algebra or geometry exams in lieu of regular eighth-grade math exams. Therefore, we assume there is limited risk of bias stemming from treatment-induced variation in students' mathematics exam types.

each student's treatment effect using their covariate values. We then used the predicted effects to portion the data into quartiles and examined average treatment effects and school and student characteristics across quartiles.

A third set of exploratory analyses relates to our survey-measured outcomes about principal leadership practices and school culture and school instructional practices. For these analyses, we used structural equation modeling (Baron and Kenny, 1986; Imai, Keele, and Yamamoto, 2010), in which we decomposed the "total" effect estimates on achievement measures into two components: the indirect effect, which captured the proportion of the total effect that was realized through a given mediator, and the direct effect of the treatment.

### *Treatment-on-the-Treated Methodology to Examine Effects of Offering the EDP and Coaching*

The experimental ITT analysis provided an unbiased estimate of the effects of offering the treatment to study-school principals. However, not all principals took up the offer, and others either left their school or discontinued participation in the EDP and coaching midstream. From a practitioner perspective, an even more important question is the impact of principals actually taking part in the program. To understand the effects of actively participating in the EDP and coaching (RQ 5), we first performed the analyses described here and then performed the quasi-experimental analysis described in the next subsection.

To understand the range of plausible effects from minimal up to complete participation in the EDP and coaching, we first conducted an instrumental variables analysis, in which we used random assignment to receive the EDP and coaching as an instrument for the receipt of treatment. We refer to this as the treatment-on-the-treated (TOT) analysis. A TOT analysis can yield unbiased estimates of the effects of compliance with the offer of treatment if we assume that the schools that never take up the treatment are unaffected by the offer.

However, a TOT analysis is complicated in a context in which schools' degree of compliance with treatment is on a sliding scale, as is the case here in which principals in the treatment group could and did partially complete the EDP and coaching. Using a TOT analysis, we could estimate the effects of treatment on any treatment-group school that participated in *any* amount of the EDP or coaching. But this would not tell us the effects among schools who fully complied with both the EDP and coaching, since it would attribute measured effects equally to partial and full compliers. Put another way, such a TOT analysis provides a lower-bound estimate of the effect of participating in the EDP and coaching.

Alternately, we could take a different approach to the TOT analysis in which we estimate the effects of treatment on any treatment-group school that participated in *all* the EDP or coaching. But this relies on the strong assumption that partially compliant schools experienced no effects of treatment. Performing the TOT analysis in this manner would provide an upper-bound estimate of the effect of participating in the EDP and coaching.

As an initial bounding exercise, we enacted both TOT analyses. The first instrumental variables analysis used receipt of *any* amount of either the EDP or coaching as the endogenous indicator of treatment receipt, which is instrumented by the random assignment indicator. The second instrumental variables analysis used receipt of the full dosage of both the 12-units of the EDP and the 60 hours of coaching as the endogenous indicator of treatment receipt, which is again instrumented by the random assignment indicator. We conducted these analyses using Stata's *xtivreg* module, using the same covariates included in the ITT model discussed above, and accounting for the clustering of students within schools with a school-level random effect. We assessed the robustness of these estimates using a conventional two-stage least-squares specification (implemented using the *ivregress* module in Stata) with cluster-robust standard errors at the school level, which yielded very similar estimates to our preferred specification.

This bounding exercise provided us with presumably unbiased estimates of the range of plausible effects of compliance with treatment: from the most minimal participation up to maximum participation in both the EDP and coaching. We used these boundaries to benchmark our second analysis, which is the quasi-experimental exploration (described in the next subsection) of our preferred definition of full compliance.

### *Quasi-Experimental Methodology to Assess Impacts in Fully Participating Schools*

As a second step to answering our second confirmatory research question (RQ 5), we analyzed the impact of the treatment when schools' principals were fully participating with the intended amount of the EDP and coaching, using nonexperimental methods. For this analysis, we identified *fully participating schools* as schools in which the principal completed at least ten (of 12) EDP units and at least 60 hours of coaching, which aligns with NISL's goals for principals' participation in the intervention. By design, this definition also coincided with the principal remaining in the school for at least two (and usually all three) school years following randomization.

Although NISL encouraged perfect attendance in the EDP and offered make-up sessions to principals, make-up sessions could be difficult for some principals to attend if they were not available sufficiently close to their schools. In addition, NISL found that most principals who attended at least ten units continued to engage with their coaching, their EDP reading and online material, and their ALPs and that a modest number of absences did not typically indicate a lack of overall engagement with the program.

We conducted an additional quasi-experimental comparison of effects by comparing outcomes of students initially assigned to schools that fully participated with those of students initially assigned to control schools whose principals did not take part in any of the EDP or coaching.<sup>18</sup> Unfortunately, this quasi-experimental analysis does not guarantee unbiased

---

<sup>18</sup> Although our quasi-experimental methodology excludes noncompliant control schools from the comparison group, we found in additional sensitivity analyses that this exclusion had no substantive impact on our results.

estimates of the effects of full compliance with treatment, because fully participating schools and principals were not a randomly selected subset of the treatment group but instead represent principals motivated to complete the program. Fully participating schools and students can differ in terms of both observable and unobserved characteristics that might influence student achievement outcomes. Nevertheless, we would expect our analysis of fully participating schools to identify effects within the plausible range of estimates provided by our TOT bounding exercise. If our quasi-experimental results were to exceed this range, that would provide evidence of potential bias in our quasi-experimental effect estimates.

To minimize the potential bias, we conducted this analysis using the same model that was used in the ITT analyses described above, and we included as many observable covariates as we possessed. Specifically, we controlled for school-level covariates, including school-level averages of baseline versions of the outcome measures and demographic characteristics (to account for school-level factors that may have affected principals' compliance decisions), and student-level covariates, including baseline versions of the outcome measures and demographics (to control for student-level factors that may be related to student mobility and variation in exposure to principals with different levels of compliance).<sup>19</sup> Per WWC standards, we also assessed whether the student-level baseline test scores were equivalent between the contrasted treatment and control schools. As shown in Table A.26 in the appendix, overall differences were smaller than WWC's 0.25 standard deviation threshold; therefore, these analyses are expected to meet WWC standards with reservations.

As a further augmentation to the modeling approach described above, we also incorporated a doubly robust estimation strategy (Bang and Robins, 2005). This involved first weighting control schools according to their observed similarities with the fully participating treatment schools and then using these weights in the second level of the two-level hierarchical linear model (given in the equations presented earlier) that explicitly controlled for the observable student and school characteristics. In this analysis, treatment schools were weighted by 1, and comparison schools were weighted by  $\frac{\hat{P}}{1-\hat{P}}$ , where  $\hat{P}$  is the estimated propensity score (Stuart, 2010).

In this case, the propensity score captures the probability of being a fully participating treatment school conditional on observable school characteristics listed above. We estimated the propensity scores using generalized boosted methods (GBM; McCaffrey et al., 2013). GBM combines boosting (i.e., iterations) and regression trees (which partition the data set into numerous regions according to the covariate values). It is data-adaptive and nonparametric. It uses many piecewise functions of the covariates and automatically selects which covariates

---

<sup>19</sup> We also explored including district indicators in these models (to account for differences in compliance patterns across districts), which yielded coefficient estimates of a similar magnitude to those from models that do not include these indicators. We used the latter as our primary specification because districts that do not have fully participating treatment or control schools do not contribute to the estimation of effects in the models with district indicators, which might therefore address a slightly different research question than RQ 5.

should be included and the best functional form, including testing all possible interactions to achieve the best balance between the treatments and comparison units. We implemented GBM using the *twang* package in Stata (Cefalu, Liu, and Martin, 2015).

Notwithstanding this doubly robust estimation strategy, there could be remaining unobserved principal-, school-, or student-level factors that are related both to our outcomes of interest and to variation in compliance rates across principals. For example, schools that are more (or less) successful at raising student achievement might be more likely to have principals who also complete the program. Therefore, these nonexperimental findings should be interpreted with caution.

### *Pooling Results Across States*

For both our ITT and our quasi-experimental analyses, our prespecified confirmatory contrasts of interest were our pooled findings across all three state samples. This allowed us the best-powered tests possible for our two primary research questions. Because different states had substantive differences in their available covariate controls, we estimated the effects of treatment separately within each state. We then pooled these state-specific estimates using fixed-effects meta-analysis methods in which each state's estimate is weighted proportional to its precision (Cooper, Hedges, and Valentine, 2009). To better understand potential heterogeneity in the effects of the intervention, we also examined state-specific effect estimates, but we consider these to be exploratory subgroup analyses.

### *Analysis of Impacts on Survey Outcomes*

When analyzing the effects of the EDP and coaching on survey-reported outcomes, we used a similar approach as we did for our analyses of student achievement outcomes, analyzing effects in an experimental, ITT framework and in a nonexperimental comparison of survey responses from fully participating schools relative to survey responses from control schools. However, because all our survey outcomes were available at the school level only, these analyses did not require a multilevel approach.<sup>20</sup> First, to partially account for potential bias related to survey nonresponse, we generated weights for the teacher and principal survey analyses that are based on the estimated likelihood of each school sampled having teacher or principal survey responses. Propensity to participate in the survey was estimated as a function of schools' characteristics and of the experience levels of the school principal at the time of random assignment. These characteristics provided limited predictive power with respect to observed response rates.

---

<sup>20</sup> In addition, we focused exclusively on responses from those principals who participated in the intervention in our nonexperimental analysis of principals' survey responses at fully participating schools, since our goal when examining principal survey responses was to explore whether the intervention had direct effects on participants. When considering teacher survey responses at fully participating schools, however, we included all fully participating schools, even in cases in which the principal who had participated in the EDP and coaching was no longer working at the school.

Nevertheless, the weights helped to somewhat better align the analytic sample and inferences with the observable characteristics of the total sample of schools invited to participate in the survey.

In both the teacher and principal survey analyses, we conducted a regression analysis that included the response weights to compare survey outcomes for the relevant treatment and comparison samples. We analyzed each principal or teacher survey factor in a separate regression and included controls for school covariates defined in the pretreatment baseline period, including the percentage of students at the school who were eligible for free lunch, the average achievement levels of students at the school in ELA and mathematics, school average absences rates, total school enrollment, and the percentage of students at the school in different racial and ethnic groups. We used these specific covariate controls for inclusion in the model because they were the most consistently associated with survey outcomes.

A key limitation of our analyses of survey measures is that, unlike in our analyses of student achievement, we lacked any baseline measure of survey-reported outcomes to include as covariates. In theory, the lack of a baseline control should affect the precision of, but should not bias, the estimated effects obtained in the experimental ITT framework. However, our analysis comparing fully participating schools with control schools is subject to greater bias and relies more on controls for observable covariates to account for preexisting differences between full compliers and other schools.

Notwithstanding our use of survey nonresponse weights, our survey results could still be biased because of survey nonresponse. We observed similar response rates for teachers and principals across treatment and control schools, which suggests that treatment itself did not bias the rates of response. Even so, our sample of responding staff and schools may not be representative of the overall study sample on unobservable dimensions. Even if survey nonresponse did not bias our comparison of treatment and control schools, it could mean that our findings do not apply equally to all types of schools in the study.

## Missing Data

A modest level of missing student and school-level observations were missing at the item level (i.e., individual variable values were missing for an observation that was otherwise present in the data). We accounted for missing data in our analytic models using a two-stage process: preliminary data cleaning followed by a formal multiple imputation of remaining missing values. In the data-cleaning phase, missing items that were informed by a logical progression observed in the data were filled in using the logical information. In particular, a student with a standard grade progression and missing grade information in one year (e.g., {5, missing, 7, 8}) would have the missing grade filled in according to the others (grade six in this example). The data were inspected for any such logical patterns that would fully explain the missing value, and such cases were filled in.

After preliminary data cleaning, less than 1 percent of the total item observations were missing.<sup>21</sup> To account for the remaining missing data, we implemented a multiple imputation through chained equations algorithm (MICE) implemented using the *mice* package (van Buuren and Groothuis-Oudshoorn, 2011) in the R statistical computing environment (R Development Core Team, 2019). The chained equations approach uses Gibbs sampling (Geman and Geman, 1984) to draw imputed values from a prediction model for each missing variable, consistent with the variable's type (e.g., continuous, discrete) and range of support, using all other available variables as the predictors. It passes through each variable with missing values iteratively, using the most recently predicted missing values for any other variables. The *mice* R package was selected because its functionality accommodated the hierarchical nature of the data in developing a prediction model for each student-level and school-level variable. The package also features diagnostic plots to inspect that the Gibbs sampling is functioning properly and that the imputed draws are reasonable given the distribution of the observed values. We drew ten imputations for each missing observation, creating ten imputed versions of the data set for analysis. All covariates used in the analytic models for our primary outcomes, including an indicator variable of treatment status, were used in the imputation process. Outcome variables were also included in the imputations, with missing outcomes reset to missing in the imputed data sets used for analyses (i.e., we do not use imputed missing outcomes in our treatment effect estimates).

Models estimating the treatment effect on each outcome of interest were run separately on each of the ten imputed data sets. A final treatment effect estimate and standard error for each outcome was then aggregated from the ten individual estimates using Rubin's (1987) method.

## Adjustments for Multiple Comparisons

For each of our confirmatory contrast research questions (RQs 4 and 5), we conducted only one test within each subject area domain (ELA or mathematics) in our pooled sample results. We treated these as tests of distinct outcomes in each domain, consistent with WWC guidelines. It was a somewhat gray area, however, regarding whether to count our tests separately across RQs 4 and 5, particularly since our samples (in this case, for treatment-assigned versus fully participating schools) were heavily overlapping, and estimated effects were likely to be highly correlated. Because each contrast considers a conceptually different research question with different practical and policy implications, we believed that it was most informative to count comparisons separately within each of our research questions.

Separately, our analyses included both state-specific exploratory analyses and analyses of impacts on various survey factors. Each of these analyses was designed to inform hypothesis

---

<sup>21</sup> We use the term *item observation* to denote a single variable observed within a single student in a single year. Most missing observations were at the student level, and the most common variables to experience missing values were baseline math and ELA scores. The missing item observation count is at the student level; if a school-level variable was missing, it was counted as missing for each student in the school.

generation around interpretation of our primary findings. For cases in which some findings related to a particular exploratory research question were statistically significant, we conducted multiple comparisons adjustments (Benjamini and Hochberg, 1995) to provide readers some guidance as to the degree of potential for false-positive findings. However, for the purposes of hypothesis generation, results that do not meet this rigorous standard but are directionally suggestive may still aid our interpretation of our primary findings. The multiple exploratory comparisons we conducted in each research question were generally unlikely to be completely independent tests; therefore, traditional adjustments for multiple comparisons may be overly conservative.

## Analytic Samples

In this section, we provide descriptive statistics about the schools, principals, teachers, and students that we analyze. We compare our analytic samples across states and across treatment and control conditions in the study. We also summarize the degree to which there was attrition of students and schools from our original study sample over time.

### *School Leader and Teacher Survey Respondent Samples*

Among the 159 principals who took our survey, which was conducted more than two school years after the treatment group started the EDP, 63 percent self-reported as white, 25 percent as African American, 9 percent as Hispanic, and 2 percent as Asian. The average number of years of leadership experience was 5.5, with 3.8 years in the current school. The average principal taught for 10.6 years before transitioning into school leadership. Characteristics of principal respondents were similar between the treatment and control schools.

Among the 2,654 teachers who took our survey, 67 percent self-reported as white, 17 percent as African American, 11 percent as Hispanic, and 4 percent as Asian. Females constituted 78 percent of the sample. The average number of years of teaching experience was 14.4, with 9.0 years in the current school. 23 percent were self-contained classroom teachers, and 71 percent were subject-specific teachers. Many teachers taught multiple subject areas, with 40 percent teaching subjects including ELA, reading, writing, or literature; 35 percent teaching mathematics, 28 percent teaching science, and 29 percent teaching social studies.

### *Descriptive Summary of Student Analytic Sample*

As previously discussed, our analysis of the effects of the EDP and coaching on student achievement outcomes focused on the initial cohort of sixth-grade students present in study schools in the first school year following randomization. For these students, we observed achievement outcomes over time and typically across their sixth, seventh, and eighth grades over the course of three school years, and we consider outcomes in study year three to be our primary test of the impacts of the intervention on student outcomes.

Table 3.4 shows the observable characteristics of students and schools, overall and by state and cohort, in our analytic sample. Overall, the students in our study sample were primarily nonwhite students (68 percent), and most of the study students came from low income backgrounds and were therefore eligible for free or reduced-price lunch or categorized as having lower socioeconomic status (72 percent). Study students looked fairly similar across states in terms of their eligibility for free lunch but differed in terms of their race and ethnicity, with more black students in study schools in State C and more Hispanic students in study schools in State A, in particular.

Relative to the nation’s public school student population as a whole, 52.3 percent of whom were eligible for free or reduced-price lunch as of school year 2016–2017, this sample reflected a more disadvantaged student population (National Center for Education Statistics, 2017). This corresponds to the i3 grant’s focus on disadvantaged schools and students.<sup>22</sup>

The average total school enrollment (across all grades served, including grades other than six through eight) in our sample of middle schools is 827 students. Schools varied in locale in each state and included both rural and urban districts. District sizes varied, with the largest district contributing more than 20 participating schools to the study. Study schools were smallest in size in State C relative to school enrollments in State A and State B.

**Table 3.4. Student and School Characteristics in Experimental Analyses Analytic Sample, Overall and by State and Cohort**

Characteristic	Overall Study Sample	State A	State B Cohort 1	State B Cohort 2	State C
Students					
Number of students	63,337	19,564	22,319	13,571	7,883
Percentage ELL	0.155	0.238	0.117	0.178	0.022
Percentage SPED	0.123	0.133	0.119	0.121	0.116
Percentage FRPL	0.683	N/A	0.658	0.684	0.757
Percentage SES	0.789	0.789	N/A	N/A	N/A
Percentage FRPL or SES	0.716	0.789	0.658	0.684	0.757
Percentage black	0.189	0.078	0.159	0.225	0.486
Percentage Hispanic	0.396	0.640	0.301	0.409	0.037
Percentage white	0.323	0.141	0.459	0.289	0.451
Percentage other race	0.092	0.142	0.080	0.078	0.025

<sup>22</sup> Because we lack statewide comparison data in two states for the achievement data in our study sample, we do not report on achievement levels of the study schools here.

Characteristic	Overall Study Sample	State A	State B Cohort 1	State B Cohort 2	State C
Schools					
Number of schools	323	102	84	61	76
School total enrollment average (all grade levels)	827	836	986	1,008	496
Baseline average absences	10.038	N/A	10.942	9.736	9.281
Percentage ELL	0.125	0.244	0.080	0.125	0.016
Percentage SPED	0.143	0.144	0.141	0.145	0.142
Percentage FRPL	0.698	N/A	0.634	0.686	0.779
Percentage SES	0.803	0.803	N/A	N/A	N/A
Percentage FRPL or SES	0.731	0.803	0.634	0.686	0.779
Percentage black	0.252	0.101	0.168	0.278	0.527
Percentage Hispanic	0.351	0.604	0.299	0.402	0.028
Percentage white	0.317	0.145	0.467	0.262	0.425
Percentage other race	0.080	0.150	0.065	0.058	0.020

NOTES: ELL = English language learner; SPED = special education; FRPL = free or reduced-price lunch; SES = lower socioeconomic status. Baseline year attendance data were unavailable in State A. State A provided data on SES status rather than FRPL. One school in State A was newly opened in school year 2015–2016 and did not have baseline data included in the table. Sample includes only students with both ELA and mathematics outcome data in year three. School and student characteristics differ somewhat because of systematic differences in the characteristics of students who attend larger versus smaller schools.

### *Descriptive Summary of Comparison Groups in Student Achievement Analyses*

Our analyses compared either students initially at randomized treatment and control schools or students initially at fully participating treatment schools and at control schools that did not cross over and receive the EDP. Table 3.5 shows descriptive statistics at baseline for each of these analytic samples.

As we would expect given random assignment to treatment and control conditions, schools appeared to be similar on most observable characteristics prior to the intervention. We found few substantive differences between the observable characteristics of treatment and control school samples overall, though control schools' average test scores were slightly higher than treatment schools at baseline.

Table 3.5 also displays student and school characteristics for the portion of the treatment group whose principals fully participated in the intended EDP and coaching program and the portion of the control group whose principals did not participate in any of the EDP. We observed relatively few differences between the treatment group schools and students as a whole and the portion of the treatment group that fully participated in the EDP and coaching. The most notable difference was slightly lower school average mathematics test scores at baseline in the fully participating school sample. Because almost all control schools were compliant with their assignment to not receive the EDP, there were no meaningful differences between the control-assigned and control-compliant samples.

Formal tests of the baseline equivalence between the fully participating treatment group and the control-compliant group, overall and for each state, are provided in Table A.26 in the appendix. As shown there, we found that baseline achievement levels in State C were meaningfully lower in the fully participating sample than in the control-compliant sample. This suggests that the schools that opted to fully participate in the treatment in State C were dissimilar from those that did not. No meaningful differences in baseline achievement were apparent in either States A or B, however.

**Table 3.5. Student and School Characteristics in Experimental Analyses Analytic Sample, Overall and by Treatment and Control Groups**

Characteristic	Overall Study Sample	Treatment-Assigned	Control-Assigned	Treatment (Fully Participating)	Control-Compliant
Students					
Number of students	63,337	32,316	31,021	13,287	29,556
Baseline average ELA scores	0.038	0.077	0.025	0.034	-0.011
Baseline average mathematics scores	0.050	0.082	0.042	0.040	0.010
Baseline average absences	7.356	N/A	7.616	6.883	7.430
Percentage ELL	0.155	0.158	0.153	0.175	0.154
Percentage SPED	0.123	0.126	0.121	0.115	0.121
Percentage FRPL	0.683	0.695	0.672	0.665	0.673
Percentage SES	0.789	0.786	0.792	0.815	0.783
Percentage FRPL or SES	0.716	0.724	0.708	0.706	0.705
Percentage black	0.189	0.199	0.178	0.202	0.176
Percentage Hispanic	0.396	0.391	0.401	0.406	0.403
Percentage white	0.323	0.318	0.329	0.293	0.328
Percentage other race	0.092	0.092	0.092	0.100	0.094
Schools					
Number of schools	323	161	162	57	155
School total enrollment average (all grade levels)	827	827	827	923	898
Baseline average ELA scores	-0.122	-0.151	-0.093	-0.171	-0.086
Baseline average mathematics scores	-0.161	-0.197	-0.125	-0.290	-0.116
Baseline average absences	10.038	10.096	9.980	10.210	9.991
Percentage ELL	0.125	0.122	0.128	0.129	0.130

Characteristic	Overall Study Sample	Treatment-Assigned	Control-Assigned	Treatment (Fully Participating)	Control-Compliant
Percentage SPED	0.143	0.144	0.142	0.133	0.143
Percentage FRPL	0.698	0.702	0.695	0.672	0.692
Percentage SES	0.803	0.805	0.802	0.806	0.795
Percentage FRPL or SES	0.731	0.734	0.729	0.707	0.724
Percentage black	0.252	0.252	0.253	0.228	0.246
Percentage Hispanic	0.351	0.348	0.354	0.386	0.311
Percentage white	0.317	0.323	0.310	0.305	0.311
Percentage other race	0.080	0.077	0.082	0.081	0.084

NOTES: ELL = English language learner; SPED = special education; FRPL = free or reduced-price lunch; SES = lower socioeconomic status. Baseline year attendance data were unavailable in State A. Achievement measures are standardized relative to our study sample and, as a consequence, do not indicate the performance of schools or students relative to their statewide peers. State A provided data on SES status rather than FRPL. One school in State A was newly opened in school year 2015–2016 and did not have baseline data included in the table.

### *Attrition from Our Analytic Sample*

At the time of randomization, 332 schools were randomized in pairs across three states, with 166 schools assigned to the treatment condition and 166 schools assigned to the control condition. However, in our analytic sample for analyses of student achievement, our sample included a total of 323 schools (161 treatment-assigned and 162 control-assigned) for which student outcome data were available in the third year following random assignment. Thus, our analytic sample for student achievement analyses included 97.3 percent of schools that were initially randomized, or an attrition rate of 2.7 percent of schools. The difference in school attrition rates across our treatment and control sample was just 0.6 percent.

We refer to the nine schools for which data are unavailable as *attritors* from our experimental study. Attritor schools, in some cases, were schools that were included in the study because of imperfect data and did not have grades spanning sixth through eighth ( $N = 1$ ), alternative schools where student test score data were unavailable at the state level ( $N = 2$ ), and schools that closed before the end of the study period ( $N = 6$ ).

At the student level, attrition was more pronounced. Students who attrited from our study sample were observed as present in study schools in the first school year following random assignment but had no achievement outcomes available as of study year three. Student attrition could occur for a variety of reasons, including when students did not take achievement exams or took only specialized versions of those exams, advanced into the ninth grade ahead of schedule, or dropped out of school entirely. However, we observe that attriting students in the majority of cases simply left our statewide analytic samples, likely because their families moved to a different state.

As shown in Table A.27, we lacked achievement outcome data for approximately 11.8 percent of students as of year three overall across all three study states. Rates of attrition were

slightly higher in control schools, with a difference of 0.4 percent. This difference in available testing data across the treatment and control students was directionally similar to the differences in the year prior to random assignment; therefore, we did not have reason to believe that the treatment induced changes in testing rates in schools.

## 4. Results

In this chapter, we organize the results by the research questions we list in Chapter 1.

### Results About Implementation

#### *RQ 1: What Proportion of Principals Who Were Offered the EDP and Coaching Participated?*

NISL intended for treatment group principals to participate in at least ten of the 12 EDP units and to receive 60 hours of coaching. We use these two metrics for our definition of *fully participating* principals. In Table 4.1, we present the numbers of principals offered both the EDP and coaching in the first row. To understand the proportion of principals who participated, we then present the participation rates in the EDP in row two and participation in coaching in row three. The final row shows the percentage of principals fully participating in *both* the EDP and coaching and that therefore met our definition of fully participating. Only 35 percent of principals in the treatment arm of the study fully participated in both the EDP and coaching. Participation rates varied substantially by state, from a low of 15 percent in State C to a high of 49 percent in State B.

**Table 4.1. Proportion of Principals Who Fully Participated in the EDP and Coaching**

Ingredients of Full Participation	Overall		State A		State B		State C	
	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>
Principals originally offered the EDP and coaching	100.0	166	100.0	52	100.0	74	100.0	40
Principals who attended ten or more units of the EDP	57.2	95	50.0	26	66.2	49	50.0	20
Principals who received 60 or more hours of coaching	37.3	62	36.5	19	50.0	37	15.0	6
Principals who received both the EDP and coaching (ten or more EDP units and 60 or more hours of coaching)	34.9	58	30.8	16	48.6	36	15.0	6

SOURCES: NISL study roster, NISL coach logs, and EDP attendance records.

#### Participation in the EDP

Table 4.2 provides more detail about EDP participation among the 166 principals at treatment-assigned schools initially invited to participate in the EDP as part of this study. As already shown in Table 4.1, we see that 57 percent of principals offered the EDP participated fully in it, in the sense of attending at least ten of its 12 units according to NISL's attendance records. Rates of participation in the EDP varied across states, with full compliance of 50 percent in State C, 50 percent in State A, and 66 percent in State B. State C had the highest rate of schools whose principals did not participate in the EDP at all (43 percent). Among partial compliers, the number of EDP units attended was distributed fairly uniformly between one and ten units.

**Table 4.2. Participation Rate in the EDP, Overall and by State**

Amount of Participation	Overall		State A		State B		State C	
	%	N	%	N	%	N	%	N
Full participation (10–12 units)	57.2	95	50.0	26	66.2	49	50.0	20
Partial participation (More than 0 and less than 10 units)	18.7	31	25.0	13	20.3	15	7.5	3
Did not participate (0 units)	24.1	40	25.0	13	13.5	10	42.5	17

SOURCES: EDP unit attendance records.

According to NISL’s tracking of LEA and principal participation in the study, the reasons for partial or nonparticipation in the EDP varied (Table 4.3). About 49 percent of noncompleting principals ( $N = 35$ ) declined of their own accord to fully participate, and another 30 percent ( $N = 21$ ) did not fully participate because of a transition out of their original school or, in one case, a school closure. The third-largest driver of partial participation and nonparticipation in the EDP, at 21 percent, was district-level decisions to exit the program ( $N = 15$ ).

Reasons for partial participation or nonparticipation in the EDP varied somewhat across states. In particular, in States A and C (which had lower EDP participation rates), the principals’ decisions to not complete the EDP were a larger proportion of noncompliance. In State B, where EDP participation rates were the highest, LEAs’ decisions to exit the study were a relatively larger driver of nonparticipation. In the state with the lowest principal participation (State C), the two main reasons were principal movement out of their school or school closure, at 50 percent ( $N = 10$ ), and principals’ decisions not to fully participate, at 50 percent ( $N = 10$ ).

**Table 4.3. Reasons for Partial Participation or Nonparticipation in the EDP**

Reason	Overall		State A		State B		State C	
	%	N	%	N	%	N	%	N
Principal declined	25.4	18	26.9	7	12.0	3	40.0	8
Principal left midstream or missed too many sessions	23.9	17	30.8	8	28.9	7	10.0	2
Principal changed schools or school closed	29.6	21	23.1	6	20.0	5	50.0	10
Entire district withdrew from the program	21.1	15	19.2	5	40.0	10	0.0	0

SOURCE: NISL study roster.

### Participation in Coaching

As shown in Table 4.4, according to coaching logs, approximately 68 percent ( $N = 113$ ) of principals assigned to receive the offer of participating in the EDP and coaching ultimately participated to some degree in coaching. This rate varied by state, with partial participation rates of around 75 percent in State A, 78 percent in State B, and 53 percent in State C.

As summarized in Chapter 2, NISL’s guideline was for 60 hours of coaching to be delivered over a 30-month period. In practice, coaching hours well exceeded that target in many instances. Of those who participated in at least some coaching, the average hours of coaching was around 68 hours, but this varied widely, with a standard deviation of 40 and a maximum of 165 hours of coaching activity. The length of coaching relationships also varied, lasting anywhere from one month to the completion of the program after 30 months. Coaching intensity, measured in average total hours per principal, was highest in State A and lowest in State C. Overall, about 37 percent ( $N = 62$ ) of principals who took part in coaching completed the intended coaching dosage.

**Table 4.4. Participation in Coaching, Overall and by State**

Level of Participation	Overall		State A		State B		State C	
	%	N	%	N	%	N	%	N
Full participation (greater than or equal to 60 hours)	37.3	62	36.5	19	50.0	37	15.0	6
Partial participation (more than 0 and less than 60 hours)	30.7	51	38.8	15	28.4	21	37.5	15
Did not participate (0 hours)	31.9	53	34.6	18	21.6	16	47.5	19

SOURCES: NISL coach logs.

Among principals who did not fully participate in coaching, many (39 percent) were principals who did not participate in the EDP at all and thus were ineligible for coaching (Table 4.5). A slightly smaller portion, 33 percent, did not fully participate in coaching because of a movement to a different school. Roughly 19 percent opted to not participate or to end coaching early by choice, and about 5 percent withdrew from coaching when their district withdrew from the program. Finally, about 5 percent of coaching relationships ended earlier than intended because of coaches mistakenly ending the program earlier than NISL had intended.

**Table 4.5. Reasons for Partial Participation or Nonparticipation in Coaching**

Reason	Overall		State A		State B		State C	
	%	N	%	N	%	N	%	N
Never began the EDP	38.8	40	39.4	13	27.8	10	50.0	17
Principal changed schools or school closed	33.0	34	36.4	12	30.6	11	32.4	11
District withdrew from the program	4.9	5	0.0	0	13.9	5	0.0	0
Opted not to fully participate	18.4	19	24.2	8	25.0	9	5.9	2
Coach ended coaching early	4.9	5	0.0	0	2.8	1	11.8	4

SOURCE: NISL study roster.

Finally, we also asked NISL state coordinators (who were involved in recruiting districts into the i3 study) what they understood to be the main reasons for districts' or principals' nonparticipation in either the EDP or coaching. They offered the following five explanations:

1. The main reason, according to two of the three states' coordinators, was principal mobility. One of these two states' coordinators explained that, because of the state's accountability system in State C, new principals had a disincentive to remain in their building for up to five years, since the potential for forced school restructuring because of low student performance could create a "black mark" on a principal's résumé at the outset of their career. Thus, the state accountability system created an incentive, particularly for new principals, to switch schools before escalating sanctions could reach the point of school restructuring.
2. A related reason was superintendent mobility. An incoming superintendent who had a different vision than their predecessor was not always willing to honor a memorandum of understanding with NISL from the previous administration.
3. In State C, which also had strong school accountability rules, the coordinator explained that superintendents were loath to allow novice principals (especially of middle schools, as compared with elementary schools) to be out of their building for 24 school days in a year to take the EDP. And some new principals in this same state shared the concern, given the accountability system.
4. More generally, one state coordinator mentioned that novice principals might not want to be out of the building for the EDP courses, since they are "putting out fires" or trying to put their stamp on the school and do not yet understand the "process of change."
5. Finally, one state coordinator had the impression that some principals or districts dropped out once they realized the level of commitment required for participation in the EDP and coaching.

### *RQ 2: Among Those Who Participated, What Were Their Perceptions of the EDP and of Coaching?*

Principals who reported on our survey that they participated in the EDP (to any degree) held highly positive views of the EDP and of NISL coaching, as shown in Table 4.6.<sup>23</sup> They most strongly endorsed coaching, with 82 percent of principals who answered this item agreeing "to a great extent" that coaching helped them improve their school and that it was more valuable than the EDP for improving their school. Virtually all reported that their coaching was focused, as intended, on the concepts of the EDP. Interestingly, despite their very strong endorsement of coaching, only half strongly agreed that their coach was very knowledgeable about school improvement work. But more than two-thirds of principals strongly agreed that they held similar views as their coach of how to improve the school.

---

<sup>23</sup> For principals assigned to the treatment group, we administered one version of the survey that included questions about their perceptions of the EDP and coaching. For principals assigned to the control group, we asked whether they participated in any EDP or NISL coaching (to help gauge crossover), but we did not also ask perception questions.

Large majorities of principals also strongly endorsed the EDP. About seven in ten principals agreed “to a great extent” that the EDP helped them lead their school better, and more than eight in ten principals reported that they would recommend the EDP to another principal friend. A lower percentage were positive about their ALP, which each principal creates during the EDP. A little more than half of principals said that the ALP helped improve their school or that they continued to implement their ALP “to a great extent” as of the time of the survey.

**Table 4.6. Perceptions of the EDP, the Action Learning Project, and Coaching**

<b>Percentage of Principal Respondents Who Agreed with the Following Statements</b>	<b>Not At All</b>	<b>Very Little</b>	<b>Somewhat</b>	<b>To a Great Extent</b>
<b>EDP</b>				
The NISL EDP courses helped me to lead my school better.	0.0	0.0	28.6	71.4
I would recommend NISL EDP to a friend who is a principal.	0.0	0.0	14.3	85.7
<b>ALP</b>				
Creating and implementing my Action Learning Project (ALP) improved my school.	4.8	2.4	35.7	57.1
I continue to implement my ALP.	2.4	9.5	33.3	54.8
<b>Coaching</b>				
My coach has helped me to improve my school.	0.0	3.0	15.2	81.8
Coaching was more valuable than EDP itself for helping me improve my school.	0.0	3.0	15.2	81.8
My coaching was focused on EDP concepts.	0.0	0.0	12.1	87.9
My coach was knowledgeable about school improvement work.	6.1	9.1	33.3	51.5
My coach and I have a similar view of what needs to improve in my school.	3.0	3.0	24.2	69.7

SOURCE: i3 winter 2019 principal survey, treatment group. This table summarizes a subset of items related to experiences with the EDP and coaching. There were 42 total respondents to questions about the EDP; 33 of these also responded to the questions about coaching. Respondents only reported on the EDP if they had personally taken part in at least some of the program and only reported on coaching if they had participated in at least some coaching.

When we considered responses separately in the individual states (results shown in Table A.28 in the appendix), we found significant differences across the three states in the overall average rating on survey items related to the experience of and use of the EDP and ALP. We did not, however, observe significant differences across the states on survey items related to coaching. State B had the most-positive average survey responses to questions related to the EDP and ALP, and State C had the least-positive responses. This tracks with the states that had the

highest and the lowest take-up rate of the EDP and coaching, as shown in Table 4.1. Differences were most pronounced for the specific items related to continued use of the ALP and about use of the four EDP concepts (distributing leadership; aligning curriculum; use of the “NISL wheel”; and use of the “all means all” concept).

At the end of the principal survey, we asked what single concept from the EDP they used most in their school, and we asked for “one specific example of how you applied that concept in your school.” We thematically coded the responses into topical categories, which we report here.

The three most-mentioned EDP concepts were as follows:

- distributing leadership in the building (ten mentions)
- how people learn (nine mentions)
- the NISL concept “all means all,” which refers to ensuring that all students have access to high-quality learning opportunities (nine mentions).

In Table 4.7, we categorized and aggregated the full set of 40 written answers to our request for an example of how the principal applied the top EDP concept in their building. The most common examples were the distribution of leadership to more individuals in the building (seven responses) and the related practice of creating an instructional leadership team (four responses). The next most-frequent examples related to how principals organized staff meetings, used data for the selection of their priorities, and mapped whether a given practice aligned with the school’s stated vision.

**Table 4.7. How Principals Applied EDP Concepts**

<b>Example of How Principal Applied the EDP Concepts in Their Building</b>	<b>Number of Principals Who Reported an Example Like This</b>
Empower staff to take responsibility / delegate and share responsibilities	7
Create instructional leadership team	4
Incorporate EDP concepts into staff group activities such as professional development or faculty meetings	4
Heterogeneously group students in instruction—also known as detracking	4
Use data to set priorities or accomplish an objective	4
Aligning practice to the school’s stated vision	3
Increase support or personalized learning to meet the needs of all students	3
Strategic thinking	2
Send teachers to participate in NISL’s Teaching for Effective Learning series	2
Acquire staff buy-in before making a change	2
Build an instructional system	2
Spend more time in classrooms as an instructional leader	1
More fairly distribute resources to students	1
Monthly instructional rounds plus all new teachers have mentors	1

SOURCE: i3 winter 2019 principal survey, treatment group. There were 40 responses to this question.

As shown in Table 4.6, 82 percent of i3 principals agreed or strongly agreed on the survey that coaching was more valuable than the EDP to help the principal improve their school. We then gleaned more information in phone interviews about principals’ perceptions of their NISL coaching. Although the results shown in Tables 4.8 and 4.9 are derived from only 17 i3 study principal interviewees, we have greater confidence in them because they comport thematically with the answers to the identical question posed to an additional 57 principals who were part of a second (SEED grant–funded) study we are conducting in parallel on the effects of the EDP and coaching. Tables 4.8 and 4.9, which are adapted from our 2019 study (Wang et al., 2019), show that principals valued their coach as a thought partner and as someone who helped them distribute leadership to others, such as lead teachers and other administrators; improve instructional leadership; and discover helpful resources.

**Table 4.8. Principals’ Responses to the Question  
“What Top Three Lessons or Ideas Have You Used the Most from NISL Coaching, if Any?”**

<b>The NISL Coach Helped Principals . . .</b>	<b>Percentage of Principals Who Reported This as a Top Lesson</b>
Brainstorm ideas/serve as thought partner	35
Embrace or create collaborative/shared leadership in their schools	18
Improve instructional leadership	18
Connect to resources	18
Think more strategically/more intentionally	12
Other	29

SOURCE: RAND phone interviews of 17 i3 grant principals.

When asked “What was the most beneficial activity you did with your coach?” the most frequent response to this open-ended question was the coach engaging in discussions and reflections with them in the role of a thought partner (see Table 4.9). Some of the principals who offered this answer explained that coaches provided a neutral, third-party opinion on the school’s challenges and solutions that principals used to guide their thinking and decisionmaking. Principals also reported that brainstorming and reflecting with their coaches helped keep them focused on their goals, particularly regarding the ALP. Principals also found it instructive when coaches conducted team meeting observations with them (41 percent) or co-observed classrooms or conducted walkthroughs with them (24 percent). Through this activity, principals learned what to look for when observing teachers as they collaborated with each other or as they worked with students in the classroom. Principals also learned strategies to improve teacher professional learning using the observations and learned how to debrief the observed teacher.

**Table 4.9. Principals’ Responses to the Question  
“What Was the Most Beneficial Activity You Did with Your Coach?”**

<b>Activity</b>	<b>Percentage of Principals Who Reported This as a Beneficial Activity</b>
Discussing/brainstorming ideas and strategies with coach as thought partner	53
Conducting team meeting observations together	41
Learning about/gathering new resources	35
Conducting classroom observations and debriefing together	24
Reviewing data together	24
Strategizing about the ALP/receiving feedback on the plan	18
Other	41

SOURCE: RAND phone interviews of 17 i3 grant principals who had a coach.

Finally, we posed the question “How can coaching be improved?” Six of the 17 principals (35 percent) had no suggested improvements. The most common answer (suggested by five principals, or 29 percent of the 17 respondents) was to increase face-to-face time in schools with coaches. The second most common answer (reported by three principals) was to increase the duration of coaching. The other suggestions (made by two principals or fewer) were to improve the coach match, with respect to geographical proximity and background.

*RQ 3: What Form Did Coaching Take, and What Were the Main Topics of Coaching?*

For this research question, we drew on coaches’ perceptions according to their logs and survey responses. (We report principal perceptions in RQ 2.) According to the electronic coach logs, a total of 41 coaches conducted sessions with principals over the span of December 2015 to June 2019. The average number of communications (whether face-to-face, email, phone, or web-based) that a coach logged with a principal was 33, with those communications totaling an average of 68 hours. The typical principal met with his or her coach a total of 11 times face-to-face, nine times by phone, and two times virtually. The balance of interactions were by email. The majority of the 113 principals had only one NISL coach during the study period, but 27 of them had two coaches, and one had three coaches during the 18–30-month period. The duration of a logged coaching session lasted anywhere from one minute (e.g., for transactional emails to schedule meetings) to ten hours. Of the communications longer than five minutes, the average duration was 137 minutes.

According to coach logs, the ALP was the most popular topic that coaches discussed with principals (57 percent of logged communications), followed by the EDP (43 percent of logged communications), followed by the NISL wheel (27 percent of logged communications), followed by the diagnostic assessments that principals complete in the first three units of the EDP (11

percent). Note that, because coaches could tag more than one topic per session, the percentages in the prior sentence add to more than 100 percent.

Across our two coach survey administrations (spring 2017 and spring 2018), a total of 33 out of the 41 NISL coaches described their coaching experiences.<sup>24</sup> The coaches were a veteran group, with 91 percent having served as a school principal in the past (for an average of nine years). Two-thirds or more had each of the following prior professional experiences: served as a coach to principals prior to the i3 study, designed professional development for principals, mentored principals, recruited principals, and worked in a school district central office.

Coaching required substantial travel time, and coaches reported spending slightly more time traveling to meet principals for face-to-face meetings (10.8 hours per month on average travel time) as interacting with them monthly (10.7 hours per month on average). This time varied substantially, with coaches reporting as little as one hour to as many as 34 hours per month for travel. Although all but two of the coaches agreed “to a great extent” that they had gotten the training they needed from NISL to do the coaching, 45 percent only agreed “somewhat” that NISL had reasonable expectations for the job. Those that agreed only somewhat tended to have longer travel times.

We asked coaches “In your view, what are the most important ingredients that make the i3 grant coaching relationship successful?” The most frequently mentioned ingredient (mentioned by 18 coaches) was a trusting and/or positive relationship between the coach and principal. Other frequent responses included the coach and principal having knowledge of the EDP, effective listening and communication, consistency of focus, the willingness of the principal to follow through, and customization of the coaching for each principal.

We concluded the coach survey with the question, “In your view, what are the greatest challenges or barriers to a successful i3 grant coaching relationship?” The most frequently mentioned answer related to the school district not supporting the coaching relationship (15 coaches mentioned this), because of an ever-changing set of district directives that did not align with the EDP, district disengagement with the EDP, district- or state-required new programs crowding out time for the principal’s ALP, unexpected and new assignments for the principal from the district, or lack of a principal career ladder for which certification in the EDP curriculum and coaching fits into leadership development. Another frequently mentioned barrier was difficulty in scheduling time for the coach meetings with the principal. A related additional frequent mention was the principal being interrupted during coach meetings, which cut them short. Finally, another frequently mentioned barrier was a lack of commitment from the principal to the coaching—i.e., “Lack of investment of time and follow-through between coaching conversations.”

---

<sup>24</sup> For those NISL coaches who responded to both surveys, we refer here to their responses from the more recent survey (spring 2018).

## Impact Results: Student Outcomes

### *RQ 4: What Is the Effect of Offering the EDP and Coaching on Student Achievement, Attendance, and Grade Progression After Three Years?*

This research question focused on the effects of being offered the EDP and coaching, rather than the effect of fully participating in them. Across our pooled three-state sample, we found no statistically significant effects on middle school students' achievement outcomes three years after principals were offered the EDP and coaching. However, we cannot rule out the possibility of very small positive or negative effects of offering the treatment that we were not statistically powered to reliably detect.<sup>25</sup> Results, shown first in Table 4.10, also indicate that there were no significant or substantive differences in overall effect sizes in ELA, mathematics, or attendance across study years. Analyses of impacts on students' on-time grade progression rates also yielded no significant effects, and these results are included in Table A.31 in the appendix.

**Table 4.10. Estimated Impact of Offering the EDP and Coaching on Student Achievement**

Sample	Study Year	Number of Schools	ELA		Mathematics		Attendance	
			Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
All three states	Year 3	323	0.007	(0.013)	0.011	(0.021)	0.140	(0.205)
	Year 2	323	0.005	(0.013)	0.014	(0.019)	0.234	(0.237)
	Year 1	324	0.015	(0.012)	0.019	(0.019)	-0.011	(0.166)
State A	Year 3	102	-0.008	(0.029)	0.026	(0.033)	-0.149	(0.367)
	Year 2	102	0.006	(0.028)	0.005	(0.032)	-0.032	(0.503)
	Year 1	103	0.028	(0.033)	-0.002	(0.034)	N/A	N/A
State B	Year 3	145	0.004	(0.014)	0.033	(0.033)	0.176	(0.205)
	Year 2	145	0.006	(0.014)	0.041 <sup>a</sup>	(0.025)	0.442	(0.275)
	Year 1	145	0.007	(0.013)	0.036	(0.023)	-0.024	(0.162)
State C	Year 3	76	-0.004	(0.035)	-0.044	(0.047)	0.665	(0.459)
	Year 2	76	0.011	(0.031)	-0.016	(0.043)	0.361	(0.341)
	Year 1	76	0.044	(0.029)	-0.036	(0.057)	0.099	(0.292)

NOTE: One study school in State A included grade six but not grades seven and eight; therefore, the school was included in Year 1 estimates only.

<sup>a</sup>  $p < 0.1$ .

We also conducted exploratory analyses of effects in each of the three states in our study, and we do not observe statistically significant effects of the offer of the intervention in any of them (as shown in Table 4.10). There was some modest heterogeneity in the results across states, with

<sup>25</sup> In addition, our experimental effect estimates do not adjust for the receipt of the EDP among five noncompliant control schools, as a result of principal transitions.

somewhat more positive (and marginally statistically significant in year two) math effect estimates in State B and somewhat more negative math effects in state C, for example. Overall, however, each state's effect estimates were consistent with our overall finding of no detectable impacts on student achievement outcomes of the offer of the EDP and coaching within a three-year time span.

In further exploratory analyses, we also examined whether there were effects on achievement exams, such as algebra 1, science, and geometry exams that are distinct to specific study states. These results are in Table A.29 of the appendix. As with our main analyses of ELA and mathematics, we found no statistically significant effects on any type of exam.

To complete our exploratory analyses, we also looked for evidence of variation in effects as a function of pretreatment characteristics of schools and their students to provide context for the estimates presented above. As described in Chapter 3, we did so using a machine-learning technique developed by Athey and Imbens and implemented by Davis and Heller (Athey and Imbens, 2015; Athey and Imbens, 2016; Davis and Heller, 2017; Davis and Heller, forthcoming). We did not detect any clear patterns of heterogeneity in the effects; therefore, we do not present these results in this report.

#### *RQ 5: What Is the Effect of Full Participation in the EDP and Coaching on Student Achievement After Three Years?*

As a first step in examining the effects of fully participating in the EDP and coaching, we considered the range of plausible estimated impacts as determined by our analysis of the effects of TOT. In Table 4.11, we report an upper-bound and a lower-bound estimate for each of the achievement and attendance outcomes. First, we report the estimated lower-bound effect of participating *to any degree* in either the EDP or coaching. Then, we report the estimated upper-bound effect of *fully participating* in the EDP and coaching when assuming that partial participation had no impact at all. These analyses set the stage for our next quasi-experimental analysis of full participation by identifying a range of plausible effects.

Overall, none of the results of our separate TOT analyses shown in Table 4.11 were statistically significant. But they reinforce the notion that we cannot rule out the possibility of smaller effects of full participation in the EDP and coaching that we may have lacked sufficient statistical power to reliably detect. The range of positive estimates on achievement is fairly wide, both overall and specifically in State B. In contrast, the range of directionally negative estimates of participation on achievement in State C is also fairly wide. For attendance, the range of estimates for States B and C are directionally positive (i.e., increased days of attendance), but they are directionally negative in State A.

Table A.31 in the appendix separately provides the same bounding estimates for grade progression outcomes, where we found no evidence of impacts. None of the estimated effects for on-time grade progression outcomes are significant, and all the estimates of impact were very close to zero.

**Table 4.11. Estimated Range of Impacts of Participating in the EDP and Coaching on Student Achievement and Attendance in Schools**

Sample	Study Year	Number of Schools	ELA		Mathematics		Attendance	
			Lower-bound Coeff.	Upper-bound Coeff.	Lower-bound Coeff.	Upper-bound Coeff.	Lower-bound Coeff.	Upper-bound Coeff.
All three states	Year 3	323	0.011	0.023	0.023	0.046	0.076	0.148
	Year 2	323	0.006	0.009	0.026	0.056	0.256	0.460
	Year 1	324	0.015	0.019	0.029	0.065	-0.047	-0.097
State A	Year 3	102	-0.012	-0.029	0.037	0.097	-0.202	-0.511
	Year 2	102	0.009	0.023	0.007	0.017	-0.044	-0.110
	Year 1	103	0.039	0.097	-0.002	-0.005	N/A	N/A
State B	Year 3	145	0.005	0.008	0.032	0.059	0.200	0.329
	Year 2	145	0.009	0.013	0.050 <sup>a</sup>	0.081 <sup>a</sup>	0.497	0.850
	Year 1	145	0.006	0.013	0.042	0.073	-0.024	-0.048
State C	Year 3	76	-0.011	-0.030	-0.086	-0.252	1.338	3.813
	Year 2	76	0.021	0.057	-0.036	-0.102	0.581	1.545
	Year 1	76	0.085	0.241	-0.071	-0.212	0.177	0.500

NOTE: One study school in State A included grade six but not grades seven and eight; therefore, the school was included in Year 1 estimates only.

<sup>a</sup>  $p < 0.1$ .

Next, we directly estimated the effects of full participation in the EDP and coaching on achievement and attendance outcomes using nonexperimental methods. In this analysis, full participation was defined as principals in the treatment group who attended ten or more of the 12 units of the EDP and who also received 60 or more contact hours with their coach. In this analysis, we compared fully participating treatment-group schools with control schools that did not have any exposure to the EDP, and we relied on student and school baseline characteristics to account for any preexisting differences between these two groups that was not caused by the treatment.

Overall, as shown in Table 4.12, we again found no statistically significant effects on middle school students' achievement and attendance outcomes after three years in our pooled three-state sample. We also examined effects across each of the three states, also shown in Table 4.12. Consistent with our experimental analyses, there was some limited evidence of heterogeneity in effects across states. As with our experimental analyses, effect estimates in State B were directionally higher in most years and statistically significant in mathematics in year one and attendance in year three. State B was also the state with the highest proportion of fully compliant schools from the original treatment-assigned group. However, neither of the effect estimates in State B were robust to an adjustment for the multiple (three) state-specific tests we conducted in each outcome domain in each year. As a consequence, we cannot rule out the possibility that one

or both are simply caused by random chance. In addition, the apparent attendance impact in state B in year three (about half a day more of school attended by students in fully participating schools) is somewhat larger than our upper bound estimate from our TOT bounding analysis, which suggests this nonexperimental estimate may be upwardly biased.<sup>26</sup>

**Table 4.12. Estimated Impact of Full Participation in the EDP and Coaching on Student Achievement and Attendance**

Sample	Study Year	Number of Fully Participating Treatment Schools	Number of Compliant Control Schools	ELA		Mathematics		Attendance	
				Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
All three states	Year 3	57	155	0.014	(0.020)	0.005	(0.030)	0.272	(0.282)
	Year 2	57	155	0.007	(0.020)	-0.009	(0.031)	-0.102	(0.344)
	Year 1	58	155	0.004	(0.015)	0.032	(0.027)	N/A	N/A
State A	Year 3	15	49	-0.014	(0.049)	-0.022	(0.044)	-0.008	(0.619)
	Year 2	15	49	-0.010	(0.045)	-0.059	(0.049)	-0.083	(0.619)
	Year 1	16	49	0.065	(0.047)	0.006	(0.051)	N/A	N/A
State B	Year 3	36	70	0.016	(0.019)	0.036	(0.044)	0.487 <sup>a</sup>	(0.227)
	Year 2	36	70	0.004	(0.020)	0.032	(0.034)	0.407	(0.291)
	Year 1	36	70	-0.007	(0.018)	0.07 <sup>a</sup>	(0.033)	-0.311	(0.202)
State C	Year 3	6	36	-0.098	(0.067)	-0.116	(0.089)	-0.577	(0.843)
	Year 2	6	36	-0.004	(0.055)	-0.078	(0.100)	-0.419	(0.575)
	Year 1	6	36	0.028	(0.062)	-0.051	(0.108)	-0.306	(0.738)

NOTE: One study school in State A included grade six but not grades seven and eight; therefore, the school was included in Year 1 estimates only.

<sup>a</sup>  $p < 0.05$ .

In States B and C, our nonexperimental estimates of the effect of full participation were, for the most part, within the bounds established by our TOT findings. However, our nonexperimental estimates in mathematics in State A were not in the plausible range indicated by our experimental findings. Our nonexperimental pooled cross-state estimates were also more negative than the range of plausible impacts implied by our TOT findings. We have less confidence in any nonexperimental estimates that are inconsistent with our experimental estimates. The two sets of results may differ because of unobserved selection biases that we were unable to control for and that could lead the nonexperimental estimates to be less accurate.

<sup>26</sup> Alternately, it is possible that, contrary to our expectations, the intervention had negative impacts on student attendance in the treatment schools that either did not participate or that did not participate fully in the EDP and coaching. This could lead to inconsistent results between experimental and nonexperimental analysis of the effects of fully participating in the EDP and coaching.

In further exploratory analyses, we also examined whether there were effects of full participation in the intervention on achievement exams, including algebra 1, science, and geometry, that are distinct to specific study states. These results are in Table A.30 in the appendix. We identified one effect that was significant at traditional thresholds ( $p < 0.05$ ) in State B on algebra test scores in year three. However, this estimate both exceeds the plausible bounds implied by our TOT analysis and was not robust to an adjustment for multiple comparisons in the domain of mathematics. We therefore cannot rule out the possibility that it may be caused by random chance.

## Impact Results: Leadership Practices and School Practices

### *RQ 6: What Is the Effect of Offering and of Full Participation in the EDP and Coaching on Principals' Leadership Practices and on School Policy, Culture, and Practices?*

In accordance with the logic model in Figure 2.1, we divide our results into two sections: (1) principal leadership practices and (2) school culture and instruction. Tables A.1 through A.26 in the appendix list each survey item that loads onto a survey factor (e.g., “the school has a strategic plan”) that we present in Tables 4.13 and 4.14. In each section, we present estimated effects of offering the EDP and coaching alongside estimated effects of full participation in both the EDP and coaching.<sup>27</sup>

#### Effects on Principal Leadership Practices

As shown in Table 4.13, we found one statistically significant effect of either offering or of full participation in the EDP and coaching out of the 11 total aspects of principal leadership that we measured.<sup>28</sup> There was also one suggestive effect, but this was only statistically significant at the  $p < 0.10$  level. The one statistically significant effect was a large, positive effect of full participation in the EDP and coaching on principals' perceptions that their school had a strategic plan. *Strategic plan* is a four-item survey scale designed to measure EDP concepts that aggregates principals' answers about (1) whether the school has a public vision statement that includes the goal of improving student achievement, (2) whether the school has a strategic plan that explicitly links to a vision statement, (3) whether teachers and administrators understand the strategic plan, and (4) whether the school has a recurring review of the strategic plan. The precise wording of the four items is included in Table A.1 in the appendix. The size of the effect (0.63 of a standard deviation) is equivalent to the difference in the median (50th percentile) principal's

---

<sup>27</sup> We also calculated TOT estimates comparable with those included in our presentation of results for RQ 5. We do not include these estimates for brevity's sake, but they are available upon request. In essence, we are less confident of any nonexperimental estimates of the effect of fully participating in the treatment that are not, at a minimum, directionally consistent with our estimated effects of offering the treatment.

<sup>28</sup> When accounting for our five hypotheses within the domain of principal leadership practices as reported by principals, we estimate a less than 10 percent chance that this is a false-positive finding.

response to this survey factor in our sample and a respondent in the 80th percentile of the range of responses for this factor. In Table A.32 in the appendix, we provide results separately by state for this factor. Estimated effect sizes were substantially larger in States A and B than in State C, noting that principal take-up of the EDP and coaching was lowest in State C.

The one marginally significant, suggestive finding was a relatively large, positive effect of full participation in the EDP and coaching on a factor measuring principals' views that the schools' curriculum was aligned and evidence-based. However, there is a reasonable chance that this result may simply be caused by random chance, especially given the large number of distinct survey factors that we examined.

**Table 4.13. Impact of Offering and of Full Participation in the EDP and Coaching on Principal Leadership Practices**

Outcome	Survey Respondent Type	Effects of Offering Treatment		Effects of Full Participation	
		Coeff.	Std. Error	Coeff.	Std. Error
Strategic plan for the school	P	0.292 <sup>a</sup>	(0.167)	0.634 <sup>b</sup>	(0.223)
Effective leader	T	0.070	(0.138)	0.293	(0.180)
School has differentiated roles for teachers	T	0.058	(0.134)	0.182	(0.177)
School has differentiated roles for teachers	P	-0.054	(0.156)	0.158	(0.228)
Teachers have input into staffing, curriculum, or professional development	T	0.120	(0.133)	0.045	(0.171)
Curriculum is aligned and evidence-based	P	0.160	(0.173)	0.411 <sup>a</sup>	(0.226)
High priority on use of student assessment data	P	0.024	(0.170)	0.057	(0.246)
School offers high-quality professional development for teachers	P	0.015	(0.158)	0.155	(0.201)
Administrators observe teachers' classrooms	T	0.155	(0.120)	0.115	(0.142)
Teachers observe other teachers' classrooms	T	-0.117	(0.113)	-0.048	(0.149)
Teachers receive actionable feedback	T	0.046	(0.127)	0.152	(0.173)

NOTE: P = principal survey; T = teacher survey. Principals from 159 schools responded to the principal survey and are used to estimate effects of offering treatment. Of these, the full participation analysis excludes 48 schools in which the principal did not fully participate, for a total of 101 schools. The corresponding sample sizes for the teacher survey are 197 total schools (each with multiple teacher respondents per school), of which 130 schools are included in the analysis of effects of full participation.

<sup>a</sup>  $p < 0.1$ .

<sup>b</sup>  $p < 0.01$ .

#### Effects on School Policy, Culture, and Practices

Among the 15 survey scales or items we analyzed about school culture, policy, or practices shown in Table 4.14, we found one highly statistically significant effect that we can be

reasonably confident was not by chance<sup>29</sup> and four suggestive effects, one of which was no longer significant at the  $p < 0.05$  level once we adjusted for multiple hypothesis testing, and the remaining three of which were only statistically significant at the  $p < 0.10$  level even prior to multiple hypothesis testing.

**Table 4.14. Impact of Offering and of Full Participation in the EDP and Coaching on School Policy, Culture, and Practices**

Outcome	Survey Respondent Type	Effects of Offering Treatment		Effects of Full Participation	
		Coeff.	Std. Error	Coeff.	Std. Error
Positive school culture	P	-0.275 <sup>a</sup>	(0.156)	-0.159	(0.191)
School is unsafe and disorderly	T	0.136	(0.108)	0.018	(0.150)
School is unsafe and disorderly	P	0.150	(0.155)	0.113	(0.207)
Teachers endorse the school	T	-0.263 <sup>b</sup>	(0.120)	-0.113	(0.146)
Educators' self-efficacy	T	0.051	(0.124)	0.216	(0.174)
Educators' self-efficacy	P	-0.078	(0.164)	-0.210	(0.227)
High-performing teachers are assigned to struggling students	P	-0.080	(0.082)	0.116	(0.113)
High-quality academics	P	-0.012	(0.170)	0.302	(0.216)
High-quality academics	T	-0.049	(0.129)	0.077	(0.172)
Students' higher-order thinking skills	P	-0.085	(0.161)	0.108	(0.215)
Principal focuses on NISL concepts of learning	P	-0.056	(0.165)	0.097	(0.227)
Personalized instruction for students	P	0.189	(0.170)	0.634 <sup>c</sup>	(0.207)
Teachers collaborate about matters of instruction and professional development	P	0.000	(0.083)	0.082	(0.112)
Teachers collaborate about matters of instruction and professional development	T	0.152	(0.138)	0.315 <sup>a</sup>	(0.187)
School assigns coaches to low-performing teachers	P	-0.071	(0.071)	-0.160 <sup>a</sup>	(0.084)

NOTE: P = principal survey; T = teacher survey. As the labels suggest, the factor "School is unsafe and disorderly" is coded such that a higher score represents greater disorder/disaffection.

<sup>a</sup>  $p < 0.1$ .

<sup>b</sup>  $p < 0.05$ .

<sup>c</sup>  $p < 0.01$ .

The robust statistically significant effect was of full participation in the EDP and coaching on principals' perceptions that instruction was personalized to students' needs. Personalized learning is another EDP core concept. We operationalized the measure of this concept with the combination of two survey items: principal perceptions that (1) teachers work together to

<sup>29</sup> In this case, a multiple hypothesis adjustment for ten tests of principal-reported factors related to the domain of school policy, culture, and practices indicates a less than 10 percent chance that this is a false positive finding.

develop short, informal assessments to routinely assess student understanding and adjust instruction and (2) teacher and administrators design units of study specifically for struggling students. Fully participating treatment school principals' ratings on the personalized items were 0.63 standard deviations higher than control principals.<sup>30</sup> This estimated effect size is equivalent to the difference in the median (50th percentile) principal's response to this survey factor in our sample and a respondent in the 84th percentile of the range of responses for this factor. In Table A.32 of the appendix, we provide results separately by state for this factor. Estimated effect sizes for this factor were substantially more positive in States A and B than in State C.

The effect that was significant at the 5 percent level but not robust to an adjustment for multiple hypothesis testing was a negative effect of offering the EDP and coaching on teachers' endorsement of their school—i.e., teacher agreement with statements about recommending their school to parents and feeling good about the direction of the school. Further, we found a smaller (but still negative) and non-statistically significant estimate on teachers' endorsement of their school in places where the principal had fully participated in the EDP and coaching. In this case, however, the effect estimates were directionally less negative in the analysis that focused on fully compliant schools, which casts some doubt as to whether this difference was truly caused by participation in the EDP and coaching.

Finally, the three marginally significant effects were (1) principals' reports about school culture (those who were offered the EDP and coaching reported lower ratings than the control group), (2) teachers' reports that they collaborate about matters of instruction and professional development (a greater percentage of teachers in schools where the principals fully participated in the EDP and coaching reported this kind of collaboration than teachers in control schools), and (3) principals' reports that they assigned coaching to low-performing teachers (fewer principals reported doing this in schools that fully participated in the EDP and coaching). However, these were only statistically significant at the 10 percent level and were not robust to adjustments for multiple hypothesis testing. Some or all of these results may simply reflect random chance.

In additional exploratory analyses, we decomposed the “total” effect estimates on achievement measures into two components: (1) the indirect effect, which captured the proportion of the total effect that was realized through a given mediator (we test two mediators: the school having a strategic plan and personalized learning for students), and (2) the direct effect of the treatment. All the indirect effect estimates were not statistically significant, which is not surprising given that most of the total effects on achievement outcomes were small and statistically insignificant. Therefore, we do not discuss these analyses further in this report.

---

<sup>30</sup> This estimated effect size is slightly higher than the upper bound we identified in our TOT analysis of this factor, which was 0.59 standard deviations, but the difference is very small.

## 5. Discussion and Conclusion

---

In this chapter, we summarize the findings, present the limitations of the analysis, and interpret the results. Drawing on implementation lessons and outcomes from this study and a second one we are conducting of the EDP and coaching, we will provide a comprehensive set of recommendations in the third and fourth report of our series on NISL’s EDP.

### Summary of Findings

Three school years after the start of the intervention, we did not detect statistically significant effects on student academic achievement, attendance, or on-time grade progression of offering or of fully participating in the EDP and paired coaching of novice middle school principals. This finding applied to our pooled analysis of 323 principals across the three states participating in the study, as well as to each of the states.

We did, however, find effects on two school practices that the EDP teaches principals to do. The first is the school having a strategic plan, and the second is the personalization of student instruction. For both, we found large positive effects in which fully participating principals in the EDP and coaching reported agreement that was substantially higher than principals from the control group. These effects were equivalent to the difference in the median (50th percentile) principal’s response to each survey factor and respondents in the 80th percentile or higher of the range of responses in our sample. We note that both of these measures are drawn from the principal survey and that we found no effects from teacher-reported school practice measures, which indicates that the EDP and coaching influenced principals’ perceptions of their practices but did not extend to teachers or did not extend to teachers to a sufficient degree in enough schools for this study to detect.

We found positive effects on two of the most proximate outcomes—principal perceptions and practices—but not on the next most proximate—teacher practices—or on the third most proximate—student achievement.

A low proportion (35 percent) of the principals who were offered the EDP and coaching fully participated in both. Participation rates varied by state—especially for coaching. In each of the three states in this study, 50–66 percent of the principals in the treatment group attended ten or more of the 12 EDP units. But as few as 15 percent of principals in State C and as many as 50 percent of principals in State B received 60 or more hours of coaching.

Despite the low participation rates, those who participated reported highly positive views of both the EDP and coaching. About seven in ten principals agreed “to a great extent” that the EDP helped them lead their school better, and more than eight in ten principals reported that they

would recommend the EDP to another principal friend. Similarly, eight in ten principals agreed to a great extent that “my coach has helped me to improve my school.”

A total of 41 coaches coached 13 principals in the study and engaged in an average of 33 communications (whether face-to-face, email, phone, or web-based) with the coached principal. The communications summed to an average duration of 68 hours. The coaches were a veteran group, with 91 percent having served as a school principal in the past (for an average of nine years) and two-thirds or more possessing relevant prior experience, such as coaching, mentoring, or recruiting principals. Coaching required substantial travel time, and coaches reported spending slightly more time traveling to meet principals for face-to-face meetings (10.8 hours per month on average travel time) as they did interacting with them monthly (10.7 hours per month on average). This varied substantially, with coaches reporting as little as one hour to as many as 34 hours per month for travel.

## Limitations

Our estimates of the effect of being offered the EDP and coaching and of full participation in the EDP and coaching each had some relevant limitations. In particular, although our experimental effect estimates were (in theory) unbiased, they focused on the effects of offering the intervention rather than of actually participating in it. As a consequence, the magnitude of the effects of full participation could not be directly estimated within the experimental framework. Also, a small number ( $N = 5$ ) of control schools received some exposure to the study-induced treatment, mostly as a result of principal transitions (four of the five instances). This rate of crossover, while low, could have had a slight impact on our estimated effects in the experimental analyses. Additionally, principal participants who switched schools could have carried any benefits of their professional development into their new school environment, and we did not make any adjustments for this, nor were we able to explore this possibility in our evaluation. Also, our experimental effect estimates may have failed to pick up on any spillover effects of the program on districtwide practices or performance, since a majority of our randomized pairs were within-district pairs.

Our quasi-experimental effect estimates could be biased in ways that we cannot control for based on which types of principals and schools were more likely to fully participate in the EDP and coaching. This bias could have shifted our effect estimates to be either more positive or more negative. For example, schools whose principals are more motivated or more able to implement the actions recommended by the EDP and their coach might also be more likely to participate fully in those activities, which could have biased our effect estimates upward. Alternatively, schools whose principals were struggling to provide effective leadership in their schools might be more likely to engage fully in the EDP and coaching as a source of support, which could have biased our effect estimates downward. To a large extent, we hoped that these differences were controlled for in our models using the baseline characteristics of schools and students that we

were able to observe, but this may or may not have been the case, and our nonexperimental results are, in some cases, not consistent with the range of plausible impacts implied by our experimental findings. In addition, concerns about potential bias were magnified in the context of our analysis of survey outcomes, in which we lacked a control for our survey outcome measures at baseline.

Our quasi-experimental effect estimates may also be biased because of differences in principal turnover. In fully participating schools, principals who remained in coaching for 60 or more hours had to remain in their school for much of the study period in order to do so, and this would not be the case in all control schools. Unfortunately, we had incomplete data on principal transitions in the control school sample, so we could not adjust for any differences in principal mobility between fully participating schools and comparison schools. If comparison schools experienced more turnover, this would have potentially biased our estimates in a positive direction.

Our survey analyses also have limitations. Survey respondents represented an incomplete sample and are likely not missing at random, which introduces challenges with respect to generalizability (i.e., if survey respondents are dissimilar from nonrespondents across the board) and to potential bias (i.e., if different types of treatment principals or teachers responded than control principals or teachers who responded). Fortunately, response rates across treatment and control samples in our surveys were within just two percentage points of each other, and we applied survey weights to correct somewhat for potential nonresponse bias. Even so, it is possible that the pattern of responses in the treatment group was partially determined by individual principals' experiences with the intervention, which would not be the case for control group responders. Finally, when analyzing the survey, we explored a large number of factors, increasing the possibility of finding effects by chance. To minimize this risk, we adjust for multiple hypotheses testing to avoid placing too much confidence in results that are not highly significant.

## Discussion

In one sense, these results are consistent with prior research, since an approximately equal number of prior studies of principal development and coaching have found effects as those that have not found any. But the results from this i3 study differ from the two prior studies *of this particular intervention*—i.e., NISL's EDP, but without paired coaching. Those two studies (Nunnery, Yen, and Ross, 2011; Nunnery et al., 2011) found positive math and/or ELA gains within three school years, and this i3 study did not. This i3 study also included a substantial amount of coaching in addition to the EDP, whereas the two prior studies only included the EDP. Why did this i3 study find no significant effects on student outcomes?

Although we cannot be certain, we believe a likely reason is lower engagement of both districts and principals in the i3 study compared with the two prior studies of the EDP. In the two

prior studies, the SEA paid for the EDP program in Pennsylvania, and a combination of the SEA and LEAs paid for the EDP in Massachusetts, rather than receiving “free” services paid for by a federal grant (as was the case in this i3 study). Further, the districts in these two prior studies offered the EDP to principals as part of a districtwide initiative, as compared with a stand-alone RCT that involved only some middle school principals in a district. In both the PA and the MA studies, principals in the participating LEAs volunteered to take the EDP, as compared with the i3 study in which superintendents informed principals they would be taking it. Finally, in the Massachusetts and Pennsylvania studies, but not in the i3 study, NISL directly involved the districts’ leadership in the same EDP content that principals were learning by including some district leaders in the same EDP program, and providing to those district staff the same materials the NISL EDP facilitators have to deliver the EDP curriculum. We cannot compare attrition and participation rates across the i3, MA, and PA studies, since the MA and PA studies only examined fully participating schools and do not report statistics about attrition. But there are some telling patterns in the i3 study that collectively lead us to conclude there was low commitment to the ED, as follows:

- The first and most important indicator of low commitment in the i3 study was the low take-up rate by principals of the EDP and coaching. Only 57 percent of principals fully participated in the EDP, and only 35 percent fully participated in both the EDP and coaching. State factors played an important role, as we will explain below: As few as 15 percent and as many as 49 percent of principals in any of the three states fully participated in both the EDP and coaching.
- Interviews with state coordinators and notes from coach logs suggest that mandating participation led to low commitment or opt-out once principals realized the time required.
- Certain state and district policies undermined principals’ full participation in the EDP and coaching. This was most true for State C, where there was the lowest principal take-up rate. In that state, according to the NISL state coordinator, a stringent state annual school grading system created an incentive for new school principals in particular to move schools before three full years elapsed lest a failing school grade (which superintendents could attribute to a principal’s three years of leadership) create a lasting black mark for their resumes. Principal mobility was highest in this state—50 percent of State C’s treatment group versus 20 percent in State B and 23 percent in State A—and was a larger driver of incomplete participation in this state than in the other two states.
- The focus on novice principals in middle schools may have contributed to lower LEA commitment to the study. Grade levels six through eight are of particular focus in state accountability systems because each one is tested (unlike in elementary schools or in high schools) and because grade eight is a gateway to high schools. The NISL state coordinator from State C told us that, anecdotally, superintendents and even some principals were hesitant to be absent from the school building for 24 school days of the year to attend the EDP, when the novice principal needed to learn their school and develop a plan, all with a short runway in middle schools especially because of annual school accountability results derived from each grade level.
- There is suggestive evidence that the matching of coaches to principals did not always align with the local LEA culture. The reported match was positive in States A and B and

negative in State C. In State B, all but one of the coaches were retired principals from that state, and this is the state where we saw the highest take-up rate of the EDP and of coaching by principals, the most-positive perceptions of the EDP and coaching, and the most-positive outcomes. But this does not necessarily imply that coaches need always be local. In State A, where principal perceptions of the EDP and coaching and state-specific outcomes fell in the middle of the range, the coaches were “national” in the sense of living out of state. The state coordinator felt that out-of-state coaches were received well by State A principals and that there was a distinct benefit to having coaches and principals come from different state or local contexts. But in State C, the state coordinator heard differently about the reception of “national” coaches. She said that State C principals felt they could not relate to an “outsider” who did not get their local context. State C is also the most rural, which might have contributed to the sense of outsiders.

Putting all these factors together, State C had (1) the least positive inputs on all five of the indicators of buy-in listed above and (2) the least positive outcomes on most outcome measures we examined. The pattern for this state in particular leads us to hypothesize that low buy-in was the main difference between the i3 study and the two prior studies of the EDP.

There are additional differences between the i3 study and the two prior ones that might also explain the discrepancy in results. The first possible reason is the difference in the intervention itself. The version of the EDP in the i3 study had one fewer unit and was delivered over the course of 12 months, compared with the Massachusetts and Pennsylvania version of the EDP that had 13 units and was offered over 15–18 months. (However, the content in both versions of the EDP was highly similar.) Also, the i3 study adds coaching, which the prior studies did not include. It is possible that the coaching had a deleterious effect that reduced a positive effect of the EDP, although we lack any evidence for this and, given principals’ positive perceptions of coaching, we do not think this is a likely explanation.

Another potentially relevant difference is that the previous studies found positive effects when comparing participating principals and schools against other schools statewide, rather than the within-district matched pairs that we compared in this study. Comparing effects with schools outside the district could upwardly bias the effects the Massachusetts and Pennsylvania studies found if those principals who volunteered for the EDP were unobservably different (e.g., more motivated) than principals from the comparison schools. Further, comparing effects across districts would capture any spillover benefits of the intervention on districtwide performance, whereas comparing effects within districts (as we did) would not allow us to detect any positive spillover effects. This is particularly possible because, in the prior studies, district leaders participated in the EDP. It is also possible that the design of the i3 within-district study could have constrained districtwide engagement to some extent, since not all of a district’s eligible principals were able to participate in the EDP at the same time.

Although low participation rates likely diluted effects of the professional development and coaching on effects overall, we did see pockets of effects in the i3 study. The main two are the

following ones that we noted above about principal perception that their school had a strategic plan and that learning was personalized to students:

- The first (the school having a strategic plan) is one that principals from case studies told us was an early stage activity they undertook after completing the EDP. One of the earliest activities several principals did among the nine case studies was to hold all-staff meetings to co-develop or refine the school vision and then prominently and repeatedly repeat that vision—e.g., painted on hallway walls, announced in daily announcements, and included in recurring communications, such as a weekly recap from principals to teachers (Wang et al., 2019). The fact of it being an earlier-stage activity and one that does not incur new monetary costs (unlike other high-cost measures, such as increasing in-school planning time for teachers) might be the reason we detected it in this i3 study.
- The second was personalized learning. This is an intensive activity by our measure that involved the development of units of study for struggling students and frequent formative assessments. These two activities go to the heart of the EDP content about principals as instructional leaders who need to empower and develop teachers' capacity to adapt their instruction in data-informed ways. In the case studies, in which we visited schools twice and interviewed staff over two years, we observed a pattern in which the alignment of teachers' instructional practices (e.g., getting in greater concert among same-grade, same-subject teachers and greater concert across grades in the same subject; figuring out what formative assessments to use; and learning how to use formative assessment data to adjust instruction) was the work of several years. We therefore hypothesize that personalized learning was a late-stage event in the three-year study period and a strong signal of the EDP penetrating the instructional core of the school.

Taken together, there are some promising signs that the EDP and coaching could have effects on substantively important school practices. The promising signs are principals' positive perceptions of the professional development and coaching, even when they did not volunteer for it. The other promising signs are evidence of changes to the two school practices of having a strategic plan and personalizing learning to students. But this study also provides the important cautionary tale that time-intensive and therefore cost-intensive interventions, such as a 24-day principal professional development program and 60 hours of one-on-one coaching of principals, is unlikely to influence student achievement when LEAs or individual principals are not fully bought in.

# Appendix. Technical Appendix

---

This appendix contains the following:

- a list of the survey items in each of the survey scales listed in Table 3.1 of the report
- differences in baseline characteristics of the fully participating treatment and compliant control groups
- information on the rate of student attrition from our analytic sample by study year three
- state-specific effect estimates in some achievement exams.

## Survey Scales from the Principal and Teacher Surveys

Tables A.1 through A.25 are a complete list of the survey scales we developed in accordance with the logic model shown in Figure 2.1 of the main report. The ones shown in Tables A.9 and A.10, which are about frequency of observing classrooms, were excluded from the factor analysis, while all other survey items shown in Tables A.1 through A.25 were included. These are the same factors that appear in Table 3.3 of the main report. The tables list the exact wording from the surveys. The titles of each factor indicate which respondent group (principals or teachers) we draw from.

**Table A.1. Principal Reports the School Has a Strategic Plan**

---

Survey Items
<i>How much do you agree with the following statements about your school?</i>
We have a clear and public vision statement that includes the goal of improving student achievement.
We have a written strategic plan that explicitly links to our vision statement.
All teachers and administrators understand the strategic plan.
We have a recurring review (that is at least annual) of the strategic plan where we use data to monitor our progress in meeting milestones.

---

SOURCE: i3 winter 2019 principal survey.

NOTE: Survey scale is 1–4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent).

**Table A.2. Teachers Report That Principal Is an Effective Leader**

---

Survey Items
<i>How much do you agree with each of the following statements about your principal's management of the school?</i>
My principal is an effective manager who makes the school run smoothly.
My principal runs meetings that are a good use of teachers' time.
My principal has clearly articulated a strategic approach to improving the school.
My principal respects the teachers.
My principal welcomes feedback from teachers and staff.
My principal delegates authority to other school and teacher leaders.
My principal routinely uses data to investigate challenges facing the school.
<i>How much do you agree with each of the following statements about your principal's instructional leadership of the school?</i>
My principal has very high academic expectations for all students, including for students who are the lowest performing.
My principal fosters a school culture where both teachers and school leaders continuously work on improving instruction.
The principal knows what's going on in my classroom.
The principal understands what skills teachers need to foster meaningful student learning.
The principal understands the subject matter that I teach.
The principal uses research about how students learn to design curriculum, assessments, or instruction.
The principal ensures that teachers who need support get intensive coaching
The principal ensures that the majority of staff meeting time is devoted to discussing curriculum and instruction rather than logistics.
Teachers have confidence in the principal as an instructional leader.

---

SOURCE: i3 winter 2019 teacher survey.

NOTE: Survey scale is 1–4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent).

**Table A.3. Principal Indicates School Has Differentiated Roles for Teachers**

---

Survey Items
<i>To what extent do you reserve the following responsibilities for your highly-skilled teachers?</i>
Assignment to low-performing students
Modelling lessons or practices for other teachers
Mentoring other teachers
Leading teacher teams
Writing student assessments
Writing curriculum

---

SOURCE: i3 winter 2019 principal survey.

NOTE: Survey scale is 1–4 (1 = none of the time; 2 = some of the time; 3 = most of the time; 4 = all of the time).

**Table A.4. Teachers Indicate School Has Differentiated Roles for Teachers**

---

Survey Items
<i>What types of leadership roles are available to teachers in your school? Check all that apply. [These are in addition to grade-level leadership or department-chair leadership.]</i>
Mentoring other teachers
Instructional coaching of other teachers
Modelling lessons for teachers
Evaluating teachers in their classrooms
Leading teacher teams (e.g., a professional learning community)
Writing student assessments
Writing curriculum
Writing / developing professional development/in-service for teachers

---

SOURCE: i3 winter 2019 teacher survey.

NOTE: For this item, we took the average value of the number of roles that teacher respondents in a given school selected. We first calculated the percentage of the eight items that a respondent said were available at the school and then took the average of those percentages among all teachers in a given school. The more roles that respondents selected, the higher the average value of differentiated roles for a given school.

**Table A.5. Teachers Indicate They Have Input into Staffing, Curriculum, or Professional Development**

---

Survey Items
<i>To what extent do teachers participate in making decisions about the following aspects of your school?</i>
Determining books and other instructional materials used in classrooms.
Establishing the curriculum and instructional approach.
Determining the scheduling and content of professional development.
Hiring new teachers.

---

SOURCE: i3 winter 2019 teacher survey.

NOTE: Survey scale is 1–3 (1 = not at all; 2 = our input is encouraged; 3 = we fully participate).

**Table A.6. Principal Ensures Curriculum Is Aligned and Evidence-Based**

---

Survey Items
<i>Thinking about the school year 2017–2018, how much do you agree with the following statements about instructional planning at your school?</i>
Curriculum, instruction, and learning materials are well-aligned across classrooms within a grade and across the different grade levels at this school.
We select curriculum based on evidence of effectiveness.
We select curriculum and plan instruction to reflect the latest research about how students learn.
My school engages in a comprehensive review of our curricula and common assessments to ensure they align with standards.

---

SOURCE: i3 winter 2019 principal survey.

NOTE: Survey scale is 1–4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent).

**Table A.7. Principal Indicates High Priority for Use of Student Assessment Data**

Survey Items
<i>What priority does your school give to the following assessment practices?</i>
Teachers of the same subject area and grade aligning their interim or informal assessments with one another.
Teachers using frequent, short, informal classroom assessments to assess individual students' educational needs.
Teachers using a range of assessment data to identify student misunderstandings.
Using interim or summative test scores to assess and respond to individual students' educational needs.

SOURCE: i3 winter 2019 principal survey.  
NOTE: Survey scale is 1–4 (1 = not a priority; 2 = low priority; 3 = medium priority; 4 = high priority).

**Table A.8. Principal Indicates School Offers High-Quality Professional Development for Teachers**

Survey Items
<i>How much do you agree with the following statements about professional development options for teachers in school year 2017–2018?</i>
Were mostly developed at my school by school leaders or teachers.
Are carried out over time to allow an iterative process of teachers applying what they learn in their classroom.
Involve teachers in reviewing student work together as a part of the professional development.
Devoted substantial time to developing units of study to support struggling students.
Devoted substantial time to techniques for assessing student understanding as part of regular classroom instruction.

SOURCE: i3 winter 2019 principal survey.  
NOTE: Survey scale is 1–4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent).

**Table A.9. Teachers Indicate That Administrators Observe Teachers' Classrooms**

Survey Items
<i>Thinking about the 2017–18 school year, approximately how often did the following people observe your classroom?</i>
The principal
An administrator who is not the principal (e.g., assistant principal, dean)
An instructional coach

SOURCE: i3 winter 2019 teacher survey.  
NOTE: Survey scale is 1–5 (1 = never; 2 = 1–2 times per year; 3 = 3–4 times per year; 4 = about once a month; 5 = about once a week).

**Table A.10. Teachers Indicate That Teachers Observe Other Teachers' Classrooms**

---

Survey Items
<i>Thinking about the 2017–18 school year, approximately how often did the following people observe your classroom?</i>
One or more teachers to aid your professional growth
One or more teachers to learn from your success or to learn about a new approach or method

---

SOURCE: i3 winter 2019 teacher survey.  
NOTE: Survey scale is 1–5 (1 = never; 2 = 1–2 times per year; 3 = 3–4 times per year; 4 = about once a month; 5 = about once a week).

**Table A.11. Teachers Say They Have Received Actionable Feedback**

---

Survey Items
<i>How much do you agree with the following statements about the instructional feedback you receive from school leaders and/or coaches?</i>
I get useful feedback on my written lesson plans.
I get useful feedback based on reviews of samples of my students' written work.
I get useful feedback on my techniques for assessing students' understanding of the material.
I get useful post-classroom observation feedback.
The feedback I receive from school leadership is both substantive and fair.
I receive the support I need to implement suggested changes.

---

SOURCE: i3 winter 2019 teacher survey.  
NOTE: Survey scale is 1–4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent).

**Table A.12. Principal Reports There Is a Positive School Culture**

---

Survey Items
<i>How much do you agree with the following statements about your school?</i>
Adults and students respect one another at this school.
Standards for student behavior are clear and consistently upheld by all teachers.
Teachers, administrators, and parents assume joint responsibility for student discipline.

---

SOURCE: i3 winter 2019 principal survey.  
NOTE: Survey scale is 1–4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent).

**Table A.13. Principal and Teachers Feel School Is Unsafe and Disorderly**

---

<b>Survey Items</b>
<i>To what extent is student learning hindered by the following behaviors in your school?</i>
Students arriving late
Students creating classroom disturbances
Students bullying others, including verbal abuse, online abuse, intimidation, or physical injury
Students intimidating or verbally abusing teachers or staff
Students using/possessing drugs and/or alcohol
Teachers arriving late
Teachers being absent
Teachers being unprepared for their lessons

---

SOURCE: i3 winter 2019 principal survey and i3 winter 2019 teacher survey.

NOTE: Survey scale is 1–4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent).

**Table A.14. Teachers Endorse Their School**

---

<b>Survey Items</b>
<i>Please mark the extent to which you disagree or agree with the following statement:</i>
“I would recommend this school to parents seeking a place for their child.”
<i>How much do you agree with the following statements about teachers at your school?</i>
Teachers feel good about the direction in which the school is heading.

---

SOURCE: i3 winter 2019 teacher survey.

NOTE: Survey scale is 1–4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent).

**Table A.15. Principal’s Sense of Educators’ Efficacy**

---

<b>Survey Items</b>
<i>How much impact can you and your teachers have on the following?</i>
Struggling students’ persistence on difficult assignments.
Students attending school.
Teaching students how to manage their behavior.
Making sure all our students are on the path to college and career readiness.

---

SOURCE: i3 winter 2019 principal survey.

NOTE: Survey scale is 1–4 (1 = no impact; 2 = a little impact; 3 = a moderate impact; 4 = a great deal of impact).

**Table A.16. Teachers' Sense of Self-Efficacy**

---

<b>Survey Items</b>
<i>How much impact can you, as a teacher, have on the following?</i>
Reluctant students' motivation to do work
Students' self-perceptions of their ability to master advanced topics.
Students' attendance at school.
Improving students' behavior.
Ensuring all of our students are on a path to college and career readiness.

---

SOURCE: i3 winter 2019 teacher survey.

NOTE: Survey scale is 1–4 (1 = no impact; 2 = a little impact; 3 = a moderate impact; 4 = a great deal of impact).

**Table A.17. Principal Believes Teachers Are Committed to High-Quality Academics**

---

<b>Survey Items</b>
<i>How much do you agree with the following statements about your school?</i>
Virtually all the teachers share a sense of urgency that they help all of their students be on the path for career or college readiness.
My teachers work to get lower-performing students get beyond remediation and onto more advanced content.
My teachers are willing to make changes to their instruction to increase student learning.
There is a system for teachers to collaborate with me and other school leaders to improve the school.
My teachers are receptive to both giving and receiving feedback about their instructional practice.

---

SOURCE: i3 winter 2019 principal survey.

NOTE: Survey scale is 1–4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent).

**Table A.18. Teachers Report the School Has High-Quality Academics**

---

Survey Items
<i>How much do you agree with the following statements about teachers in your school?</i>
Teachers share a sense of urgency around helping all students be college and career ready.
Teachers systematically consider revising their units of study from one year to the next.
<i>How much do you agree with the following statements about curriculum and instructional design at your school?</i>
My school regularly reviews our curricula and our common assessments to ensure they are aligned with standards.
My school designs or revises curricula to reflect the latest knowledge on how students learn.
Teachers and administrators design units of study specifically for struggling students.
Teachers work together to establish a shared vision for what constitutes high quality student work.
<i>How much do you agree with the following statements about the level of academic challenge in your own classroom(s)?</i>
Students use data and/or text references to support their ideas.
Students provide and receive constructive feedback to and from their peers.
Students explain how they reach conclusions when there is conflicting evidence.
Students in my classrooms could explain to a visitor how they worked through a challenging idea.
Students synthesize and apply information to analyze complex problems.
<i>How much do you agree with the following statements about your professional development (PD) experiences from school year 2017–18? Select “not applicable” if you did not work at this school in 2017–18.</i>
They helped me to improve the quality of my instruction.
They were part of a sustained and coherent program, rather than being short-term or ad-hoc.
I reviewed student work with other teachers as a part of my professional development.
In the PD, we devoted substantial time to developing units of study to support struggling students
In the PD, we devoted substantial time to techniques for assessing student understanding as part of regular classroom instruction.
<i>What priority does your school give to the following academic activities?</i>
Ensuring that all students, including lower-performing students, understand the concepts behind the material they are learning
Ensuring that lower-performing students get beyond remediation and onto more advanced content.
Eliciting and then addressing students’ prior knowledge when introducing a new topic.
Routinely using short, informal assessments to both assess student understanding and then adjust instruction.
Teachers aligning their instruction across a subject.
Teachers aligning their instruction vertically across grade levels.
Providing students with examples of high-quality work for students to refer to when revising their own work.

---

SOURCE: i3 winter 2019 teacher survey.

NOTE: Survey scale is 1-4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent; or 1 = not a priority; 2 = low priority; 3 = medium priority; 4 = high priority).

**Table A.19. Teachers Focus on Students' Higher Order Skills**

---

Survey Items
<i>How much do you agree with the following statements about performance expectations for your students in the most recent school year (2017–2018)?</i>
Teachers have a shared vision for what constitutes high quality student work.
Teachers ensure that curriculum and assessments require students to solve complex problems.
Teachers encourage students to build on each other's ideas during discussion.
Teachers identify samples of student work to help students understand the school's expectations for learning.
Teachers ensure instruction in the current grade prepares students for the higher-level work they will encounter in subsequent grades.
Teachers review state and district standards in comparison to standards from high-performing states and countries.
Teachers review state and district standards in comparison to 21st century workplace demands.

---

SOURCE: i3 winter 2019 principal survey.

NOTE: Survey scale is 1–4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent).

**Table A.20. Principal Indicates Instruction Is Personalized to Students**

---

Survey Items
<i>Thinking about the school year 2017–2018, how much do you agree with the following statements about instructional planning at your school?</i>
Teachers work together to develop short, informal assessments to routinely assess student understanding and adjust instruction.
Teacher and administrators design units of study specifically for struggling students.

---

Source: i3 winter 2019 principal survey.

NOTE: Survey scale is 1–4 (1 = not at all; 2 = very little; 3 = somewhat; 4 = to a great extent).

**Table A.21. Principal Prioritizes NISL's Concepts of Learning**

---

Survey Items
<i>What priority does your school give to the following activities?</i>
Devoting significant time in a majority of staff meetings to discuss instruction (e.g., pedagogy to foster critical thinking, support of struggling students, formative assessment).
Providing the resources (such as professional development, coaching, mentoring, or teacher teams) to help teachers to be more reflective about their instructional practices.
Encouraging teachers to first elicit and address students' preconceptions when introducing a new topic.
Ensuring students: (1) possess factual knowledge, (2) use their knowledge to recognize patterns or connect ideas, and (3) can apply their knowledge in unfamiliar situations.
Focusing students on their own learning process—i.e., surface their preconceptions and take what they currently know as the starting point for further learning—to help students control and direct their own learning efforts.
Students using data, texts, or original sources to support their ideas.
All or almost all students can clearly articulate what concepts they are trying to learn.

---

SOURCE: i3 winter 2019 principal survey.

NOTE: Survey scale is 1–4 (1 = not a priority; 2 = low priority; 3 = medium priority; 4 = high priority).

**Table A.22. Principal Indicates Teachers Collaborate About Matters of Instruction and Professional Development**

---

**Survey Items**

---

*How often do teachers in your school engage in the following activities with other teachers?*

- Meet to design courses or curriculum.
- Meet to develop or revise components of school improvement plans.
- Meet to provide feedback to teachers based on observations of their practice.
- Meet to discuss what they observe after a lead teacher models a lesson or instructional practice.
- Meet to plan professional development sessions.

---

SOURCE: i3 winter 2019 principal survey.  
 NOTE: Survey scale is 1–4 (1 = never; 2 = 1–4 times a year; 3 = at least monthly; 4 = at least weekly).

**Table A.23. Teachers Indicate They Collaborate About Instruction**

---

**Survey Items**

---

*How often do teachers engage in the following activities with other teachers?*

- Meet with teachers in the same grade or the same subject.
- Meet with teachers or administrators to design courses or curriculum.
- Meet to ensure instruction in the current grade prepares students for the higher-level work they will encounter in subsequent grades.
- Meet to review and discuss student data (e.g., periodic student assessments, student work, attendance).
- Meet to monitor progress toward school improvement plan milestones.
- Meet to plan professional development sessions.
- Meet to develop formal and informal assessments

*Approximately how many minutes per week—on average—are built into your school’s schedule for teachers to collaborate with one another on lesson planning and instructional decisions? Do not include time allocated to coaching, mentoring or your individual planning.*

---

SOURCE: i3 winter 2019 teacher survey.  
 NOTE: Survey scale is 1–4 (1 = never; 2 = 1–4 times per year; 3 = at least monthly; 4 = at least weekly; or 1 = none; 2 = 30 minutes or less per week; 3 = 31–60 minutes per week; 4 = 61 or more minutes per week).

**Table A.24. Principal Has Autonomy to Deploy School Resources**

Survey Items
<i>To what extent do you or your leadership team have control over the following?</i>
Changing the class/bell schedule to be more responsive to student learning needs.
Directing additional resources to support struggling students such as assigning high-performing teachers for low-performing students
Determining the scheduling and content of teacher professional development programs.
Hiring new full-time teachers
Reassigning or removing teachers
Deciding how the school budget will be spent

SOURCE: i3 winter 2019 principal survey.  
NOTE: Survey scale is 1–4 (1 = no control; 2 = a little control; 3 = a fair amount of control; 4 = a great amount of control)

**Table A.25. Principal Indicates They Have Curricular Autonomy**

Survey Items
<i>To what extent do you or your leadership team have control over the following?</i>
Setting academic standards for students
Selecting curricula
Selecting summative assessments for students (beyond state-mandated ones)
Teachers have a role in selecting curricula, materials, and/or the instructional approach.

SOURCE: i3 winter 2019 principal survey.  
NOTES: Survey scale is 1–4 (1 = no control; 2 = a little control; 3 = a fair amount of control; 4 = a great amount of control).

## Descriptive Statistics, Sensitivity Analyses, State-by-State Results

Table A.26 provides formal tests of the baseline equivalence between the fully participating treatment group and the control-compliant group, overall and for each state. Table A.27 shows the rates of missing achievement outcome data for students. Table A.28 shows principals' responses to questions about their experiences with the EDP, by state. Table A.29 shows the effects of offering the EDP and coaching on achievement exams that are distinct to specific study states, and Table A.30 shows the effects of full participation in the EDP and coaching on those exams. Table A.31 shows the effects of the EDP and coaching on student grade progression. In Table A.32, we provide results separately by state for two factors: (1) the effect of full participation in the EDP and coaching on principals' perceptions that their school had a strategic plan and (2) the effect of full participation in the EDP and coaching on principals' perceptions that instruction was personalized to students' needs

**Table A.26. Standardized Treatment Versus Control Group Differences on Baseline Test Scores, by Sample Used for Analyses of Full Participation in EDP and Coaching**

	Study Year	Sample for Outcome: ELA Test Score		Sample for Outcome: Mathematics Test Score	
		Covariate: Baseline ELA Test Score	Covariate: Baseline Math Test Score	Covariate: Baseline ELA Test Score	Covariate: Baseline Math Test Score
All three states	Year 3	-0.020	-0.070	-0.013	-0.009
	Year 2	-0.023	-0.070	-0.028	-0.069
	Year 1	-0.015	-0.067	-0.019	-0.070
State A	Year 3	-0.041	-0.030	-0.042	-0.031
	Year 2	-0.049	-0.032	-0.049	-0.031
	Year 1	-0.027	-0.025	-0.026	-0.027
State B	Year 3	0.014	-0.021	0.014	-0.022
	Year 2	0.008	-0.028	0.011	-0.024
	Year 1	0.014	-0.021	0.007	-0.026
State C	Year 3	-0.072	-0.256	-0.072	-0.252
	Year 2	-0.060	-0.234	-0.066	-0.239
	Year 1	-0.072	-0.248	-0.073	-0.248

NOTE: Standardized differences were calculated by dividing the difference in the mean of a given covariate between the treatment and control groups used in the analysis of a given outcome by the standard deviation of that outcome. The differences were calculated for each state separately. The full sample differences were then calculated by weighting state-specific differences by the number of schools used in the analysis for that state.

**Table A.27. Percentage of Initially Randomized Students Without Achievement Outcomes in Year Three, by Subject and Assignment Group**

Subject	Overall	Treatment	Control	Difference
ELA Year 3	11.8	11.6	12.0	-0.4
Mathematics Year 3	11.8	11.6	12.0	-0.4

NOTE: Percentage of students reported is from the sample of students present in study schools in year one and in grade six in year one across all three study states.

**Table A.28. Mean Responses Overall and by State for Selected Survey Questions About Participating Principals' Experiences with the EDP**

Survey Questions	Overall		Test of Group Differences (P-Value)	State A		State B		State C	
	Mean	S.D.		Mean	S.D.	Mean	S.D.	Mean	S.D.
Average of all 11 questions related to principals' views of and usage of the EDP.	3.38	(0.41)	0.006 <sup>c</sup>	3.28	(0.39)	3.59	(0.35)	3.17	(0.43)
I continue to implement my ALP.	3.40	(0.77)	0.026 <sup>b</sup>	3.27	(0.59)	3.76	(0.56)	3.00	(1.05)
I use teams to help facilitate distributed leadership.	3.33	(0.75)	0.082 <sup>a</sup>	3.00	(0.85)	3.59	(0.62)	3.40	(0.70)
I have worked to align curriculum and assessments to high performance standards.	3.21	(0.81)	0.067 <sup>a</sup>	2.87	(0.74)	3.53	(0.80)	3.20	(0.79)
I use the "NISL wheel" to decide how to change the systems at my school.	2.88	(0.83)	0.090 <sup>a</sup>	2.93	(0.70)	3.12	(0.78)	2.40	(0.97)
I routinely apply the idea of "all means all" to make decisions at my school.	3.40	(0.73)	0.005 <sup>c</sup>	3.20	(0.77)	3.82	(0.39)	3.00	(0.82)

NOTE: Results shown from treatment group principals who participated in at least some of the EDP. Response options for questions ranged from 1 ("not at all" / "not a focus") to 4 ("to a great extent" / "adopted or increased to a great extent"). Results are shown for the average across all 11 questions related to participants' views of the EDP and for those specific questions for which differences were at least marginally significant ( $p < 0.1$ ) in an analysis testing for variance across state.

<sup>a</sup>  $p < 0.1$ .

<sup>b</sup>  $p < 0.05$ .

<sup>c</sup>  $p < 0.01$ .

**Table A.29. Impact of Offering the EDP and Coaching on Student Achievement on State-Specific Exams**

Sample	Study Year	Exam	Number of Schools	Estimates	
				Coeff.	Std. Error
State B	Year 3	Mathematics	145	0.025	(0.036)
	Year 3	Algebra 1	145	0.045	(0.048)
	Year 3	Geometry	125	0.072	(0.055)
	Year 2	Mathematics	145	0.030	(0.029)
	Year 2	Algebra 1	129	0.089	(0.060)
State C	Year 3	Science	76	-0.001	(0.045)

NOTE: Algebra 1 was administered to a subset of eighth-grade students in all schools in State B and in most schools in grade seven. Geometry was administered to a subset of eighth-grade students in most schools in State B. No results are statistically significant at the  $p < 0.1$  level or lower.

**Table A.30. Impact of Full Participation in the EDP and Coaching on Student Achievement on State-Specific Exams**

Sample	Study Year	Exam	Number of Fully Participating Treatment Schools	Number of Compliant Control Schools	Estimates	
					Coeff.	Std. error
State B	Year 3	Mathematics	36	70	0.044	(0.053)
	Year 3	Algebra 1	36	70	0.133 <sup>a</sup>	(0.061)
	Year 3	Geometry	33	59	0.104	(0.076)
	Year 2	Mathematics	36	70	0.042	(0.035)
	Year 2	Algebra 1	33	63	0.086	(0.080)
State C	Year 3	Science	6	36	-0.013	(0.086)

NOTE: Algebra 1 was administered to a subset of eighth-grade students in all schools in State B and in most schools in grade seven. Geometry was administered to a subset of eighth-grade students in most schools in State B. *Mathematics* here refers to scores from the regular standardized mathematics exam, which was distinct from the algebra 1 and geometry exams.

<sup>a</sup>  $p < 0.05$

**Table A.31. Impact of the EDP and Coaching on Student Grade Progression**

Sample	Number of Treatment Schools	Number of Control Schools	Effects of Offering Treatment		Effects of Any Participation in Treatment		Upper Bound of Effects of Full Participation in Treatment	
			Coeff.	Std. Error	Coeff.	Std. Error	Coeff.	Std. Error
All three states	161	162	0.000	(0.000)	0.000	(0.001)	-0.000	(0.002)
State A	51	51	0.000	(0.001)	0.000	(0.001)	0.000	(0.002)
State B	73	72	-0.000	(0.001)	-0.000	(0.002)	-0.000	(0.003)
State C	37	39	-0.010	(0.008)	-0.020	(0.019)	-0.056	(0.058)

NOTE: Grade progression is measured as the rate at which sixth graders in the first year following treatment were in grade eight as of the third year following treatment. Estimates from nonexperimental analyses are available upon request.

**Table A.32. State-by-State Estimates of the Effects of EDP and Coaching on Selected Survey Factors**

Factor	State	Effects of Offering Treatment		Effects of Full Participation	
		Coeff.	Std. Error	Coeff.	Std. Error
Strategic plan for the school	State A	0.377	(0.346)	1.419 <sup>b</sup>	(0.519)
	State B	0.543 <sup>a</sup>	(0.286)	0.463	(0.318)
	State C	0.164	(0.297)	0.007	(0.871)
Personalized instruction for students	State A	0.314	(0.331)	1.043 <sup>b</sup>	(0.454)
	State B	0.562 <sup>b</sup>	(0.250)	0.889 <sup>c</sup>	(0.240)
	State C	-0.098	(0.300)	-0.0607	(0.673)

NOTE: This table includes the two survey scales for which we found highly statistically significant effects in our pooled three-state analyses. Both of these survey scales are from the principal survey.

<sup>a</sup>  $p < 0.1$ .

<sup>b</sup>  $p < 0.05$ .

<sup>c</sup>  $p < 0.001$ .

## References

---

- Alwin, Duane. F., "The Use of Factor Analysis in the Construction of Linear Composites in Social Research," *Sociological Methods and Research*, Vol. 2, No. 2, November 1973, pp. 191–212.
- Athey, Susan, and Guido Imbens, *Machine Learning Methods for Estimating Heterogeneous Causal Effects*, Stanford, Calif.: Stanford University, Graduate School of Business, 2015.
- , "Recursive Partitioning for Heterogeneous Causal Effects," *Proceedings of the National Academy of Sciences*, Vol. 113, No. 27, July 5, 2016, pp. 7353–7360.
- Bang, Heejung, and James M. Robins, "Doubly Robust Estimation in Missing Data and Causal Interference Models," *Biometrics*, Vol. 61, No. 4, December 2005, pp. 962–973.
- Barnett, Bruce G., and Gary R. O'Mahony, "Mentoring and Coaching Programs for the Professional Development of School Leaders," in Jacky Lumby, Gary Crow, and Petros Pashiardis, eds., *International Handbook on the Preparation and Development of School Leaders*, New York: Routledge, 2008, pp. 232–262.
- Baron, Reuben M., and David A. Kenny, "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations," *Journal of Personality and Social Psychology*, Vol. 51, No. 6, 1986, pp. 1173–1182.
- Benjamini, Yoav, and Yosef Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 57, No. 1, 1995, pp. 289–300.
- Bloom, Gary, Duff L. Danilovich, and Janet Fogel, "Passing the Baton: A New Program from ACSA and the New Teacher Center at UC Santa Cruz Is Improving the Way a New Generation of Site Leaders Is Prepared and Supported," *Leadership*, Vol. 35, No. 1, 2005, pp. 30–33.
- Bollen, Kenneth A., and Richard Lennox, "Conventional Wisdom on Measurement: A Structural Equation Perspective," *Psychological Bulletin*, Vol. 110, No. 2, September 1991, pp. 305–314.
- Bossi, Mike, "Does Leadership Coaching Really Work?" *Leadership*, Vol. 38, No. 1, 2008, pp. 31–36.
- Camburn, Eric M., Ellen Goldring, Henry May, Jonathan Supovitz, Carol Barnes, and James P. Spillane, *Lessons Learned from an Experimental Evaluation of a Principal Professional*

- Development Program*, Philadelphia, Pa.: Consortium for Policy Research in Education, 2007.
- Cefalu, Matthew, Shuangshuang Liu, and Craig Martin, *Toolkit for Weighting and Analysis of Nonequivalent Groups: A Tutorial on the TWANG Commands for Stata*, Santa Monica, Calif.: RAND Corporation, TL-170-NIDA, 2015. As of March 16, 2020: <https://www.rand.org/pubs/tools/TL170.html>
- Comrey, Andrew L., and Howard B. Lee, *A First Course in Factor Analysis*, 2nd ed., Hillsdale, N.J.: Lawrence Erlbaum Associates, 1992.
- Cooper, Harris, Larry V. Hedges, and Jeffrey C. Valentine, eds., *The Handbook of Research Synthesis and Meta-Analysis*, 2nd ed., New York: Russell Sage Foundation, 2009.
- Cronbach, Lee J., *Research on Classrooms and Schools: Formulation of Questions, Design, and Analysis*, Stanford, Calif.: Stanford University, Stanford Evaluation Consortium, July 1976.
- Davis, Jonathan M. V., and Sara B. Heller, “Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs,” *American Economic Review*, Vol. 107, No. 5, May 2017, pp. 546–550.
- , “Rethinking the Benefits of Youth Employment Programs: The Heterogenous Effects of Summer Jobs,” *Review of Economics and Statistics*, forthcoming.
- Desimone, Laura M., and Katie Pak, “Instructional Coaching as High-Quality Professional Development,” *Theory into Practice*, Vol. 56, No. 1, 2017, pp. 3–12,
- Geman, Stuart, and Donald Geman, “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No. 6, November 1984, pp. 721–741.
- George W. Bush Institute, *Principal Talent Management According to the Evidence: A Review of the Literature*, Dallas, Tex., October 2016.
- Goff, Peter, J. Edward Guthrie, Ellen Goldring, and Leonard Bickman, “Changing Principals’ Leadership Through Feedback and Coaching,” *Journal of Educational Administration*, Vol. 52, No. 5, 2014, pp. 682–704.
- Goldring, Ellen, Andrew Porter, Joseph Murphy, Stephen N. Elliot, and Xiu Cravens, “Assessing Learning-Centered Leadership: Connections to Research, Professional Standards, and Current Practices,” *Leadership and Policy in Schools*, Vol. 8 No. 1, 2009, pp. 1–36.
- Grissom, Jason A., and James R. Harrington, “Investing in Administrator Efficacy: An Examination of Professional Development as a Tool for Enhancing Principal Effectiveness,” *American Journal of Education*, Vol. 116, No. 4, August 2010, pp. 583–612.

- Herman, Rebecca, Susan M. Gates, Aziza Arifkhanova, Mark Barrett, Andriy Bega, Emilio R. Chavez-Herrerias, Eugeniu Han, Mark Harris, Katya Migacheva, Rachel Ross, Jennifer T. Leschitz, and Stephani L. Wrabel, *School Leadership Interventions Under the Every Student Succeeds Act: Evidence Review*, Santa Monica, Calif.: RAND Corporation, RR-1550-3-WF, 2017. As of March 16, 2020:  
[https://www.rand.org/pubs/research\\_reports/RR1550-3.html](https://www.rand.org/pubs/research_reports/RR1550-3.html)
- Herrmann, Mariesa, Melissa Clark, Susanne James-Burdumy, Christina Tuttle, Tim Kautz, Virginia Knechtel, Dallas Dotter, Claire Smither Wulsin, and John Deke, *The Effects of a Principal Professional Development Program Focused on Instructional Leadership: Appendices*, Washington, D.C.: U.S. Department of Education, October 2019.
- Hobson, Andy, *Mentoring and Coaching for New Leaders: Full Report*, Nottingham, United Kingdom: National College for School Leadership, 2003.
- Hox, Joop J., Mirjam Moerbeek, and Rens van de Schoot, *Multilevel Analysis: Techniques and Applications*, 3rd ed., New York: Routledge, 2017.
- Huff, Jason, Courtney Preston, and Ellen Goldring, “Implementation of a Coaching Program for School Principals: Evaluating Coaches’ Strategies and the Results,” *Educational Management Administration and Leadership*, Vol. 41, No. 4, 2013, pp. 504–526.
- Imai, Kosuke, Luke Keele, and Teppei Yamamoto, “Identification, Inference and Sensitivity Analysis for Causal Mediation Effects,” *Statistical Science*, Vol. 25, No. 1, 2010, pp. 51–71.
- Jacob, Robin, Roger Goddard, Minjung Kim, Robert Miller, and Yvonne Goddard, “Exploring the Causal Impact of the McREL Balanced Leadership Program on Leadership, Principal Efficacy, Instructional Climate, Educator Turnover, and Student Achievement,” *Educational Evaluation and Policy Analysis*, Vol. 37, No. 3, September 2015, pp. 314–332.
- Klar, Hans W., Kristin Shawn Huggins, Parker M. Andreoli, and Frederick C. Buskey, “Developing Rural School Leaders Through Leadership Coaching: A Transformative Approach,” *Leadership and Policy in Schools*, March 21, 2019.
- Lackritz, Anne D., Maria Cseh, and Donald Wise, “Leadership Coaching: A Multiple-Case Study of Urban Public Charter School Principals’ Experiences,” *Mentoring and Tutoring: Partnership in Learning*, Vol. 27, No. 1, 2019, pp. 5–25.
- Lee, Omaira Z., *The Leadership Gap: Preparing Leaders for Urban Schools*, dissertation, University of Southern California, August 2010, Ann Arbor, Mich.: ProQuest, 3418101, 2010.
- Lüdtke, Oliver, Alexander Robitzsch, Ulrich Trautwein, and Mareike Kunter, “Assessing the Impact of Learning Environments: How to Use Student Ratings of Classroom or School

Characteristics in Multilevel Modeling,” *Contemporary Educational Psychology*, Vol. 34, No. 2, April 2009, pp. 120–131.

McCaffrey, Daniel F., Beth Ann Griffin, Daniel Almirall, Mary Ellen Slaughter, Rajeev Ramchand, and Lane F. Burgette, “A Tutorial on Propensity Score Estimation for Multiple Treatments Using Generalized Boosted Models,” *Statistics in Medicine*, Vol. 32, No. 19, 2013, pp. 3388–3414.

Miller, Robert James, Roger D. Goddard, Minjung Kim, Robin Jacob, Yvonne Goddard, and Patricia Schroeder, “Can Professional Development Improve School Leadership? Results from a Randomized Control Trial Assessing the Impact of McREL’s Balanced Leadership Program on Principals in Rural Michigan Schools,” *Educational Administration Quarterly*, Vol. 52, No. 4, 2016, pp. 531–566.

Murphy, Joseph F., Ellen B. Goldring, Xiu C. Cravens, Stephen N. Elliott, and Andrew C. Porter, “The Vanderbilt Assessment of Leadership in Education: Measuring Learning-Centered Leadership,” *Journal of East China Normal University*, Vol. 29 No. 1, 2011, pp. 1–10.

Muthén, Bengt, and David Kaplan, “A Comparison of Some Methodologies for the Factor Analysis of Non-Normal Likert Variables,” *British Journal of Mathematical and Statistical Psychology*, Vol. 38, No. 2, 1985, pp. 171–189.

National Center for Education Statistics, “Table 204.10. Number and Percentage Of Public School Students Eligible for Free or Reduced-Price Lunch, by State: Selected Years, 2000–01 Through 2014–15,” Washington, D.C., March 2017. As of March 16, 2020: [https://nces.ed.gov/programs/digest/d16/tables/dt16\\_204.10.asp](https://nces.ed.gov/programs/digest/d16/tables/dt16_204.10.asp)

Nunnery, John A., Steven M. Ross, Shanan Chappell, Shana Pribesh, and Elizabeth Hoag-Carhart, *The Impact of the NISL Executive Development Program on School Performance in Massachusetts: Cohort 2 Results*, Norfolk, Va.: Old Dominion University, Center for Educational Partnerships, July 2011.

Nunnery, John A., Cherng-Jyh Yen, and Steven M. Ross, *Effects of the National Institute for School Leadership’s Executive Development Program on School Performance in Pennsylvania: 2006–2010 Pilot Cohort Results*, Norfolk, Va.: Old Dominion University, Center for Educational Partnerships, March 2011.

Organisation for Economic Co-Operation and Development, *TALIS 2008 Technical Report: Teaching and Learning International Survey*, Paris, 2010. As of April 16, 2020: <https://www.oecd.org/education/school/44978960.pdf>

Pedhazur, Elazar J., and Liora Pedhazur Schmelkin, *Measurement, Design, and Analysis: An Integrated Approach*, student ed., Mahwah, N.J.: Lawrence Erlbaum Associates, 1991.

- R Development Core Team, “The R Project for Statistical Computing,” R version 3.5.3 (Great Truth), 2019. As of March 16, 2020:  
<https://www.R-project.org/>
- Raudenbush, Stephen W., Andres Martinez, and Jessaca Spybrook, “Strategies for Improving Precision in Group-Randomized Experiments,” *Educational Evaluation and Policy Analysis*, Vol. 29, No. 1, March 2007, pp. 5–29.
- Rubin, Donald B., *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley, 1987.
- Ryu, Ehri, and Stephen G. West, “Level-Specific Evaluation of Model Fit in Multilevel Structural Equation Modeling,” *Structural Equation Modeling*, Vol. 16, No. 4, 2009, pp. 583–601.
- Silver, Michael, Chad R. Lochmiller, Michael A. Copland, and Ann Marie Tripps, “Supporting New School Leaders: Findings from a University-Based Leadership Coaching Program for New Administrators,” *Mentoring and Tutoring: Partnership in Learning*, Vol. 17, No. 3, 2009, pp. 215–232.
- Strong, M., A. Barrett, and G. Bloom, “Supporting the New Principal: Managerial and Instructional Leadership in a Principal Induction Program,” presented at the Annual Meeting of the American Educational Research Association, Chicago, Ill., 2003.
- Stuart, Elizabeth A., “Matching Methods for Causal Inference: A Review and a Look Forward,” *Statistical Science*, Vol. 25, No. 1, 2010, pp. 1–21.
- Tabachnick, Barbara G., and Linda S. Fidell, *Using Multivariate Statistics*, Boston, 4th ed., Mass.: Allyn and Bacon, 2001.
- van Buuren, Stef, and Karin Groothuis-Oudshoorn, “mice: Multivariate Imputation by Chained Equations in R,” *Journal of Statistical Software*, Vol. 45, No. 3, December 2011, pp. 1–67.
- Wang, Elaine Lin, Heather L. Schwartz, Monica Mean, Laura Stelitano, and Benjamin K. Master, *Putting Professional Learning to Work: What Principals Do with Their Executive Development Program Learning*, Santa Monica, Calif.: RAND Corporation, RR-3082-DOED, 2019. As of February 4, 2020:  
[https://www.rand.org/pubs/research\\_reports/RR3082.html](https://www.rand.org/pubs/research_reports/RR3082.html)
- Warren, Susan R., and Virginia E. Kelsen, “Leadership Coaching: Building the Capacity of Urban Principals in Underperforming Schools,” *Journal of Urban Learning, Teaching, and Research*, Vol. 9, 2013, pp. 18–31.
- What Works Clearinghouse, *Standards Handbook, Version 4.1*, Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, January 2020a. As of April 16, 2020:

<https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf>

———, “WWC Review of This Study,” review of McREL Balanced Leadership Program School Leadership Review Protocol, webpage, March 2020b. As of March 11, 2020: <https://ies.ed.gov/ncee/wwc/Study/81452>

Wise, Donald, “School Leadership Coaching: What Does It Look Like,” *International Journal of Educational Leadership Preparation*, Vol. 5, No. 1, January–March 2010, pp. 1–6.

Wise, Donald, and Blanca Cavazos, “Leadership Coaching for Principals: A National Study,” *Mentoring and Tutoring: Partnership in Learning*, Vol. 25, No. 2, 2017, pp. 223–245.

Wise, Donald, and Marc Hammack, “Leadership Coaching: Coaching Competencies and Best Practices,” *Journal of School Leadership*, Vol. 21, No. 3, May 2011, pp. 449–477.

WWC—*See* What Works Clearinghouse.

In this report, part of a series on professional development for school principals, the authors analyze the effects of a large-scale implementation of the National Institute for School Leadership’s (NISL’s) Executive Development Program (EDP) and paired coaching for middle school principals in three states. The EDP is a widely used principal professional development program that previously has been shown to have a positive influence on student achievement outcomes. For this study, NISL-certified coaches offered at least 60 hours of one-on-one coaching to principals.

The implementation of the EDP and coaching spanned three states, 332 schools, and 118 school districts. The study examined the implementation of the EDP and coaching, the perceptions of participants, and the impacts of the professional development. The authors considered both the impacts of the offer of and the full participation in the EDP and coaching on student academic outcomes and on school practices, as measured by principal and teacher surveys.

\$34.00

ISBN-10 1-9774-0512-6  
ISBN-13 978-1-9774-0512-8



[www.rand.org](http://www.rand.org)