Emerging Leaders Program: Impacts on Students, Teachers, and Leaders in Three Sites

December 2019

Prepared for:
Brenda Neuman-Sheldon
Executive Director, Evaluation
New Leaders
bneuman-sheldon@newleaders.org

Prepared by:
SRI International
Nicole L. Arshan
Katrina G. Laguarda
Jared Boyce
Andrea D. Beesley
Rebecca L. Goetz
Xin Wei
Elise Levin-Guracar

SRI Project Number: P23751

SRI Education™
A DIVISION OF SRI INTERNATIONAL

Suggested citation:

Arshan, N. L., Laguarda, K. G., Boyce, J., Beesley, A. D., Goetz, R. L., Wei, X., Levin-Guracar, E. (December 2019). *Emerging Leaders program: Impacts on students, teachers, and leaders in three sites.* Menlo Park, CA: SRI International.

This report was developed by SRI International with funds provided by New Leaders under a Supporting Effective Educator Development grant from the U.S. Department of Education (grant #U367D150015). However, these contents do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the U.S. Federal Government.

Contents

List of Exhibits

Acknowledgments

Carrying out a large randomized control trial is a significant undertaking, and many individuals made critical contributions to the study reported here. We are indebted to the Emerging Leaders, teachers, principals, and district staff members who agreed to participate in this independent evaluation and who responded to multiple data collection requests. In particular, we thank the district leaders who championed the study, offered essential advice and guidance, and provided data for the analysis of student outcomes: Kevin Barlow and Gary Scheib of the Arlington Independent School District; Dr. Matthew Weber, Jill Rhodes-Pruin, Hope DeLeon, Jacqueline Shapiro, and Theresa Urrabazzo of the San Antonio Independent School District; and Dr. Angela Whitelaw, Dr. Cynthia Alexander-Mitchell, Jeffrey Shive, and Amanda Jordan of Shelby County Schools. Donya Logan and Kay Lair served ably as the study's local site research coordinators.

New Leaders research staff Brenda Neuman-Sheldon and Marianna Valdez provided invaluable advice and support throughout the study, from the earliest stages of study design to final reporting. New Leaders national program staff members contributed to the development of a program logic model and fidelity measures, collaborated on the administration and scoring of the program's leadership assessments, and collected data on program implementation for the research team. They included Alison Johnson, Kaitlyn Mittan, and Jenna Zelenetz. Additional national staff who provided critical support for the study included Gabe Scheck. New Leaders regional program staff members deepened our understanding of program operations and served as liaisons for the study in participating districts. They included Joi Kidd-Stamps, Carol Johnson, Barbara Beaver, Marsha Hollis, Velia Soto, and Delores Brown.

This report is the culmination of nearly 4 years of evaluation research by a large team at SRI International. We would like to thank current and former SRI colleagues who contributed in myriad ways to the conception, design, and execution of this study in all of its phases: Naa Ammah-Tagoe, Ashley Campbell, Erin Harless, Harold Javitz, Hannah Kistler, Patrik Lundh, Mary McCracken, Veronica Madrigal, Mark Sieber, and Marjory Wechsler. We also appreciate the contributions of Katie Roberts, Mimi Campbell, and Lauren Cassidy to the editing and production of this report. We thank Ryoko Yamaguchi for her support as an evaluation technical assistance provider.

The authors assume full responsibility for any errors or omissions.

Executive Summary

In 2017–18, New Leaders partnered with SRI Education to undertake a randomized control trial of the Emerging Leaders program in three sites: Arlington Independent School District and San Antonio Independent School District in Texas and Shelby County Schools in Tennessee. The Emerging Leaders program was implemented largely as designed and had a positive, statistically significant impact on participants' data-driven instruction leadership knowledge. This large impact on leadership knowledge led to few measured impacts on the instructional practice of teachers on Emerging Leaders participants' instructional teams. The program had positive impacts on the math achievement of some subgroups of students. Impacts on overall math achievement were mediated by (i.e., operated through) program impacts on participants' leadership knowledge and by teachers' participation on instructional teams. The Emerging Leaders program had no measured impact on students' English language arts (ELA) achievement. Supplemental analyses suggest that these differences in student achievement impacts may have been driven by differences in how data-driven instruction was enacted by math- and ELA-focused instructional teams.

New Leaders launched the Emerging Leaders program in 2011 to develop the instructional leadership skills of teachers, instructional coaches, and assistant principals to support more robust school leadership pipelines in partner districts and to train teacher leaders to guide and support instructional improvement in their schools. New Leaders selects candidates for Emerging Leaders through a rigorous admissions process. During the1-year program, Emerging Leaders participants lead school-based instructional teams that they identify in consultation with principals, drawing on the program's training and job-embedded assignments to guide their work. Emerging Leaders participants coach instructional teams to set goals, engage in regular data-driven instructional (DDI) cycles, design corrective instruction to address student misconceptions and instructional gaps, and monitor student progress. The program is therefore designed to develop the leadership skills of Emerging Leaders participants, expand the use of DDI strategies by instructional team members, and, ultimately, increase the achievement of students taught by instructional team members (Valdez, Broin, & Carroll, 2015). By the end of the year, Emerging Leaders participants are expected to demonstrate proficiency on the program's target leadership concepts, signaling readiness to assume greater formal responsibility.

Study Design

In spring 2017, New Leaders recruited 112 Emerging Leaders candidates in the three study school districts. All candidates successfully completed New Leaders' admissions process and worked with their principal to identify an instructional team to lead during the program year (e.g., 4th grade ELA, Algebra I). The research team then randomly assigned these candidates to two groups: Group I (treatment) participated in the Emerging Leaders program in 2017–18, the randomized control trial (RCT) year, and Group II (control) delayed participation until 2018–19.

The study estimated program impact on Emerging Leaders participants' leadership skills, as measured by an assessment of DDI leadership knowledge. The research team also assessed impacts on the practice of teachers identified for instructional teams using a daily log with questions about instructional planning, use of student data, corrective instruction, and student efficacy strategies. We used state assessment data in math and ELA to measure impacts on the students of instructional team members in the subject aligned to the instructional team's identified focus. We report impact estimates either as differences in predicted percentages (for binary outcomes) or as a standardized Cohen's *g* (for continuous outcomes). We interpret statistical significance at the $p < .05$ level, although we also describe marginally significant results at the $p < .1$ level as "approaching but not meeting" significance.

In addition to these impact estimates, we examined the implementation of the Emerging Leaders program in the RCT districts, drawing on program data, interviews, and observations of team meetings. We also investigated the extent to which Group I Emerging Leaders participants continued to use the leadership skills and strategies they developed in the program in the year after the RCT via a survey of program alumni, supplemented by interviews.

Emerging Leaders Program Implementation

To support interpretation of the program's impacts on participants, teachers, and students, we studied Emerging Leaders program implementation in the RCT districts to learn whether the program was delivered with fidelity to the New Leaders model and where there were specific successes or challenges.

The Emerging Leaders program was largely implemented as designed.

The program met (and often exceeded) most of the thresholds New Leaders set for fidelity of implementation, including the recruitment of larger-than-average cohorts of qualified candidates. Nearly all Emerging Leaders participants (93%) completed more than 80 hours of training and other program requirements, including summer intensive training, seven learning cycle meetings, and four job-embedded assignments. Emerging Leaders participants also gave high ratings to both the relevance of this training and the quality of facilitation provided by local program directors.

Implementation fell short of the New Leaders design in some ways, however. Approximately 40% of participants did not receive the level of personalized one-on-one coaching the program called for, in part because local program directors spent more time than expected on individualized follow-up with some program participants to ensure that they completed required assignments. Further, although most Emerging Leaders participants successfully completed all program requirements, very few (13%) demonstrated leadership proficiency at the end of the program (defined as scores of "proficient" or higher on at least 3 of 7 target leadership concepts, as measured by New Leaders assessments). These proficiency levels were defined by New Leaders to signal participants' program success and readiness for greater leadership; proficiency level thresholds for program fidelity were set in line with prior cohorts' performance at program completion. New Leaders implemented a new scoring process in the RCT year, however, which led to nationwide drops in proficiency rates across Emerging Leaders sites relative to prior years. As a result, it is difficult to determine whether these low proficiency rates signal a significant decline in participants' success in the program compared with past cohorts, or simply reflect the change in scoring procedures.

Instructional team members described their DDI work as more structured, rigorous, and effective than the DDI work they had attempted in the past, although teams sometimes failed to move beyond initial steps in reviewing and analyzing student work.

Instructional teams engaged in regular DDI cycles and used Emerging Leaders program tools and protocols to frame their discussions. Although all the instructional teams we observed reviewed student data or samples of student work, the instructional teams varied in the extent to which they were able to develop actionable insights from the data. Some teams had difficulty following the protocol and completing a full analysis in the allotted time (often only 20–30 minutes) and as a result never engaged in discussion of student misconceptions or plans for corrective instruction. Other teams formulated hypotheses about student performance that were not grounded in the data (e.g., special education needs, recent bad weather). Similarly, in discussions about instruction, instructional team members often shared suggestions for new strategies without a strong rationale grounded in the data, rather than specifically focusing on diagnosing and addressing gaps in instruction that perpetuated student misconceptions.

When observed instructional teams did engage in planning for corrective instruction, math-focused teams identified a specific misconception or an academic skill that required additional instructional support; the observed ELA teams were unlikely to make this connection.

The research team saw evidence of corrective instruction planning in math-focused instructional teams that aligned to program expectations. As they reviewed assessment data, math teachers quickly

generalized from specific assessment items to categories of math problems and underlying math skills, pinpointing gaps in students' understanding by examining where their problem-solving process had broken down. Building on these insights, math teachers generated strong plans for corrective instruction, tightly aligned to the specific student misconceptions they identified in the data. In contrast, ELA teams we observed framed student misconceptions in terms that did not generalize beyond the specific assessment question or assignment under discussion (focusing, for example, on test-taking strategies). We caution that these observations were not representative of all instructional teams. Observations were neither sampled nor scheduled at random, and a minority of math- or ELA-focused teams engaged in corrective instruction planning while being observed. However, where observation teams did see planning for corrective instruction, the distinction between disciplines was clear and appeared to be facilitated by the math teams' greater insight into students' problem-solving processes.

Impacts on DDI Leadership Knowledge and Instructional Team Practices

We measured the program's impact on Emerging Leaders participants' DDI leadership knowledge using a written assessment of proficiency in DDI that the New Leaders staff developed as part of the larger suite of leadership skills assessments given to Emerging Leaders participants. To assess the program's impact on teachers—both those Emerging Leaders who were themselves classroom teachers and members of instructional teams—we designed a daily log to capture Emerging Leaders-aligned instructional planning behaviors, including reviewing assessment data and student work, planning for corrective instruction, and implementing strategies to promote student self-efficacy. The log also included a measure of teacher efficacy.

The Emerging Leaders program had a positive and statistically significant impact on all three facets of DDI leadership knowledge measured.

Emerging Leaders were asked to complete New Leaders' DDI assessment in spring of the RCT year; these assessments were scored blind to treatment status to ensure unbiased results. Group I Emerging Leaders participants (treatment) outscored Group II candidates (control) on all three facets of DDI leadership knowledge measured: using multiple forms of data to drive student achievement (g = .95, p < .01), leading a team through a DDI cycle (g = .90, p < .01), and understanding of efficacy concepts (g = .94, p < .01). There were no differences in impacts between Emerging Leaders assigned to math- and ELA-focused instructional teams.

These impact estimates demonstrate that although few program participants met New Leaders' standards for leadership proficiency (including on this DDI leadership knowledge assessment), the program did increase participants' DDI leadership knowledge relative to candidates who had not yet completed training.

The Emerging Leaders program had a positive impact on two corrective instruction practices among teachers in math-focused instructional teams. There were no notable impacts on teachers' planning practice when teachers were analyzed without respect to team focus.

Overall, teachers on instructional teams in Group I (treatment) and Group II (control) had generally comparable weekly instructional planning behaviors. However, Group I teachers on math-focused instructional teams displayed substantially larger, positive impacts on two key facets of corrective instruction. That is, they were more likely than Group II teachers to review assessment data looking for students' incorrect answers (52% Group I; 24% Group II, p < .001) and more likely to select new

instructional strategies for revisiting past content (84%; 65%, $p < .05$). The program did not have these effects on teachers in ELA-focused teams.

Impacts on Students

Finally, we examined the extent to which the Emerging Leaders program affected student learning, both overall and by student subgroup, and the extent to which these impacts were mediated by teachers' regular participation in instructional team work and by Emerging Leaders participants' DDI leadership knowledge.

The Emerging Leaders program's overall impact on math achievement was positive and approached (but did not meet) the threshold for statistical significance. There were positive, statistically significant impacts on the math achievement of female students, Latinx students, English language learners, and economically disadvantaged students.

Although the estimated impact of the Emerging Leaders program on overall student achievement in math, as measured by state assessments, was positive, it approached but did not meet the standard threshold for statistical significance ($g = .20$, $p < .1$) (Exhibit ES-1). However, the program had positive, statistically significant impacts on the math achievement of four student subgroups: female students ($g = 0.22$, $p < .05$), Latinx students ($g = 0.32$, $p < .05$), English learners ($g = 0.41$, $p < .01$), and economically disadvantaged students ($g = 0.26$, $p < .05$). For a student in the middle of the statewide math achievement distribution, these impacts translate into gains of 9 percentile points for female students (that is, moving from the 50th to the 59th percentile in the statewide math achievement distribution), 13 percentile points for Latinx students, 16 percentile points for English learners, and 9 percentile points for economically disadvantaged students. The program had no statistically significant impact on subgroups of students who were African American, White, in the highest or lowest performing quartiles, or who received special education services.

Program impacts on math achievement were mediated in ways that support Emerging Leaders' theory of action: program impacts operated both through students having teachers regularly attend instructional team meetings and through Emerging Leaders participants' DDI leadership knowledge.

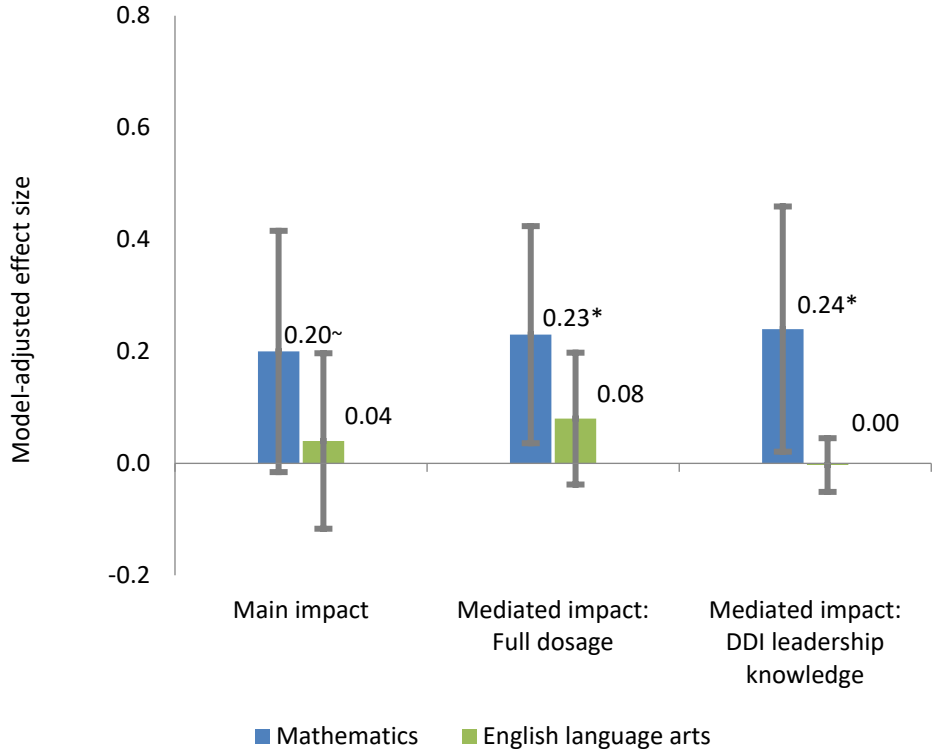
Students' experience of the full dosage of the Emerging Leaders program (meaning that their Emerging Leader remained in the district and successfully completed the program, their teacher attended most instructional team meetings, and they remained in their teacher's classroom for a full year) led to a statistically significant positive impact on math achievement ($g = 0.23$, $p < .05$) (Exhibit ES-1). An effect of this size is equivalent to 9 percentile points for students at the middle of the statewide math achievement distribution (that is, moving from the 50th to the 59th percentile). Similarly, Emerging Leaders participants' gains in DDI leadership knowledge produced corresponding gains in student math achievement ($g = 0.24$, $p < .05$), an effect also equivalent to 9 percentile points on the distribution of state achievement. These mediating impacts suggest that greater engagement with the Emerging Leaders program drove the potential program effects on math achievement.

The Emerging Leaders program had no measured effect on student ELA achievement.

The estimated effect of the program on students' state ELA assessments was close to zero ($g = .04$, $p > .05$), and there were no statistically significant program impacts in ELA for any student subgroups. Similarly, there was no measured impact for students who received the full dosage of the program ($g = .08$, $p > .05$), nor were program effects meditated by impacts on Emerging Leaders participants' DDI leadership knowledge ($g = .00$, $p > .05$) (Exhibit ES-1). Both the small estimated effect size and lack of

statistical significance on both overall and mediation effects suggests the program did not have an impact in ELA in this study.

Exhibit ES-1. Emerging Leaders Impacts on Student Achievement: Main Impacts and Impacts Mediated by Students' Experience of the Program and Participants' DDI Leadership Knowledge



~p < .1, *p < .05, **p < .01. Grey brackets represent the 95% confidence interval of the estimated effect size.
Source: State English language arts and math assessments, 2017–18.
Samples: For ELA main impact, 33 instructional teams and 5,446 students. For math main impact, 28 instructional teams and 6,317 students. For ELA mediated impact (full dosage), 30 instructional teams and 4,986 students. For math mediated impact (full dosage), 28 instructional teams and 5,875 students. For ELA mediated impact (DDI leadership knowledge), 23 instructional teams and 3,765 students. For math mediated impact (DDI leadership knowledge), 23 instructional teams and 4,805 students.

Sustaining Leadership Gains After the Emerging Leaders Year

We examined whether leadership gains made during the Emerging Leaders program year continued into the next year (2018–19), when Emerging Leaders were no longer directly supported by the program.

Emerging Leaders aspired to positions with greater leadership opportunities, and nearly half moved to new jobs or schools after completing the program.

By 2018–19, almost half the Emerging Leaders (47%) had moved to a different job, a different school or district, or both. The subset who had changed jobs (33%) had typically moved into roles with greater or different leadership responsibility, such as principal, assistant principal, instructional coach, district-level instructional support staff, or other school administrator. Emerging Leaders saw themselves as future

administrators, if not immediately as principals. Some Emerging Leaders who did not find new positions were disappointed and reported that they planned to look for new positions outside their districts. Emerging Leaders across the board experienced a decreased sense of self-efficacy in the year after their training, when they no longer had access to the scaffolding and support of the Emerging Leaders program.

In the year after the program, many Emerging Leaders faced challenges in finding opportunities for instructional leadership but continued to use program tools and strategies that were well aligned with their current responsibilities and the preexisting practice, structures, and policies of their schools and districts.

Broadly speaking, Emerging Leaders were using some of the strategies and tools they learned in the program that aligned well with their responsibilities and the preexisting practice, structures, and policies of their schools. Of all tools and strategies, Emerging Leaders that we interviewed were most likely to report engaging in data analysis for DDI and planning corrective instruction. Most assistant principals and coaches were also using Emerging Leaders tools and strategies for engaging in difficult conversations, describing this as one of the most valuable aspects of their training.

Emerging Leaders who were assistant principals found themselves with little time or opportunity for meaningful instructional leadership because of their management and discipline responsibilities. Of the Emerging Leaders surveyed who were teachers, half were continuing to lead an instructional team for instructional planning in the year after the program, a significant drop from the year before. Despite these challenges, Emerging Leaders believed that their experiences in the program were relevant to their work in the first post-program year, particularly those who had changed roles in 2018–19.

Conclusions and Implications

As New Leaders continues to improve the Emerging Leaders program and develop other training for instructional leadership, program staff may want to consider the following implications of this study's findings:

- Instructional teams engaged in DDI work benefit from trained leadership to guide them, protected time for teams to engage in this work, and strong tools to provide structure to their practice.
- Despite Emerging Leaders' similar effects on the DDI leadership knowledge of participants leading math- and ELA-focused teams, these leaders drove different changes in both their instructional teams' instructional planning practice and student achievement in the instructional teams' classrooms, suggesting that ELA-focused instructional teams may require different or additional supports to improve student achievement.
- The Emerging Leaders program may provide a model for scaffolded leadership development that begins before assuming the principalship and operates via structured, job-embedded assignments with a focus on individual elements of leadership. Given that the structure, schedule, and responsibilities of different professional roles (e.g., teachers, coaches, and assistant principals) provide different affordances for instructional leadership training and development, any such training should be aligned to the needs of professionals' current responsibilities.

Chapter 1: Introduction and Study Design

New Leaders launched the Emerging Leaders program in 2011 to improve the instructional leadership skills and capacities of teacher leaders, instructional coaches, and assistant principals. Districts that partner with New Leaders to offer the Emerging Leaders program typically have two closely related goals: (1) to develop a more robust principal pipeline by identifying and preparing candidates for assistant principal positions, principal residency programs (either the New Leaders Aspiring Principals program or a local program), or the principalship itself; and (2) to train a cadre of teacher leaders who can effectively support schools' instructional improvement efforts. In this 1-year program, Emerging Leaders work to develop instructional leadership skills, including skills related to corrective instruction through training embedded in the leadership of an instructional team in their school.

In 2016, the U.S. Department of Education awarded New Leaders a Supporting Effective Educator Development (SEED) grant to implement and rigorously evaluate the Emerging Leaders program. In 2017–18, the New Leaders staff leveraged these funds to undertake a rigorous evaluation of the impact of the Emerging Leaders program in three sites: Arlington Independent School District in Texas, San Antonio Independent School District in Texas, and Shelby County Schools in Tennessee. New Leaders' research partner, SRI Education (the research team), designed and carried out a randomized control trial (RCT) to assess the program's impact on participant knowledge and skills, teacher practice, and student achievement. As part of the RCT, the research team evaluated the implementation of the Emerging Leaders program in the RCT sites, including the extent to which the program was delivered with fidelity to New Leaders' design.

This report presents findings on the program's implementation and impact on Emerging Leaders participants, instructional team members, and students in classrooms led by instructional team members. It also describes participants' leadership roles and use of program tools and strategies in the year after their participation in the program, to investigate the extent to which they sustained the leadership practices they acquired during the Emerging Leaders program year.

Study Design

In preparation for the launch of the RCT during the 2017–18 school year, New Leaders recruited Emerging Leaders candidates through a rigorous admissions process in the three RCT sites. Each candidate identified an instructional team that he or she would lead as part of the program during the application process.[1] In July 2017, before the beginning of the 2017–18 program year, the research team randomized the Emerging Leaders candidates to Group I (participation in the Emerging Leaders program in 2017–18) or Group II (delayed participation until 2018–19). Emerging Leaders, together with their principals, identified instructional team focus (e.g., 3rd grade math) and teachers prior to randomization; the research team verified instructional team teacher and student rosters for early fall 2017. This RCT offered the opportunity to demonstrate Emerging Leaders participants' effectiveness with the strongest possible evidence because the two randomized groups should be similar on both observable (e.g., years of experience) and unobservable (e.g., motivation) characteristics.

[1] To facilitate the student impact portion of the study, participants were encouraged to lead a math or ELA team in a tested grade, although principals were given ultimate authority to assign teams according to their schools' needs.

Research Questions

To examine implementation of the Emerging Leaders program across the three sites, the evaluation addressed the following research questions:

- Was the Emerging Leaders program implemented with fidelity?
- To what extent and in what ways did implementation vary within and across sites? Under what conditions?
- To what extent did participants' leadership development under Emerging Leaders differ from the experiences of the control group?

In addition, the evaluation examined the impacts of program implementation on Emerging Leaders participants, instructional team members, and the students taught by these instructional team members. The impact research questions were the following:

- What was the impact of the Emerging Leaders program on participants' leadership development, as measured by New Leaders' assessment of DDI (data-driven instruction) leadership knowledge?
- What was the impact of the Emerging Leaders program on instructional teams' instructional planning practices, as measured by a daily instructional planning log?
- What was the impact of the Emerging Leaders program on the math and ELA (English language arts) achievement of students taught by instructional team members, as measured by state assessment data?

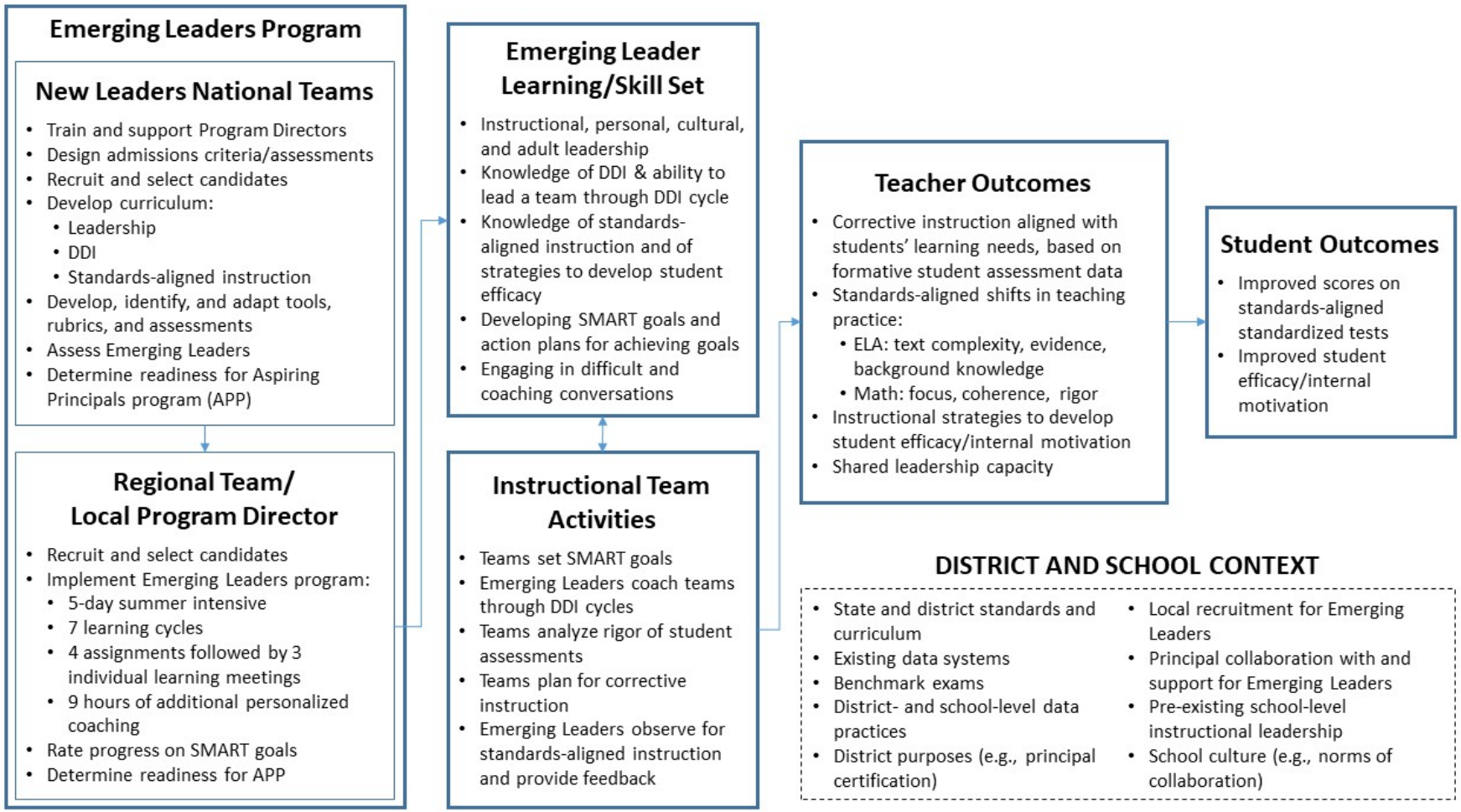
Last, the evaluation explored the extent to which the leadership gains made during the Emerging Leaders year (2017–18) continued into the next year (2018–19), when Emerging Leaders participants were no longer directly supported by the program. Relatedly, we explored whether the program met the participants' and the participating districts' expectations with respect to contributing to professional advancement and leadership pipelines. The sustainability research questions were the following:

- To what extent have Emerging Leaders alumni taken on new leadership roles in 2018–19, both formal and informal? In what ways has the Emerging Leaders program prepared them for these new roles?
- In what ways are alumni currently using the skills, strategies, practices, and tools that they acquired in the Emerging Leaders program?
- What are the varied ways in which the Emerging Leaders program has met sites' expectations and/or contributed to the development of a robust school leadership pipeline in each site?

Emerging Leaders Program Logic Model

The Emerging Leaders program logic model (Exhibit 1) offers a visual representation of the assumptions underlying the design of the yearlong program. The research team developed this logic model from a systematic review of program artifacts and in consultation with New Leaders program staff. The logic model presents the New Leaders theory of action about how program inputs (e.g., participant recruitment, selection, training, coaching, tools, assessments) support the development of participants' leadership skills and the work of instructional teams, which in turn lead to changes in teachers' practice and student achievement. We designed the evaluation to rigorously test this logic model and its underlying assumptions by choosing or, as necessary, developing, measures to assess whether the program was implemented as designed and had the intended impacts on participants, instructional team members, and students.

Exhibit 1. Emerging Leaders Program Logic Model



The boxes titled “Emerging Leaders Program” in the leftmost column of the logic model show the key components of the program that have been designed by national program staff and delivered in each of the RCT sites by regional program staff. At the national level, program components include training and support for local program directors in each RCT site; rigorous admissions criteria and baseline assessments of DDI leadership knowledge and skills developed by a national assessment team;[2] and program curriculum, readings, tools, assessments, and rubrics. At the local level, Emerging Leaders program directors facilitate a 5-day summer intensive course and seven monthly learning cycles[3] during the academic year. Local program directors also provide one-on-one coaching and feedback on program assignments that require submission of written artifacts and videotaped recordings of instructional team meetings. Program directors coach participants to lead their teams in setting SMART goals[4] for student achievement. For the program’s final assignment, participants review their teams’ progress against these SMART goals in a presentation to New Leaders staff, other Emerging Leaders participants, and local school and district leaders. Together, these program components ground participants in the theory and practice of DDI and prepare them to coach, motivate, and give feedback to the teachers on their instructional teams. In Chapter 2, we describe the extent to which the Emerging Leaders program was implemented with fidelity to this design in the three study sites. We also briefly describe the differences between the training provided to Emerging Leaders participants and the training they were likely to receive absent receipt of Emerging Leaders training.

Emerging Leader participants who complete the program as designed are expected to develop the leadership skills represented in the box in the second column of the logic model titled “Emerging Leader Learning/Skill Set.” They include seven target leadership competencies under four leadership domains (personal, instructional, cultural, and adult/team) assessed via assignments that are administered and scored by a nationally normed Assessor Corps. They include the ability to lead a team through the DDI cycle, application of standards-aligned instructional knowledge, skills for coaching other teachers, and habits of reflective practice and continuous improvement. The development of these leadership competencies in participants is the first, most direct effect of the Emerging Leaders program. In Chapter 3, we provide estimates of the program effects on participants’ instructional leadership knowledge.

During the program year, New Leaders expects that Emerging Leaders participants will practice and hone new leadership skills in the context of work with instructional teams, represented in the box titled “Instructional Team Activities.” Participants lead a team of two to five teachers, chosen in consultation with their principals, to set SMART goals, assess the rigor of assessments, engage in DDI cycles that deepen their understanding of student learning, specifically student misconceptions and gaps in skills, and plan corrective instruction to address those misconceptions. Teams share ideas for corrective instruction (that is, planning for the use of new instructional strategies) as well as strategies for developing student self-efficacy and ownership of their learning. Although instructional team work is a required component of the program and Emerging Leaders participants receive assignments that they cannot complete without the participation of an instructional team, New Leaders staff have limited influence over whether or when instructional teams meet and how they use their time together (although New Leaders does require principals to commit to supporting participants work with instructional teams). Thus, the quality and quantity of instructional teamwork depends almost entirely on the initiative and skill of the Emerging Leader and on the support provided to the team by the building principal. As part of our

[2] A national assessment team leads the admissions assessment process in all Emerging Leaders sites and evaluates all application materials. In some local sites, program directors or district administrators have discretion to admit candidates who do not meet all admissions criteria.
[3] Each learning cycle includes a web-based tutorial, an in-person meeting, personalized coaching and work on a job-embedded assignment.
[4] Emerging Leaders SMART (Specific, Measurable, Ambitious, Results-oriented, Time-bound) goals set specific targets for achievement and growth on state or local benchmark assessments by the end of the program year.

description of the Emerging Leaders program implementation in Chapter 2, we describe the instructional teams' work, along with barriers and facilitators to this work.

New Leaders' theory of action holds that the work of instructional teams will support teachers in planning instruction that is more responsive to individual student needs, based on formative assessment data, better aligned to state standards, and more effective in building student efficacy (the "Teacher Outcomes" box). Although New Leaders expects that these shifts will take place during the program year, they are more distal outcomes compared with the development of participants' own leadership skills, because the program reaches teachers only through program participants (with the exception of Emerging Leaders who are themselves teachers and are expected to make changes to their own instructional practice at the same time that they are coaching others to do the same). We provide estimates of the program's impact on instructional planning in Chapter 3.

These teacher outcomes are, in turn, expected to lead to improved student achievement on standardized standards-aligned assessments (the "Student Outcomes" box). Prior impact studies of the New Leaders Aspiring Principals Program (APP) have found that it can take as long as 3 years to realize an impact on students from placing a newly trained principal in a school building (Gates et al, 2019). However, because New Leaders views the Emerging Leaders program as more directly targeted at teachers and their instruction, its theory of action holds that student achievement will improve even during the training year (Valdez, Broin, & Carroll, 2015). We provide estimates of program impact on student achievement in Chapter 4. In Chapter 5, we describe the sustainability of Emerging Leaders training following the training year.

Data Sources

The data sources for the evaluation of program implementation, leadership outcomes, and instructional outcomes were as follows.

- **Extant program data.** New Leaders collected an array of diverse data on program participants in the delivery of the program, and we drew on all these to assess program fidelity. These data included the Emerging Leaders national community of practice sign-in sheets, baseline assessments used for admissions decisions, New Leaders admissions ratings, program director coaching logs, end-of-learning-cycle surveys, learning cycle attendance logs, assignment results, the New Leaders end-of-year participant survey, and the New Leaders assessment scores.
- **Interviews and observations of instructional team meetings.** Approximately 7 months into the program year, we interviewed a sample of Emerging Leaders participants, instructional team members, principals, and district stakeholders in each of the RCT sites. Interviews addressed the implementation of the Emerging Leaders program, the work of the instructional teams and alignment with the program logic model, and respondents' perceptions of the program.
- **End-of-year Emerging Leader survey, Groups I and II.** The research team administered a survey to compare leadership development and engagement with instructional teams during the 2017–18 school year in both treatment and control conditions. Group I Emerging Leaders participants also identified teachers who participated regularly on Emerging Leaders-led instructional teams to support exploratory analyses of impacts on teachers who actively participated in the program.
- **New Leaders' DDI assessment.** New Leaders administered an assessment of DDI to all Emerging Leaders applicants before program enrollment. Participants worked independently using online materials and submitted their responses to be scored centrally by New Leaders staff. For this study, the instructional assessment was administered again to both Group I and Group II Emerging Leaders at the end of the school year to measure changes in understanding and use of DDI.

- **Daily instructional logs.** To measure changes in instructional practice, we developed a tool to capture teachers' daily planning and teaching. The instructional log was administered to both Group I and Group II Emerging Leaders instructional teams (inclusive of the Emerging Leader, if they lead a classroom) daily for 1 week at baseline and for two separate 1-week windows in the spring to measure changes in the ways teachers planned and delivered instruction.
- **Student achievement data.** To measure changes in student achievement, we collected extant student achievement data in the form of state assessments and district benchmark assessments, where available. The state assessments, the Texas STAAR and the Tennessee TNReady, were collected from all three participating sites. The district benchmark assessment, the NWEA MAP (Measures of Academic Progress), was collected from two participating sites; the third did not use it.

The remaining chapter sections describe the samples of districts, Emerging Leaders, instructional team members, and students recruited into this study to provide context for the analyses that follow. Formal tests of baseline equivalence are specific to the analytic sample of each outcome (i.e., the participants, instructional team members, and students who remained in the sample at the study's conclusion) and vary by analyses; these are provided as supporting documentation in the Technical Appendix.

Recruitment and Randomization

In spring 2017, Emerging Leaders staff in each of the three districts recruited larger than average cohorts of program candidates to facilitate the randomization process. Emerging Leaders candidates in each of these districts were required to go through the same application and selection process as candidates in non-RCT sites. The Emerging Leaders program is designed and staffed regionally to serve up to 35 program participants per cohort in each local site. Therefore, a double cohort of Emerging Leaders candidates recruited for the study (so that half could be assigned to treatment and half to control) could have enrolled up to 210 participants (70 per site; 105 in treatment total). Local staff were able to recruit larger than average cohorts in each district, though they did not meet recruitment goals, ultimately enrolling 112 candidates which is the same number of participants that would have been enrolled in a normal program year.

One substantial barrier to meeting recruitment goals was New Leaders' history in these districts: as described below, the Emerging Leaders program had been offered in two of the RCT sites for many years. Many of the assistant principals, instructional coaches, and teacher leaders in these sites who would have been good candidates for the program had already completed it, narrowing the pool of willing and eligible candidates.

By early fall 2017, six Emerging Leaders (two from Group I and four from Group II) withdrew from the study because they left the district or took another position within the district that would have prevented them from completing the program. The actual numbers enrolled in the program were about the same as in a normal program year but about half of the numbers planned for the RCT, with 106 candidates at baseline (56 in treatment). As a result, study sample sizes and the study's power to detect effects were both reduced.

Before randomization, the research team and the New Leaders staff began with an outreach, education, and data collection process, holding webinars and contacting all Emerging Leaders candidates and their principals. Through this process, the research team ensured that the Emerging Leaders candidates understood the research and randomization procedure and collected names of instructional team members who had already been identified, as well as plans for instructional teams in which team members had not yet been identified, including focal grades and subjects.

The research team then split the sample of Emerging Leaders into randomization blocks by district, focal subject, focal grade, and, in selected cases, school improvement status and feeder pattern. We randomized half of each block into treatment (Group I) and half into control (Group II, who received no Emerging Leaders training in 2017–18 but instead received it in 2018–19).

The research team communicated these randomization decisions in summer 2017, enabling local New Leaders staff to begin training Group I before the start of the 2017–18 school year. In fall 2017, we confirmed the membership of each instructional team, including the names of instructional team members identified after randomization, collected baseline instructional log data from both groups, and documented shifts in the instructional team sample between summer and fall 2017.

No new Emerging Leaders participants were included in the study sample after randomization.

Early Joiner Sample for Instructional Team Members and Students

Between randomization and instructional team verification in fall 2017, 32% of the instructional team members left the initial sample; Group I lost more instructional team members and added more new instructional team members than Group II (Exhibit A-2 in the appendix).[5] Much of the turnover on Group I instructional teams was due to changes in program participants' job assignments between randomization in early summer and the beginning of the school year. As Emerging Leaders participants changed schools, changed roles, or adopted schedules that did not accommodate the originally intended teams, they recruited new teachers to participate on the instructional teams that they would lead as part of the program. In all analyses, we use the sample of instructional team members as defined at the time of baseline data collection in fall 2017. In What Works Clearinghouse (WWC) terminology, such a sample is said to include "early joiners," study subjects who entered the sample after randomization but early in the cycle of program implementation.

To identify students enrolled in an instructional team members' classroom, we collected roster data from each district for the teachers in the early joiner sample. We requested districts provide us with student enrollment in teachers' classrooms at the date schools reported formal enrollment to the state. In the two Texas districts, we used the date of the state accountability count (the last Friday in October 2017). In Shelby County, we used the "snapshot" date (the 40th day of the 2017–18 school year). Students were included in the assigned sample if they were in a tested grade and subject (Chapter 4 has more detail on tested grades and subjects). Students were analyzed according to their assigned condition as of the formal enrollment date (an intent-to-treat framework).

Assigned Sample Sizes for Attrition Calculations

In keeping with WWC standards, the research team provides cluster-level and individual attrition for each study outcome in later chapters of this report; attrition varied by outcome based on response rates to different data collection instruments. We identify the assigned sample in all attrition estimates as including 112 Emerging Leaders (the number at the time of randomization in spring 2017) and 350 instructional team members (the number at the time of instructional team verification in early fall 2017).

[5] Group II Emerging Leaders were not asked to meet with their instructional teams during the RCT year; they were instructed to continue "business as usual.". However, the research team needed to confirm and update instructional team rosters to support data collection from teachers in both treatment and control groups for the RCT.

Samples

As delineated in the program's logic model (Exhibit 1), the Emerging Leaders program works by developing participants' DDI leadership knowledge and skills to change teachers' instructional practice, thereby driving impacts on student achievement. Reflecting this theory of action, our study design depends on nested samples of students within teachers within instructional teams, which are in turn led by Emerging Leaders program participants. In this section, we describe the samples of districts, Emerging Leaders, and students, including sample eligibility and characteristics. We provide overall numbers for parsimony, as few differences existed between Group I and Group II. All descriptive statistics provided in this chapter are also presented by group membership in the Technical Appendix.

Districts

New Leaders recruited three districts to participate in the RCT—Arlington Independent School District (AISD), San Antonio Independent school District (SAISD), and Shelby County Schools (SCS). Although all three districts serve high populations of low-income students of color (Exhibit 2), several other important contextual factors vary across the districts.

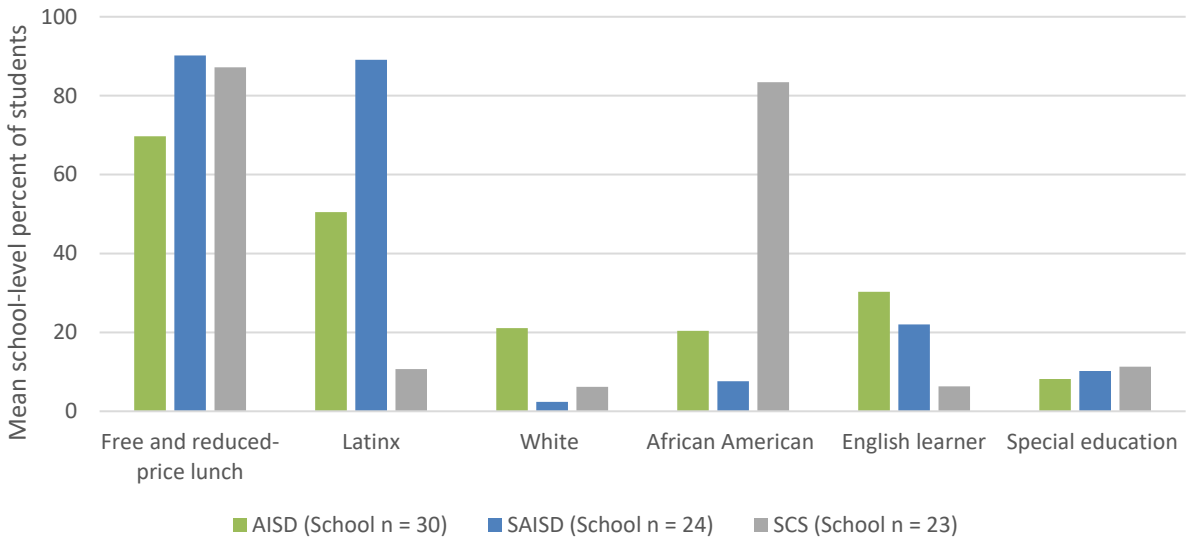
SCS is the largest school district in Tennessee, formed by the 2013 merger of Memphis City Schools and the surrounding countywide system, Shelby County Schools. SCS has a long history with New Leaders, dating back to 2005–06 when the first Memphis City Schools candidates enrolled in New Leaders' national Aspiring Principals program. In 2012–13, Memphis City Schools partnered with New Leaders to offer Emerging Leaders as part of its local leadership development pipeline, a partnership that has continued through the start of the study. The RCT year, 2017–18, marked the sixth cohort of participants to enroll in Emerging Leaders. SCS actively partners with a number of national organizations, and Emerging Leaders was one of many instructional leadership development programs operating in the district during the RCT year. SCS is majority African American and serves relatively few Latinx students, white students, or English learners, compared with the Texas districts in the study. Tennessee adopted a revised set of academic standards in 2017–18, known as the Tennessee Academic Standards for math and ELA. Because these new standards were based on the Common Core State Standards (CCSS), albeit with some significant revisions, regional program staff were able to implement the Emerging Leaders program without modifications.

AISD, in the Dallas suburbs, has partnered with New Leaders since 2012–13, offering the Emerging Leaders program as part of its principal pipeline (all Emerging Leaders candidates must have their administrative credential to be eligible for the program in this district). The RCT year marked the fifth cohort of Emerging Leaders candidates to enroll in the program. Approximately half of the district's students are Latinx, a fifth are white, a fifth are African American, and 5% are Asian. About a third of AISD students are English learners. Texas has not adopted the CCSS), meaning that local Emerging Leaders program directors adapted some program materials to better align them with Texas Essential Knowledge and Skills (TEKS) standards.

SAISD is a relatively compact urban district surrounded by several close-in suburban districts serving similar student populations. Because of their geographic proximity, SAISD competes with these other districts in the local labor market for principals, assistant principals, and teachers; and for this reason, the district's talent development efforts are particularly focused on retaining highly effective educators by offering multiple avenues for career advancement within the district. SAISD offered the Aspiring Principals program in 2016–17; the RCT year was SAISD's second year of partnership with New Leaders but its first experience with the Emerging Leaders program. District leaders expected that the Emerging Leaders program would help them to identify and prepare assistant principals and school implementation

specialists to enter SAISD's principal residency program. At the same time, they expected that Emerging Leaders would also prepare master teachers and other highly effective classroom teachers to become coaches or school implementation specialists. SAISD's student population is almost entirely Latinx, and a fifth of its students are English learners. Like AISD and the rest of Texas, SAISD uses TEKS.

Exhibit 2. School-Level Demographic Characteristics, Fall 2017 Baseline Sample



Source: 2016–17 school enrollment and demographic data accessed via state websites.

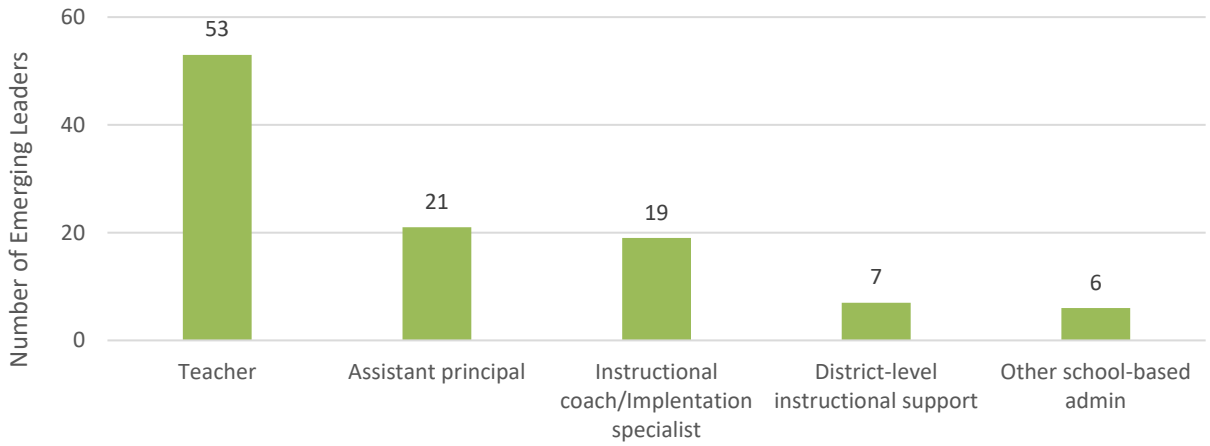
Emerging Leaders Participants

Of the 106 Emerging Leaders in the program at the beginning of the 2017–18 school year, 80% were women and 72% were people of color.[6] The average Emerging Leader participant had 17 years of experience. Exactly half of Emerging Leaders participants (53 of 106) were themselves classroom teachers (Exhibit 3). Of the remainder, 21 were assistant principals, 19 were school- or district-based instructional coaches or implementation specialists,[7] 7 provided district-level instructional support (e.g., content area advisors or curriculum specialists), and 6 were other school-based administrators (testing coordinators and other program coodinators). All of the Emerging Leaders participants interviewed for this study were teachers, assistant principals, or instructional coaches/implementation specialists during the RCT year, and findings based on interview data reflect the experience of participants in those roles.

[6] More descriptive statistics are in the Technical Appendix, with numbers disaggregated by Group I and II; where the two groups are similar, the overall descriptive numbers are provided for parsimony. The sample sizes in these report exhibits vary somewhat because of a small number of missing cases or Emerging Leaders participants who declined to allow their data to be used for research purposes.

[7] Implementation specialists are similar to instructional coaches in SAISD; they support teachers in implementing district curriculum and instructional strategies.

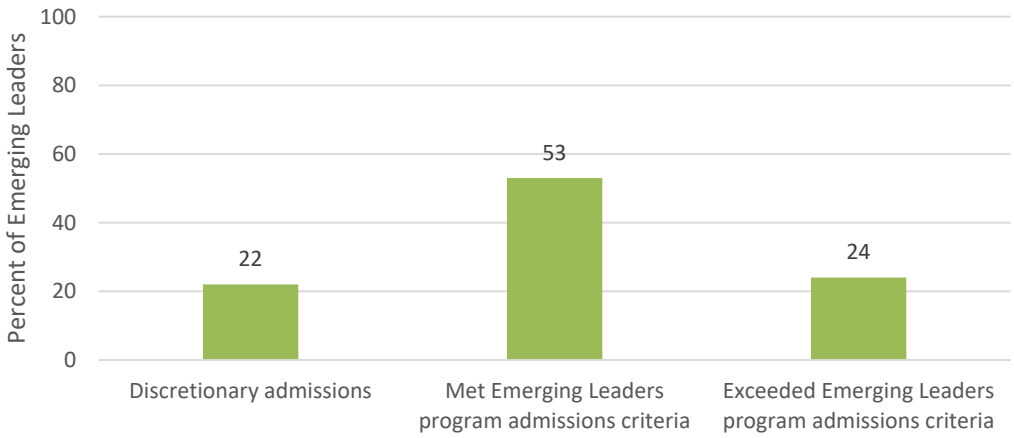
Exhibit 3. Emerging Leader Job Assignments, Fall 2017 Baseline Sample



n = 106
Source: Emerging Leaders program enrollment forms and fall 2017 instructional team verification.
Note: The "Instructional coach/Implementation specialist" category included school-based coaches but also PAR (peer assistance and review) coaches, instructional coaches who worked across schools, and special education coaches.

New Leaders admissions staff had the option to accept some candidates for the program who approached but did not meet New Leaders admissions criteria ("discretionary admissions" in Exhibit 4). In total, 22% of Emerging Leaders candidates entered the program in this category. The proportion of discretionary admissions in the RCT sites was consistent with the proportion of discretionary admissions in Emerging Leaders sites nationwide (18% across 16 sites), suggesting that New Leaders did not alter its admissions criteria to accommodate the recruitment demands of the RCT. Candidates in each admissions category were evenly distributed across Groups I and II.

Exhibit 4. Emerging Leaders Admissions Criteria, Fall 2017 Baseline Sample

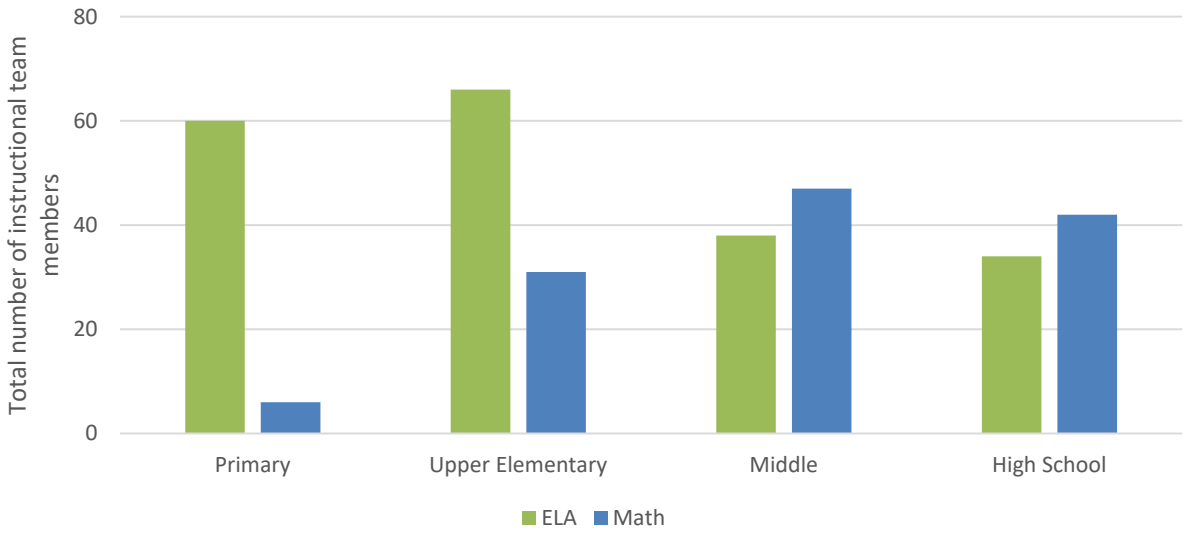


n = 106
Source: New Leaders application data.
Note: Emerging Leaders candidates met program admissions criteria if they scored at least a 2 ("approaching proficient") on 4 of 6 leadership concepts, and as low as 1.5 on no more than 2. Emerging Leaders candidates exceeded program admissions criteria if they scored at least a 2 ("approaching proficient") on 6 of 6 leadership concepts.

Instructional Team Members

Exhibit 5 presents the number of Emerging Leaders participants and instructional team members in the baseline study sample by subject and grade level. Local New Leaders staff's requests that program participants and their principals designate an ELA or math focus for their instructional team were largely effective—only 26 of 350 instructional team members taught a different subject. In math, more than two-thirds of the study sample taught in secondary schools (middle and high schools), while in ELA, a large majority taught in elementary schools.

Exhibit 5. Instructional Team Members by Grade Level and Subject Area, Fall 2017 Baseline Sample



n = 324
Source: New Leaders application data and fall 2017 instructional team verification.
Note: Does not include 26 teachers of subjects other than ELA and math.

Report Overview

The remainder of this report is divided into five chapters. Chapter 2 presents findings on the implementation of the program in the RCT sites, including the leadership development experiences of Group II Emerging Leaders (the control group) during the RCT year. Chapter 3 presents findings on the impact of the program on leadership outcomes and teacher practice. Chapter 4 presents findings on impacts on student achievement. Chapter 5 presents findings on sustaining the leadership practices and supporting sites' leadership pipeline. The report concludes with a discussion of implications.

Chapter 2: Emerging Leaders Program Implementation

- The program met (and often exceeded) most of the thresholds New Leaders set for fidelity of implementation, including the recruitment of larger-than-average cohorts of qualified candidates and these candidates' attendance at training and completion of program assignments.
- The program did not meet New Leaders' stated goals for delivery of personalized coaching and for participants' demonstration of leadership proficiency at the end of the program. The latter was likely due to changes in scoring procedures implemented by New Leaders during the RCT year.
- Instructional coaches were often best positioned to meet program expectations related to team leadership because Emerging Leaders program requirements aligned well with their existing responsibilities.
- Instructional team members described their DDI work as more structured, rigorous, and effective than the DDI work they had attempted in the past, though teams sometimes failed to move beyond initial steps in reviewing and analyzing student work.
- When observed instructional teams did engage in planning for corrective instruction, math-focused teams identified a specific misconception or an academic skill that required additional instructional support; the observed ELA teams did not make this connection.

To test the Emerging Leaders program theory of action, we begin with a careful examination of program implementation. Examination of the implementation allows for a well-contextualized interpretation of estimated program impacts and information to guide future program investments and implementation. Defining and measuring implementation fidelity provide insight into whether impacts stemmed from the program (as designed by New Leaders and represented in the program logic model or from an intervention) that, as it unfolded in the program sites, deviated significantly from New Leaders' original vision. Measures of fidelity also help to define those elements of the intervention that must be replicated to achieve the same impacts with future Emerging Leaders cohorts. They can also help program staff assess and revise the assumptions that originally informed the program's theory of action.

The fidelity measures developed for this evaluation focused on those aspects of the program over which New Leaders had the greatest control, namely New Leaders' delivery of various elements of the program (the boxes in the leftmost column of the logic model; Exhibit 1) and the performance of Emerging Leaders participants (the boxes in the second column of the logic model). The program's theory of action makes a series of assumptions about the impact of the Emerging Leaders program on instructional teams, about the impact of instructional teams' DDI work on teacher practice, and about the impact of those changes in teacher practice on student achievement. As New Leaders have no direct touch points with instructional team members, the data previously available to understand the regularity and quality of their engagement had previously been limited. To supplement New Leaders' program fidelity data, the research team visited each RCT site to conduct interviews and observe instructional team meetings. These interviews focused on better understanding the reasons for variations in program fidelity across sites and across instructional teams and on describing teams' work and the extent to which teachers and others perceived the Emerging Leaders program work as a departure from their previous practice.

Finally, to better understand what would have happened to Group I participants in the absence of the Emerging Leaders program, the research team collected data on the experiences of Group II Emerging Leaders during the RCT year. We sought to determine whether the control group had participated in other leadership development programs in 2017–18, whether those leadership development programs resembled the Emerging Leaders program, and whether Group II Emerging Leaders had engaged with their designated instructional teams despite requests that they refrain from doing so as members of the control group. Group II Emerging Leaders who received training very similar to Group I's may have had similar outcomes, diluting the impacts attributed to the Emerging Leaders program. Thus, understanding to what degree the experience of the control group under status quo conditions may have been similar to the experience of the treatment group provides context for understanding the practical significance of the program's impacts.

Data Sources on Emerging Leaders Program Implementation

- New Leaders 2017–18 program data, including Emerging Leaders national community of practice sign-in sheets, baseline assessments used for admissions decisions, New Leaders admissions ratings, program director coaching logs, end-of-learning-cycle surveys, learning cycle attendance logs, assignment results, New Leaders end-of-year participant survey, and New Leaders assessment scores
- 60 interviews with Emerging Leadership participants, instructional team members, principals, district leaders, and program staff members in the three RCT sites
- Observations of 10 instructional team meetings
- End-of-year survey of Emerging Leaders participants in Group I (treatment) and Group II (control) administered in May and June 2018. All Emerging Leaders participants still employed in the study sites in spring 2018 received the survey and 87 completed it, a response rate of 78% (84% treatment, 70% control).

Implementation Fidelity

In consultation with program staff, we identified six key components of the Emerging Leaders program, each one aligned with a program input or activity as shown in the Emerging Leaders program logic model (Exhibit 1).

- Training and support of local program directors
- Recruitment and selection of Emerging Leaders candidates
- Regional implementation of training (e.g., relevance, quality, delivery of personalized coaching)
- Emerging Leaders participants' completion of training
- Emerging Leaders participants' demonstration of leadership proficiency
- Emerging Leaders participants' work with instructional teams (e.g., leading meetings and coaching team members)

For each of these key components, we developed one or more indicators to be measured using program data collected by New Leaders (Appendix Exhibits A-10 and A-11 provide more detail on data sources, individual thresholds, and site-level thresholds for fidelity of implementation). Exhibit 6 shows the percentage of program participants meeting each fidelity indicator and whether the sites met the threshold for fidelity of implementation during the RCT year. Taken together, these results summarize program implementation and the extent to which it was consistent with the assumptions underlying the New Leaders' theory of action for the program.

Exhibit 6. Emerging Leaders Program Fidelity of Implementation, 2017–18

Definition	Individual-level threshold	% Meeting Threshold	Meets Fidelity?
COMMUNITY OF PRACTICE Program directors attend and facilitate trainings	Attendance at 14 community of practice meetings or national trainings during the program year	100	✓
Key Component 2: Recruitment & Selection			
HIGHLY QUALIFIED CANDIDATES New Leaders recruit Emerging Leader candidates who meet rigorous selection criteria	Candidates demonstrate "approaching proficient" on assessments scored with the Emerging Leaders rubric, earning at least a 2.0 on the majority of concepts and as low as a 1.5 on no more than 2 concepts	78	✓
Key Component 3: Regional Implementation of Training			
RELEVANCE OF PROGRAMMING Emerging Leaders participants find trainings useful	An average rating of "agree" or above across a bank of questions on perception of programming relevance	92	✓
QUALITY OF FACILITATION Emerging Leaders participants find the quality of facilitation high	A rating of "agree" or above on question on perception of facilitation quality	95	✓
ONE-ON-ONE COACHING Emerging Leaders participants receive personalized coaching	At least 7 hours[8] of personalized coaching, not including time spent in learning meetings	57	X
Key Component 4: Emerging Leaders Completion of Training			
LEARNING CYCLES Participants attend all induction, intensive, & learning cycle sessions	Emerging Leader participants attend (or make up) induction, intensive, and all seven learning cycle sessions	93	✓
JOB-EMBEDDED ASSIGNMENTS Participants complete assignments and attend associated learning meetings	Emerging Leader participants complete all four assignments and attend all three learning meetings, as indicated by having a valid score for each assignment and learning meeting	100	✓
Key Component 5: Emerging Leaders Demonstration of Leadership Proficiency			
LEADERSHIP SKILLS Participants demonstrate leadership proficiency	Scores of "proficient" (3.0 on a scale of 1.0–4.0) or higher on at least 3 of 7 target concepts and at least 2.0 on the remaining target concepts, as assessed by New Leaders	13	X
READINESS FOR ASPIRING PRINCIPALS Participants demonstrate readiness for the Aspiring Principals Program	Scores of "proficient" (3.0 on a scale of 1.0–4.0) or higher on at least five of seven target concepts and growth on the remaining 2 concepts, as assessed by New Leaders	2	X
Key Component 6: Instructional Team Work			
TEAM MEETINGS Participants led instructional team meetings regularly	Emerging Leader participants lead a team meeting at least 10 times during year	95	✓
OBSERVATIONS Participants observe team members and give feedback	Emerging Leader participants observe and provide feedback to team members at least 8 times	68[9]	--

[8] The program logic model assumes 9 hours of personalized coaching in addition to 3 hours spent in learning meetings. The fidelity threshold is set at 7 hours to account for incomplete record-keeping.
[9] Because the program was close to meeting the threshold for this indicator as measured by the end-of-cycle surveys, and because of the measurement error associated with this data source, we cannot say with confidence whether the program met (or failed to meet) the fidelity threshold.

The program met (and often exceeded) most of the thresholds New Leaders set for fidelity of implementation, including the recruitment of larger-than-average cohorts of qualified candidates and these candidates' attendance at training and completion of program assignments.

On components aligned with the leftmost column of the program logic model (New Leaders' development of local program directors, recruitment and selection of highly qualified candidates, and delivery of Emerging Leaders training), the program met and often exceeded the thresholds New Leaders set for fidelity of implementation. Perhaps most significant, the New Leaders national admissions staff succeeded in recruiting larger-than-average cohorts of qualified program candidates to satisfy the sample requirements of the RCT while maintaining minimum standards for the number of candidates who fell short on some admissions criteria: 78% of Emerging Leaders candidates across the three sites were "highly qualified," meeting all criteria for admission.

Attrition from the program was low; of the 58 Group I Emerging Leaders who enrolled in the program in summer 2017, five dropped out because they moved to other districts or moved into new positions with time demands that precluded their participation in the Emerging Leaders program. Nearly all of the remaining 53 Emerging Leaders participants (93%) completed more than 80 hours of training and other program requirements, including summer intensive training, seven learning cycle meetings, and four job-embedded assignments. In addition, almost all Emerging Leaders participants gave high ratings to both the relevance of the training regional offices provided and the quality of facilitation. Emerging Leaders participants led instructional team meetings as required by the program, and the majority observed members of their instructional teams and provided feedback as often as expected.

The program fell short of benchmarks for delivery of personalized one-on-one coaching, in part because local program directors spent more time than expected on individualized follow-up with a relatively small number of Emerging Leaders participants who needed extra support to complete the program.

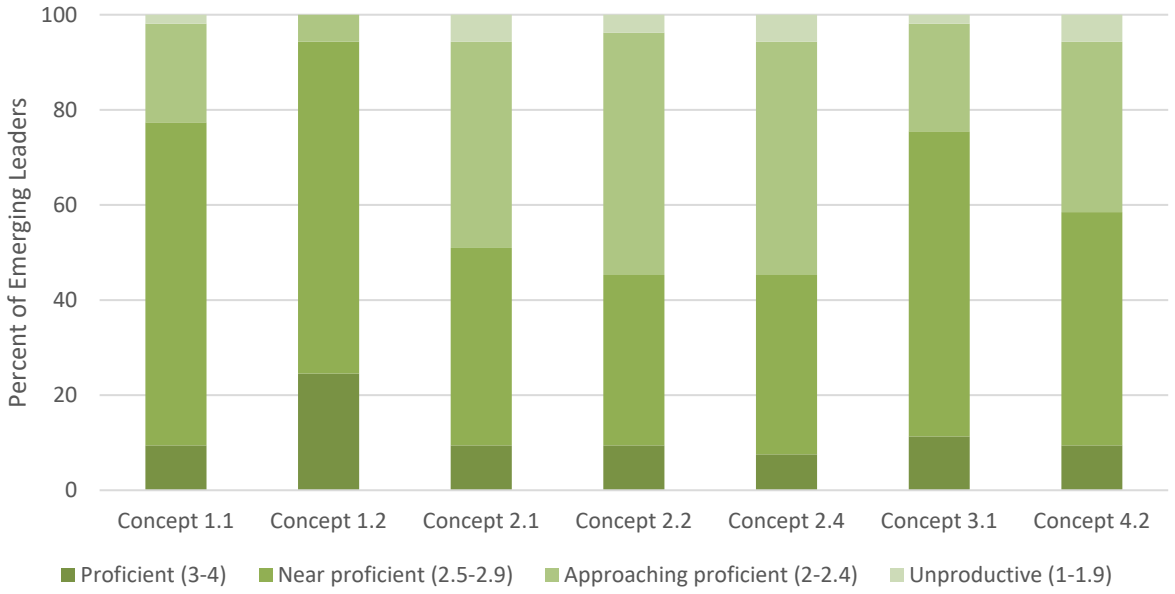
Approximately 40% of participants did not receive the level of personalized one-on-one coaching the program called for (appendix Exhibit A-12). The overall percentage masks significant variation by site (appendix Exhibit A-10). Nearly all Emerging Leaders participants received the expected level of coaching at one site, compared with 59% in a second site and just 6% in the third. In interviews, all three regional program directors explained that they focused their coaching efforts on program participants who needed the most support. In particular, program directors spent considerable time following up with participants who had not completed program assignments or who were in danger of dropping out of the program, in order to maintain cohort size and prevent attrition from the RCT. These triage decisions explain some of the variation on this indicator across individual sites.

Personalized coaching is intended to support participants as they practice new leadership skills in the context of their daily responsibilities, complete assignments, address roadblocks and challenges, and reflect on their growth. Coaching is the primary means by which Emerging Leaders participants learn to adapt and apply the program's skills and tools to their daily work and their specific school contexts; it is also the means by which they receive the feedback critical for development of personal leadership skills (that is, the ability to self-reflect to continuously improve performance) (Valdez, Broin, and Carroll, 2015). Thus, the shortfall in coaching hours for a large portion of participants may be a shortcoming of the program as delivered in the RCT sites, in particular when considering Emerging Leaders' ability to sustain the work outside of the structures and supports of the training year.

The program did not meet New Leaders' stated goals for demonstration of leadership proficiency at the end of the program, likely due to changes in scoring procedures implemented by New Leaders during the RCT year.

Although Emerging Leaders participants successfully completed all program requirements, very few (13%) demonstrated leadership proficiency at the end of the program, as determined by a set of written and performance assessments designed by New Leaders and scored by New Leaders' Assessor Corps, a small group of specially trained, nationally normed staff responsible for scoring assessments from all program sites. These proficiency levels were defined by New Leaders to signal participants' program success and readiness for greater leadership; program fidelity thresholds (that is, the percentage of participants expected to score proficient on at least 3 or 5 of 7 target leadership concepts) were set in line with prior cohorts' performance at program completion. By the end of the program, the New Leaders national team expected that 80% of participants would demonstrate leadership proficiency and that 30% would demonstrate readiness for the Aspiring Principals Program, according to the thresholds shown in Exhibit 6. In 2017–18, just 13% and 2% of Emerging Leaders met these benchmarks, respectively. Most Emerging Leaders participants (72%) failed to demonstrate proficiency on any of the seven target leadership concepts assessed by the New Leaders national Assessor Corps, although many were near proficient (2.5–2.9 on a scale of 1–4, where 3.0 and higher is "proficient") (Exhibit 7).

Exhibit 7. Distribution of Scores on Assessments of Seven Target Leadership Concepts, 2017–18



n = 53.
Source: New Leaders leadership assessments, 2017–18.
Note: Concept 1.1: Reflective Practice and Continuous Improvement; Concept 1.2: Communication and Interpersonal Relationships; Concept 2.1: Pedagogy and Instructional Strategies; Concept 2.2: Data Driven Instruction; Concept 2.4: Standards-Based Planning - Curriculum, Assessments, and Scope & Sequence; Concept 3.1: Urgency and Efficacy; Concept 4.2: Leadership Development.

The most likely explanation for these low proficiency rates lies in the new scoring procedures implemented by New Leaders during the RCT year. Historically, program directors scored assessments

locally. In 2017–18, New Leaders established a new national Assessor Corps with responsibility for scoring assessments from all Emerging Leaders sites. The Assessor Corps was intended to establish greater consistency and reliability of scores across program sites. The Assessor Corps appears to have been more stringent in its interpretation of the Emerging Leaders assessment rubrics than local program directors had been in the past: across all program sites, Emerging Leaders proficiency rates at program completion dropped by 25 percentage points between 2016–17 and 2017–18. Such a remarkable drop in proficiency rates seems unlikely to be driven by a drop in the achievement of Emerging Leaders participants, especially given that—at least in the RCT sites—most implementation measures point to an otherwise successful program implementation.

These lower scores from the Assessor Corps may represent an artificial depression of scores within the RCT sites if the corps applied the rubrics more strictly than the Emerging Leaders national team intended. Alternatively, it may be that in prior years local program directors—who had relationships with participants and had seen many of them put forth substantial effort to complete the program—gave participants the "benefit of the doubt" in applying the rubrics to participant practice, thereby inflating scores. Regardless, the change in scoring process makes it difficult to interpret the RCT cohort's proficiency rates relative to a threshold.[10] These more consistent scoring procedures, however, should provide more reliable data for measurement of program impacts discussed in the next chapter.

The fidelity indicators developed for this evaluation were not intended to be exhaustive or to cover every aspect of Emerging Leaders program implementation. Rather, they measured key program components that the New Leaders staff believed were critical for its success, and they summarized the extent to which the program, as delivered to participants in each of the three RCT sites, conformed to New Leaders' expectations. In the remainder of this chapter, we look beyond the summary statistics represented in the fidelity indicators to examine particular successes and challenges to program implementation during the RCT year.

Successes and Challenges to Program Implementation During the RCT Year

Interviews with Emerging Leaders program directors, participants, instructional team members, and principals, as well as observations of a small sample of instructional team meetings, offered insight into specific successes and challenges to program implementation in the RCT sites. These included successes and challenges related to participant recruitment and selection; Emerging Leaders program training; and variations in the quality of instructional team work.

Emerging Leaders Participant Recruitment and Selection

The challenge of recruiting a larger-than-normal program cohorts may have undermined the perception of Emerging Leaders' selectivity within districts.

New Leaders recruited larger-than-normal cohorts of Emerging Leaders candidates that met implementation fidelity requirements (i.e., over 70% of candidates met all admissions criteria) in each RCT site. Recruiting these larger cohorts to satisfy the design requirements of the RCT (so that half of all admitted candidates could be randomized to treatment and half to control) posed a challenge for the New Leaders staff and district leaders. One site did not agree to join the study until April 2018, well after the

[10] We recommend the national team carefully review the instruments, rubrics, and artifacts used in scoring to ensure they accurately reflect expectations for higher score points. This review team may find it helpful to ask local program directors to identify candidates who scored lower than expected for the sake of reviewing assignments that were likely to have received a higher score in prior years.

recruiting season had begun. At another site, regional New Leaders staff members originally misunderstood the requirement for the double cohort. In both these sites, the recruiting process was rushed as staff scrambled to vet candidates and the admitted cohorts were not as large as initially planned. Although these candidates met New Leaders' stated program requirements, the pressure to fill out the study sample may have contributed to a widespread perception among key stakeholders that the candidates were less qualified than those in prior years. In one site, district administrators expressed some disappointment at the results of their own recruitment efforts, and the program director reported that New Leaders staff had accepted virtually all program applicants.

Many of the concerns with the quality of candidates may have been driven by district leadership and candidates' contextual understanding of the program. In two of the districts, where New Leaders had been operating for several years, respondents noted the district's brightest stars had already completed the program. New Leaders launched in these districts in response to pent up demand for programs that would prepare promising candidates to move immediately into assistant principal or principal positions. By the RCT year, district staff reported that this pent up demand had largely been met, and that they and New Leaders staff had recruited from a pool of candidates who were typically in earlier stages of the leadership development trajectory. Further, in reflecting on the candidates completing Emerging Leaders, several district staff seemed to confound the program with New Leaders' Aspiring Principals Program (discussed in more detail in the next subsection). Given that Emerging Leaders trains leaders at a variety of points in the leadership pipeline (including e.g., current teachers), but Aspiring Principals supports candidates in their transition to the principalship, this comparison may have cast an unfair light on the current Emerging Leaders cohorts.

Further, program directors worried that the need to recruit larger-than-average cohorts for the RCT may have diminished the perception of selectivity within the RCT cohort. Program directors in all three sites noted that they spent more time and effort in the RCT year than in the past chasing down candidates for late assignments and other program requirements. For prior cohorts, the perception of highly selective admissions and the implicit threat of being dropped from the program midstream led to a sense of urgency and a culture of healthy competition within the cohort, according to program directors. They reported this dynamic was notably absent in the RCT sites, believing that the participants knew the program was undersubscribed.

Principals and district administrators frequently expressed confusion or contradictory views about the purpose of the Emerging Leaders program; in some cases, this confusion led to unrealistic expectations or ineffective support for participants.

The Emerging Leaders program is designed to improve the instructional leadership skills and capacities of teacher leaders, instructional coaches, and assistant principals. This is the primary program goal (Valdez, Broin, & Carroll, 2015), although the most successful participants also complete the program prepared to enter Aspiring Principals or another principal residency program. Each of the three study districts had adopted the Emerging Leaders program in the hopes that that it would contribute to their local principal pipeline, although neither New Leaders nor its district partners expected that all program completers would go on to become principals. The district leaders who had made the decision to adopt the program noted that it was equally important to develop instructional leadership skills among those leaders who intended to remain in the classroom or to serve other teachers as instructional coaches.

These dual program goals were often unclear to participants and school leaders, however, leading to unrealistic expectations about program content, how the program would advance participants' careers, or confusion about who should participate Participants and their principals described the primary purpose of the program variously as professional development for participants in their current roles, preparation for a principal residency program (e.g., Aspiring Principals), or a stepping stone to an assistant principal

placement in the next year. Principals reported different approaches to evaluating Emerging Leaders candidates from their schools, reflecting these different perspectives on program goals. Some principals reported that they endorsed applications from all candidates, not wanting to serve as gatekeeper for a high-quality professional development opportunity. Other principals considered candidates much more carefully and were much more selective in their endorsements, signing off only on those they believed were ready for a school leadership role.

Those principals and district leaders who viewed the Emerging Leaders program primarily as a principal development program were more critical of participants' skills and of the effectiveness of the overall recruitment and selection process. Prior experience with the Aspiring Principals program in SCS and SAISD tended to exacerbate the criticism. In those sites, district and school personnel were often unable to distinguish between the Emerging Leaders program and Aspiring Principals in interviews, referring to both programs as simply "New Leaders." District and school staff members commonly held expectations that did not align with the instructional focus of the Emerging Leaders program. For example, district and school staff cited the lack of training on budgeting, resource management, and human resources as a significant failing of Emerging Leaders. In addition, district and school staff often assessed the growth of the participants and the success of the program relative to whether they were ready for the principalship, rather than whether they were effective as a leader of an instructional team.

Emerging Leaders Program Training

Many Emerging Leaders participants reported making strides in their adult leadership skills in response to constructive critical feedback. Emerging Leaders regional program directors tended to triage participants for personalized coaching, meaning that some candidates received more time and support than expected, while others did not receive the minimum number of hours promised by the program.

Emerging Leaders participants consistently reported they received specific, actionable feedback from their instructional teams, school leaders, and/or Emerging Leaders program staff (typically in the context of learning meetings). The feedback focused on inter- and intrapersonal skills, including how participants communicated with the instructional teams, participants' comfort level with difficult conversations, and ways that participants could shift their behavior and mind-set to assume a leadership role on their instructional team. For example, participants who were teachers regularly struggled with the transition from peer to leader; this struggle led to some instructional teams making slow progress because of a lack of organization and direction. Emerging Leaders participants were aware of the specific interpersonal skills they needed to work on and viewed program assignments and protocols as opportunities for growth in these skills. In particular, both participants and instructional teams reported the difficult conversation protocol was a significant support for everyone involved to improve skills in being appropriately direct, providing critical feedback, and receiving critical feedback. Participants, instructional teams, and Emerging Leaders program directors shared similar perceptions of how specific participants had grown as leaders, suggesting open communication about participants' growth and positive outcomes from the interpersonal feedback provided to participants.

Regional program directors encountered several challenges to providing individualized coaching during the RCT year, and these experiences help to explain why some program participants did not receive the 12 hours[11] of personalized coaching consistent with the Emerging Leaders program design. Although

[11] The Emerging Leaders program agreement specifies 12 hours of personalized coaching for each Emerging Leader, including the time spent in learning meetings (approximately 3 hours total). The fidelity threshold reported in the previous chapter is set at 7 hours, taking into account the time spent in learning meetings (reported separately), and the fact that the hours logged in the coaching tracker often did not include time spent on phone calls or other informal coaching interactions.

program directors as a group provided the expected total number of coaching hours across all three sites, these hours were unevenly distributed among individual participants. Program directors reported that they spent much more time checking in with participants about basic program requirements during the RCT year than they had in the past, when participants who missed learning meetings or assignment deadlines might have been allowed to drop out of the program more readily. Thus, they sometimes reduced coaching time with participants who were already performing well in the program, in order to redirect that time to participants who required extensive scaffolding, support, or encouragement to complete the program successfully. In addition, program directors reported that the time they spent preparing to facilitate monthly learning cycle meetings (a responsibility previously assigned to specialized facilitators who traveled among sites) detracted from the time they spent coaching.

Two of the program directors also worked remotely from the sites they served, traveling to the districts for one week each month. This limited time on site reduced their flexibility in scheduling (and rescheduling) face-to-face coaching sessions. When scheduled meetings with busy educators inevitably fell through, program directors often resorted to making them up by phone, rather than waiting until their next week on site; they did not always record these informal interactions in the program's coaching tracker. For all of these reasons, program directors reported that they had been forced to triage their face-to-face coaching time, prioritizing participants who had the greatest need for additional support as one way of compensating for the lack of time available for personalized coaching overall. Among the Emerging Leaders participants we interviewed, reflections on coaching centered on required learning meetings. Participants had little to say about the coaching they may have received outside the context of formal learning meetings, which may reflect that these meetings were infrequent for some participants.

Alignment of Instructional Team Work to Existing Roles and Practices

Instructional teams were most likely to thrive in schools with an established culture of data use and strong norms supporting collaboration among teachers.

The most effective instructional teams we observed were embedded in higher-capacity schools. Instructional teams at schools with existing schedules for collaborative planning time or for professional learning communities were more successful at meeting regularly. Emerging Leaders participants at schools with a positive culture of regular and routine classroom observations were more likely to be on track to complete the required instructional team observations, and instructional teams with prior use of data used more Emerging Leaders protocols and pushed their analysis and usage of the data to deeper levels (e.g., to identify student misconceptions, design corrective instruction plans to address them, and follow through on those plans). Program participants in schools where these practices were absent faced greater challenges in finding reliable times to hold instructional team meetings, gaining access to classrooms to conduct observations (specifically when observations were used mainly for formal teacher evaluations), and developing basic data literacy among team members. Additionally, Emerging Leaders participants in schools without an existing DDI culture faced greater challenges obtaining support from school leaders to complete their program work.

Emerging Leaders participants who were instructional coaches were often best positioned to meet program expectations related to team leadership because of the flexibility in their schedules and because program requirements aligned well with their day-to-day responsibilities.

Nearly all Emerging Leaders participants met the minimum program expectations for leading team meetings (at least 10 over the course of the year), and a majority met the threshold for conducting observations of team members (at least eight times over the course of the year, or about monthly). Participants varied widely in the amount of contact they had with instructional team members, both in team meetings and in observations. Among the Emerging Leaders participants we interviewed, some met

with their teams several times a week and led team meetings at least weekly. Other Emerging Leaders had difficulty meeting as often as planned and convened on a less-than-optimal schedule (squeezing in short meetings before or after school, for example). Emerging Leaders participants faced similar challenges scheduling time to observe and provide individualized feedback to team members.

Instructional coaches were best positioned to incorporate the Emerging Leaders work into their normal responsibilities because observations and coaching were directly related to their positions. While they struggled to balance their overall increased workload, the flexibility of their schedules and alignment of their responsibilities facilitated their ability to complete—or even exceed-- program requirements. Among the respondents we interviewed, assistant principal participants were the most likely to struggle with competing demands on their time and regularly reduced their desired level of Emerging Leaders work for the benefit of other responsibilities. Those participants with full-time teaching responsibilities were the most likely to struggle with program requirements due to the lack of flexibility in their schedules. These participants were typically either able to hold regular meetings or conduct observations; which requirement they were able to fulfill depended on whether they shared common planning time with their instructional team. Of the schools in the site visit sample, we did not encounter any examples of school leaders who adjusted participants' schedules or reduced participants' job responsibilities to make time for their work on Emerging Leaders.

Instructional Team Implementation of DDI

Both Emerging Leaders participants and teachers reported the DDI work they had done together in Emerging Leaders teams was more structured, rigorous, and effective than the DDI work they had attempted in the past.

Instructional teams came to Emerging Leaders' DDI work with varying levels of experience analyzing data or planning instruction collaboratively in teams. Those teams with no formal training in DDI or in districts without established norms for professional learning community (PLC) work were especially likely to report that Emerging Leaders was more structured, rigorous, and effective than what they had done in the past.

Instructional teams consistently described several ways that Emerging Leaders participants supported and guided them toward improving their instructional practices related to student data. These improvements were relative to the existing capacities of each instructional team, such that teams with less data-driven instructional capacity saw different improvements (e.g., fully completing a DDI cycle) than teams with greater capacity (e.g., identifying and addressing specific student math misconceptions).

Instructional team members provided examples of improvements, including identifying specific students who needed additional support or intervention, identifying and addressing specific content knowledge gaps, authoring intentional distractors on assessments to surface misconceptions, conducting rigor analyses, and scheduling and following through on reteaching key content. As one elementary Emerging Leader described the difference,

> *Last year we took the data and said, "This is where [the students] are," and left it there. It's really bad because… we may speak to [the data], but there's no plan. [This year], we had a vision, that this is where we are going to go …. Now, what do you do next? This program gives you the next steps.*

In addition to crediting Emerging Leaders participants for supporting their growth, instructional teams cited the structure of the DDI cycles, regular meetings, and related activities as protecting the time required for engaging in the collaborative work and creating accountability for changing their practices.

Emerging Leaders program directors assessed most instructional teams as meeting or making significant progress toward their SMART goals for student achievement; in many cases, however, we were unable to assess how ambitious team SMART goals actually were.

SMART goals are intended to guide DDI work by setting a common goal to improve student achievement that then enables teams to focus and align their work. In their end-of-year review of participants' instructional team progress, Emerging Leaders program directors reported that 43% of instructional teams met their SMART goals and an additional 38% of instructional teams made significant progress toward these goals. To ensure that SMART goals would be well aligned with the impacts on student outcomes to be measured in this evaluation, New Leaders staff members and program directors reviewed SMART goals together to ensure that each goal identified a specific student outcome measure, set performance goals for *all* students in the grade or class, and set ambitious goal(s) for growth, consistent with New Leaders' guidance. One SMART goal that met each of these criteria read as follows:

> *By March 2018, high school students enrolled in Unified Geometry will grow from 24% to 34% proficiency in geometry performance as measured by the NWEA MAP assessment. The remaining 66% of students will be approaching mastery with a [MAP] score of at least 225.*

Although Emerging Leaders participants and their program directors most likely developed a shared understanding of program SMART goals as they reviewed them together in learning meetings, our own independent review found that most were not readily interpretable without significant additional information. More than half set a target for growth but did not specify the baseline; for example, goals may have set a target as "10% growth" without specifying whether that growth referred to an additional 10% of a student's starting score, 10% of a student's target score, or 10% of the gap between the starting score and the target score. A small number of goals were restricted in scope in some way: setting targets for only some groups of students or for narrowly defined achievement domains (e.g., setting goals for an isolated standard, such as phonemic awareness). It is possible that our external review – which did not include discussion with instructional team members—missed important facets of how teams used these goals. For example, narrowly defined goals may have been initial goals for teams who then moved onto a second SMART goal once they accomplished the first. Those teams with unclear quantitative goals may still have found that the focus on a particular standard or subgroup of students provided enough clarity to support the team's needs. Our review suggests that New Leaders may want to review this particular facet of training to ensure it meets participants' needs.

Instructional teams engaged in regular DDI cycles and used the Emerging Leaders program tools and protocols to frame their discussions, though they did not consistently move beyond initial steps in reviewing and analyzing student work.

Participants in all the observed instructional team meetings reviewed student data at some point during the meeting. The instructional teams varied in detail they used in discussing the data. For example, one observed team meeting included a focused collaborative conversation on formative assessment data and its aligned standards using the ATLAS "Looking at Data" protocol to identify specific student skill gaps and misconceptions. In other team meetings, instructional team members brought student data to review, but they spent the majority of the meeting discussing non-content-related reasons why students' scores might be low (e.g., the timing of the test, either just before or after a holiday or day off), explaining why particular individual students may not have scored well (e.g., identifying specific students with special education needs), or discussing their preferred classroom polling app or other pieces of technology.

Many of the instructional teams we observed had difficulty completing a full data analysis protocol or corrective instructional action planning protocol in the time allotted for the team meeting. Of the 10 team meetings we observed (six ELA teams and four math teams), four included extended discussions of

student misconceptions and instructional strategies to address or correct those misconceptions. A fifth meeting focused on development of a common assessment and included a brief discussion of answer choices that would serve as distractors to surface student misconceptions. The remaining teams either did not attempt to review assessment data or student work, or they got entangled in unproductive processes (e.g., discussions of color-coding assessment data) and ran out of time before they could make meaning from the tables in front of them, let alone glean actionable insights into their students' progress. Although most instructional team meetings that we observed allocated time for discussing instruction, only a few designed corrective instruction plans that identified a specific content standard, selected new instructional strategies, designated subgroups of students to receive corrective instruction, or set a timeline for the corrective instruction to occur. More often, teams discussed corrective instruction more broadly without establishing a clear action plan. For example, multiple teams identified a need for corrective instruction based on assessment results but did not set a clear timeline for when and how corrective instruction would occur.

When the observed instructional teams did engage in planning for corrective instruction, math-focused teams identified a specific misconception or an academic skill that required additional instructional support; observed ELA teams did not make this connection.

In the two math-focused instructional team meetings that engaged in corrective instruction, math teachers quickly generalized from specific assessment items to categories of math problems (e.g., finding the area of composite figure, using slope to find the distance between two points on a coordinate plane), and from there considered the strategies that students used to solve those problems. In some instances, math teachers had presented students with an algorithm for solving a particular kind of problem. This enabled teachers to easily pinpoint the source of student error by examining incorrect answers or observing where large numbers of students had gotten stuck and abandoned the question, because the steps students were expected to take were so well understood by the teachers. They could also easily identify pre-requisite knowledge and skills. In some cases, teachers identified instances where their own reliance on algorithms in teaching had limited their students' understanding of the math they had been trying to teach (see text box). Building on these insights, teachers generated strong plans for corrective instruction, tightly aligned to the specific student misconceptions they identified in the data.

Examples of Student Misconceptions and Next Steps

Math Instructional Team Meetings

In a **high school math** team meeting, teachers noted that students had scored poorly on an item asking them to find the length of a line segment plotted on a coordinate plane, if the distance from one end of the line segment to a point somewhere on the line is given. Students had calculated the slope of the line incorrectly, had started at the wrong end point when calculating ratios, or had found the length of the unknown segment but then had failed to add it to the known segment to get the full length of the line. After some discussion, teachers agreed that students' poor understanding of slope was a critical misconception. Near the end of the discussion, teachers had this exchange:

> *Teacher 1: I started [teaching this problem] very traditionally, as in step 1, step 2.… I was going, these are the points, take the ratio and find the distance ratio. Because I wasn't being so creative with it, students were more inclined to memorize the steps [without understanding the significance of the slope in the problem, a concept they had covered in an earlier unit].*

> *Teacher 2: I hear what you're saying. You showed steps without any instruction other than the steps. A high-impact strategy when you go back to correct is to provide … some way where you can start with this and then say, "Ok, how about everyone take 2 minutes and figure out how to do it?" That way you can address misconceptions when they have come to consensus. Give them something where they can see how one step leads to the next.*

An **elementary math** team discussed a problem relating the area of rectangles to properties of addition and multiplication. Students had been given a six-sided composite figure (a rectangle and a square joined together) with the lengths of only some sides given and were asked to find the area of the figure. In discussion, teachers identified the central misconception as students' assumption that they only needed to find one area (rather than decomposing the figure into a rectangle and a square and solving for the missing sides). Teachers discussed the rigor of the item, based on the number of steps required, as well as strategies for helping students identify known and missing information and plan to solve multi-step problems:

> *Teacher 1: That procedure would be to know when to divide into two figures. All the children got that 11 times 4 area, but the missing information was in the square, so you had to determine that before you start. I want to go over to where it says planning and evidence. Use what you have. First, you have to know that you have missing information there. I have to plan—decompose into two different figures. Now, I have to find the area for two plain figures. That takes a lot of planning.*

> *Teacher 2: Also, for this it takes three operations. It is three operations because you have to add, subtract, multiply, then add again… and in a certain order, like you said, using reasoning and planning… In our reasoning steps now, we have to think about increasing the rigor of each item.*

For corrective instruction, this teacher planned a week-long spiraling review, beginning with finding the area of simple figures, and helping students talk through their plans for solving multi-step problems. As evidence of progress, she noted, "Today [students had] no problem at all, and it was the same thing. But I heard more talk. I heard more talk."

By contrast, in the two ELA-focused instructional team meetings that engaged in corrective instruction, the ELA teachers we observed struggled to articulate generalizable insights about student misconceptions or understanding. These teachers were more likely to focus on their own thought processes as they considered specific test questions, possible sources of misunderstanding in the questions themselves, and test-taking strategies more generally, without generalizing to the underlying reading or writing skill being assessed (see text box). As a result, teachers had fewer ideas for corrective instruction, grounded in the data, that might have developed students' skills and understanding.

Examples of Student Misconceptions and Next Steps

ELA Instructional Team Meetings

An **elementary ELA team** reviewed an item assessing a standard on using a glossary or dictionary to determine the meaning of unknown words. Students were asked to use context clues from the passage and choose the correct meaning of "bank" as it appeared in the reading. A teacher shared how she explained to her students how to eliminate the meanings of "bank" that did not fit the context:

> *So, "bank" could mean many different things. It's how it is used. So, it's that context again. Okay, Bessie Coleman was the first African American to fly a plane. … So, "she learned how to bank turns and loop the loop" would have to do with some of these barnstormers doing these tricks. So, the other [answer choices], the heaping, sound like something you would do to put things in a place or something like that… So, if they can eliminate these actions, the heap, and the build and surround, that doesn't fit in this context. This is about movement … "turns" [in the dictionary definition] really helps, the word "turns."*

After this reflection, however, the team did not identify the misconception that had prevented students from replicating this kind of thought process. The team discussed several more multiple choice assessment items in this way, without drawing any conclusions about what students' thought processes might have been. When the discussion moved on to instruction, this teacher described general test-taking strategies, rather than ways to help students think strategically about using context clues to discern the meaning of unfamiliar words:

> *We have the acronym S-T-A-A-R, and the T is think about the question and underline the hot words or the important words…. A is analyze your answer choices and eliminate the silly answers, the ones that don't fit. And the second A… is to always prove. So, they have to write their paragraph number.*

The Emerging Leader on this team, who was also a classroom teacher, proposed experimenting with a tool for vocabulary instruction that would help students identify word structure, appositives, and context clues to identify meaning. The other teachers on the team did not commit to using this new tool, however.

A **middle school ELA team** met to review samples of student work, discuss misconceptions, and make plans for corrective instruction. In one classroom, students had read Martin Luther King's "I Have a Dream" speech and annotated examples of repetition as a rhetorical device and examples of imagery, using two different highlighter colors. No examples of repetition were highlighted on the sample of student work circulated among the group. Teachers identified the following examples of student misconceptions or errors related to annotation:

- Annotating too much (e.g., one student had turned in a paper where all of the text was highlighted except for names and pronouns)
- "Coloring" rather than using highlighting in a meaningful way
- Repeating the words of the text verbatim in marginal notations (e.g., one student wrote "MLK had a dream" in the margin of her paper)

Teachers shared strategies for providing feedback to students, including sharing examples of poor work from other classes. In their discussion, teachers focused on the characteristics of the student work (e.g., overuse of highlighting), but did not reflect on the reasons why students might find it difficult to select only key words or phrases for highlighting or to recognize any of the repetition in the passage or its function as a rhetorical device. As a result, the group failed to gain much, if any insight, into student misconceptions and the kinds of corrective instruction that might best address those misconceptions, as grounded in the evidence contained in the student work being discussed.

In planning corrective instruction, ELA teachers were also much more likely to discuss plans for providing students with additional instruction, looking to tutoring programs, resource teachers, or opportunities for additional small group instruction to supplement the regular ELA block. These additional resources are often offered in ELA but not in math, especially in elementary schools (Bitterman, Gray, and Goldring, 2013). Several ELA teams with such services available to them discussed how to coordinate logistics and identify students for small group instruction and tutoring groups. These discussions may have had an opportunity cost, however, as they appeared to distract teachers from discussing changes to their own instruction that might also have helped their students.

Treatment-Control Contrast

In this section, we transition from describing the Group I implementation of the Emerging Leaders program to contrasting the leadership development experiences of Group I and Group II. Examining the treatment-control contrast was a key element in this rigorous evaluation as it established the extent to which there were meaningful differences in leadership development experiences or activities between the treatment group (Group I) and the control group (Group II). If meaningful differences existed between the groups, as we would expect, then we can more confidently claim that any impacts on instructional practices or student outcomes did indeed arise from the Emerging Leaders program.

A majority of Group II Emerging Leaders participated in some kind of intensive leadership development program in 2017–18. Although these programs were about as time intensive as the Emerging Leaders program, they were much less likely to have a strong focus on data analysis and corrective instruction strategies.

Although New Leaders asked Group II Emerging Leaders to refrain from joining any other leadership development programs in 2017–18, most did not heed that request. More than half (55%; Exhibit A-21) of Group II Emerging Leaders participated in some form of leadership training in 2017–18, such as a graduate program, principal certification, instructional coach training, or another leadership development program (Exhibit A-24). Group II's leadership development programs met with similar frequency (Exhibit A-25).

These alternative leadership development programs differed from the Emerging Leaders program in their focus on some key content and objectives, however. Emerging Leaders in Group I were much more likely than those in Group II to report that analyzing data to identify student misconceptions, setting student achievement goals, creating a team vision and goals, aligning instruction with state standards, and planning corrective instruction were a critical focus of their leadership development program (Exhibit A-26).

Group I Emerging Leaders were more likely to have led instructional teams on a regular basis in 2017–18.

As part of the admissions process, and before randomization, New Leaders asked each Emerging Leaders candidate to identify an instructional team that he or she would lead during the program year. New Leaders asked Group II Emerging Leaders to engage in "business as usual" practice during the RCT year. That is, New Leaders did not explicitly ask Group II candidates to lead their identified teams during the 2017–18 year. However, leading the team may have been an existing responsibility of their current position (for example, as a grade-level lead, department chair, or instructional coach). As expected, Group I Emerging Leaders were much more likely to lead their instructional teams in the kinds of tasks required by the Emerging Leaders program on a regular basis during the RCT year. Group II Emerging Leaders also engaged in this work, albeit less frequently. For example, 78% of Group I Emerging Leaders

reported leading instructional teams biweekly or more often, compared with 42% of Group II participants (Exhibit A-20). Many Group II Emerging Leaders engaged with their identified instructional teams sporadically; about half reported occasionally leading teams in analysis of student data and lesson planning (every few months or less often).

Group I Emerging Leaders were much more likely to report significant growth in instructional leadership skills during the RCT year; they also believed they were better prepared to lead a school.

Despite Group II's participation in other leadership development experiences, Group I Emerging Leaders were much more likely to report their leadership skills had improved "to a great extent" during the 2017–18 school year. Approximately 70% of Group I Emerging Leaders reported they had gained the skills necessary to lead an instructional team, analyze student data, and direct a team to plan corrective instruction, compared with just 13–22% of Group II Emerging Leaders (Exhibit A-26). Similarly, 71% of Group I Emerging Leaders reported they gained the skills necessary to lead a school compared with 30% of Group II.

Summary and Discussion

New Leaders' theory of action for the Emerging Leaders program assumes a long chain of influence, from the program as delivered by national and local teams, through Emerging Leaders participants and their developing leadership skills, to teachers, and finally, to students. The program reaches classrooms only through the work of Emerging Leaders participants and the teachers on the instructional teams that they lead.

The Emerging Leaders program was delivered by New Leaders largely as designed, including recruitment of a larger-than-normal cohort and supporting nearly all candidates through the designated training and assignments. Further, the evidence suggests that the Emerging Leaders program was, in fact, a significant departure from the status quo for Group I. Although a majority of Group II Emerging Leaders also participated in a leadership development program of some kind during the RCT year, Group I Emerging Leaders were much more likely to lead instructional teams and were also more likely to report significant growth in leadership skills. We are therefore assured that this RCT is a meaningful test of the Emerging Leaders logic model.

Our study of the Emerging Leaders implementation did provide several insights into meaningful barriers and facilitators of program implementation. First, New Leaders' coaches were unable to deliver nine hours of personalized coaching to all Emerging Leaders participants as intended. This coaching is intended to support participants in adapting program strategies and tools to the particular circumstance of their own schools and teams and as such, plays a potentially critical role in helping participants maximize the benefits of their training. During the RCT year, the New Leaders coaches found themselves triaging their supports to focus on those candidates most in need. Coaches may have been hampered by barriers to this coaching, including physical distance and increased program responsibilities. But New Leaders would be well served to examine the role of this work within the context of their logic model: do they think all Emerging Leaders participants benefit from coaching, or is coaches' time best spent supporting the needs of struggling candidates?

We also noted that the frequency of instructional team work, which depended largely on the nascent leadership skills of program participants, varied significantly. Our interviews and observations of instructional team meetings offered a window on that variation. New Leaders own data on instructional teams, gathered via end-of-cycle surveys, is more limited. Because instructional team work is such a

critical component of the program, New Leaders may want to consider what other kinds of data collection (e.g., a log of instructional team work that includes a record of tools and protocols used) could give program staff better insight into team activity.

Participants' experience of the Emerging Leaders program—most significantly, their ability to engage in instructional team work—differed depending on whether participants were teachers or held some other role (instructional coaches, assistant principals). New Leaders may consider whether adapting the program to take into account to the constraints and affordances of these specific roles may help participants integrate new knowledge and skills into their practice more quickly.

In addition, although our sample was small, the kinds of work we observed on ELA-focused instructional teams compared with math-focused teams suggests that the program's approach to DDI might play out very differently in each discipline, with math-focused teams finding it easier to abstract from a single question into a larger skill. New Leaders may consider whether to provide additional training and supports to Emerging Leaders to practice these skills across disciplines, as they are likely to be in position to coach more than their single subject.

We turn next to a discussion of the program's effects on Emerging Leaders participants and on their instructional team members.

Chapter 3: Impacts on Leaders and Instructional Team Practices

- Emerging Leaders had a positive and statistically significant impact on all three facets of DDI leadership knowledge measured: using multiple forms of data to drive student achievement, leading a team through a DDI cycle, and building understanding of efficacy concepts.
- The Emerging Leaders program had a positive impact on two corrective instruction practices for teachers in math-focused instructional teams. There were no notable impacts on teachers' planning practice when teachers were analyzed without respect to team focus.

The previous chapter explored various aspects of Emerging Leaders program implementation: Although New Leaders delivered the program largely as designed, both the quantity and quality of the DDI work carried out by instructional teams varied, depending participants' leadership skills and on the constraints of their roles, school context, and instructional team focus. This chapter shifts from a focus on program implementation to adult outcomes, first for Emerging Leaders participants, and then for members of the instructional teams. We examine the program's impact on participants' DDI leadership knowledge as a measure of the instructional leadership capacity that participants develop in the program. We then investigate whether Emerging Leaders impacts instructional team members' instructional practices in alignment with the program's theory of action.

Impacts on DDI Leadership Knowledge

A central focus of the Emerging Leaders program is the use of data to inform instruction. New Leaders staff developed an assessment of DDI leadership knowledge to assess candidates' knowledge of key skills for leading a team of instructors through a DDI cycle. Our primary measure of leadership skills is New Leaders' Data Driven Instruction assessment, an externally validated and reliable instrument (α=.81; Goff & Hyun, 2014), which provides impact data on three key facets of candidates' DDI leadership knowledge.

- Using multiple forms of data to drive student achievement (Leadership Competency 2.2a)
- Leading a team through a DDI cycle (Leadership Competency 2.2b)
- Building understanding of efficacy concepts (Leadership Competency 3.1c)

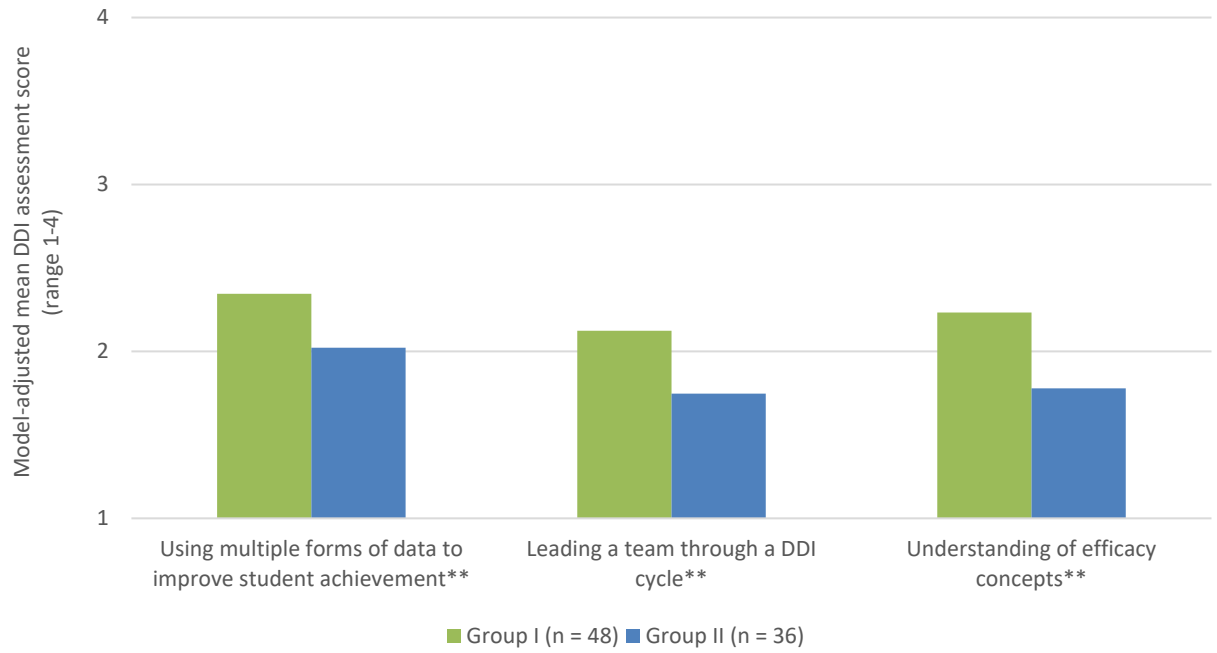
Data Sources and Methods

- New Leaders data-driven instruction assessment was administered to all Emerging Leaders at baseline as part of the application process (spring 2017) and again to Group I and Group II Emerging Leaders in February/March 2018.
- The DDI was scored on a 1–4 scale. Half points were allowed, creating a 7-point scale.
- New Leaders national assessor corps members double-scored each outcome assessment blind to treatment condition, achieving inter-rater reliabilities of .79 – .90, depending on the measure.
- At follow-up, 84 of 112 Emerging Leaders completed DDI assessments, resulting in overall attrition of 25% and differential attrition of 16%. The final analytic sample, after attrition, remained equivalent at baseline.
- Impacts on the DDI were measured using ordinary least squares (OLS) models, controlling for baseline DDI scores and other key covariates (grade level, subject, district, school characteristics, and New Leaders assessor).
- Graphs provide model-adjusted mean scores for Group I and Group II Emerging Leaders (i.e., predicted values).
- The estimated treatment effect, in raw scores, is the difference between the predicted DDI scores for Group I and Group II. This raw difference is translated into an effect size (g) to provide context for the size of the effects relative to the distribution of scores.
- The Technical Appendix provides additional detail on the measure, impact models, and analysis.

The Emerging Leaders program had a positive and statistically significant impact on all three facets of DDI leadership knowledge measured: using multiple forms of data to drive student achievement, leading a team through a DDI cycle, and building understanding of efficacy concepts.

The DDI assessment is part of the larger suite of assessments given to Emerging Leaders participants to assess their progress and proficiency on leadership. Group I Emerging Leaders outscored Group II candidates on all three facets of DDI leadership knowledge measured by the DDI assessment (Exhibit 8): using multiple forms of data to drive student achievement (2.34 Group I; 2.20 Group II, $g = .95$, $p < .001$), leading a team through a DDI cycle (2.12; 1.75, $g = .90$, $p < .001$), and understanding of efficacy concepts (2.23; 1.78, $g = .94$, $p < .001$). As noted in the previous chapter, very few candidates scored in the proficient range on this assessment (Exhibit 7), although New Leaders has designed these assessments to set the bar for proficiency high and instituted new scoring procedures in 2017–18 that likely depressed scores across all Emerging Leaders sites. Within these overall low scores, however, Group I participants outscored Group II candidates by a substantial margin. While the measured difference between groups on the 4-point scale used to score the DDI assessment is fairly small, it translates into nearly a full standard deviation difference.

Exhibit 8. Impacts on DDI Leadership Knowledge Competencies



~$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$.
Source: DDI assessment scores, spring 2018.
Note: Assessment score range: 1 = unproductive; 2 = approaching proficient; 3 = proficient; 4 = advanced.

The Emerging Leaders program was equally effective at improving participants' DDI leadership knowledge across subject areas. Program impacts were also similar across participants' roles with the school.

We ran a series of subgroup analyses to test whether the program worked differently for different groups of program participants. In particular, we compared participants leading math-focused instructional teams compared with participants leading ELA-focused teams and participants who were themselves teachers with those who held other positions. None of these sets of estimated impacts revealed differences between the two subgroups being compared, indicating that the New Leaders program's impacts were consistent across all groups. Complete estimates of each subgroups' effects on the DDI are available in appendix Exhibits A-51 and A-52.

We turn next to the impacts of the Emerging Leaders program on instructional team members to examine whether these impacts on DDI knowledge translate into consistent differences in teachers' instructional practice.

Impacts on Teacher Instructional Practice

The instructional log asked instructional team members with their own classrooms to document their daily instructional planning behaviors on a given day. We designed the log to measure instructional practices related to DDI cycles as conceptualized in the Emerging Leaders program. We defined instructional planning broadly, including any work teachers do to prepare to teach. In addition to completing written (or unwritten) lesson plans, these tasks might include reviewing standards, curricula, or pacing guides;

reviewing formative or interim assessment data (e.g., exit tickets, benchmark assessment data, etc.); reviewing student writing, classwork, or homework; setting learning targets; and assigning students to groups. Most of the items included on the logs focused on the planning work a teacher executed that day. A smaller set of items asked teachers to document instructional activities designed to develop student efficacy and ownership of their growth and achievement. The log had four main topics:

- **General instructional planning activities:** These items asked teachers to record the amount of time they spent planning, whether they planned with others (and with whom), and what activities they completed during daily instructional planning, including whether they reviewed assessment data or samples of student work.
- **Review of assessment data or student work:** These items were designed to assess whether teachers had adopted the data analysis strategies that Emerging Leaders participants had been trained to engage in with their teams in their own daily instructional planning. Specifically, the logs asked what types of data or student work samples teachers had reviewed in daily planning, whether teachers reviewed disaggregated data, and whether they reviewed incorrect answers to individual questions or samples of student work in order to identify student misconceptions or misunderstandings. These last two steps—review of incorrect answers and identification of student misconceptions—is a key element of the DDI cycle as reflected in Emerging Leaders program tools and training (New Leaders, 2016a).
- **Corrective instruction action planning:** These items were designed to assess whether teachers had adopted the corrective instruction planning strategies that are a key element of the Emerging Leaders DDI cycle as a regular part of their own daily instructional planning. Log items asked whether teachers planned to revisit and reteach past content and skills, whether they selected instructional strategies specifically designed to address student errors or to convey the material in a new manner, whether they grouped students for corrective instruction based on data, and whether they planned to reassess students' understanding after corrective instruction (New Leaders, 2016b).
- **Activities to develop student efficacy:** Program participants lead instructional teams in setting goals for student achievement and in engaging students in pursuit of those goals. Emerging Leaders participants were trained in strategies for developing student efficacy and ownership over their growth and achievement (New Leaders, 2016b) and coached their teams in using these strategies in their classrooms. The logs assessed whether teachers had adopted these strategies as a regular feature of their instruction. Log items asked teachers whether their students had engaged in the following activities in class that day: revising their work in response to feedback, reviewing assessment data, setting goals for their performance on future assessments, and conferencing with their teachers about those goals**.**

In addition to the instructional log questions, we asked teachers to respond to two sets of survey questions:

- **Teacher self-efficacy measures:** Teachers rated their own effectiveness on two survey scales replicated from earlier research on the relationship between instructional leadership and teacher self-efficacy (Bellibas et al., 2017; Klassen et al, 2009). The scales measured teachers' self-efficacy related to implementing alternative instructional strategies and efficacy related to engaging students in learning.

Data Sources and Methods

- The research team developed a daily instructional planning log for this study. Notably, these logs measure only frequency, not quality, of teacher instructional planning activity. They will, therefore, identify shifts in how teachers on instructional teams spend planning time but do not measure any improvements in the quality of planning. The appendix contains a detailed description of log items and tests of reliability.
- Administered to all Emerging Leaders instructional team members in Group I and Group II, with team membership confirmed at baseline (6 weeks after the beginning of the 2017–18 school year). This "early joiner" sample includes teachers who joined teams after randomization but early in the implementation of the Emerging Leaders program.
- Teachers completed instructional planning logs daily for 1 week at baseline (early fall 2017) and for 2 separate weeks at follow-up (spring 2018).
- Although outcome data was collected for 12 days of instructional planning (Sunday–Friday of 2 weeks), these data were collapsed into a binary (i.e., 0/1 or yes/no) value for each week. The data in the model therefore represent whether teachers engaged in each practice over the week and sample *n*'s represent weeks, not teachers.
- At follow-up, 313 teachers on 97 Emerging Leaders teams completed logs, resulting in cluster-level attrition of 13% overall (7% differential) and teacher-level attrition of 17% overall (3% differential).
- Impacts on the teacher instructional log were measured using hierarchical linear models (HLM), controlling for baseline log scores and other key covariates (grade level, subject, school-level demographics, and district) while accounting for clustering of log weeks within teachers and teachers within instructional teams.
- Binary outcomes were estimated with logit link function, with results presented as predicted percentage of weeks in which teachers did each type of instructional activity for ease of interpretation.
- The Technical Appendix provides additional detail on impact models and analysis.

Instructional logs have been found to be valid, reliable substitutes for observational data on teachers' practices and time usage in the classroom (Gallagher, Arshan, & Woodworth, 2017; Gallagher et al. 2012; Rowan & Correnti, 2009; Rowan, Jacob, & Correnti, 2009). The log asked teachers to document instructional planning and class activities on that day, maximizing the likelihood they would accurately remember and report planning and instruction. Each daily log also included two different teacher efficacy survey items to provide the full scale over the course of the week. Exhibit 9 shows the 21 primary teacher outcomes measured via these daily logs (results for other teacher outcome measures less closely aligned to the Emerging Leaders theory of action are included in the appendix). As some practices may have diminishing returns or even be counterproductive when done multiple times in the same week, we analyzed whether the Emerging Leaders program had an impact on whether Group I teachers were more likely to engage in a particular activity at any point during a single week, rather than whether they engaged in these activities on a larger proportion of days.

Exhibit 9. Primary Teacher Outcomes Measured

<table>
<tr>
<td>General Instructional Planning Activities
(instructional planning today, via daily log)
• Total planning time (minutes)
• Collaborated with others in planning (e.g. teacher team, informal group, instructional coach, resource teacher, aide)
• Reviewed student work or assessments</td>
<td>Corrective Instruction Action Planning
(instructional planning today, via daily log)
• Decided to revisit content or skills covered in previous lessons
• Selected new instructional strategies to address specific student errors and/or strategies that were different from the way skills or content was previously taught
• Grouped students based on assessment data
• Planned to re-assess students using a quiz, test, or writing sample
• Planned to address gaps in skills not previously taught</td>
</tr>
<tr>
<td>Review of Assessment Data or Student Work
(instructional planning today, via daily log)
• Reviewed student progress against goals
• Reviewed disaggregated assessment data (e.g., by class period, by student sub-group, by standard)
• Reviewed item-level scores
• Examined incorrect answers
• Identified errors or misunderstandings in assessment data or samples of student work</td>
<td>Activities to Develop Student Efficacy
(student activities in today's class, via daily log)
• Corrected, revised, or improved work in response to feedback (their own or a classmate's)
• Reviewed assessment results
• Set goals for next assessment
• Made an action plan
• Conferenced with teacher about progress or goals
• Completed post-assessment reflection sheet</td>
</tr>
<tr>
<td colspan="2">Teacher Self-Efficacy Survey Measures
(teacher self-assessment on a 5-point scale, via one-time survey)
• Instructional efficacy (e.g., craft good questions, implement a variety of assessment strategies, provide alternate explanations, implement alternative instructional strategies)
• Student engagement efficacy (e.g., motivate students who show low interest, get students to believe they can do well, help students value learning, assist families)</td>
</tr>
</table>

Impacts on Instructional Practice Among Math and English Language Arts Teachers

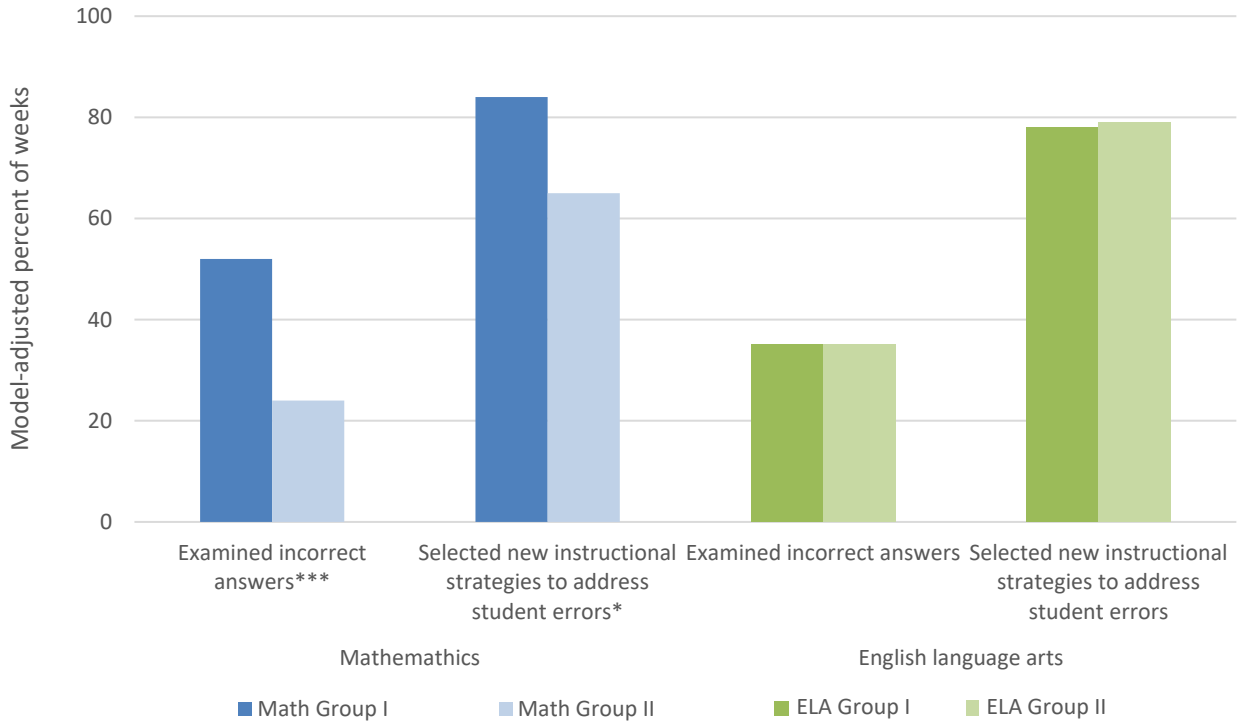
As described in the previous chapter, our observations of math- and ELA-focused instructional teams suggested that the DDI cycle and corrective instructional action planning played out differently in math-focused instructional teams compared with ELA-focused teams. Thus, we examined impacts on teacher practice among math and ELA teachers separately, using a similar process to the subgroup estimations described using the DDI data above. In these analyses we considered whether the Emerging Leaders program had an impact on math and ELA teachers, and whether those impacts differed by subject.[12]

[12] Such subgroup analyses can lead to a high risk of finding a statistically significant result by chance (i.e., Type 1 error or false positive), especially when working with an instrument with as many potential outcomes as the instructional log. To minimize the risks of such a result we took two steps. First, we only ran subgroup analyses for the instructional log questions we deemed most closely related to the program's logic model; in the technical appendix we refer to these as the "instructional log primary outcomes." Second, in this section we only report on findings if (1) an impact in one of the subgroups has a *p*-value of less than .05, indicating a statistically significant result for that subgroup and (2) the difference in impacts between the subgroups has a *p*-value of less than .05, indicating that the difference is also statistically significant. In this section, we do not report marginally significant results (those with a p value between .1 and .05).

The Emerging Leaders program had a positive impact on two key practices among math teachers: examining incorrect answers and selecting new instructional strategies to address student errors and misconceptions.

Relative to math teachers on Group II instructional teams, math teachers on Emerging Leaders instructional teams were more likely to review assessment data for students' incorrect answers (52% Group I; 24% Group II, p < .001, Exhibit 10), whereas there was no impact for ELA teachers (35%; 35%, *p* > .05). Math teachers on Emerging Leaders instructional teams were also more likely to select new instructional strategies for revisiting past content (84% Group I; 65% Group II, p < .05), whereas there was no similar impact for ELA teachers (78%; 79%, *p* > .05).

Exhibit 10. Impacts on Math Teachers' Instructional Practices: Reviewing Assessment Data and Corrective Instruction Action Planning



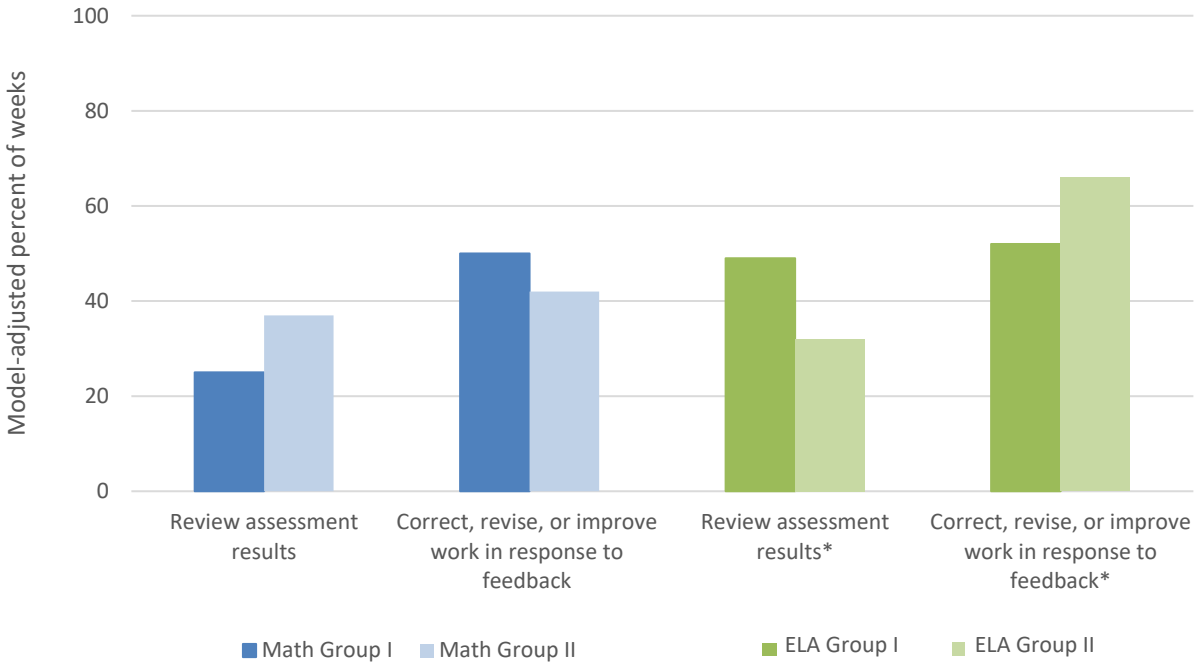
*p < .05, **p < .01, ***p < .001.
Source: SRI instructional log, spring 2018.
Note: The exhibit shows 2 of 10 possible teacher outcomes in the areas of reviewing assessment data and corrective instruciton action planning (items highlighted in green). There were no measurable impacts on other outcomes for math teachers.
Sample: For ELA, 59 instructional teams, 302 log weeks. For math, 45 instructional teams, 249 log weeks.

Group I English language arts teachers were more likely to ask students to review their standardized assessment data and less likely to ask students to revise their work in response to feedback, compared with Group II English language arts teachers. Math teachers demonstrated the opposite pattern.

ELA teachers in Emerging Leaders instructional teams were more likely to ask students to review assessment data (49% Group I; 32% Group II, *p* < .05, Exhibit 11) and less likely to revise their own work,

(52%; 66%, $p < .05$), as shown in Exhibit 11. There were no statistically significant differences for math-focused instructional team members, though results trended in the opposite direction, with non-significant post estimates indicating math teachers on Emerging Leaders instructional teams were, if anything, were less likely to have students review assessment data (25%; 37%, $p < .05$) and more likely to revise their work (50%; 42%, $p < .05$).

Exhibit 11. Impacts on ELA and Math Teachers' Instructional Practices: Activities to Develop Student Efficacy



*$p < .05$, **$p < .01$, ***$p < .001$.
Source: SRI instructional log, spring 2018.
Sample: For ELA, 59 instructional teams, 302 log weeks. For math, 45 instructional teams, 249 log weeks.

Impacts on Instructional Practice of Emerging Leaders Program Participants

The Emerging Leaders program had different instructional impacts on the Emerging Leaders participants who were still teaching relative to the other instructional team members. There are two possible explanations for this differential impact on program participants compared with other teachers. First, Emerging Leaders participants' direct exposure to New Leaders training may have increased the likelihood that they would change their own instructional practices compared with instructional team members who received training and coaching under a train-the-trainers model.

Another possible explanation for the strong impacts among program participants is that Group I Emerging Leaders had a better understanding of and commitment to the program's objectives than members of instructional teams and a better understanding of the kinds of responses to the daily instructional logs that would reflect desirable behavior. As with most self-reported data, the small, positive differences on these measures could reflect an unconscious bias toward socially desirable responses among program participants.

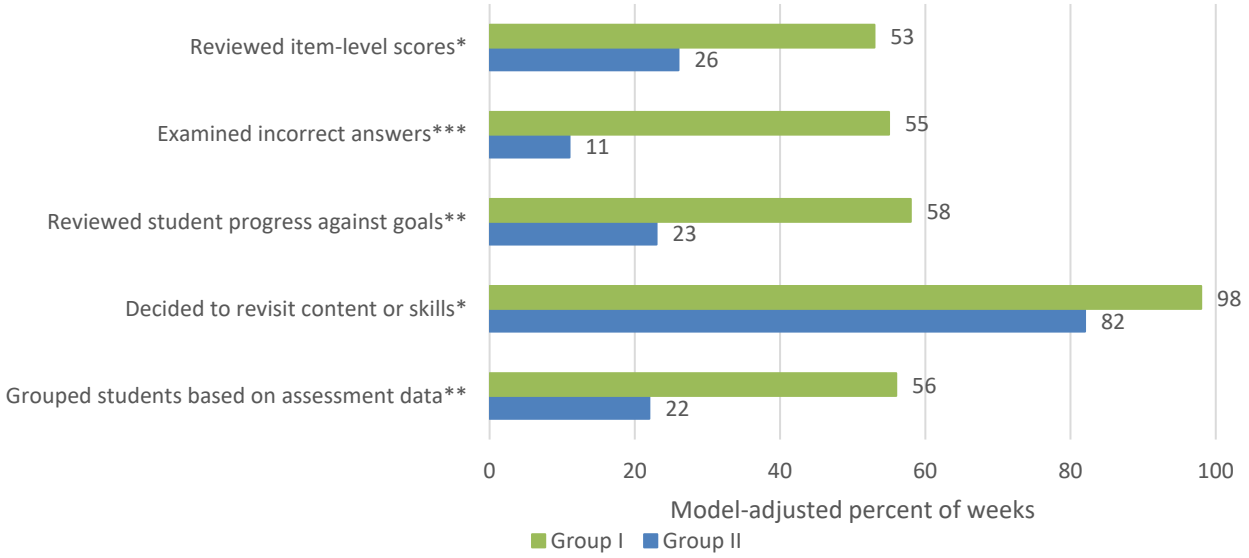
The Emerging Leaders program had several statistically significant positive impacts on the instructional practices of Emerging Leaders who were teachers themselves.

As with math teachers, these impacts showed up in two areas measured by the instructional logs: review of assessment data and corrective instruction action planning (Exhibit 12). Relative to Group II Emerging Leaders who were themselves teachers:

- Group I Emerging Leaders-were more likely to review assessment data specifically looking at students' incorrect answers (55%; 11%, $p < .001$), whereas there was no similar effect for instructional team members who were not Emerging Leaders (39%; 32%, $p > .05$).
- Group I Emerging Leaders-were more likely to review students' progress against learning goals when reviewing assessment data (58%, 23%, $p < .01$), whereas there was no similar effect for instructional team members who were not Emerging Leaders (52%; 49%, $p > .05$).
- Group I Emerging Leaders-were more likely to decide to revisit content covered in prior lessons (98%; 82%, $p < .05$), whereas there was no similar effect for instructional team members who were not Emerging Leaders (95%; 95%, $p > .05$).
- Group I Emerging Leaders-were more likely to group students based on assessment data (56%; 22%, $p < .01$), whereas there was no similar effect for instructional team members who were not Emerging Leaders (34%; 35%, $p > .05$).

Although there were impacts on Emerging Leaders who were themselves teachers, we did not find any statistically significant effects of the Emerging Leaders program on the instructional practices of teachers who were not Emerging Leaders themselves.

Exhibit 12. Impacts on Emerging Leaders Program Participants: Reviewing Assessment Data and Corrective Instruction Action Planning



*$p < .05$, **$p < .01$, ***$p < .001$.
Source: SRI instructional log, spring 2018.
Note: The exhibit shows 5 of 10 possible teacher outcomes in the areas of reviewing assessment data and corrective instruction action planning (items highlighted in green). There were no measurable impacts on other outcomes for Emerging Leaders participants who were also teachers.
Sample: 45 instructional teams, 86 log weeks.

Overall Impacts on Instructional Practice

We began this section by examining instructional impacts for math teams and ELA team separately. We now present impacts of Emerging Leaders on the instructional practices of all instructional team members, using the entire study sample; these results include both the teachers lead by Emerging Leaders and Emerging Leaders who are also classroom teachers.

The Emerging Leaders program had limited impacts on teacher outcomes when analyzed without respect to subject focus or Emerging Leaders participation.

Group I teachers spent an average of 33 additional minutes on instructional planning each week, over and above the time spent by teachers in the control group (252 minutes; 219 minutes, $p < .08$, Exhibit 13), although this difference approached (but did not meet) the threshold for statistical significance. On average, teachers on Emerging Leaders instructional teams were more likely to examine students' incorrect answers when reviewing assessment data (42% Group I; 30% Group II, $p < .05$). They were also more likely to examining question-/item-level scores (43%; 34%, $p < .1$), although this difference approached (but did not meet) the threshold for statistical significance.[13] Teachers were generally comparable across groups on other teacher outcome measures, including corrective instructional action planning, activities to develop student efficacy, and teachers' own ratings of self-efficacy were small and not statistically significant.

Exhibit 13. Emerging Leaders Program Impacts on Teacher Outcomes

Teacher Outcome	Group I Mean (% weeks)	Group II Mean (% weeks)	Sig.
General Instructional Planning Practices			
Total weekly planning time (in minutes)	252[a]	219[a]	~
Collaborated with others in planning	97	95	
Reviewed student work or assessments	89	86	
Review of Assessment Data or Student Work			
Reviewed student progress against goals	53	46	
Reviewed disaggregated assessment data (e.g., by class period, by student sub-group, by standard)	67	65	
Reviewed item-level scores	43	34	~
Examined incorrect answers	42	30	*
Identified errors or misunderstandings in assessment data or samples of student work	68	64	
Corrective Instruction Action Planning			
Decided to revisit content or skills covered in previous lessons	96	93	
Selected new instructional strategies to address specific student errors and/or strategies that were different from the way skills or content was previously taught	79	74	
Grouped students based on assessment data	38	34	
Planned to re-assess students using a quiz, test or writing sample	51	43	

[13] In this report, we use the standard p < .05 threshold to determine statistical significance. Under this standard, these results and all others we describe as "approaching (but not meeting) the threshold for statistical significance" are considered marginally significant at p < .10.

Teacher Outcome	Group I Mean (% weeks)	Group II Mean (% weeks)	Sig.
Planned to address gaps in skills not previously taught	37	31	
Activities to Develop Student Efficacy			
Corrected, revised, or improved work in response to feedback (their own or a classmate's)	51	55	
Reviewed assessment results	38	33	
Set goals for next assessment	8	9	
Made an action plan	87	85	
Conferenced with teacher about progress or goals	82	80	
Completed post-assessment reflection sheet	60	68	
Teacher Self-Efficacy			
Instructional efficacy (scale of 0–4)	3.17[b]	3.23[b]	
Student engagement efficacy (scale of 0–4)	3.43[b]	3.45[b]	

~$p < 0.10$, $*p < .05$. $**p < .01$. $***p < .001$.
Note: Group means are reported as the proportion of log weeks. For example, Group I teachers reviewed student work in 89% of log weeks, on average, controlling for other variables. [a]Total weekly planning time is reported in total number of minutes spent in instructional planning. [b]Teacher efficacy is reported on a Likert scale from 0–4.
Sample: 97 instructional teams, 590 log weeks.

Summary and Discussion

The Emerging Leaders program had large, statistically significant impacts on the three facets of DDI knowledge measured: using multiple forms of data to drive student achievement, leading a team through a DDI cycle, and understanding of efficacy concepts. The program was equally effective for different subgroups of Emerging Leaders participants: both teachers and others (coaches/assistant principals), and leaders of both math- and ELA-focused teams. Our interpretation of these findings was tempered by the low overall scores on the DDI assessment: although Group I outscored Group II Emerging Leaders on each of these three skills, few Group I participants scored in the proficient range. However, these low scores may be due in part to new scoring procedures adopted in 2017–18.

The Emerging Leaders program had large, statistically significant impacts on two key facets of math teachers' corrective instruction: examining incorrect answers and selecting new instructional strategies to address student errors or misconceptions. Further, the program shifted the use of student work in different ways by subject area, with ELA teams becoming more likely to have students review standardized assessments and less likely to revise their own work, while math teachers trended in the opposite direction. These differing log impacts by subject area are consistent with differences in DDI work observed in math and ELA instructional team meetings, where we observed math and ELA instructional teams interpreting corrective instruction differently.

When analyzed without regard to subject area, the Emerging Leaders program had few notable impacts on teachers' self-reported instructional practice, and no impact on teachers' self-efficacy.

Shifts in teacher practice are difficult to measure and the instructional logs designed for this study had two significant limitations. First, teachers in both treatment and control were already engaged in many of the practices measured by the instructional logs at baseline, leaving little opportunity for the program to shift the frequency of these practices among teachers on Emerging Leaders teams. To highlight one example, about two-thirds of teachers in Group I and Group II reported that they reviewed assessment data at least

weekly to identify student errors or misunderstandings, a key element of the DDI cycle as designed by the Emerging Leaders program. Teachers on Emerging Leaders teams may have reviewed assessment data more carefully or arrived at better insights about student misconceptions, but instructional logs are best suited to measure low-inference frequency measures. Second, some of the items included on the logs did not have high rates of interrater reliability, suggesting that the measures may have been too noisy to detect small but still meaningful differences in teacher practice (the appendix has more detail on reliability testing).

We next turn to the program's effects on student academic achievement in math and ELA.

Chapter 4: Impacts on Student Achievement

- The Emerging Leaders program's overall impact on math achievement was positive and approached (but did not meet) the threshold for statistical significance. There were positive, statistically significant impacts on the math achievement of female students, Latinx students, English language learners, and economically disadvantaged students.
- Program impacts on math achievement were mediated in ways that support the Emerging Leaders program theory of action: program impacts operated both through students having teachers regularly attend instructional team meetings and through Emerging Leaders participants' DDI leadership knowledge.
- The Emerging Leaders program had no measured effect on student English language arts achievement.

In Chapter 3, we saw that the Emerging Leaders program had substantial impacts on participants' DDI leadership knowledge, leading to modest impacts on instructional team practices, which included the examination of incorrect answers and selection of new instructional strategies to address student errors, and key shifts to the math teachers' corrective instruction practices. In this chapter, we continue the exploration of the impacts of the Emerging Leaders program by moving to the final element of the logic model, effects on student learning.

We first present the main impact estimates: those on all students taught by Emerging Leaders instructional team members. To unpack and further explore the program's impacts, we also present several exploratory estimates to help understand how the impacts may function. We begin by presenting impacts on student subgroups to understand whether the overall lack of program effects may hide meaningful impacts for particular groups of students and next turn to impacts on district benchmark assessments. While there were problems with these benchmark assessment data—most notably, one district did not systematically collect these data and was therefore excluded from the analysis—these results help us understand whether program effects may be detected with the data that teachers are more likely to use in their regular instructional practice. Once we present a clear picture of the program's direct effect on student achievement, we describe whether either students' exposure to the program or Emerging Leaders participants' growth in leadership skills mediated the program's impact on student achievement. The methods used in these analyses are summarized below. Full analysis details are in the Technical Appendix.

Data Sources and Methods

- *State assessments of math*: STAAR (TX) or TNReady (TN) grades 3–8 and end-of-course exams in Algebra I.
- *State assessments of English language arts*: STAAR (TX) or TNReady (TN) grades 3–8 and end-of-course exams in ninth-grade English I and 10th grade English II.
- *Benchmark assessment data (SAISD and SCS only)*: grades 2–11 NWEA MAP.
- State assessment scores were standardized within grades using the relevant statewide mean and standard deviation. MAP scores were standardized within grades and within districts using national means and standard deviations.
- We include the student's own baseline score for both math and ELA from the prior year as baseline controls. In SAISD and SCS we supplement the baseline state assessment data with the 2nd grade NWEA scores for the 3rd grade students; as these data are unavailable in AISD, we omit 3rd grade teaching teams from the analysis for this district only.
- Students were identified in fall 2017 from student rosters of teachers confirmed as members of Emerging Leaders instructional teams at baseline (6 weeks after the beginning of the 2017–18 school year). Such a sample includes "early joiners," or teachers and students who entered the study sample after randomization but early in the implementation of the Emerging Leaders program.
- Our analytic sample includes all students assigned to treatment who completed the aligned state assessment in spring 2018. This sample includes 6,317 students of teachers on 28 math teams and 5,446 students of teachers on 33 ELA teams. These analytic samples meet the WWC requirements for a low-attrition RCT under liberal assumptions. The Technical Appendix provides additional detail on attrition rates.
- Impacts on student outcomes were estimated using 2-level HLM models, clustering students within instructional teams and controlling for student and school characteristics.
- The Technical Appendix provides additional detail on impact models and analysis.

Primary Impacts on Math and English Language Arts Achievement

We selected state assessments in math and ELA to be the main outcomes for the analyses because they are the most policy-relevant outcomes for districts and schools. State test scores are frequently used in education evaluation studies given that they are consistently available starting in grade 3, as required for federal accountability. Moreover, the Emerging Leaders program trained participants to lead instructional teams in conducting standards-based rigor analyses of lesson plans and instructional materials. Participants worked with teachers to more deeply understand what content was required for students to demonstrate proficiency of specific state content standards and to design and teach learning activities aligned with the required content. State assessments are designed to measure the content of their content standards, making them well suited to assess whether the Emerging Leaders program creates learning gains for students. While the specific content of state tests varied somewhat by state, we adjusted the state assessment data to make cross-state comparisons meaningful (May, Perez-Johnson, Haimson, Sattar, & Gleason, 2009). The treatment coefficients are standardized and can be interpreted as an effect size (see text box).

Interpreting the Size of Student Impacts

In this report, we present the estimated magnitude of student impacts in effect sizes, as estimated by our statistical models. Presenting the results in effect sizes allows us to use 95% confidence intervals to display not just the estimated impact of findings but the relative certainty with which we estimate these effects. The standardized treatment coefficient is interpreted as the treatment effect size in accordance with WWC effect size reporting guidance for state-normed tests.

Educators and policymakers often find post hoc translations of effect sizes into other metrics more easily interpretable. Percentile growth provides the best performing of these translations (Baird & Pane, 2019). In addition to relying on fewer assumptions than other post hoc translations, percentile points are used by WWC to translate effect sizes into an "improvement index." This translation therefore allows for comparison to other rigorous evaluations within the field.

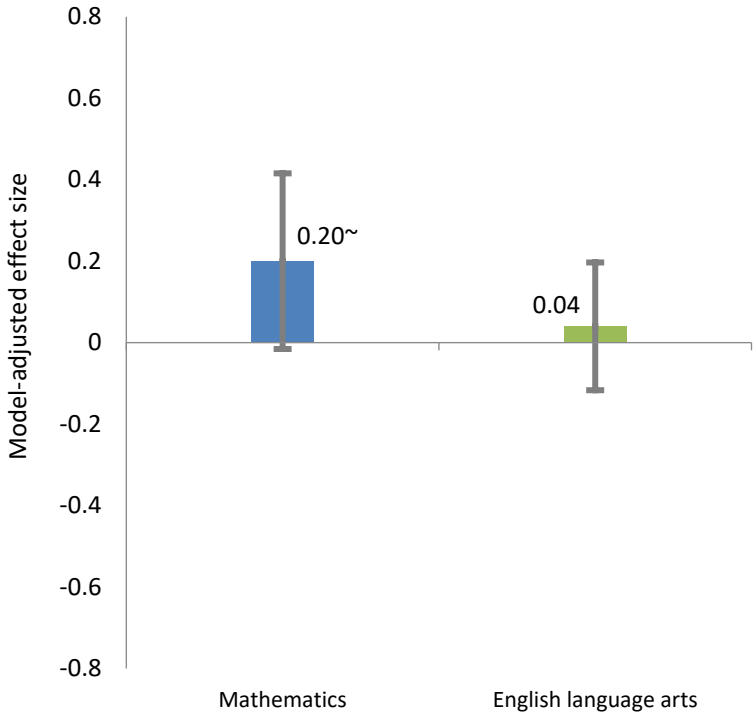
A 1 percentile point impact indicates that the median control student in our study at the 50th percentile of statewide achievement would have moved, on average, to the 51st percentile of statewide achievement. We provide percentile translations for all student impact estimates with statistically significant ($p < .05$) findings on state assessments, our primary outcome measure. We include both these percentile translations for statistically significant findings and improvement indices for all student impact results in the Technical Appendix (Exhibits A-86–A-88).

The Emerging Leaders program's overall impact on math achievement was positive and approached (but did not meet) the threshold for statistical significance.[14]

The Emerging Leaders program had a positive impact on student math achievement on state standards-aligned assessments which approached but did not meet the standard threshold for statistical significance ($g = .20$, $p < .1$, Exhibit 14). That is, although the size of the measured effect was promising, it was estimated imprecisely enough that it may have arisen by chance. The error bars on the impact estimates shown in Exhibit 14 represent this imprecision; they show a 95% confidence interval for the estimated effect size.

[14] In this report, we use the standard $p < .05$ threshold to determine statistical significance. Under this standard, results we describe as "approaching (but not meeting) the threshold for statistical significance" are considered marginally significant at $p < .10$.

Exhibit 14. Estimated Treatment Effects on State Assessment Outcomes



~p < .1, *p < .05, **p < .01, ***p < .001. Grey brackets represent the 95% confidence interval of the estimated effect size.
Source: District data, state assessment outcomes, 2017-18 school year.
Samples: For math, 28 instructional teams and 6,317 students. For ELA, 33 instructional teams 5,446 students.

The Emerging Leaders program had no measurable impact on students' English language arts achievement on state assessments.

There was no measurable impact of the Emerging Leaders program on students' ELA achievement (Exhibit 14). In addition to being statistically nonsignificant, the estimated effect of the program on ELA was close to zero (g = .04, p > .05), indicating that the Emerging Leaders program as implemented in these three districts is unlikely to have had meaningful impacts on students' ELA achievement.

Impacts by Student Subgroup

High-quality corrective instruction requires that teachers employ instructional practices, provide resources, and create learning opportunities tailored to the needs of the students receiving the instruction and requires different pedagogical approaches for different groups of students (Bellert, 2015; Guskey, 2007, 2016; Marzano, 2010). Accordingly, we recognized the possibility that Emerging Leaders practices may benefit some student subgroups more than others and investigated it.

We analyzed state assessments related to different subgroups of students based on student demographics. We began by analyzing student outcomes according to the major demographic characteristics of race/ethnicity (African American, Latinx, White) and gender (female, male). We then analyzed student outcomes according to the common student status designations of English learner

status, special education status, economically disadvantaged status, and high- and low-performing status.[15]

We investigated student subgroups using a within-subgroup approach. This means, for example, that we analyzed female student outcomes to answer the question, "Does the Emerging Leaders program have a measurable impact on female students?" without comparing these effects with those of male students. We did so because, unlike with the adult subgroups, not all student subgroups had a clear single control group.

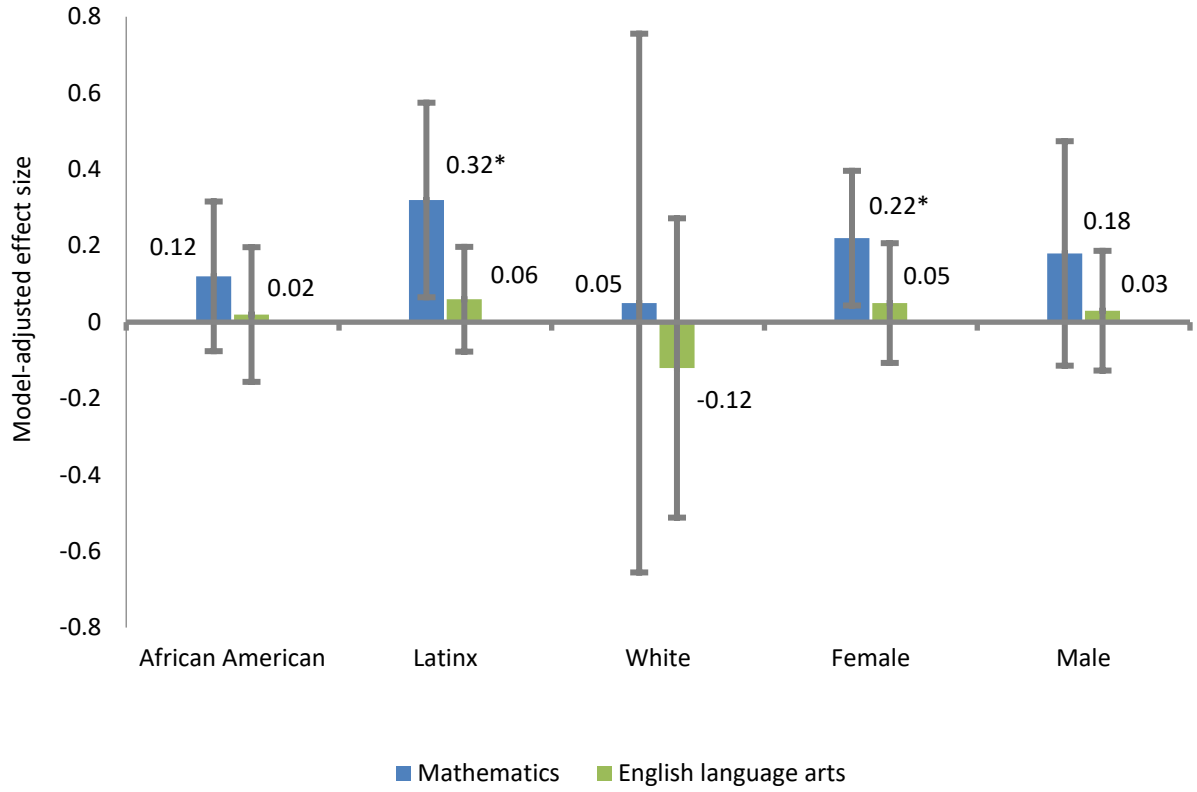
The Emerging Leaders program had positive effects on math achievement for students who were female, Latinx, English learners, or economically disadvantaged.

Several student subgroups experienced positive effects from the Emerging Leaders program in math. The Emerging Leaders program had positive effects on math achievement for female students ($g = 0.22$, $p < .05$), Latinx ($g = 0.32$, $p < .05$), English learners ($g = 0.41$, $p < .01$), and economically disadvantaged students ($g = 0.26$, $p < .05$) (Exhibits 15 and 16). For a student in the middle of the statewide math achievement distribution, these impacts translate into gains of 9 percentile points for female students (moving from the 50th to the 59th percentile), 13 percentile points for Latinx students, 16 percentile points for English learners, and 9 percentile points for economically disadvantaged students. The program had no measurable impact on the math achievement of students who were African American, receiving special education services, male, or high or low performing.

The imprecision in our estimates of the overall impact of the Emerging Leaders program on math achievement may be the result of the variation in the program's impact across student subgroups. Student subgroup impacts for math ranged from $g = -0.02$ ($p > .05$) for special education students to $g = 0.41$ ($p < .01$) for English learners. The size of this range is much larger than it is for the subgroup impacts in ELA. The imprecision of the overall estimate in math was most likely due—at least in part—to the differences in the impact estimates across groups of students.

[15] High-performing students are students in the highest quartile of baseline student achievement. Low-performing students are students in the lowest quartile of baseline student achievement.

Exhibit 15. Estimated Treatment Effects by Student Subgroup

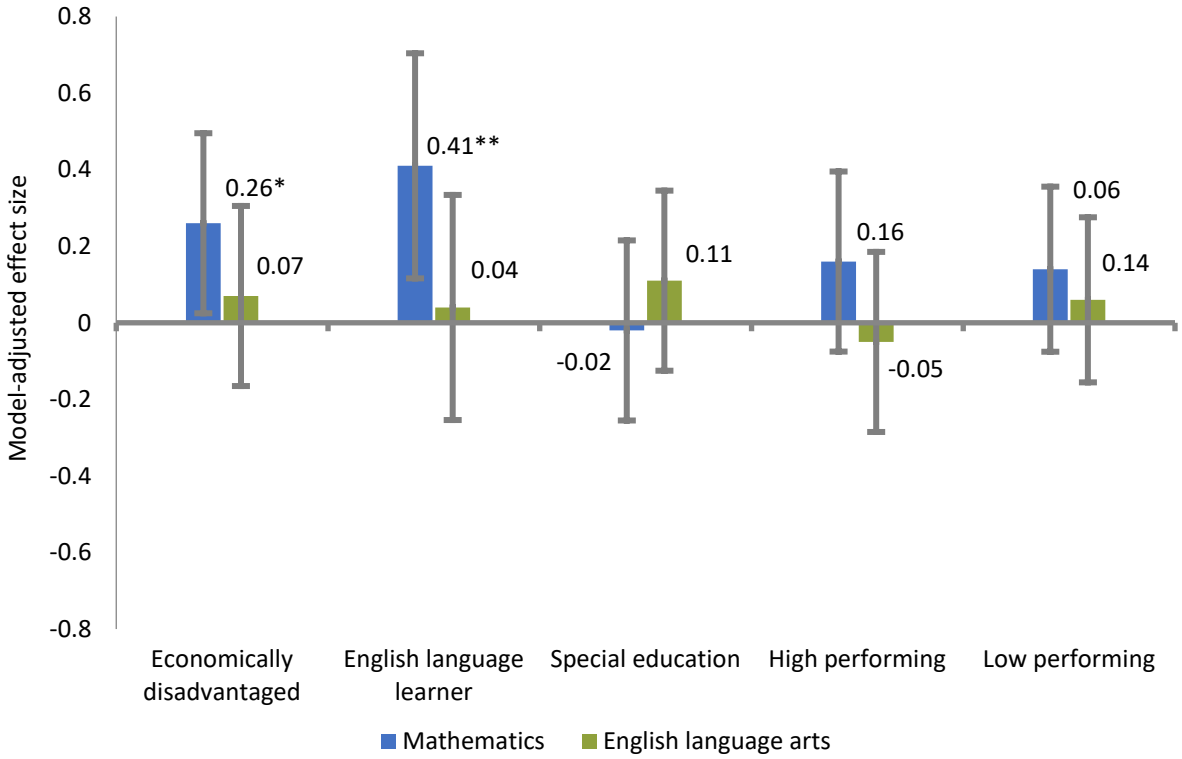


~*p* < .1, **p* < .05, ***p* < .01, ****p* < .001. Grey brackets represent the 95% confidence interval of the estimated effect size.

Source: District data, state assessment outcomes, 2017–18 school year.

Samples: For African American students, 27 instructional teams in math and 3,181 students, 31 instructional teams in ELA and 1,907 students. For Latinx students, 27 instructional teams in math and 2,370 students, 32 instructional teams in ELA and 2,805 students. For White students, 21 instructional teams in math and 497 students, 24 instructional teams in ELA and 387 students. For female students, 28 instructional teams in math and 3,047 students, 32 instructional teams in ELA and 2,663 students. For male students, 28 instructional teams in math and 3,270 students, 33 instructional teams in ELA and 2,783 students.

Exhibit 16. Estimated Treatment Effects by Student Subgroup



~$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$. Grey brackets represent the 95% confidence interval of the estimated effect size.

Source: District data, state assessment outcomes, 2017-18 school year.

Note: "Low performing" is the bottom quartile of the achievement distribution at baseline, and "High performing" is the top quartile.

Samples: For economically disadvantaged students, 28 instructional teams in math and 3,440 students, 33 instructional teams in ELA and 3,711 students. For English learner students, 25 instructional teams in math and 932 students, 28 instructional teams in ELA and 1,295 students. For special education students, 28 instructional teams in math and 584 students, 31 instructional teams in ELA and 467 students. For high-performing students, 27 instructional teams in math and 1,480 students, 32 instructional teams in ELA and 1,493 students. For low-performing students, 28 instructional teams in math and 1,644 students, 33 instructional teams in ELA and 1,248 students.

The Emerging Leaders program had no measurable impact on any student subgroup in English language arts.

We did not find a measurable impact of the Emerging Leaders program on ELA achievement for any of the student subgroups we analyzed (Exhibits 15 and 16). The estimated subgroup effects ranged from $g = -0.12$ ($p > .05$) for White students through $g = 0.11$ ($p > .05$) for special education students and generally fell close to the estimated overall effect of 0.04. We interpret this collection of subgroup analyses to confirm the overall impact finding that the Emerging Leaders program did not have a measurable impact on student achievement in ELA in this study.

Impacts on Student Achievement on District Benchmark Assessments

Two of the three participating sites used the NWEA MAP assessment in ELA and math as their district benchmark assessments: SAISD and SCS. The NWEA MAP data provided us with the opportunity to investigate student achievement across state lines using identical outcome measures, unlike the state assessment data. AISD did not use the NWEA MAP assessment, so this site is not included in the analyses presented here.

The student data for the NWEA MAP analyses contained significant baseline differences between the treatment students and control students which were not present in the state assessment data. We weighted the NWEA MAP data appropriately to establish baseline equivalence before conducting impact analyses. However, it is still possible that unobserved factors may have contributed to any measured impact of the Emerging Leaders program. These limitations did not exist for the main impact analysis based on state assessment data: baseline equivalence was established, and all three participating sites were represented. We therefore provide these results as supplemental findings to our main impact analyses.

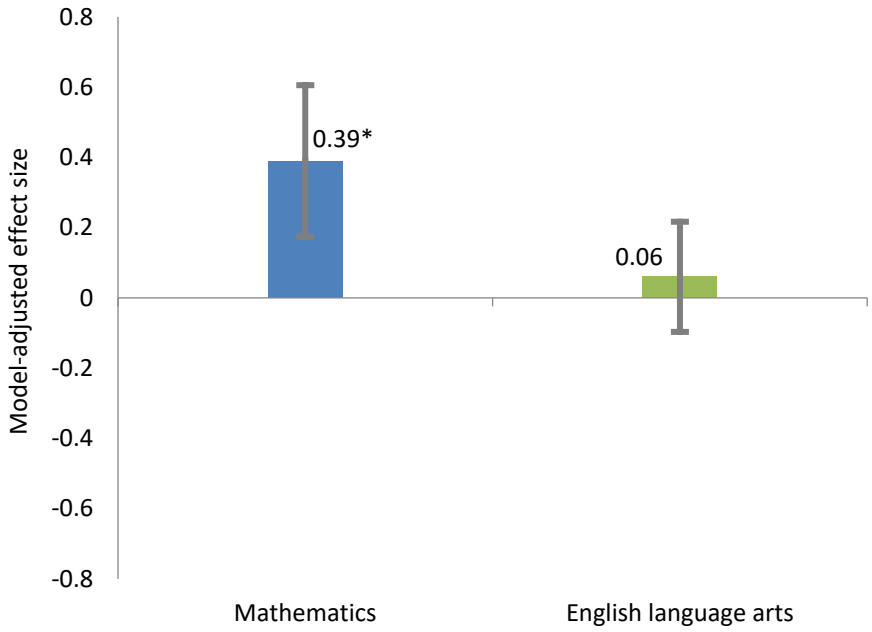
For the two sites with data, the Emerging Leaders program had a statistically significant positive effect in math achievement on district benchmark assessments.

We found a positive impact for the Emerging Leaders program on students' math achievement on the district benchmark assessments ($g = 0.39$, $p < .05$, Exhibit 17) for the two sites with NWEA MAP data. This was similar in direction to the main finding from the state assessments ($g = 0.20$, $p < .10$) and exceeds the threshold for statistical significance.

The Emerging Leaders program had no measurable impact on students in ELA on district benchmark assessments.

We did not find a measurable impact for the Emerging Leaders program on students' ELA achievement for the two participating sites with NWEA MAP data ($g = .06$, $p > .05$, Exhibit 17). These estimates are similar in scale and significance to the overall main impact findings on state assessments ($g = .04$, $p > .05$), indicating that the Emerging Leaders program as implemented is unlikely to have had meaningful impacts on students' ELA achievement.

Exhibit 17. Estimated Treatment Effects on District Benchmark Outcomes



~$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$. Grey brackets represent the 95% confidence interval of the estimated effect size.
Source: NWEA MAP assessment data from SAISD and SCS, 2017–18.
Samples: For ELA district benchmark assessments, 29 instructional teams and 3,370 students; for math district benchmark assessments, 22 instructional teams in ELA and 4,388 students.

Relationship Between Elements of the Emerging Leaders Logic Model and Student Achievement

Emerging Leaders' theory of action posits an indirect, or mediated, relationship between program activities (training, coaching, and assessments) and student outcomes: The program develops instructional leaders who in turn coach teachers who in turn affect students. Our main impact analyses provide evidence as to whether the offer of participation in an Emerging Leaders team impacts students achievement. This model provides the strongest possible test of the program logic model, as it accounts for any differences between the two groups in who would choose to participate in the work if given the chance. However, it does not provide evidence as to the extent to which key facets of the work (e.g., engagement in the teams, Emerging Leaders' success in the assignments) directly relate to student outcomes.

Mediation analyses explore the extent to which a middle factor, or "mediator," affects the relationship between an initial change and an outcome (Baron & Kenney 1986; MacKinnon, Fairchild & Fritz, 2007). We used mediation analysis to explore whether two middle elements of the Emerging Leaders program logic model affected whether students benefited from the program. First, we analyzed whether the "full dosage" of the Emerging Leaders program affected student outcomes. Experiencing the full dosage refers to three measurable activities: (1) Emerging Leaders participants remained employed in the district and, if in Group I, successfully completed the Emerging Leaders program, (2) teachers on instructional teams attended all or most of the instructional team meetings, and (3) students completed a full year of instruction in the classroom led by one of these teachers. If any one of these did not occur, students

would not have had the full opportunity to receive the potential impact of the Emerging Leaders program. We consider this analysis to represent the treatment-on-treated impact of the Emerging Leaders program on student achievement, as these students were best positioned to experience instruction from teachers who were themselves best positioned to fully experience Emerging Leaders data-driven instruction practices led by participants who successfully completed the program.

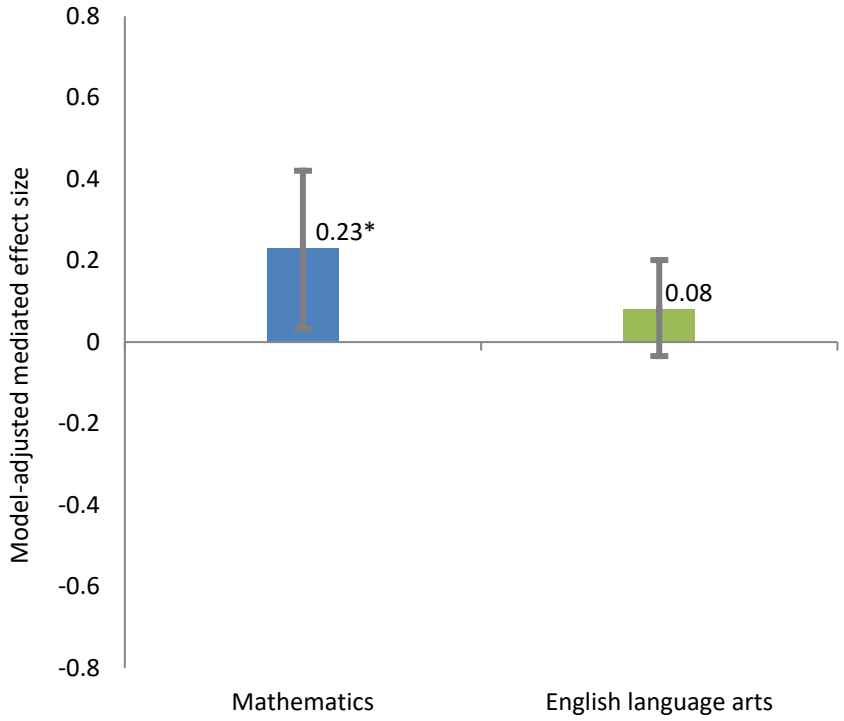
Second, we analyzed whether Emerging Leaders participants' increased DDI leadership knowledge as measured by New Leaders' DDI assessment affected student outcomes. The theory of action of the Emerging Leaders program is that Emerging Leaders participants will lead instructional team meetings that guide members through identifying student misconceptions, collaboratively planning and implementing corrective instruction and then observing and coaching instructional team members to improve their classroom instruction. These instructional leadership activities require Emerging Leaders participants to have significant DDI leadership knowledge, which is measured by the New Leaders DDI assessment. Our prior analysis of the impacts of the Emerging Leaders program on the participants themselves found a positive impact of the program on participants' DDI leadership knowledge.[16] If the Emerging Leaders program theory of action is accurate, we would expect to see higher levels of DDI leadership knowledge cause positive effects in student achievement. In other words, we would say that participants' DDI leadership knowledge mediates the impact of the Emerging Leaders program on students. We explored participants' DDI leadership knowledge through a mediation analysis of whether students of Group I or Group II instructional teams led by participants with greater DDI leadership knowledge outperformed students of Group I or Group II instructional teams led by participants with lesser DDI leadership knowledge.

Emerging Leaders had a positive, statistically significant impact on math achievement students who experienced the full dosage of the Emerging Leaders program.

We found a positive impact for the Emerging Leaders program on the math achievement (g = 0.23, p < 0.05, Exhibit 18) of students who experienced the full dosage of the Emerging Leaders program. An effect of this size is equivalent to 9 percentile points (that is, moving from the 50th to the 59th percentile on the statewide math achievement distribution). Unlike the main finding on math achievement (g = 0.20, p < .10), this mediation estimate exceeds the threshold for statistical significance. This finding indicates that greater engagement with the Emerging Leaders program across all actors (Emerging Leaders participants, instructional team members, classroom students) has a meaningful influence on the program's impact on student math achievement.

[16] In Chapter 3, we analyzed three facets of DDI leadership knowledge separately. For parsimony, in this chapter we have averaged the three scores and standardized the resulting averages (the Technical Appendix provides supporting detail).

Exhibit 18. Estimated Treatment-on-Treated Effect for Students Who Experienced the Full Dosage of the Emerging Leaders Program



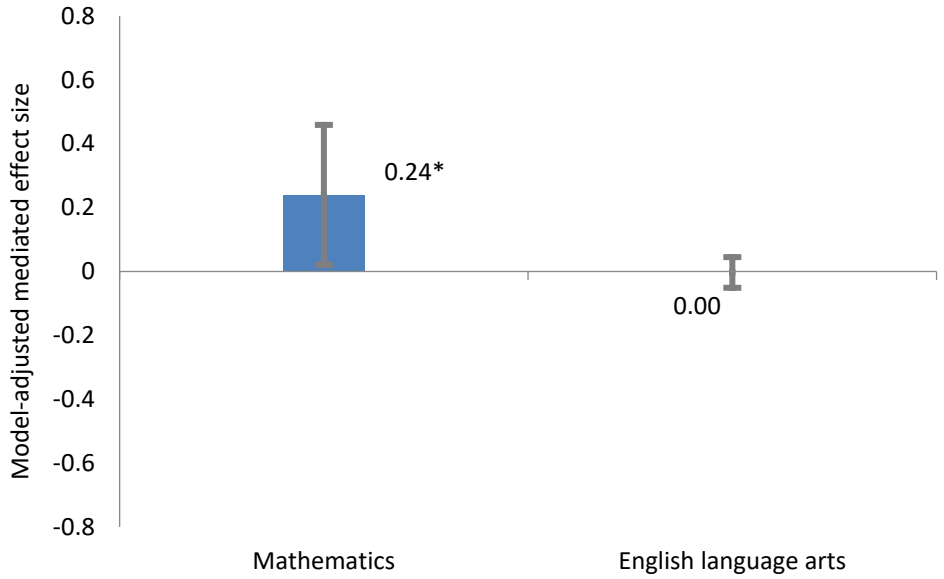
~p < .1, *p < .05, **p < .01, ***p < .001. Grey brackets represent the 95% confidence interval of the estimated effect size.
Source: District data, state assessment outcomes, 2017–18 school year; New Leaders program data, DDI assessment, 2017–18 school year.
Samples: For ELA state assessments, 30 instructional teams and 4,986 students. For math state assessments, 28 instructional teams and 5,875 students. The Technical Appendix provides additional detail on the sample.

Increases in participants' DDI leadership knowledge lead to statistically significant improvements in student math achievement.

Exhibit 19 illustrates the mediated treatment effect of increased Emerging Leaders participants' knowledge on student achievement. These mediated impacts can be interpreted as the effect of scoring 1 standard deviation higher on the DDI assessment on student achievement. As the overall impact of the program on participants' leadership knowledge is approximately 1 standard deviation, the estimated impact roughly translates to an effect size.

We found a positive and statistically significant mediating impact of DDI leadership knowledge on students' math achievement (g = 0.24, p < .05) for Emerging Leaders participants. An effect of this size is equivalent to 9 percentile points (that is, moving from the 50th to the 59th percentile on the statewide math achievement distribution). Put another way, increases in participants' DDI leadership knowledge lead to greater student math achievement.

Exhibit 19. Estimated Mediated Treatment Effect of DDI Leadership Knowledge.



~p < .1, *p < .05, **p < .01, ***p < .001. Grey brackets represent the 95% confidence interval of the estimated effect size.
Source: District data, state assessment outcomes, 2017–18 school year; New Leaders program data, DDI assessment, 2017–18 school year.
Samples: For ELA state assessments, 23 instructional teams and 3,765 students. For math state assessments, 23 instructional teams and 4,805 students. The Technical Appendix provides additional detail on the sample.

Interpreting a mediated effect size required the combination of two estimates: the impact of the program on participants' DDI leadership knowledge and the mediated effect of DDI leadership knowledge on student achievement. The mediated effect size of 0.24 means that each 1 standard deviation increase in DDI score (the measure of DDI leadership knowledge) would result in a corresponding increase of 0.24 effect size in math. Given the estimated effect sizes of .90–.94, we can say that this estimated effect is similar to the main finding on math achievement (g = 0.20, p < .10), although estimated with greater precision such that it achieved statistical significance.

Among students who experienced the full dosage of the Emerging Leaders program, there was no measurable impact on ELA achievement. Increased DDI leadership knowledge as measured by the DDI assessment also had no mediating effect on students' ELA achievement on state assessments.

The Emerging Leaders program had no measurable impact on the ELA achievement (g = .08, p > .05, Exhibit 18) of students who experienced the full dosage of the program. Similarly, we did not find a measurable mediating effect of DDI knowledge on students' ELA achievement (g = .00, p > .05, Exhibit 19). This further reinforces the overall finding, namely that Emerging Leaders did not have a measurable impact on student achievement in ELA.

Summary and Discussion

Our main impact analysis examined the extent to which the Emerging Leaders program had direct effects on student achievement, estimating the impact of enrolling a participant in the Emerging Leaders program

on the math and ELA achievement of students taught by members of the participants' instructional team. This is the most relevant question for district leaders who are considering implementing the Emerging Leaders program to drive impacts on student achievement within the training year.

The Emerging Leaders program had generally positive but inconsistent effects on math achievement. Its overall impact was positive and marginally significant, and the program had positive and statistically significant impacts on student math achievement for several student subgroups—those who were female, Latinx, English learners, or economically disadvantaged. The program also had positive, statistically significant impacts on student math achievement for students who experienced the full dosage of the program and for students within teams lead by Emerging Leaders with greater impacts on DDI leadership knowledge. In contrast, we found no evidence of a programmatic effect on student achievement in ELA, on either the state or district benchmark assessments, for any student subgroup, or evidence of a mediated effect through greater program engagement or success.

Looking beyond the present study, we compared our findings with other evaluation studies of DDI programs. Broadly speaking, education research literature lacks many well-implemented rigorous evaluations of DDI programs. Despite long-standing practitioner interest in DDI (Halverson et al., 2007; AASA, 2002), there is a scant body of causal evidence to examine its effectiveness on student achievement (Hamilton et al., 2009). This study (with Cavalluzzo et al., 2014 and Gleason et. al., 2019) is only the third RCT to estimate the effects of instructional teams working through a structured DDI process. Both prior studies found null impacts of the DDI work on student achievement, making this study particularly relevant for educators looking to use data to drive student achievement gains. While these results suggest that the different patterns in program impacts between math and ELA may have arisen due to the specifics of sampling or implementation in our study, we have also been able to examine differences on the programs' impacts on math- and ELA-focused Emerging Leaders and instructional team members to contextualize these results.

The differences in student achievement do not appear to be driven by differences in the DDI leadership knowledge of the Emerging Leaders leading math-focused instructional teams, as there was no statistically significant difference in impacts on DDI leadership knowledge between these two groups of Emerging Leaders ($p < .05$ for all three measures). Emerging Leaders' DDI leadership knowledge seem to translate into different practices within math- and ELA-focused instructional teams, however, leading to a positive impact on math teachers on two key facets of corrective instruction: reviewing assessment data looking for students' incorrect answers (52% Group I; 24% Group II, $p < .001$) and selecting new instructional strategies for revisiting past content (84%; 65%, $p < .05$). These differences in instructional practices were relatively large and there were no comparable impacts on ELA-focused instructional team practice. ELA-focused instructional teams, in contrast, were more likely to have students review their own assessment data (49%; 32%, $p < .05$) and less likely to have students review and revise their own work (52%; 66%, $p < .05$). These differences in impacts by subject area also emerged during our observations of instructional team meetings: The ELA teams that we observed engaging in corrective instruction did not generalize beyond the specific assessment question or assignment under discussion, whereas math teams were more likely to identify the skill associated with the assessment question and could use the incorrect answer to understand students' misconceptions. Despite the fact that Emerging Leaders participants and teachers reported that the DDI work they had done together in Emerging Leaders teams was more structured, rigorous, and effective than the DDI work they had attempted in the past, it may be that DDI—or at least DDI as enacted by Emerging Leaders—operates differently in math and ELA classrooms.

We next turn to an analysis of the program's sustainability following the training year.

Chapter 5: Sustaining Leadership Gains After the Emerging Leaders Year

- Emerging Leaders aspired to positions with greater leadership opportunities, and nearly half moved to new jobs or schools after completing the program.
- In the year after the program, Emerging Leaders faced challenges in finding opportunities for instructional leadership because of competing demands on their time and because of the leadership opportunities available in their current positions.
- In their jobs, Emerging Leaders still continued to use program tools and strategies that were well aligned with their current responsibilities and the preexisting practice, structures, and policies of their schools and districts.

During the Emerging Leaders year (2017–18), the RCT examined program impacts on Emerging Leaders, instructional team members, and students in alignment with the program's logic model. In this chapter, we move beyond this logic model to better understand how the training executed during the RCT year shaped participants' career steps, their leadership of instructional teams, and districts' leadership pipelines in the year following the program (2018–19). Following the completion of all Emerging Leaders training, were participants' ambitions for career advancement realized? Did they have opportunities in their jobs to lead instructional teams as they were trained to do? Did they use the tools and strategies they learned in the program, and were they also influencing others to use them? To what extent did district administrators feel the Emerging Leaders program was providing them with a robust leadership pipeline, and how did they envision the role of the program in the future?

Data Sources and Methods

- Emerging Leaders follow-up survey of participants in Group I, administered in February/March 2019. All 53 Emerging Leaders still employed in the study sites in winter 2019 received the survey and 43 completed it, a response rate of 81%.
- 44 interviews with Emerging Leadership participants, instructional team members, principals, district leaders, and program staff conducted in March 2019.
- The Technical Appendix provides additional detail on analysis of qualitative and survey data.

Leadership Opportunities Following Emerging Leaders Training

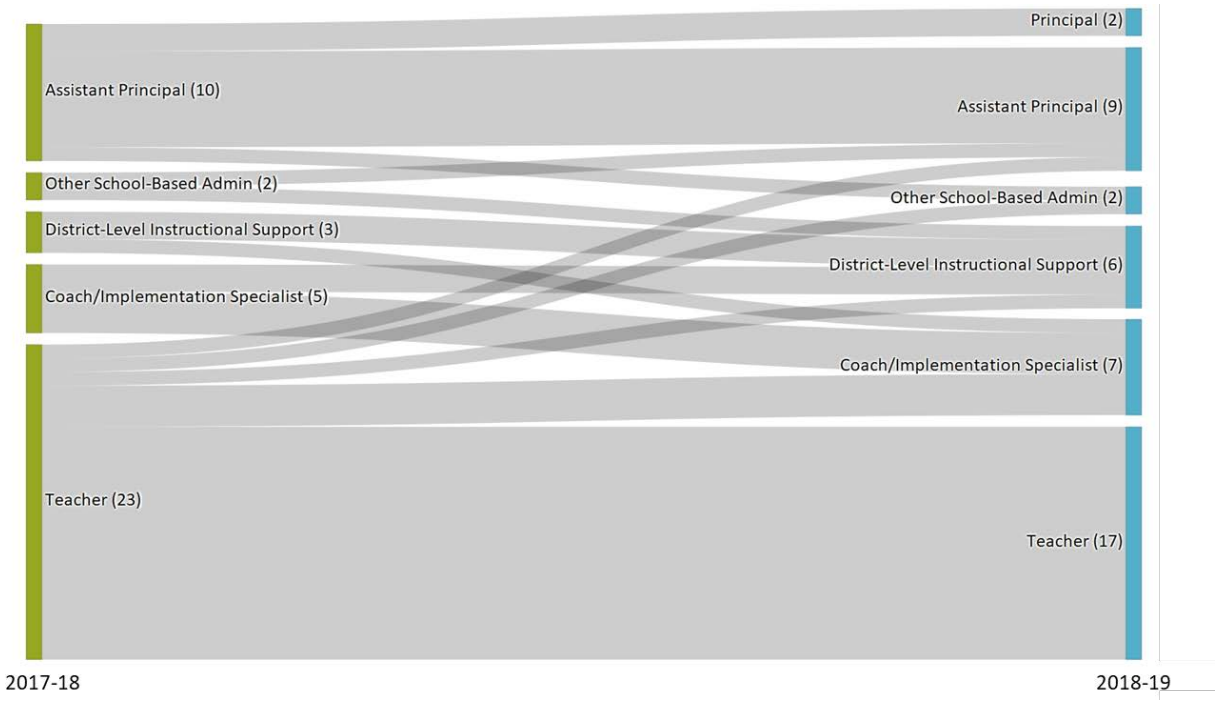
Emerging Leaders participants apply to and complete the program to meet a variety of professional goals, including a desire to advance to more formal professional leadership roles. In this section, we examine Emerging Leaders' professional aspirations and the extent to which they took on new roles following training.

Emerging Leaders aspired to positions with greater leadership opportunities, and nearly half moved to new jobs or schools after completing the program. Some moved into positions with more formal leadership opportunity; many coaches made lateral moves to other coaching or instructional support positions.

In the year following training, almost half the Emerging Leaders (47%) were in a different job, a different school or site, or both, than in the RCT year (Exhibit 20). The subset who changed jobs (33%) typically

moved into roles with greater or different leadership responsibility, such as principal, assistant principal, instructional coach, district-level instructional support staff, or other school administrator. The general pattern across the districts was that new school- or district-level instructional support staff/coaches came from the teaching ranks, and new principals were former assistant principals. Movement patterns among other positions varied, with a significant number of lateral moves from one kind of instructional support role to another. No Emerging Leaders who changed job titles became teachers. A small majority (54%) of Emerging Leaders surveyed remained in the same job and school/site between 2017–18 and 2018–19, including most of those who started out as teachers (*n* = 17 of 23, 74%).

Exhibit 20. Emerging Leaders Who Changed Positions Between 2017–18 and 2018–19



Source: Emerging Leaders follow-up survey, winter 2019.

Some participants described themselves as having greater career ambitions as a result of their Emerging Leaders training.

The Emerging Leaders program raised some participants' expectations that they were ready for and would be able to get a different position afterward. For example, an Emerging Leader who became a principal said, "If I hadn't gone through Emerging Leaders, I think my career would just be assistant principal from here on out." An Emerging Leader who moved from teacher to coach said:

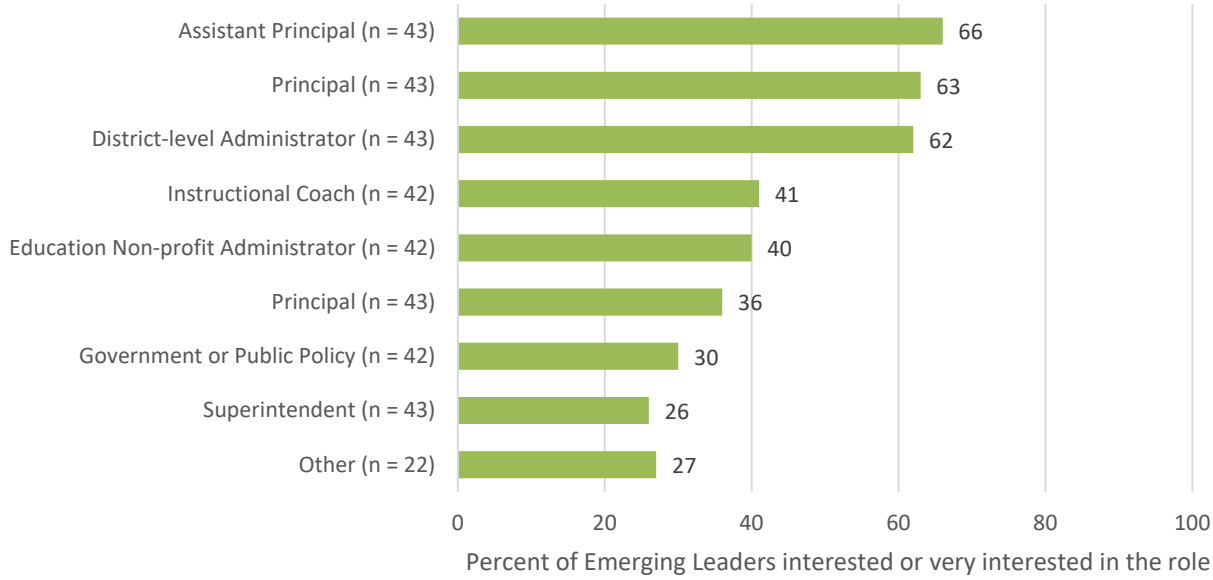
> *I like the leadership opportunity I was given my first year….To be honest, without Emerging Leaders, I wouldn't have even thought about [becoming a coach]….It really gave me the confidence and really showed me my leadership skills that I just didn't really know in the way that I had.*

Emerging Leaders did see themselves as future administrators, if not immediately as principals. When asked about their interest in moving into a variety of roles in the short- or long-term future, Emerging Leaders were most likely to express interest in becoming assistant principals, principals, and district-level administrators (Exhibit 21). Of the 10 assistant principals who completed the program, two became

principals in 2018–19. On the survey, 12% of Emerging Leaders reported that they were "completely ready" to be principals, even though this was not an immediate expectation, as each of the RCT sites required completion of a principal residency program prior to assuming the principalship. The majority (52%) reported that they were "quite ready" (52%), consistent with their aspiration to become assistant principals as a next step (Exhibit 22). Emerging Leaders believed they were most prepared for team and personal leadership and least prepared for the operational leadership aspect of the principal position, which is not a focus of the Emerging Leaders program.

Exhibit 21. Jobs to Which Emerging Leaders Aspired

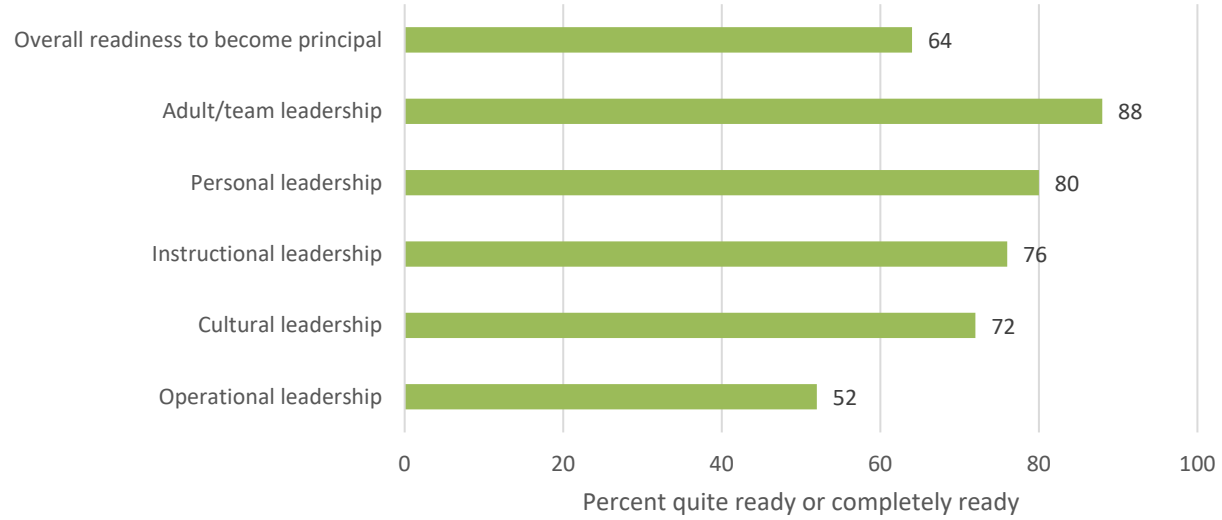
Q: What kinds of leadership roles interest you? (Mark all that apply.)



Source: Emerging Leaders follow-up survey, winter 2019.
Note: Excludes those who responded that they already serve or have served in the role.

Exhibit 22. Emerging Leaders' Perceptions of Readiness for the Principalship

Q: [If not already a principal] To what extent do you feel ready to become a principal?



n = 25.
Source: Emerging Leaders follow-up survey, winter 2019.

Some Emerging Leaders expressed frustration at not being considered for either a principal residency program or an open assistant principal position at the end of the Emerging Leaders year.

Some Emerging Leaders who did not find new positions were disappointed. For example, an Emerging Leader who had been and remained in a teaching position said, "They [the district] say there is nothing available. That's always that, and I've already waited 8 years." One Emerging Leader reported that the encouragement received from the Emerging Leaders program staff contrasted with the lack of encouragement received from the district:

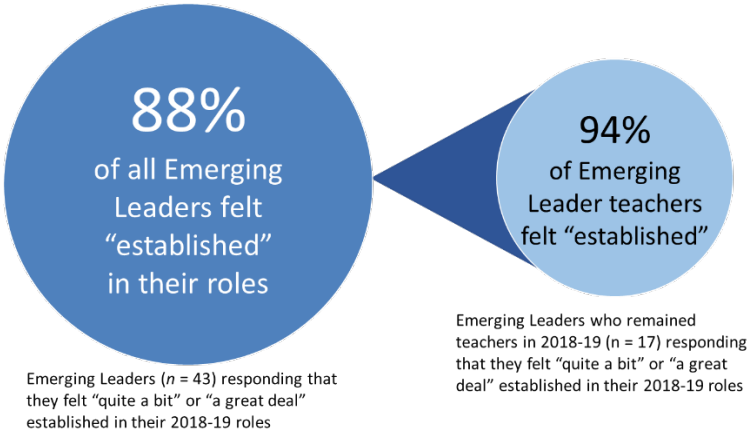
> *My plans were to become a principal. Not to speak ill of the district, but I don't know what they're thinking because [my Emerging Leaders coach] kept saying, "You're ready, you're ready, you're ready." I decided not to apply because I had some indication that it wasn't going to happen.*

In interviews, Emerging Leaders who did not receive hoped-for position changes after the Emerging Leaders year were likely to say that they would look for positions outside the district.

Emerging Leaders experienced a drop in efficacy in the year after their training.

In the year after their training, Emerging Leaders no longer had access to the scaffolding and support of the Emerging Leaders program. In that year, Emerging Leaders generally reported that they were "established" in their roles, particularly those who remained teachers (Exhibit 23). Despite this level of comfort, the 2019 follow-up survey data revealed a statistically significant drop in all areas of leadership efficacy (reporting that they were "extremely effective") between 2017–18 and 2018–19. This was true for Emerging Leaders whether or not they were in formal leadership positions in 2018–19 (Exhibit 24 shows responses from those who reported having leadership responsibilities; response patterns for the full sample were similar).

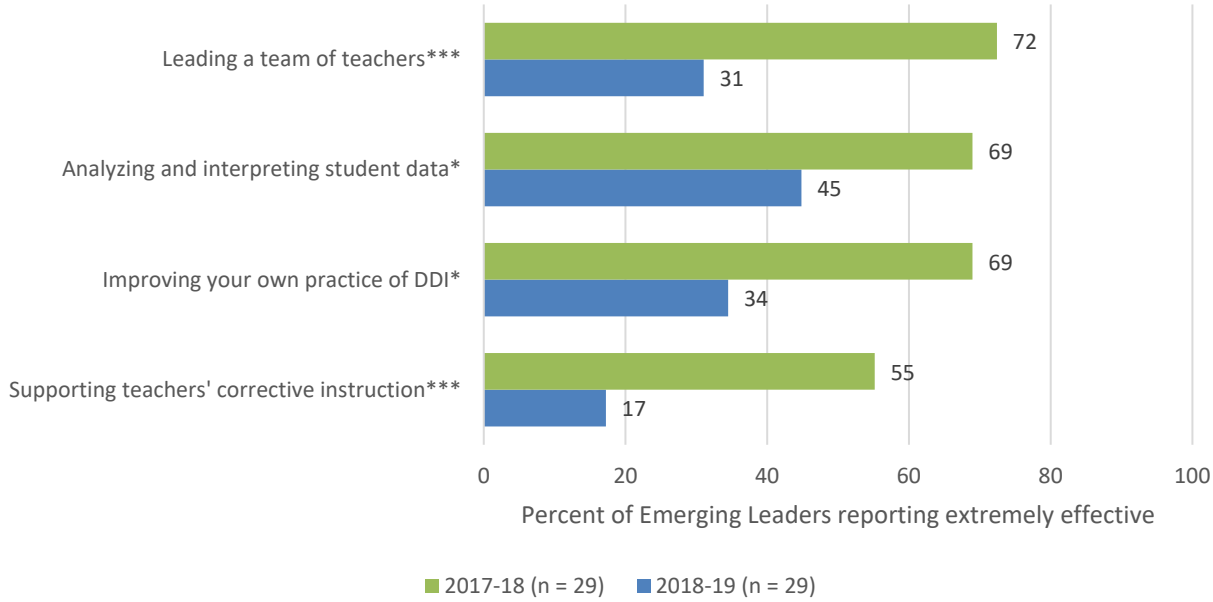
Exhibit 23. Emerging Leaders Reporting They Were "Established" in Their 2018–19 Roles



Source: Emerging Leaders follow-up survey, winter 2019.

Exhibit 24. Emerging Leaders Self-Reported Leadership Efficacy, 2017–18 and 2018–19

Q: This school year, how effective do you feel as a leader in each of the following roles?



~$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$.
Source: Emerging Leaders follow-up survey, winter 2019 and SRI end-of-year leadership survey, spring 2018.
Note: Includes only Emerging Leaders who responded to the survey in both years and who held formal leadership roles in 2018–19.

District representatives and managers' comments supported the importance of efficacy for leaders. In commenting on Emerging Leaders who were not considered ready for the principalship, they most often cited a need for them to develop greater confidence. One manager said, "[The Emerging Leader] gives feedback, but I think it's not quite, it's not natural just yet." Another principal of an Emerging Leader said:

It's the confidence. It's the taking initiative of different tasks. The risk-taking behavior is needed. As a principal, you are faced with multiple issues throughout the day, some of which staff know about, some of which staff don't know about and don't need to know about. They are constantly dealing with issues, and I can't constantly ask my supervisor, "What do you think about this, what do you think about that?" I can't do that. That's something that [the Emerging Leader] needs to improve on in order to be able to be, I think, principal quality.

The lower 2018–19 self-efficacy scores were not significantly correlated with Emerging Leaders' program proficiency scores. Therefore, the Emerging Leaders' self-efficacy was not related to their knowledge of leadership strategies. However, their lower sense of self-efficacy may limit their willingness and ability to apply this DDI leadership knowledge, particularly in situations that did not provide them with formal leadership roles.

Sustainability of Emerging Leaders Work Following the RCT Year

In addition to formal title changes, we expected that Emerging Leaders would continue to look for opportunities to leverage the skills, strategies, and tools they trained on during the RCT year. In this section, we examine how Emerging Leaders perceived opportunities to leverage their training and the extent to which they used particular skills, strategies, and tools when given the chance.

Opportunities to Lead Instructional Teams

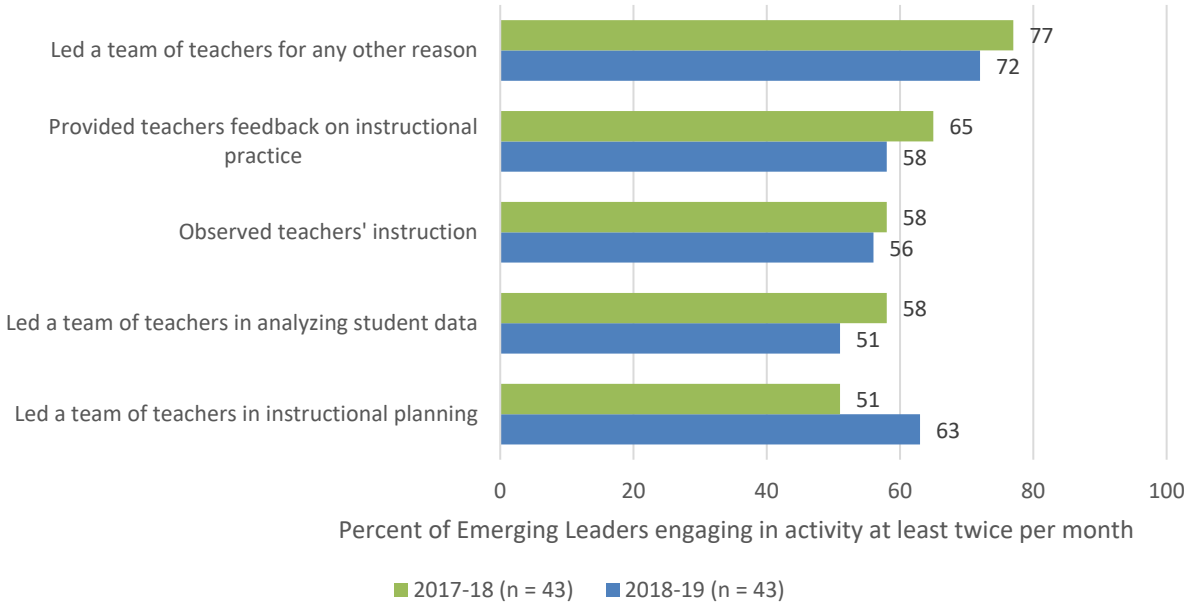
Leading instructional teams gives Emerging Leaders the clearest opportunity to leverage Emerging Leaders program skills, strategies, and tools. We begin with an examination of participants opportunities to lead instructional teams and any facilitators or barriers to this work.

Emerging Leaders described their instructional team leadership opportunities similarly during the RCT year and the year following their training.

In the 2018–19 survey, Emerging Leaders overall reported leading instructional teams about as often as they had in 2017–18 (Exhibit 25; differences between 2017–18 and 2018–19 were not statistically significant).

Exhibit 25. Emerging Leaders' Teacher Leadership Opportunities in 2017–18 and 2018–19

Q: How often did you do each of the following during the 2017–18 or 2018-19 school year? Please consider all of the teachers in your building, not just the teachers on your Emerging Leaders team.



~$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$.
Source: Emerging Leaders follow-up survey, winter 2019 and SRI end-of-year leadership survey, spring 2018.
Note: Includes only those Emerging Leaders who responded to a survey in both years.

About half the Emerging Leaders we interviewed were still leading teams in the year after the program, and about half were providing individual coaching (some were doing both and some neither). However, the distribution of leadership opportunities was uneven across the three districts in the study. In one district, all but one interviewed Emerging Leader was leading a team, for example, but in another district, only one Emerging Leader was consistently doing so.

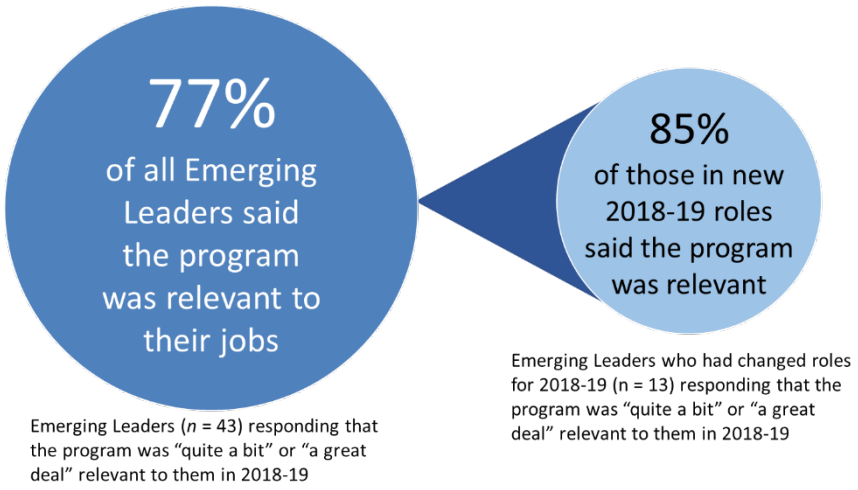
Emerging Leaders saw their training as relevant to their ongoing work and were excited to apply their coaching and team leadership skills.

Emerging Leaders believed that their experiences in the program were relevant to what they were doing in the first post-program year. In the survey, 77% answered that the Emerging Leaders program was "quite a bit" or "a great deal" relevant to their role in the year after the program, and among those who had changed roles for 2018–19—changes that were associated with greater or different leadership responsibility—85% gave that answer (Exhibit 26).

Emerging Leaders particularly appreciated and wanted to apply what they had learned about coaching and working with teams. For example, one instructional coach said:

I'm still using the skills that they taught me and trying to implement [them] with my other grade levels. For me, because I've been a coach, I've always tried to change or help people change individually to help them be successful. Whereas, to me, Emerging Leaders taught me how to work with teams, and to me that impacts more.

Exhibit 26. Emerging Leaders Found the Program Relevant



Emerging Leaders (*n* = 43) responding that the program was "quite a bit" or "a great deal" relevant to them in 2018-19

Emerging Leaders who had changed roles for 2018-19 (n = 13) responding that the program was "quite a bit" or "a great deal" relevant to them in 2018-19

Absent the direct structures and supports of the training (e.g., protected time, job-embedded assignments), Emerging Leaders working as teachers or assistant principals faced more barriers to applying their training than those working in other roles.

Assistant principals struggled to sustain the instructional leadership focus of the Emerging Leaders program following the training year. The difficulty that assistant principal Emerging Leaders had in finding opportunities for instructional leadership without the structures of the Emerging Leaders training is notable because, in the survey, the assistant principal was a role to which the greatest number of Emerging Leaders participants aspired (Exhibit 21). When Emerging Leaders did become assistant principals, they struggled to make time or opportunity for meaningful instructional leadership. In interviews, Emerging Leaders who were serving as assistant principals described how the teachers they were assigned to formally evaluate were not the ones they led and coached in professional learning communities (PLCs. Additionally, they noted that their duties overseeing discipline and operational issues crowded out instructional leadership. For example, one assistant principal Emerging Leader described teacher leadership challenges:

> *I evaluate them, but I am not responsible for their PLCs.…So, I actually work with the eighth-grade science team in their PLC, but I don't supervise them. It's something I've been fighting all year because it makes no sense to me at all. I don't understand and nobody seems to be able to get me to understand how I can support teachers when I'm not actually working with them. That makes no sense to me.*

Another assistant principal Emerging Leader found little time to support instruction:

> *So, as an assistant principal here, one of the bigger things that I do all day is discipline....The part that I really like the most but I don't get to do enough of is just being an instructional leader, going in the classroom and observing, coaching teachers, being in PLCs and helping to support the leaders and the teachers in the PLCs.*

Another Emerging Leader assistant principal struggled to stay focused on instructional leadership in the new position without the structure of the Emerging Leaders program and in the face of serious challenges and daily emergencies that commanded most of the available time and attention:

> *[A]t least last year, in Emerging Leaders, I had my homework…it kind of kept me on track. Here it's kind of like you lose your way and once everything starts coming up and happening....I already had to put two trespass warnings on two parents....We have a behavior unit here and we didn't have it fully staffed for half the year, and so that was taking up a lot of our time and parents were…calling the cops on each other.*

Even in cases where instructional leadership was a stated component of the assistant principal job, the reality was often different, according to one Emerging Leader:

> *I am supposed to be in the classroom, I think it's three times a week, and I'm lucky if I make it once. It's bad. It's really bad. I went to two classes yesterday, which was phenomenal for me. Handling all the discipline, the parents coming up, wanting to talk about whatever happened….I know one of the books that we used in Emerging Leaders had a sample schedule…but it's not at all the reality here on my campus…because we have these radios and they are calling us constantly.*

Like the Emerging Leader assistant principals, Emerging Leaders who remained teachers reported challenges in finding formal or informal opportunities for instructional leadership. While the overall cohort of Emerging Leaders reported similar leadership opportunities in the year following training to those in the RCT year, this finding did not hold for the subsample of Emerging Leaders who remained teachers. Only half (53%) of these Emerging Leaders reported that they were still leading teams in instructional planning twice a month or more often, compared with 91% of these same respondents the year before. In settings that lacked formalized opportunities to serve as a teacher leader, Emerging Leaders did not often report finding informal opportunities either. Of the Emerging Leader teachers we interviewed, only one was continuing to lead an instructional team in the year after the program. Another continued supporting new teachers as a mentor, but this was a position that predated Emerging Leaders program participation. One teacher team member explained that an Emerging Leader's leadership opportunity ended with the program, at the end of the previous year:

> *Last year, [the Emerging Leader] had an assigned Emerging Leaders group that she worked with, it was fifth-grade reading. She worked with them once a week, and she didn't this year because it wasn't part of the requirement because she is finished with the program.*

In interviews, Emerging Leaders who remained teachers seldom mentioned finding informal opportunities for leading their colleagues. However, some were able to take on additional leadership responsibility by volunteering when a need arose. For example, one Emerging Leader teacher recalled:

> *Then the principal's like, "Oh my gosh. We've got to make sure we're keeping up with the data, and there's nobody there to hold them accountable like the dean was…. We need somebody to help these teams that are not being successful." And I was like, "OK, well, I can take them on, and then start trying to implement some of the [Emerging Leaders strategies]."*

Use of Emerging Leaders Skills, Strategies, and Tools

Broadly speaking, Emerging Leaders were using the strategies and tools they learned in the program. Implementation was constrained by available opportunities for instructional leadership.

Teachers learned from Emerging Leaders' coaching.

Qualitative data indicated that teachers who worked with Emerging Leaders largely found their coaching and team leadership helpful. One teacher mentioned "the intentional meetings and really making sure that we are also focusing on that goal and not getting off track, and I think it has helped so much these last 2

years, improving my students' learning." Another teacher described the coaching support as building her confidence:

> *[The Emerging Leader] is not giving me the answers….She says, "What do you think the problem is? What do you think you can do about it?"….getting me to come up with the answer to my own questions.*

Another teacher mentioned how the Emerging Leaders coaching made it easy for her to ask for help:

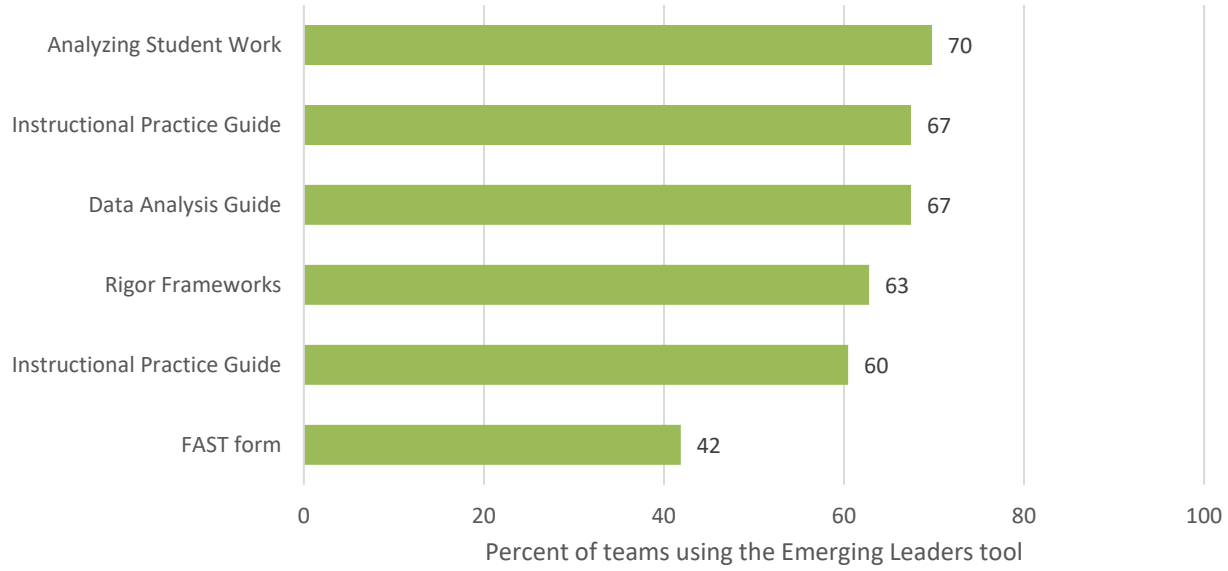
> *[The Emerging Leader] has had my back, too…. She always was there…. She was just so good that I want to be just like her…. She's a very good leader. I feel like I've learned so much.…The actual experience with her and not feeling that loneliness or being so scared to ask questions. She's very open to taking in our feedback.*

Emerging Leaders most commonly used strategies and tools for data analysis and corrective instruction and were less likely to report using strategies and tools for goal-setting and rigor analysis.

Of the program strategies, Emerging Leaders were most likely to report doing data analysis for data-driven instruction and doing and supporting others in corrective instruction. Accordingly, the most commonly reported Emerging Leaders program tools were those related to data analysis and corrective instruction: Analyzing Student Work, the Corrective Instruction Action Plan (CIAP), and the Data Analysis Guide (Exhibit 27).

Exhibit 27. Emerging Leaders Program Tool Use, 2018–19

Q. To the best of your knowledge, are instructional teams in your school currently using any of the following Emerging Leaders tools, or tools similar to Emerging Leaders tools? (Select all that apply)



n = 43.
Source: Emerging Leaders follow-up survey, winter 2019.

Emerging Leaders in all three districts said that data use was a priority. Specific data analysis practices varied between schools. Most Emerging Leaders who discussed supporting DDI used a

calendar and analysis approach that was largely consistent with the Emerging Leaders program approach. In SAISD, for example, Emerging Leaders were reviewing data with their instructional teams in set cycles. Implementing DDI was sometimes challenging because of lack of timely access to data, poorly timed district assessments, fear among teachers about being judged by student data, uneven data literacy among teachers, lack of data in noncore subjects, and lack of alignment between data sources and resources for responsive instruction. Several Emerging Leaders described having to go through the district data office for student data that were available only in limited ways at limited times rather than being able to pull data directly when they wanted it. Another had to work through issues with data availability before it was possible to use data to guide instruction: "We've just had to sit down with student work, and that's something again where I would say it took us a little bit to even get there. We were having technical glitches within district."

In the year after the program, some Emerging Leaders reported working on data analysis with teachers in a more informal way than they had during the Emerging Leaders program. One teacher team member noticed the difference, and said:

> *Last year, we were more looking at test questions and the analysis of the reports of the test together, and this [year] is more, "Hey, how is it going? What's strong this year versus last year?" It's more of a conversation than a sit-down and looking at everything as a whole.*

Emerging Leaders across all schools and districts reported supporting corrective instruction. As most teachers have students who need additional help with difficult concepts, Emerging Leaders found corrective instruction to be one of the most popular program strategies to implement. One Emerging Leader who provided coaching to a group of schools was able to spread corrective instruction to faculty in all of them, according to the instructional support manager:

> *[The Emerging Leader] has truly been a proponent of corrective instruction. She's made sure that the entire team knew about corrective instruction. She's made sure that teachers that she supported knew about it….Immediately when people say "reteaching," [the Emerging Leader] is the one that always says, "No, it's called corrective instruction. Because if you're reteaching, you're teaching it the same way that you taught at first. With corrective instruction, you want to correct whatever that instruction was and provide it in a different way for the students."*

Not all Emerging Leaders doing corrective instruction were using the CIAP tool; in some cases, Emerging Leaders were helping teachers address student misconceptions in a way that was more informal than what they learned in the program.

Emerging Leaders from each participating district reported using SMART goals, but the level of implementation varied between districts. Rigor analysis was seldom mentioned in interviews with Emerging Leaders or their teams.

Emerging Leaders at each district discussed using SMART goals with teachers. Their implementation of SMART goals varied across districts, partly because the three districts took different approaches to SMART goal use. In SCS, for example, SMART goals were often set by the district and sometimes only for tested grades or subjects; in the other districts, SMART goals were set at the school or grade level or left up to teachers and teams to set. Some Emerging Leaders took the fairly generic SMART goals they got from the district and set smaller, more actionable goals within them that instructional teams could track throughout the year. Other Emerging Leaders seemed frustrated by the different approaches to SMART goals that they were expected to implement or were not able to see value in implementing SMART goals. Emerging Leaders may not always have labeled the goals they set with teachers as SMART goals: One teacher team member said she was not familiar with SMART goals but

went on to describe how, with the Emerging Leader, she set goals that that were specific, measurable, actionable, and so forth.

In interviews, few Emerging Leaders reported doing rigor analysis. AISD had its own rigor analysis protocol, which the few Emerging Leaders who did rigor analysis said they used instead of the Emerging Leaders program rigor frameworks. In the other two districts, very few Emerging Leaders said that they used the rigor analysis framework. Therefore, rigor analysis did not appear to be an area of strategic focus for the three districts participating in the study.

Those in leadership roles appreciated strategies and tools for difficult conversations.

Most Emerging Leaders who were in leadership or coaching positions described using what they had learned in the program to engage in difficult conversations. Across schools and districts, they said that this was one of the most valuable aspects of the Emerging Leaders program, some calling it "life changing" and "super necessary." Coaches, for example, reported using the difficult conversation framework in such situations as addressing legal compliance with special education plans and in supporting teachers when their school was designated as in need of improvement. An Emerging Leader liked using the FAST form for feedback meetings because it let teachers know what to expect from the conversation, which reduced teachers' fear and led to more productive dialogue. One Emerging Leader said that these difficult conversations had helped her see past her own frustrations with some teachers to understand that the teachers did want to learn and grow. An Emerging Leader who was new to a leadership position described how the program helped her listen to others' perspectives:

> *Framing difficult conversations is something that I really, truly, value because we're adults and it's not like talking to a child. And so, I have to respect what they say, and in respecting what you say, that means you have to be heard. And then at the same time you have to hear me. I really hadn't had a chance to do before I got into this position. And I think that had I not had Emerging Leaders, I might've gotten just shoved in there like, "Oh, figure it out."*

Emerging Leaders' use of tools for difficult conversations was influenced by the norms of their districts and schools and the affordances of their positions. Alone or with colleagues, some Emerging Leaders were able to instantiate the use of tools from the program when aligned with district needs or established practices. For example, one Emerging Leader said, "We use the FAST form, everybody, because that's something that [another Emerging Leader] and I brought back from Emerging Leaders, and it aligns with what they want to use in the district." Another was able to implement the IPG and get departmental colleagues to use it as well.

The benefits of the support for difficult conversations appeared to be limited to the minority of Emerging Leaders whose roles had formalized opportunities or expectations for these interactions.

Emerging Leaders did not uniformly use the same skills, strategies, and tools across all settings. Rather, the extent to which they sustained their work and the ways in which they implemented these skills, strategies, and tools were shaped by their current roles and school environments.

Emerging Leaders were adapting program strategies and tools to the preexisting traditions and affordances of their settings. If they were still teaching, Emerging Leaders used the tools and strategies in their own instruction and influenced others to use them if they saw an opportunity to do so. In all roles, evidence of informal leadership was not strong: Emerging Leaders grappled with how to assert authority that they were not formally given. And for some schools and districts, Emerging Leaders strategies represented a substantial shift from previous practices. Some Emerging Leaders arrived at new positions ready to implement Emerging Leaders practices but found that teachers and administrators were not completely receptive. One Emerging Leader described using adaptive leadership to tailor her approach:

> *Adaptive pieces…have been huge this year…. At the beginning of last year, I was like, "This is great and we're going to be able to do this, and we're going to rock it, and I think we should work on this, this and this." Pretty quickly, I had to step back and go, "You need to slow down. Just because you're on board and ready to go doesn't mean that everybody else is."*

Emerging Leaders saw their selective implementation of program components as consistent with the guidance they had been given. For example, one Emerging Leader told us that they had been encouraged to take this "do what works" approach in the Emerging Leaders program:

> *[T]here was a huge variety of protocols and we tried different ones all the time, and I really appreciated [Emerging Leaders program staff] telling us, "Take the ones that work for you. Try the other ones, because in practice, it might work better than you think it will." I really appreciated that you could customize it for yourself and for your team.*

District Perceptions of Program Impact and Longevity

Emerging Leaders focus on instructional leadership aligns to district needs for leadership pipeline development.

At the district level, administrators reported that they had wanted the Emerging Leaders program to improve their current and potential leaders' instructional leadership capacity, providing a perspective for local needs that aligns well to the Emerging Leaders logic model. One administrator explained that assistant principals within their district did not spent their time on instructional leadership and so—absent Emerging Leaders—were not ready to implement school leadership that "translated into student achievement increases." An assistant superintendent said that they wanted leaders who could "lead with the desire to change outcomes for students" and "support teachers around what is necessary in order for them to be successful themselves" in improving student learning.

District administrators perceived that overall the Emerging Leaders program had contributed positively to their leadership pipelines by providing additional instructional leadership capacity. One district administrator, who had previously been concerned about principal candidates' instructional leadership capacity, said:

> *[W]e promoted a lot of people out of the [Emerging] Leaders program. Not everybody, because fit is a real important thing for us, and not everybody who goes through the program masters what you're looking for, but it has made a difference. Our pipeline is actually pretty good right now. We've been able to fill a lot of our roles with people from the inside. The preparation of our assistant principals for the principalship is much stronger.*

Districts did not forecast a long-term need for a program with the scope and intensity of Emerging Leaders.

Looking ahead, district leaders were uncertain about the growth of the Emerging Leaders program in their districts. In interviews, leaders in two districts were concerned about the limited capacity of a cohort model. One district mentioned that the Emerging Leaders program may not be renewed in the future because the district had "outgrown" a cohort model, and instead they were working on developing internal capacity for leadership professional development through a badging system. This district leader said, "We really want to go bigger than the principal and the assistant principal. We want to make sure that everybody feels like they're being grown and developed and there's a systematic way in which we do that." Another district intended to play a more active, directive role in working with professional development providers and planned to offer a portfolio of options to employees.

Summary and Discussion

Emerging Leaders were changed by their experience in the program. They had a new set of strategies and tools and the desire to implement them in positions in which they could provide instructional leadership. In the next year, when they were no longer supported by the program, many of the Emerging Leaders worked at new sites, new jobs, or both. Those who changed job titles typically moved into roles with greater or different leadership responsibility. Whether Emerging Leaders remained in their jobs or moved, their intention to implement program strategies and tools encountered the conditions of the real world, in which opportunities for instructional leadership varied. In many cases, Emerging Leaders wanted to focus on instructional leadership more than their roles required or allowed them to. Their sense of effectiveness in their jobs dropped from where it had been during the program year. In the year after the program, Emerging Leaders' implementation of program tools and strategies varied, heavily influenced by the possibilities for instructional leadership they found in their roles. Emerging Leaders adapted to their schools' contexts by choosing the program strategies and tools that seemed most aligned with their responsibilities and least in conflict with existing practices. They implemented these tools and strategies generally with greater informality than during the program year. District leaders had hoped that the Emerging Leaders program would increase district instructional leadership capacity and agreed that it had generally done so.

Chapter 6: Conclusions and Implications

Implementation of New Leaders' Emerging Leaders program in the three districts was executed largely as designed; the program met and often exceeded the thresholds New Leaders set for fidelity of implementation, including the recruitment of qualified program candidates and those candidates' completion of most program activities. Implementation fell short of the New Leaders design in several ways, however. Approximately 40% of participants did not receive the level of personalized, one-on-one coaching specified in the program design. Further, although Emerging Leaders participants successfully completed all required elements of the program, very few (13%) demonstrated leadership proficiency as defined by New Leaders—though these proficiency rates may have been driven by changes in the scoring processes implemented during the RCT year. Nevertheless, the program had an impact on all three facets of participants' DDI leadership knowledge measured by the DDI assessment: using multiple forms of data to drive student achievement, leading a team through a DDI cycle, and understanding of efficacy concepts.

While impacts on this DDI assessment were consistent across different subgroups of Emerging Leaders participants, these impacts translated inconsistently into impacts on instructional team members' practice and student achievement. The Emerging Leaders program had a positive impact on math teachers on two key facets of corrective instruction, shifted ELA teachers away from having students revise their own work and towards looking at their own assessment data, and had several positive impacts on the instructional practice of Emerging Leaders participants who were themselves teachers.

These differential impacts on instructional practice by subject matter appear to have translated into differences in student achievement impacts. Impacts on students' math achievement were promising. While we did not find an overall impact on student math achievement, we did find that the Emerging Leaders program had a statistically significant impact on student achievement in math for female, Latinx, low-income, and English learner students. Further, we found that impacts on math achievement were mediated by impacts on Emerging Leaders participants' DDI leadership knowledge and student exposure to participating teachers and Emerging Leaders. We found no impacts of the Emerging Leaders program on students' ELA achievement, even for students whose teachers participated regularly on Emerging Leaders instructional teams or whose teams were led by program participants who experienced the greatest impacts on DDI leadership knowledge.

Notably, although this study tested impacts on students during the training year, the program's purpose is not solely to improve achievement of students in that year. Rather, New Leaders developed the Emerging Leaders program to improve participants' instructional leadership skills, develop districts' principal pipelines and the distributed leadership capacity of schools and lay the groundwork for improved student achievement in the medium- and long-term. We were assessing a more proximal outcome by measuring the programs' impacts on students in the immediate short-term, as Emerging Leaders participants were training on these new skills via job-embedded performance tasks.

When examining the program's sustainability following the training year, we found that program graduates aspired to positions with greater leadership opportunities, and nearly half moved to new jobs or schools after completing the program. In the year after the program, Emerging Leaders faced challenges in finding opportunities for instructional leadership but continued to use program tools and strategies that were well aligned with their current responsibilities.

This evaluation may provide valuable lessons learned for districts, schools, and programs engaging in data-driven instruction or developing a leadership pipeline, in addition to lessons specific to the Emerging Leaders program. We discuss implications for each of these in turn.

Implications for Data-Driven Instruction

Instructional teams engaged in DDI work benefit from trained leadership to guide them, protected time for teams to engage in this work, and strong tools to provide structure to their practice.

Both Emerging Leaders participants and teachers reported the DDI work they had done together in Emerging Leaders teams was more structured, rigorous, and effective than the DDI work they had attempted in the past. Instructional teams cited the structure of the DDI cycles, regular meetings, and related activities as protecting the time required for engaging in the collaborative work and creating accountability for changing their practices. These perspectives put into context our observational findings, where we noted that while instructional teams engaged in regular DDI cycles and used the Emerging Leaders program tools and protocols to frame their discussions, they sometimes failed to move beyond initial steps in reviewing and analyzing student work. Further, these instructional teams were intended to drive toward the execution of SMART goals for student achievement; in many cases, however, we were unable to assess how ambitious team SMART goals actually were. Together, these improvements suggests that the complexities of DDI work benefit from trained leadership, protected time, and strong tools to provide structure.

Despite Emerging Leaders' similar effects on the DDI leadership knowledge of participants leading math- and ELA-focused teams, these leaders drove different changes in both their instructional teams' instructional planning practice and student achievement in the instructional teams' classrooms, suggesting that ELA-focused instructional teams may require different or additional supports to improve student achievement.

Using the New Leaders corrective instruction framework in the study districts and classrooms, we found that DDI may operate very differently in math and ELA. Members of math-focused instructional teams displayed substantially more positive impacts on two key facets of corrective instruction: examining incorrect answers and selecting new instructional strategies for revisiting past content. Our observations of instructional team practice supported these quantitative findings: ELA teams that discussed specific assessment items focused on the specifics of the test question, whereas math teams saw the question as measuring a skill with specific steps that could be broken down to more easily understand students' misconceptions. Correspondingly, these differential approaches to corrective instruction led to different impacts on student achievement.

Further, we understand that even in math instruction, DDI may operate differently for different subgroups of students. In some cases, this differential impact may work to narrow the achievement gap for some student historically underserved by math instruction: girls, Latinx students, and English learners. However, special education students, another historically underserved group, did not benefit from the program. Not only were the impacts for these students not statistically significant, but the point estimates also suggested near-zero, or even negative, impacts.

Implications for Leadership Development

The Emerging Leaders program may provide a model for scaffolded leadership development that begins before assuming the principalship and operates via structured job-embedded

assignments. Given that the structure, schedule, and responsibilities of different professional roles (e.g., teachers, coaches, and assistant principals) provide different affordances for instructional leadership training and development, any such training should be aligned to the needs of professionals' current responsibilities.

Prior studies of New Leaders' Aspiring Principals Program provided some of the most rigorous causal estimates of leadership programs; these impacts were not estimated until the third year of implementation (Gates et al., 2019; Gates et al., 2014). These prior studies were centered on principal leadership and included all facets of leading a school. In contrast, the Emerging Leaders program trained leaders on a smaller subset of those skills—primarily instructional leadership—before they took on other new responsibilities. In doing so, Emerging Leaders participants drove some positive impacts on student achievement during even in this initial training year. This kind of smaller-scale training before assuming the principalship may provide candidates with a scaffolded opportunity to develop skills over time. A leadership development pipeline could thus be assembled so that leaders are positioned to achieve positive student impacts from the onset of assuming the principalship.

This kind of pipeline would need to be assembled with an eye toward the affordances of different roles. The job-embedded nature of Emerging Leaders training appeared to lend itself most naturally to roles with available and flexible time—specifically, assistant principals and coaches. At the same time, assistant principals who completed the program struggled to use their newly developed instructional leadership skills without the additional structures and support the program provided. Providing assistant principals with job-embedded assignments to develop operational leadership may provide training better aligned with the role, at least in the districts studied.

Finally, we found that the clarity of goals among the staff designing and executing the leadership development, district staff, principals, and participants would help align expectations. In this study, program expectations and outcomes were not consistently clear to district staff or participants. The misalignment of program goals and individuals' expectations led to confusion or disappointment on the part of some participants, their principals, and even the district leadership.

Implications for New Leaders Program Development and the Emerging Leaders Program in Particular

New Leaders is currently developing instructional leadership programs tailored to individual district needs in addition to offering the Emerging Leaders program as originally designed. Given that this would have been our primary recommendation, we instead suggest elements of district context for leadership development that we consider crucial to tailoring these Emerging Leaders programs and support.

Consider breaking New Leaders' job-embedded training materials into a multiyear, microcredentialed badging system. Such a system could provide flexibility in leadership development over multiple years and before assuming the principalship. Further, the badging system could provide a clear indication to a leader's supervisors of the leader's skill set.

We discussed the promise of developing leaders throughout their careers and before they assume the principalship, the affordances of different roles to support the training and development of key leadership skills, and the need to communicate clearly throughout the system about the expectations regarding training. A microcredentialed badging system could help ease each of these barriers. Further, it would help districts track and manage their leadership pipelines by providing them data on where leaders in the district are positioned relative to assuming a principalship. New Leaders already has an extensive set of job-embedded training materials that cover other aspects of leadership (e.g., operational leadership,

cultural leadership) from the Aspiring Principals program. These preexisting materials may provide a strong basis for a series of smaller stand-alone microcredentials.

Emerging Leaders participants would benefit from more explicit coaching to help them adapt the training, skills, and tools to their own local contexts and roles.

Throughout this study, we saw several instances of New Leaders participants struggling to adapt the training, skills, and tools to their own local contexts and roles. Some of these struggles may have been due to structural issues that cannot reasonably be overcome (e.g., a lack of common planning time with their instructional team members). But we also saw instances where participants may, with more training, have found ways to align the tools and materials with their own contexts (or adapted the tools to their newly developed skill set). Some of these struggles may have been from a lack of coaching, which is the primary means by which Emerging Leaders participants learn to adapt and apply the program's skills and tools to their daily work and their specific school contexts; it is also the means by which they receive the feedback critical for development of personal leadership skills (that is, the ability to self-reflect to continuously improve performance) (Valdez, Broin, and Carroll, 2015). We encourage New Leaders to revisit the resources dedicated to one-on-one coaching and ensure that coaches offer explicit guidance on adapting Emerging Leaders program strategies and tools to local contexts in ways that are faithful to the program. Coaching may also help prepare Emerging Leaders find opportunities for informal teacher leadership regardless of their positions.

***

We caution that while this was a rigorously conducted RCT, it has limitations: The generalizability of all findings and the small sample number of clusters in the student outcomes models are particularly worth noting. Thus, we caution that these results provide answers in one specific setting and context and with the particular populations and implementation used, but they should not be taken as the definitive work on these issues. We encourage future researchers to build on this work, by not examining just the impacts of leadership and DDI programs, but also using the variation seen in this study's results to guide design and data collection to determine why results vary between subjects and types of students.

References

Allison, P.D. (2001). *Missing Data.* Thousand Oaks, CA: Sage Publications.

American Association of School Administrators. (2002). Using data to improve schools: What's working. Arlington, VA. Retrieved from https://eric.ed.gov/?id=ED469227

Boyce, J., & Bowers, A. J. (2018). Different levels of leadership for learning: Investigating differences between teachers individually and collectively using multilevel factor analysis of the 2011–2012 Schools and Staffing Survey. *International Journal of Leadership in Education, 21*(2), 197–225.

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association, 90*(430), 431–442.

American Association of School Administrators. (2002). *Using data to improve schools: What's working.* Arlington, VA. Retrieved from https://eric.ed.gov/?id=ED469227

Baird, Matthew D., & Pane, John F. (2019) Translating standardized effects of education programs into more interpretable metrics. *Educational Researcher, 48*(4), 217–228.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173.

Bellert, A. (2015). Effective re-teaching. *Australian Journal of Learning Difficulties*, *20*(2), 163–183.

Bellibas, M. S., Liu, Y., & Liu, Y. (2017). Multilevel analysis of the relationship between principals' perceived practices of instructional leadership and teachers' self-efficacy perceptions. *Journal of Educational Administration, 55*(1), 49-69.

Bitterman, A., Gray, L., and Goldring, R. (2013). Table 6: Number and percentage of schools that had staff with specialist or coaching assignments. *Characteristics of Public and Private Elementary and Secondary Schools in the United States: Results From the 2011–12 Schools and Staffing Survey (NCES 2013–312)*. U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved July 13, 2019 from http://nces.ed.gov/pubsearch.

Cavalluzzo, L., Geraghty, T. M., Steele, J. L., Holian, L., Jenkins, F., Alexander, J. M., & Yamasaki, K. Y. (2014). *Using data to inform decisions: How teachers use data to inform practice and improve student performance in mathematics. Results from a randomized experiment of program efficacy*. Arlington, VA: CNA Corporation.

Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, *33*(3), 378–398.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20(*1), 37–46.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation, 14*(20), 1-11.

Dunn, E. C., Masyn, K. E., Jones, S. M., Subramanian, S. V., & Koenen, K. C. (2015). Measuring psychosocial environments using individual responses: An application of multilevel factor analysis to examining students in schools. *Prevention Science, 16(*5), 718-733.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378.

Gallagher, H. A., Arshan, N., & Woodworth, K. (2017). Impact of the National Writing Project's college-ready writers program in high-need rural districts. *Journal of Research on Educational Effectiveness*, *10*(3), 570–595.

Gallagher, H. A., Woodworth, K. R., Wang, H., Bland, J. A., Bosetti, K. R., Cassidy, L., & Shields, P. M. (2012). *National evaluation of Writing Project school partnerships: Final report*. Menlo Park, CA: SRI International.

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

Gates, S. M., Hamilton, L. S., Martorell, P., Burkhauser, S., Heaton, P., Pierson, A., Baird, M. D., Vuollo, M., Li, J. J., Lavery, D. C., Harvey, M., & Gu, K. (2014). *Preparing Principals to Raise Student Achievement: Implementation and Effects of the New Leaders Program in Ten Districts*. Santa Monica, CA: RAND Corporation.

Gennetian, L. A., Morris, P. A., Bos, J. M., & Bloom, H. S. (2005). Constructing Instrumental Variables from Experimental Data to Explore How Treatments Produce Effects. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (p. 75–114). Russell Sage Foundation.

Gleason, Philip, Sarah Crissey, Greg Chojnacki, Marykate Zukiewicz, Tim Silva, Sarah Costelloe, & Fran O'Reilly. (2019*). Evaluation of Support for Using Student Data to Inform Teachers' Instruction* (NCEE 2019-4008). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goff, P., & Hyun, S. (2014). *Data-driven instruction: An analysis of instrument validity*. Madison, WI: University of Wisconsin-Madison.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6(*4), 430.

Guskey, T. R. (2007). The rest of the story. *Educational Leadership*, *65*(4), 28–35.

Guskey, T. (2016). How classroom assessments improve learning. In M. Scherer (Ed.), *On formative assessment: Readings from educational leadership (pp. 3-13)*. Alexandria, VA: Association for Supervision & Curriculum Development/.

Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2007). The new instructional leadership: Creating data-driven instructional systems in school. *Journal of School Leadership, 17*(2), 159–194.

Hamilton, L., Halverson, R., Jackson, S. S., Mandinach, E., Supovitz, J. A., & Wayman, J. C. (2009). *Using student achievement data to support instructional decision making. IES practice* guide (NCEE 2009-4067). Washington, D.C,: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Klassen, R. M., Bong, M., Usher, E. L., Chong, W. H., Huan, V. S., Wong, I. Y. F., & Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology, 34(*1), 67-76.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics,* 363–374.

Light, R. J. (1971). Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychological Bulletin, 76*(5), 365.

Marzano, R. J. (2010). Reviving reteaching. *Educational Leadership*, *68*(2), 82–83.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*, 593–614.

May, H., Perez-Johnson, I., Haimson, J., Sattar, S., & Gleason, P. (2009). *Using state tests in education experiments: A discussion of the issues* (NCEE 2009-013). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

New Leaders. (2016a). *Emerging Leaders FY 17 Cycle 1 workbook: Readings, resources, and materials*. New York, NY: New Leaders.

New Leaders. (2016b). *Emerging Leaders FY 17 Cycle 2 workbook: Readings, resources, and materials*. New York NY: New Leaders.

Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research, 151*(1), 53–79.

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). What to Do when Data Are Missing in Group Randomized Controlled Trials. NCEE 2009-0049. National Center for Education Evaluation and Regional Assistance.

Roth, P. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology, 47*(3), 537-560. doi:10.1111/j.1744-6570.1994.tb01736.x

Rouse, C. E., & Krueger, A. B. (2004). Putting computerized instruction to the test: A randomized evaluation of a “scientifically based” reading program. *Economics of Education Review, 23*(4), 323–338. doi: http://dx.doi.org/10.1016/j.econedurev.2003.10.005

Rowan, B., & Correnti, R. (2009). Studying reading instruction with teacher logs: Lessons from the study of instructional improvement. *Educational Researcher*, *38*(2), 120–131.

Rowan, B., Jacob, R., & Correnti, R. (2009). Using instructional logs to identify quality in educational settings. *New Directions for Youth Development*, *2009*(121), 13–31.

Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain, A. (2013). Effects of a data-driven district reform model on state assessment outcomes. *American Educational Research Journal*, *50*(2), 371–396.

Valdez, M., Broin, A., & Carroll, K. (2015.) Untapped: Transforming teacher leadership to help students succeed. New York, NY: New Leaders.

Technical Appendix

In 2017–18, funded by a Supporting Effective Educator Development (SEED) grant from the U.S. Department of Education, New Leaders undertook a rigorous evaluation of the impact of the Emerging Leaders program in three sites: Arlington Independent School District (AISD) in Texas, San Antonio Independent School District (SAISD) in Texas, and Shelby County Schools (SCS) in Tennessee. New Leaders' research partner, SRI Education, designed and conducted a randomized control trial (RCT) to assess four things: (1) the extent to which the Emerging Leaders program was delivered with fidelity to the New Leaders design in the RCT sites, (2) the program's impact on participants' DDI leadership knowledge and skills and teachers' instructional practice (3) the program's impact on student achievement, and (4) the extent to which the effects of the program were sustained the year following the program when participants were no longer directly supported. This technical appendix provides supporting detail for the analysis described in the report *Emerging Leaders Program: Impacts on Students, Teachers, and Leaders in Three Sites*.

Districts

Exhibit A-1 shows school-level student demographics for study schools in the three RCT sites.

Exhibit A-1. Student Demographic Characteristics in Study Schools

	All Three Districts	AISD	SAISD	SCS
% Free and reduced-price lunch	81.3	69.7	90.2	87.2
(*SD*)	(19.5)	(23.4)	(8.7)	(14.5)
% English learner	20.5	30.3	22.0	6.3
(*SD*)	(19.2)	(24.1)	(9.7)	(7.8)
% Special education	9.7	8.2	10.2	11.3
(*SD*)	(3.7)	(2.9)	(3.1)	(4.6)
% African American	35.2	20.4	7.6	83.4
(*SD*)	(34.6)	(11.3)	(12.1)	(15.6)
% Latinx	50.7	50.5	89.1	10.7
(*SD*)	(35.4)	(23.9)	(12.7)	(11.1)
% White	10.8	21.1	2.4	6.2
(*SD*)	(16.1)	(20.9)	(2.8)	(8.4)
% Asian	2.7	5.3	0.4	1.6
(*SD*)	(4.3)	(5.6)	(0.9)	(2.4)
% Native American	0.4	0.4	0.1	0.7
(*SD*)	(0.4)	(0.2)	(0.1)	(0.4)
School *n*	**77**	**30**	**24**	**23**

Note. One school in AISD did not have data on school-level characteristics; therefore, the descriptive statistics for AISD do not include that school.

Randomization and Baseline Samples: Emerging Leaders Participants and Instructional Team Members

In preparation for RCT launch in 2017–18, New Leaders staff identified 112 Emerging Leaders candidates who successfully completed the application process for the program and met its selection criteria in AISD, SAISD, and SCS. Before randomization, the research team collected names of instructional team members who had already been identified, as well as plans for instructional teams in which team members had not yet been identified, including focal grades and subjects. The research team then split the sample of Emerging Leaders plus their instructional teams into randomization blocks by district, focal subject, focal grade, and, in selected cases, school improvement status and feeder pattern. We randomized half of each block into treatment (Group I) and half into control (Group II, who received no Emerging Leaders training in 2017–18 and instead received it in 2018–19).

The blocked randomization process was conducted to achieve equal numbers of Emerging Leaders participants within ELA and math because an even division of the number of candidates maximizes statistical power (provides the smallest possible minimum detectable effect size), although small imbalances in numbers (as seen in some of the tables) have minimal impact on statistical power. Slightly uneven numbers of instructional team members between the two groups arose by chance in some instances and are not believed to have impacted the statistical analysis.

The research team communicated these randomization decisions in summer 2017, enabling the local New Leaders staff to begin training Group I before the start of the 2017–18 school year. In fall 2017, we confirmed the membership of each instructional team, including the names of instructional team members identified after randomization and those who had shifted over the summer (due to, e.g., staffing changes), collected baseline instructional log data from both groups, and documented shifts in the sample between summer and fall 2017.

Early Joiner Samples of Instructional Team Members and Students

At randomization in summer 2017, the sample comprised 112 Emerging Leaders candidates and their instructional teams: 58 in Group I and 54 in Group II (Exhibit A-2). Between randomization and baseline data collection in fall 2017, the beginning of the RCT school year, the following changes occurred (Exhibit A-2):

- 6 Emerging Leaders (2 from Group I and 4 from Group II) withdrew from the study.
- 111 instructional team members (63 from Group I and 48 from Group II) left the study, an attrition rate of 32% and differential attrition of 8% (i.e., 36% in Group I compared with 28% in Group II).

To identify students enrolled in an instructional team members' classroom, we collected roster data from each district for the teachers in the early joiner sample. We requested districts provide us with student enrollment in teachers' classrooms at the date schools reported formal enrollment to the state. In the two Texas districts, we used the date of the state accountability count (the last Friday in October 2017). In Shelby County, we used the "snapshot" date (the 40th day of the 2017–18 school year). Students were included in the assigned sample if they were in a tested grade and subject (Chapter 4 has more detail on tested grades and subjects). Students were analyzed according to their assigned condition as of the formal enrollment date (an intent-to-treat framework).

Attrition Calculations

In keeping with What Works Clearinghouse (WWC) standards, the research team provides attrition for each outcome, which varied based on response rates to different data collection instruments. Attrition

tables for each outcome measure are provided in the sections of this technical appendix along with analyses of those outcomes. In these tables, we use the randomized sample of Emerging Leaders (58 from Group I and 54 from Group II). Consistent with our use of the early joiner teacher sample, we also use this sample to calculate attrition (183 from Group I and 167 from Group II).

Exhibit A-2. Samples at Randomization and Fall 2017 Baseline Movement

	Emerging Leaders Participants			Instructional Team Members		
	Group I	Group II	Total	Group I	Group II	Total
Number at randomization	58	54	112	180	170	350
Left study from randomization to baseline (%)	3	7	5	36	28	32
Stayer (%)	97	93	95	64	72	68
Number of early joiners	0	0	0	68	44	112
Baseline sample size in fall 2017	56	50	106	183	167	350

Note. Early joiners are subjects who entered the sample after randomization but early in the cycle of Emerging Leaders program implementation.

No new Emerging Leaders participants were added to the participant sample after randomization. The sample of instructional team members was limited to Emerging Leaders teams as defined at the time of baseline data collection in fall 2017. In WWC terminology, such a sample is said to include “early joiners,” study subjects who entered the sample after randomization but early in the cycle of program implementation.

Much of the turnover in instructional teams for Group I was due to changes in job assignments, as program participants changed schools, changed roles, or adopted schedules that did accommodate the originally intended teams. While this could potentially introduce bias in the sample, the researchers collected data on instructional teams’ school assignments and schedules from the Emerging Leaders program at the end of the year.

Exhibit A-3 presents the number of Emerging Leaders participants and instructional team members in the baseline sample by subject in fall 2017. Local New Leaders regional staff members coached participants and their principals to designate an ELA (English language arts) or math focus for their instructional team—only 7 of the 106 remaining Emerging Leaders participants had an instructional team focused outside these two subjects. Emerging Leaders participants working with ELA instructional teams were more likely to be in K-5 elementary grades (42) than other grades (19) whereas participants working with math teams were less likely to be in K-5 (13) than other grades (25).

Exhibit A-3. Fall 2017 Baseline Sample, by Subject and Grade-Level Focus

	Emerging Leaders Participants			Instructional Team Members		
	Group I	Group II	Total	Group I	Group II	Total
ELA						
Primary (K–2)	9	9	18	35	25	60
Upper elementary (3–5)	13	11	24	39	27	66
Middle school (6–8)	5	4	9	17	21	38
High school (9–12)	4	6	10	12	22	34
Total	**31**	**30**	**61**	**103**	**95**	**198**
Math						
Primary (K–2)	1	1	2	4	2	6
Upper elementary (3–5)	7	4	11	17	14	31
Middle school (6–8)	5	8	13	21	26	47
High school (9–12)	9	3	12	27	15	42
Total	**22**	**16**	**38**	**69**	**57**	**126**
Other Subjects						
Primary (K–2)	-	-	-	-	-	-
Upper elementary (3–5)	-	1	1	-	3	3
Middle school (6–8)	1	1	2	3	5	8
High school (9–12)	2	2	4	8	7	15
Total	**3**	**4**	**7**	**11**	**15**	**26**
Grand Total	**56**	**50**	**106**	**183**	**167**	**350**

Emerging Leaders Descriptive Statistics at Baseline

The baseline descriptive statistics on Emerging Leaders participants in Exhibits A4–A-8 demonstrate the similarity of Group I and Group II at randomization. The sample sizes in these tables vary somewhat because of a small number of missing cases or Emerging Leaders participants who declined to allow the data to be used for research purposes. Emerging Leaders participants in Groups I and II were similar in gender, race/ethnicity, years of experience, and performance on the Emerging Leaders application criteria. Group I Emerging Leaders participants were more likely than those in Group II to have been teaching (Exhibit A-4).

Exhibit A-4. Emerging Leaders Demographics, Fall 2017 Baseline Sample

	Group I	Group II	Total
% Female	79	81	80
% Non-white	75	69	72
Emerging Leaders *n*	*53*	*48*	*101*

Source: New Leaders application data. Total *n* represents the number of Emerging Leaders at baseline who provided consent for New Leaders to share this data

Exhibit A-5. Emerging Leaders Experience, Fall 2017 Baseline Sample

		Group I	Group II	Total
Average years' total experience	Mean	17	18	17
	SD	(9)	(8)	(8)
Emerging Leaders *n*		*54*	*48*	*102*

Source: New Leaders application data. Total *n* represents the number of Emerging Leaders at baseline who provided consent for New Leaders to share these data

Exhibit A-6. Emerging Leaders Assessment Scores, Fall 2017 Baseline Sample

		Group I	Group II	Total
Use of multiple forms of data (2.2a)	Mean	1.99	1.89	1.94
	SD	(0.50)	(0.55)	(53)
DDI leadership (2.2b)	Mean	2.20	2.12	2.16
	SD	(0.33)	(0.36)	(0.34)
Analysis of rigor and CCSS alignment (2.4a)	Mean	2.25	2.18	2.22
	SD	(0.38)	(0.34)	(0.36)
Emerging Leaders *n*		*53*	*49*	*102*

Source: New Leaders application data. Total *n* represents the number of Emerging Leaders at baseline who provided consent for New Leaders to share these data

Exhibit A-7. Met Emerging Leaders Admissions Criteria, Fall 2017 Baseline Sample

	Percent of Emerging Leaders		
	Group I	Group II	Total
Exceeded Emerging Leaders criteria	30	18	24
Met Emerging Leaders criteria	52	55	53
Discretionary admissions	17	27	22
Emerging Leaders *n*	*56*	*50*	*106*

Source: New Leaders application data. Total *n* represents the number of Emerging Leaders at baseline who provided consent for New Leaders to share these data.

Exhibit A-8. Emerging Leader Job Assignments, Fall 2017 Baseline Sample

	Number of Emerging Leaders		
	Group I	Group II	Total
Teacher	30	23	53
Assistant principal	12	9	21
Instructional coach/implementation specialist	6	13	19
District-level instructional support	5	2	7
Other school-based administrator	3	3	6
Emerging Leaders *n*	*56*	*50*	*106*

Source: New Leaders application data, fall 2017 instructional team verification.

Implementation Fidelity

In consultation with New Leaders staff, we identified six key components of the program, each aligned with one of the four logic model elements (first four boxes shown in Exhibit 1 of the report). For each key component, we developed one or more indicators to measure implementation fidelity using data the New Leaders program collects (Exhibit A-9). New Leaders collects an extensive array of data on program participants as part of normal operations, and we drew on all these to assess program fidelity—Emerging Leaders national community of practice sign-in sheets, baseline assessments used for admissions decisions, New Leaders admissions ratings, program director coaching logs, end-of-learning-cycle surveys, learning cycle attendance logs, assignment results, New Leaders end-of-year participant survey, and the New Leaders assessment scores.

Exhibit A-9. Emerging Leaders Fidelity Indicators Aligned with Logic Model

Logic Model Element	Key Component	Fidelity Indicator
New Leaders National Team(s)	1. Training and support of local program directors	Community of practice (COP)
	2. Recruitment and selection of Emerging Leaders candidates	Highly qualified candidates
Regional Team/Local Program Director	3. Regional implementation of training	Relevance of programming Quality of facilitation One-on-one coaching
	4. Emerging Leaders completion of training	Learning cycles Job-embedded assignments
Emerging Leader Learning/ Skill Set	5. Emerging Leaders demonstration of leadership proficiency	Leadership skills Readiness for aspiring principals
Instructional Team Activities	6. Instructional teamwork	Team meetings Observations

We set thresholds defining fidelity for each individual study participant on each indicator, as well as district-level thresholds defining fidelity of implementation in each RCT site (Exhibit A-10). Implementation fidelity was calculated for each individual site and then aggregated up to the sample level. At the sample level (across all three sites), we counted the overall sample as achieving fidelity if all three individual sites met the district-level threshold for fidelity. If one or more sites did not meet that threshold, the sample overall did not attain fidelity. Site-level outcomes were similar for all indicators except one-on-one coaching, which a single district met (Exhibit A-11).

By the end of the 2017–18 program year, of the 58 Emerging Leaders participants originally enrolled in the program and assigned to Group I, 53 remained in the study, along with their associated instructional teams. These 53 Emerging Leaders constituted the sample for each of the fidelity measures. One important exception is the indicator on admission of highly qualified candidates, which included all Emerging Leaders selected for the program pre-randomization (and later assigned to either Group I or Group II).

Exhibit A-10. Emerging Leaders Fidelity of Implementation Definitions, Thresholds, and Indicators by District

Conceptual	Operational	Individual-Level Threshold	District-Level Threshold	Percent Meeting Fidelity AISD	SAISD	SCS	Meets Fidelity?[a]
Key Component 1: Training and Support of Program Directors							
Community of Practice (COP)	National office hosts COP and facilitation trainings	Emerging Leaders program directors attend 14 COP meetings or national trainings during the program year	Regional program director attends required meeting	100%	100%	100%	✓
Key Component 2: Recruitment & Selection							
Highly Qualified Candidates	New Leaders recruits Emerging Leader candidates who meet rigorous selection criteria	Emerging Leaders candidates meet program admissions criteria by demonstrating "approaching proficient" on assessments scored with the Emerging Leaders rubric, by earning at least a 2 on the majority of concepts and as low as a 1.5 on no more than 2 concepts	**High fidelity**: 100% of candidates met criteria **Fidelity**: At least 70% of candidates met criteria	70%	82%	83%	✓
Key Component 3: Regional Implementation of Training							
Relevance of Programming	Emerging Leaders find trainings useful	An average of "agree" or above across a bank of questions on perception of programming relevance	80% of Emerging Leaders	89%	94%	94%	✓
Quality of Facilitation	Emerging Leaders find the quality of facilitation to be high	Rating of "agree" or above on question on perception of facilitation quality	80% of Emerging Leaders	84%	100%	100%	✓
One-on-one Coaching	Emerging Leaders receive personalized coaching	At least 7 hours of personalized coaching,[b] not including time spent in learning meetings	80% of Emerging Leaders	95%	59%	6%	X
Key Component 4: Emerging Leaders Completion of Training							
Learning Cycles	Attendance at all induction, intensive, and learning cycle sessions	Emerging Leader attends (or makes up) induction, intensive, and all 7 learning cycle sessions	80% of Emerging Leaders	80%	100%	100%	✓

Conceptual	Operational	Individual-Level Threshold	District-Level Threshold	Percent Meeting Fidelity			Meets Fidelity?[a]
				AISD	SAISD	SCS	
Job-Embedded Assignments	Completion of assignments and participation in associated learning meetings	Emerging Leader completes all 4 assignments; attends all 3 learning meetings, as indicated by having a valid score for each assignment and learning meeting	80% of Emerging Leaders	100%	100%	100%	✓
Key Component 5: Emerging Leader Demonstration of Leadership Proficiency							
Leadership Skills	Participants demonstrate leadership proficiency	Scores of "proficient" (3.0) or higher on at least 3 of 7 target concepts and scores of at least 2.0 on the remaining target concepts	80% of Emerging Leaders	5%	18%	19%	X
Readiness for Aspiring Principals	Participants demonstrate leadership proficiency and readiness for Aspiring Principals program	Scores of "proficient" (3.0) or higher on at least 5 of 7 target concepts and growth on the remaining 2 concepts	30% of Emerging Leaders	5%	0%	0%	X
Key Component 6: Instructional Teamwork							
Team Meetings	Emerging Leader participants lead teacher team meetings regularly	Emerging Leader leads a team meeting at least 10 times during year	80% of Emerging Leaders	88%	100%	88%	✓
Observations	Emerging Leader participants observe teacher team members to give feedback	Emerging Leader observes and provides feedback to team members at least 8 times	80% of Emerging Leaders	63%	69%	75%	--[c]

[a] The program achieved fidelity of implementation if each of the three sites met the district-level threshold for fidelity. (One indicator included a threshold for high fidelity, which was not met.)

[b] The Emerging Leaders program agreement specifies 12 hours of personalized coaching for each Emerging Leader, including the time spent in learning meetings (approximately 3 hours total). Starting with 9 hours as the threshold for this indicator (not including time spent in learning meetings), the target shown here (7 hours) takes into account the fact that the hours logged in the coaching tracker often did not include time spent on phone calls or other informal coaching interactions.

[c] Because the program was close to meeting the threshold for this indicator as measured by the end-of-cycle surveys, and because of the measurement error associated with this data source (appendix), we cannot say with confidence whether the program met (or failed to meet) the fidelity threshold.

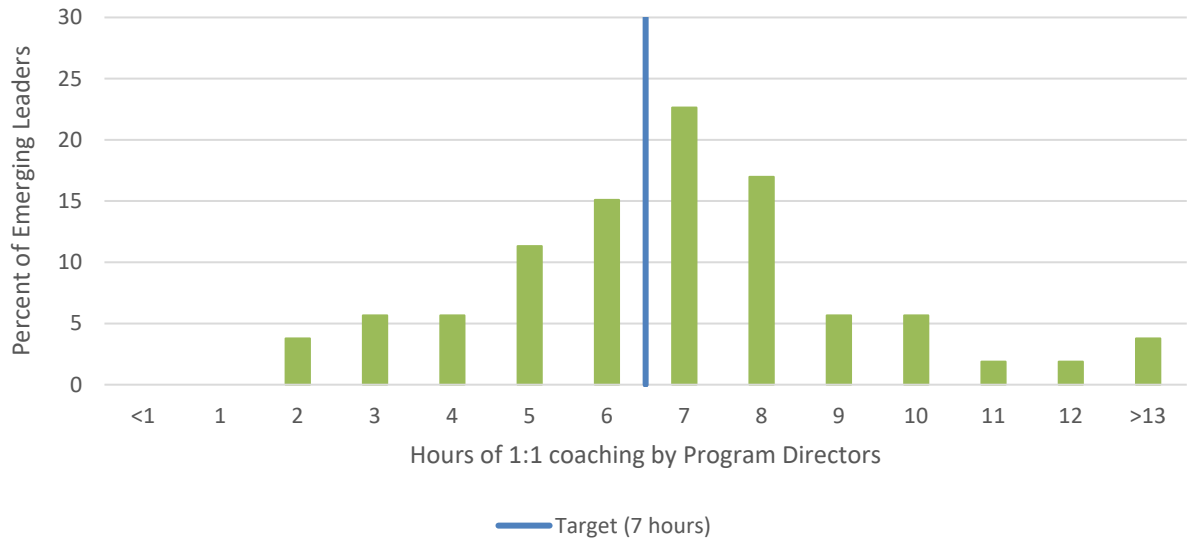
Exhibit A-11. Emerging Leaders Fidelity of Implementation Definitions, Data Sources, and Notes on Analysis

Conceptual	Operational	Data Source	Population	Analysis Notes
		Key Component 1: Training and Support of Program Directors		
Community of Practice (COP)	National office hosts COP and facilitation trainings	COP and training sign-in sheets	Regional program directors in the 3 RCT sites	Count of the number of trainings each director attended based on New Leaders' sign-in sheets.
		Key Component 2: Recruitment & Selection		
Highly Qualified Candidates	New Leaders recruits Emerging Leader candidates who meet rigorous selection criteria	New Leaders application data	All Emerging Leaders applicants	Candidates met the selection criteria by demonstrating "approaching proficient" on admissions assessments using the Emerging Leaders rubric. Candidate's applications must earn a 2 on the majority of the 6 concepts assessed and they may earn as low as a 1.5 on at most 2 concepts.
		Key Component 3: Regional Implementation of Training		
Relevance of Programming	Emerging Leaders find trainings useful	End-of-year program survey item Q1a	End-of-year program survey respondents	A Q1a scale was calculated as a simple mean of 5 survey items. The indicator was then calculated for each site as the percent of respondents where that mean score was greater than or equal to 3.
Quality of Facilitation	Emerging Leaders find the quality of facilitation to be high	End-of-year program survey item Q1b	End-of-year program survey respondents	A Q1b scale was calculated as a simple mean of 3 survey items. The indicator was then calculated for each site as the percent of respondents where that mean score was greater than or equal to 3.
One-on-one Coaching	Emerging Leaders receive personalized coaching	Coaching tracker	Emerging Leaders participants who did not formally withdraw from the program	This indicator was calculated using a coaching tracker provided by New Leaders which included each meeting and its duration in minutes. For each participant the time was summed across all meetings to get the total amount of coaching received. The indicator for each site was then calculated as the percent of respondents where the total amount of coaching was at least 7 hours (420 minutes).
		Key Component 4: Emerging Leaders Completion of Training		
Learning Cycles	Attendance at all induction, intensive, and learning cycle sessions	Learning cycle	Emerging Leaders participants who did not formally withdraw from the program	Participants passed this indicator if they were listed in the New Leaders attendance roster as having either "Attended" or "Made Up" the induction session, all 4 days of intensive sessions, and all 7 learning cycle sessions.
Job-Embedded Assignments	Completion of assignments and participation in	Emerging Leaders assessment team data	Emerging Leaders participants who did not formally withdraw from the program	Participants passed this indicator if they received a score for each assignment and learning cycle session, based on the New Leaders assessment data.

	associated learning meetings			
Key Component 5: Emerging Leader Demonstration of Leadership Proficiency				
Leadership Skills	Participants demonstrate leadership proficiency	Emerging Leaders assessment team data	Emerging Leaders participants who did not formally withdraw from the program	Participants met this indicator if they received a weighted concept score of 3 or higher ("proficient") on at least 3 of the 7 target concepts, and a score of 2 or higher ("approaching proficient") on at least 2 concepts.
Readiness for Aspiring Principals	Participants demonstrate leader-ship proficiency and readiness for Aspiring Principals program	Emerging Leaders assessment team data	Emerging Leaders participants who did not formally withdraw from the program	Participants met this indicator if they received a weighted concept score of 3 or higher ("proficient") on at least 5 of the 7 target concepts and demonstrated growth on the remaining 2 concepts. Participants demonstrated growth if their scores increased in each subsequent assignment in which the concept was assessed.
Key Component 6: Instructional Teamwork				
Team Meetings	Emerging Leader participants lead teacher team meetings regularly	Learning cycle session post-surveys	Emerging Leaders with sufficient data, read analysis notes	Because the surveys were anonymous, responses were linked across cycles using a non-case-sensitive combination of 4 questions included on each survey to generate a respondent ID (e.g., “What is the first letter of the town where you grew up?”). Fuzzy matching was used to link responses where 3 of the 4 ID questions matched. Where there were multiple responses for a given respondent in a single cycle, one response was dropped at random. Where respondents were missing surveys for 1, 2, or 3 cycles (of 7), the missing cycles were imputed using the mean of the existing surveys. Where respondents were missing surveys for 4 or 5 cycles (of 7), if the existing surveys responses were consistent (within 2), the missing cycles were imputed using the mean of the existing surveys. If they were not consistent, those respondents were dropped from the analysis. Respondents with only 1 cycle were dropped from the analysis. Once the sample was identified and missing data imputed, both indicators were calculated by summing the number of team meetings and observations across the 7 cycles for each respondent. The team meetings indicator for each site was calculated as the percent of respondents who reported leading at least 10 meetings during the year. The observations indicator for each site was calculated as the percent of respondents who reported observing teacher team members at least 8 times during the year.
Observations	Emerging Leader participants observe teacher team members to give feedback	Learning cycle session post-surveys	Emerging Leader with sufficient data, read analysis notes	

Approximately 40% of Emerging Leaders program participants did not receive the level of personalized, one-on-one coaching called for in the implementation fidelity threshold (appendix Exhibit A-12). Program directors logged a total of 386 individualized coaching hours, or an average of 7.3 hours per participant. Although 40% of participants received less coaching than expected, a substantial number of participants received more.

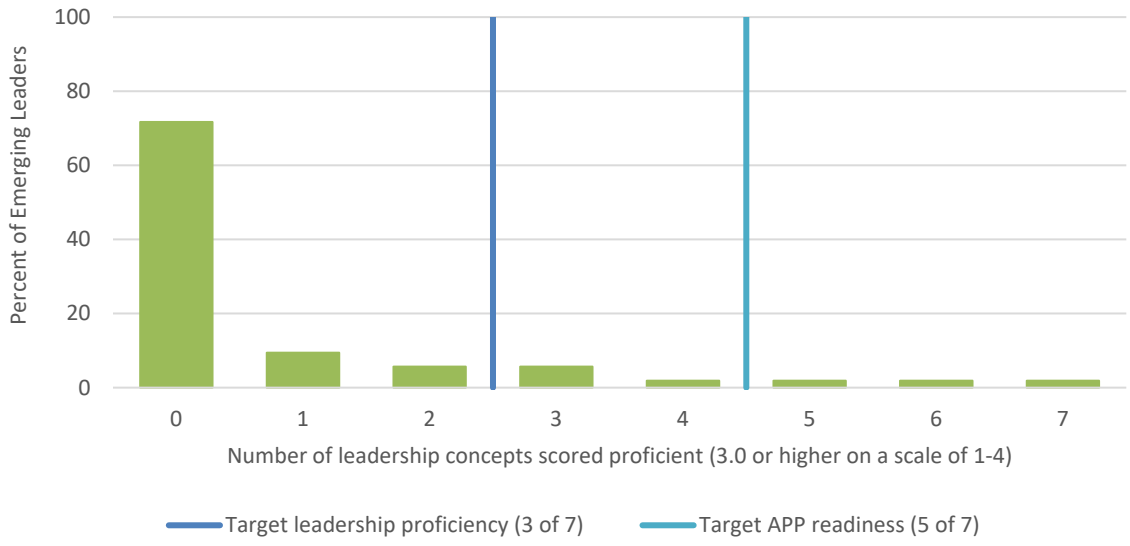
Exhibit A-12. Personalized Coaching for Group I Emerging Leaders



n = 53
Source: Emerging Leaders program director coaching logs.
Note: The Emerging Leaders program agreement specifies 9 hours of personalized coaching in addition to 3 hours spent in learning meetings. The fidelity threshold is set at 7 hours to account for incomplete record-keeping.

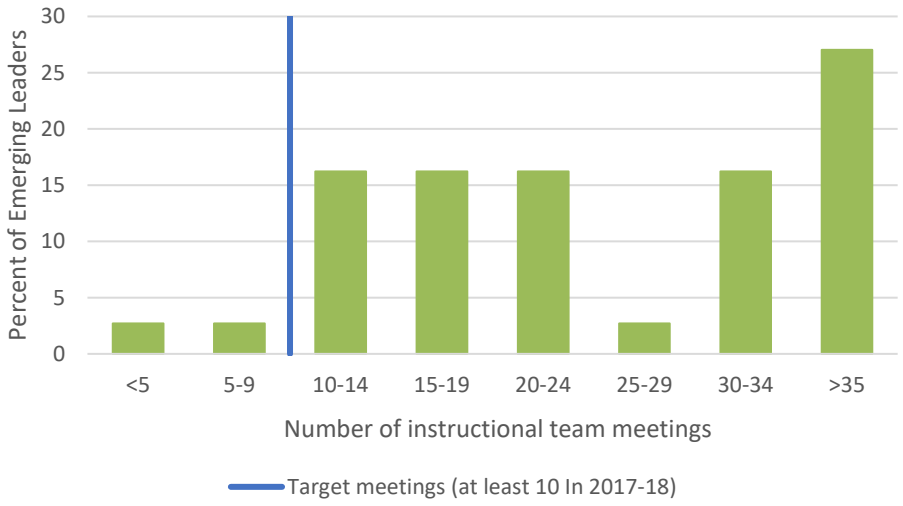
The majority of Emerging Leaders participants (72%) failed to demonstrate proficiency on any of the seven target leadership concepts addressed by the New Leaders national Assessor Corps (Exhibit A-13), with "proficient" being a score of at least 3.0 on a scale of 1.0–4.0, including half-points. Only 13% of Emerging Leaders participants demonstrated proficiency on at least three target concepts, the minimum threshold set for demonstrating leadership proficiency at the conclusion of the Emerging Leaders program, and only 6% demonstrated proficiency on five of the seven, the minimum required to determine readiness for New Leaders' principal residency program, Aspiring Principals Program (APP) (in addition to growth on the remaining two concepts).

Exhibit A-13. Emerging Leaders' Proficiency on Seven Target Leadership Concepts



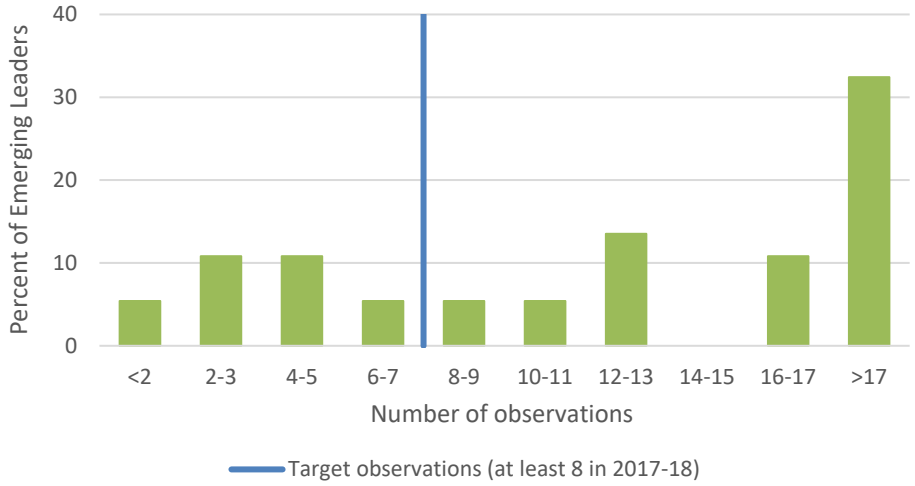
n = 53.
Source: New Leaders leadership assessments.

Exhibit A-14. Emerging Leaders Instructional Team Meetings, 2017–18



n = 37.
Note: Because responses were anonymous, some could not be linked across cycles. The sample shown here includes only those Emerging Leaders with at least four linked end-of-cycle surveys. Data for missing cycles were imputed (appendix for detail).
Source: Emerging Leaders end-of-cycle surveys, 2017–18.

Exhibit A-15. Emerging Leaders' Observation of Instructional Team Members, 2017–18



n = 37.
Note: Because responses were anonymous, some could not be linked across cycles. The sample shown here includes only those Emerging Leaders with at least four linked end-of-cycle surveys. Data for missing cycles were imputed.
Source: Emerging Leaders end-of-cycle surveys, 2017–18.

Qualitative Methods

Site Visits During the RCT Year

In late winter 2018, the research team visited each of the three RCT sites to interview Emerging Leaders participants, instructional team members, principals, district administrators, and New Leaders staff and to observe instructional team meetings. We chose a convenience sample for all qualitative data collection, maximizing the number of instructional team meetings the team members could observe and the number of Emerging Leaders they could interview while on site. In total, the team interviewed 60 respondents in 16 schools and observed 10 instructional team meetings (Exhibit A-16). We requested an interview with every Emerging Leader participant whose instructional team was meeting during the site visit window and ultimately interviewed 17 of them, approximately one-third of the Group I cohort.

Exhibit A-16. Interview Sample and Observations During Spring 2018 Site Visit

	AISD	SAISD	SCS	Total
Interviews				
Principals	4	4	4	12
Emerging Leaders participants	5	7	5	17
Instructional team members	5	9	9	23
District administrators	0	2	1	3
New Leaders program directors and senior directors	2*	2*	2	5
Total	16	24	21	60
Instructional team meetings observed	3	2	5	10

*Includes a single respondent who served as senior director of program implementation at both Texas sites.

The Emerging Leaders logic model was the organizing framework for the interview protocols; each section of the protocol addressed a key component of the program as appropriate for each respondent type. Respondents described how the Emerging Leaders program supported the leadership development pipeline in each partner district and participants' own growth and development as leaders; their experiences with the program and the elements that were most useful or relevant to their work; the work of the instructional teams and the factors that facilitated or hindered their collaboration; changes in teachers' practice that resulted from their experience with Emerging Leaders-style rigor analysis, data-driven instruction (DDI) cycles, and corrective instruction action planning; and the perceived outcomes of the program for participants, teachers, and students. Site visits took were late February and early March 2018, approximately 7 months into the program year.

Data from the interviews were synthesized in structured debrief guides aligned with the protocols and research questions to compare responses within and across respondent types, levels, and organizations and to present evidence—concrete examples, illustrative quotations, and documents—of common themes as well as divergent perspectives in each site. We subsequently used these within-site findings to design a framework for systematic cross-site analysis. We followed the key topics in the structured debriefing guides to examine any consistencies, common factors supporting implementation, range in implementation, and factors explaining differences in implementation across all sites. Themes that cut across the districts and schools were identified in cross-case meetings and then verified through careful review of interviews and debrief summaries to identify all supporting and contradictory evidence. Validated themes were included in the report.

The analysis of interviews and team meeting observations focused specifically on respondents' perceptions of the Emerging Leaders program, on explaining the variation in participants' experiences, and on describing and evaluating the instructional teams' work. The findings selected for reporting were those most relevant to the interpretation of findings presented in other sections of this report and address Emerging Leader recruitment and selection, leadership development, the work of the instructional teams, and evidence of student outcomes.

Group I Emerging Leaders Follow-up Site Visits

The research team conducted follow-up site visits at all three RCT sites in spring 2019 to learn from Emerging Leaders program alumni about their experiences after the program, with a particular focus on any changes in their positions or leadership roles and their ongoing use of Emerging Leaders program tools, approaches, and protocols. We interviewed Emerging Leaders program alumni's principals (or supervisors or managers for alumni no longer working in schools) and instructional team members (for alumni who continued to lead instructional teams) to gather additional perspectives on alumni's leadership roles and whether the Emerging Leaders program was successful in preparing them for the principalship. We interviewed district administrators and regional Emerging Leaders program directors to understand how the Emerging Leaders program operated within the district context and the extent to which it was meeting district expectations for growing the leadership pipeline.

The sampling approach prioritized deeply understanding the Emerging Leaders program alumni's different experiences. We selected an intentionally broad sample of Emerging Leaders program alumni to interview for qualitative analysis. The qualitative analysis was complemented by the alumni survey data, which provided a representative understanding of Emerging Leaders program alumni experiences. We used alumni survey data to inform the alumni selection process and considered several factors of variation for the sampling: (1) whether and how frequently alumni reported meeting with the teacher teams they led, (2) alumni's current position and whether they were in a new position since completing the Emerging Leaders program, and (3) alumni's perception of the training they received in the Emerging

Leaders program. In total, we interviewed 17 Emerging Leaders program alumni in 16 schools, about one-third of the Group I cohort (Exhibit A-17).

Exhibit A-17. Interview Sample During the Spring 2019 Site Visit

	AISD	SAISD	SCS	Total
Emerging Leaders alumni	6	6	5	17
Principals	5	3	5	13
Instructional team members	3	4	3	10
District administrators	1	0	1	2
New Leaders program directors and senior directors	1	1	1	2*
Total	16	14	15	44

*Includes a single respondent who served as senior director of program implementation at both Texas sites.

The following sustainability research questions were developed based on the Emerging Leaders program logic model to explore the extent to which the Emerging Leaders program supported leadership pipeline growth and created a lasting change in instructional leadership practice:

1. To what extent have Emerging Leaders alumni taken on new leadership roles in 2018–19, both formal and informal? In what ways has the Emerging Leaders program prepared them for these new roles?
2. In what ways are alumni currently using the skills, strategies, practices, and tools that they acquired in the Emerging Leaders program?
3. In what ways have alumni's DDI leadership knowledge, skills, and self-efficacy grown as a result of their participation in the Emerging Leaders program? In what ways have their leadership skills continued to grow in 2018–19?
4. With the benefit of hindsight, how do Emerging Leaders alumni reflect on their experience in the program?
5. What are the varied ways in which the Emerging Leaders program has met sites' expectations and/or contributed to the development of a robust school leadership pipeline in each site?
6. How have school and district contexts of support and existing practices influenced Emerging Leaders alumni leadership demonstration, application of and growth in knowledge and skills, and the Emerging Leaders program's impact on the school leadership pipeline?

We created semi structured interview protocols to investigate the research questions for each group of interviewees: Emerging Leaders program alumni, principals or supervisors, instructional team members, district administrators, and Emerging Leaders program directors. The research team conducted the interviews in person when possible and remotely when needed. The interviews were audio-recorded and transcribed for analysis.

The transcript analysis proceeded in five stages.

1. We analyzed each individual transcript to capture data related to the sustainability research questions.
2. We drafted summary statements for each research question within each cluster of Emerging Leaders program alumni, principals or supervisors, and teaching teams to understand instructional leadership practice from multiple perspectives and within the school context.

3. We drafted summary claims for each research question within each of the three participating sites by analyzing across summary statements of the clusters within each site.
4. We drafted summary claims for each research question according Emerging Leader alumni positions: summary claims for Emerging Leader alumni who were principals or assistant principals, alumni who were instructional coaches or in similar school-based support roles, and alumni who were teachers. This indicated whether there was any pattern based on the positionality or leadership roles of the alumni.
5. We looked across the site summary claims and the alumni positionality summary claims to identify overall patterns, differences, and outliers for each research question.

The result of this process was a draft of the overall sustainability claims for this study.

Survey Methods

End-of-Year Emerging Leaders Survey

The research team developed an end-of-year survey for the Emerging Leaders participants in Groups I and II. The survey was identical for both groups and was designed to (1) capture participation in leadership development programs during the 2017–18 school year and the extent to which those programs included similar activities or content to the Emerging Leaders program, (2) assess the degree to which Group II participants engaged with instructional teams in 2017–18 (both leading team meetings and coaching colleagues), and (3) collect Emerging Leaders' self-reports of their leadership skill growth during 2017–18.

The survey asked the teachers to indicate any school leadership or instructional leadership development programs they participated in during the 2017–18 school year, gauge the intensity of the programs, and rate the importance of various characteristics across the programs. The survey also asked them to indicate the frequency of various activities in the leadership programs they participated in, including attending in-person trainings, receiving feedback on their leadership of teachers, and demonstrating skills through assignments. Finally, the teachers were asked to indicate the extent to which they believed their skills had grown in such areas as leading a team of teachers, improving their knowledge of data-driven instruction, and gaining the skills they would need to lead a school.

The survey asked Emerging Leaders participants to report on their own coaching activities, such as how often they observed the instruction of teachers on their team and how many times during the 2017–18 school year they provided feedback (both formally or informally) to team members. They were also asked about their leadership of instructional team meetings, including reviewing and analyzing student data and leading the lesson planning process.

All Group I and Group II Emerging Leaders in the baseline sample received an email link to the electronic survey in May 2018. They were given 2 weeks to fill out the 10-minute survey and received a modest financial incentive for completion. We followed up with phone calls to nonrespondents to encourage their participation.

The response rate was 78% (84% for Group I and 70% for Group II) (Exhibit A-18).

Exhibit A-18. End-of-Year Emerging Leaders Survey Response Rates, Spring 2018

	Group I	Group II	Total
Emerging Leaders participants	58	54	112
Respondents	49	38	87
Response rate (%)	84	70	78
Attrition (i.e., non-response rate) (%)	*16*	*30*	*22*
Differential attrition (%)			*14*

Key elements of treatment-control contrast were participation in leadership development programs during the RCT year, engagement with instructional teams, and self-reported growth in leadership skills. Exhibits A-19–A-26 summarize the survey results.

Exhibit A-19. Teacher Leadership Roles, 2017–18

Q: Do you hold any of the following roles in your school building? (mark all that apply)

		No (%)	Yes (%)	*n*
Grade level or department chair	Group I	76	24	50
	Group II	66	34	38
Instructional team leader	Group I	54	46	50
	Group II	66	34	38
Other	Group I	80	20	50
	Group II	74	26	38
None of the above	Group I	74	26	50
	Group II	76	24	38

Source: SRI end-of-year leadership survey, spring 2018.
~$p < .1$,*$p < .05$, **$p < .01$, ***$p < .001$.

Exhibit A-20. Group I and Group II Team Leadership, 2017–18

Q: How often did you do each of the following during the 2017–18 school year? Please consider all of the teachers in your building, not just the teachers on your Emerging Leaders team.

		Never	Once or twice this school year	Every few months	Monthly	About twice per month	Weekly or more often	Mean		*n*
Led a team of teachers for any reason	Group I	2	2	2	16	22	55	5.2	***	49
	Group II	5	8	24	21	13	29	4.2		38
Led a team of teachers in the lesson planning process	Group I	4	2	14	27	20	33	4.6	*	49
	Group II	18	16	18	11	13	24	3.6		38
Led a team of teachers analyzing student achievement data	Group I	2	4	2	33	29	31	4.7	**	49
	Group II	18	18	11	16	16	21	3.6		38
Observed teachers' instruction	Group I		10	21	15	13	42	4.5		48
	Group II	5	13	11	11	3	58	4.7		38
Provided teachers w/ feedback on their practice	Group I		10	10	16	14	49	4.8		49
	Group II	11	16	13	5	8	47	4.3		38

Source: SRI end-of-year leadership survey, spring 2018.
Scale: 1 = Never, 2 = Once or twice this school year, 3 = Every few months, 4 = Monthly, 5 = About twice per month, 6 = Weekly or more often.
~$p < .1$,*$p < .05$, **$p < .01$, ***$p < .001$.

Exhibit A-21. Group I and Group II Leadership Development, 2017–18

Q: Did you participate in any school leadership or instructional leadership development programs in 2017–18?

		No (%)	Yes (%)	*n*
Any school leadership or instructional leadership development programs	Group I	29	71	49
	Group II	45	55	38

Source: SRI End-of-year leadership survey, spring 2018.
~$p < .1$,*$p < .05$, **$p < .01$, ***$p < .001$.

Exhibit A-22. Group I and Group II Leadership Development Programs, 2017–18

Q: In which program(s) did you participate? (mark all that apply)

		No (%)	Yes (%)		*n*
Emerging Leaders Program	Group I	6	94	***	35
	Group II	79	21		19
NTC Instructional Coach training	Group I	97	3		35
	Group II	84	16		19
Relay Residency Program or M.A.T.	Group I	100	0		35
	Group II	100	0		19
Teacher Leadership Academy	Group I	100	0		35
	Group II	100	0		19
TNTP (The New Teacher Project)	Group I	100	0		35
	Group II	100	0		19
Another program offered by district	Group I	94	6	**	35
	Group II	63	37		19
Master's, doctorate, or other graduate program	Group I	86	14	*	35
	Group II	58	42		19
Principal certification, district training, admin internship	Group I	89	11		35
	Group II	84	16		19

Source: SRI end-of-year leadership survey, spring 2018.

Note. Sample includes only those respondents who reported that they had participated in a leadership development program on the prior question.

~*p* < .1,**p* < .05, ***p* < .01, ****p* < .001.

Exhibit A-23. Frequency of Participation in Leadership Development Programs, 2017–18

Q: For the leadership program(s) in which you participated, approximately how often did you participate in program activities?

	Percent of Emerging Leaders							
	Never	Once or twice this school year	Every few months	Monthly	About twice per month	Weekly or more often	Mean	*n*
Group I	0	0	3	40	14	43	5.0	35
Group II	10	5	14	33	5	33	4.2	21

Source: SRI end-of-year leadership survey, spring 2018.

Scale: 1 = Never, 2 = Once or twice this school year, 3 = Every few months, 4 = Monthly, 5 = About twice per month, 6 = Weekly or more often.

~*p* < .1,**p* < .05, ***p* < .01, ****p* < .001.

Exhibit A-24. Focus of Leadership Development Programs, 2017-18

Q: Across the leadership program(s) you participated in this past year, how important were each of the following elements?

		Less Than Critical (%)	Critical (%)		n
Creating a vision and goals	Group I	34	66	*	35
	Group II	70	30		20
Setting student achievement goals	Group I	26	74	***	35
	Group II	76	24		21
Analyzing data to identify misconceptions	Group I	18	82	***	34
	Group II	67	33		21
Differentiating instruction	Group I	40	60		35
	Group II	67	33		21
Aligning instruction to state standards	Group I	40	60	**	35
	Group II	81	19		21
Planning corrective instruction	Group I	49	51	*	35
	Group II	81	19		21
Rigor analysis	Group I	57	43		35
	Group II	62	38		21
School budgeting	Group I	86	14		35
	Group II	86	14		21
Hiring and staffing	Group I	82	18		34
	Group II	81	19		21
How to lead other adults	Group I	46	54		35
	Group II	67	33		21
How to plan agendas and lead meetings	Group I	57	43		35
	Group II	76	24		21

Source: SRI end-of-year leadership survey, spring 2018.
~$p < .1$,*$p < .05$, **$p < .01$, ***$p < .001$.

Exhibit A-25. Components of Leadership Development Programs, 2017-18

Q: For the leadership program(s) in which you participated, what was the frequency of the following activities?

		Percent of Respondents							
		Never	Once or twice this school year	Every few months	Monthly	About twice per month	Weekly or more often	Mean	*n*
Trainings (in-person or virtual)	Group I	3	0	3	71	17	6	4.2	35
	Group II	10	5	14	38	14	19	4.0	21
Demon-strate skills through artifacts	Group I	0	0	0	51	29	20	4.7	35
	Group II	19	5	5	24	14	33	4.1	21
Receive feedback on leadership of teachers	Group I		6	34	31	17	11	3.9	35
	Group II	24	14	5	29	14	14	3.4	21
Have a coach observe a team meeting	Group I	29	17	29	11	9	6	2.7	35
	Group II	33	24	14	19	0	10	2.6	21

Source: SRI end-of-year leadership survey, spring 2018.

Scale: 1 = Never, 2 = Once or twice this school year, 3 = Every few months, 4 = Monthly, 5 = About twice per month, 6 = Weekly or more often.

~$p < .1$,*$p < .05$, **$p < .01$, ***$p < .001$.

Exhibit A-26. Self-Reported Growth in Leadership Skills, 2017–18

Q: To what extent do you feel you've grown in each of the following skills during the 2017–18 school year?

		Less than a great deal (%)	A great deal (%)		*n*
Gained the skills necessary to lead teacher team	Group I	29	71	***	35
	Group II	78	22		23
Improved knowledge of data-driven instruction	Group I	31	69	***	35
	Group II	87	13		23
Improved ability to analyze student data	Group I	31	69	***	35
	Group II	87	13		23
Increased ability to support corrective instruction	Group I	46	54	*	35
	Group II	78	22		23
Gained skills needed to lead a school	Group I	29	71	**	35
	Group II	70	30		23

Source: SRI end-of-year leadership survey, spring 2018.

~$p < .1$,*$p < .05$, **$p < .01$, ***$p < .001$.

Emerging Leaders Follow-up Survey

The research team developed a follow-up survey for Emerging Leaders participants in Group I to learn about the roles they held the year after their participation in the program and to assess the degree to which they engaged with instructional teams in 2018–19 (both leading team meetings and coaching colleagues) as well as roles they aspire to.

The survey asked the participants to indicate how often then engaged in leadership activities in their current role and how effective they feel in particular leadership roles aligned with the Emerging Leaders program. The survey also asked Group I Emerging Leaders participants how established and effective they feel in their professional assignment. They were asked about roles they are interested in moving into in the future and what skills they feel they need to develop for those roles. In particular, Emerging Leaders participants were asked how ready they feel to become a principal (if they are not already in that role).

The survey asked participants to report on the spread of Emerging Leaders at their site, both in terms of colleagues engaged in the program and the adoption of Emerging Leaders practices and tools.

All Group I Emerging Leaders participants received an email with a link to the electronic survey in January 2019. Participants were given 3 weeks to fill out the 15-minute survey and received a modest financial incentive for completion. We sent follow-up emails and conducted follow-up phone calls to nonrespondents to encourage their participation.

The response rate was 81% (Exhibit A-27). Exhibits A-28–A-45 summarize the survey results.

Exhibit A-27. Emerging Leaders Follow-up Survey Response Rates, Winter 2019

	Group I
Emerging Leaders	53
Respondents	43
Response rate (%)	81

Exhibit A-28. Teacher Leadership Roles, 2018–19

Q: If you are currently a classroom teacher, do you hold any of the following roles in your school?

	No (%)	Yes (%)	*n*
Grade level or department chair	67	33	12
Instructional team leader	100	0	12
Other	92	8	12
None of the above	42	58	12
Not currently a classroom teacher	92	8	12

Source: Emerging Leaders follow-up survey, winter 2019.
Note: Sample includes only those who were teachers in 2018–19.

Exhibit A-29. Emerging Leaders' Teacher Leadership Opportunities, 2018–19

Q: How often have you done each of the following so far in the 2018-19 school year?

	Percent of Respondents						
	Never	Infrequently	Monthly	About twice per month	Weekly or more often	Mean	*n*
Led a team of teachers in instructional planning	14	9	14	16	47	3.7	43
Led a team of teachers in analyzing student data	12	12	26	33	19	3.3	43
Led a team of teachers for any other reason	5	12	12	21	51	4.0	43
Observed teachers' instruction	12	21	12	2	53	3.7	43
Provided teachers feedback on instructional practice	12	19	12	5	53	3.7	43

Source: Emerging Leaders follow-up survey, winter 2019.
Scale: 1 = Never, 2 = Infrequently, 3 = Monthly, 4 = About twice per month, 5 = Weekly or more often.
Note. Sample limited to those Emerging Leaders participants who responded to this item in both 2017–18 and 2018–19.

Exhibit A-30. Emerging Leaders' Self-Reported Efficacy, Among Those Who Were Classroom Teachers, 2018–19

Q: If you are currently a classroom teacher, how effective do you feel using data in your own instructional planning?

	Percent of Respondents						
	Not at all effective	Hardly at all effective	Somewhat effective	Quite effective	Extremely effective	Mean	*n*
Using data in instructional planning	0	9	18	36	36	4.0	11

Source: Emerging Leaders follow-up survey, winter 2019.
Scale: 1 = Not at all effective, 2 = Hardly at all effective, 3 = Somewhat effective, 4 = Quite effective, 5 = Extremely effective.
Note. Of those who are teachers in 2018–19.

Exhibit A-31. Emerging Leaders' Self-Reported Leadership Efficacy, 2018–19

Q: This school year, how effective do you feel as a leader in each of the following roles?

	Percent of Respondents						
	Not at all effective	Hardly at all effective	Somewhat effective	Quite effective	Extremely effective	Mean	*n*
Leading a team of teachers	0	5	14	45	36	4.1	42
Improving your own practice of DDI	0	7	12	52	29	4.0	42
Analyzing and interpreting student data	0	7	7	45	40	4.2	42
Supporting teachers' corrective instruction	0	5	31	45	19	3.8	42
Coaching others through DDI cycles	0	2	26	43	29	4.0	42
Using tools and protocols from ELP	0	7	19	57	17	3.8	42
Engaging in difficult conversations	0	2	26	45	26	4.0	42
Exercising adaptive leadership	0	5	19	48	29	4.0	42

Source: Emerging Leaders follow-up survey, winter 2019.

Scale: 1 = Not at all effective, 2 = Hardly at all effective, 3 = Somewhat effective, 4 = Quite effective, 5 = Extremely effective.

Exhibit A-32. Emerging Leaders Self-Reported Leadership Efficacy, Among Those in Formal Leadership Roles, 2018–19

Q: This school year, how effective do you feel as a leader in each of the following roles?

	Percent of Respondents						
	Not at all effective	Hardly at all effective	Somewhat effective	Quite effective	Extremely effective	Mean	*n*
Leading a team of teachers	0	0	12	42	46	4.4	26
Improving your own practice of DDI	0	4	15	54	27	4.0	26
Analyzing and interpreting student data	0.0	4	8	42	46	4.3	26
Supporting teachers' corrective instruction	0	0	35	39	27	3.9	26
Coaching others through DDI cycles	0	0	27	42	31	4.0	26

	Percent of Respondents						
	Not at all effective	Hardly at all effective	Somewhat effective	Quite effective	Extremely effective	Mean	*n*
Using tools and protocols from ELP	0	4	15	58	23	4.0	26
Engaging in difficult conversations	0	0	19	50	31	4.1	26
Exercising adaptive leadership	0	0	19	46	35	4.2	26

Source: Emerging Leaders follow-up survey, winter 2019.

Scale: 1 = Not at all effective, 2 = Hardly at all effective, 3 = Somewhat effective, 4 = Quite effective, 5 = Extremely effective.

Exhibit A-33. Correlations Between Emerging Leaders' Self-Reported Efficacy in 2018–19 and Emerging Leaders Program Leadership Concept Scores in 2017–18

	Emerging Leaders Program Target Leadership Concept Scores						
Self-reported leadership efficacy	1.1: Reflective Practice and Continuous Improve-ment	1.2: Communi-cation and Inter-personal Relation-ships	2.1: Pedagogy and Instructional Strategies	2.2: Data-Driven Instruction	2.4: Standards-Based Planning	3.1: Urgency and Efficacy	4.2: Leadership Develop-ment
Leading a team of teachers	0.1	0.3	0.0	0.1	-0.1	0.1	0.2
Improving your own practice of DDI	0.1	0.2	0.0	0.0	-0.1	0.0	0.1
Analyzing and interpreting student data	0.1	0.2	0.1	0.1	0.0	0.1	0.2
Supporting teachers' corrective instruction	0.0	0.2	0.0	0.0	0.0	0.0	0.0
Coaching others through DDI cycles	0.1	0.1	-0.1	0.0	-0.1	0.1	0.1
Using tools and protocols from ELP	0.3	0.2	0.2	0.2	0.2	0.2	0.3
Engaging in difficult conversations	0.2	0.3	0.0	-0.1	0.0	-0.1	0.0
Exercising adaptive leadership	0.3	0.2	0.2	0.2	0.2	0.2	0.4

Source: Emerging Leaders follow-up survey, winter 2019; and New Leaders Leadership Proficiency Assessments, 2017–18.

Exhibit A-34. Emerging Leaders' Self-Reported Effectiveness in Current Assignment, 2018–19

Q: Thinking about your professional assignment…

	Percent of Respondents						
	Not at all	Not very much	Some	Quite a bit	A great deal	Mean	*n*
All respondents							
How established are you in your current role?	5	2	5	37	51	4.3	43
How effective do you feel right now in your current role?	2	7	7	49	35	4.1	43
How relevant was Emerging Leaders training for your current role?	2	5	16	26	51	4.2	43
Subsample of respondents who changed roles in 2018–19							
How established are you in your current role?	8	8	8	31	46	4.0	13
How effective do you feel right now in your current role?	0	8	8	46	38	4.2	13
How relevant was Emerging Leaders training for your current role?	0	0	15	15	69	4.5	13
Subsample of respondents who were classroom teachers in 2018–19							
How established are you in your current role?	6	0	0	29	65	4.5	17
How effective do you feel right now in your current role?	6	12	6	35	41	3.9	17
How relevant was Emerging Leaders training for your current role?	6	6	18	35	35	3.9	17

Source: Emerging Leaders follow-up survey, winter 2019.

Scale: 1 = Not at all, 2 = Not very much, 3 = Some, 4 = Quite a bit, 5 = A great deal.

Exhibit A-35. Emerging Leaders' Participation in Leadership Development Programs, 2018–19

Q: Are you participating in any school leadership or instructional development programs during 2018–19?

	No (%)	Yes (%)	*n*
Any school leadership or instructional leadership development programs	77	23	43

Exhibit A-36. Leadership Development Programs, 2018–19

Q: In which program(s) are you participating? (Select all that apply)

	No (%)	Yes (%)	*n*
New Leaders Aspiring Principals Program	100	0	10
Relay MAT (Master of Arts in Teaching)	100	0	10
Other	70	30	10
Master's, doctorate, or other graduate program	60	40	10
Principal certification	60	40	10
Administrative internship	100	0	10
District leadership training	100	0	10
Another leadership development program	70	30	10
Instructional coach training	100	0	10
TNTP	100	0	10
Teacher Leadership Academy	100	0	10

Source: Emerging Leaders follow-up survey, winter 2019.

Exhibit A-37. Jobs to Which Emerging Leaders Aspired, 2018–19

Q: How interested are you in moving into one of the following roles, either in the short term or the long term?

	Percent of Respondents						
	Not at all interested	Not very interested	Somewhat interested	Interested	Extremely interested	Mean	*n*
Assistant principal	7	10	17	17	48	3.9	29
Instructional coach	15	22	22	15	26	3.2	27
Principal	10	13	15	30	33	3.6	40
District-level administrator	2	10	26	26	36	3.8	42
Principal supervisor	21	29	14	24	12	2.8	42
Superintendent	42	23	9	7	19	2.4	43
Education non-profit administrator	17	26	17	19	21	3.0	42
Government or public policy	28	28	15	13	18	2.7	40
Other	59	9	5	0	27	2.3	22

Source: Emerging Leaders follow-up survey, winter 2019.

Scale: 1 = Not at all interested, 2 = Not very interested, 3 = Somewhat interested, 4 = Interested, 5 = Extremely interested.

Note. Excludes those who responded that they already serve or have served in the role.

Exhibit A-38. Emerging Leaders' Perceptions of Readiness for the Principalship, Among Those Reporting That They Are Interested or Extremely Interested in It, 2018-19

Q: [If not already a principal] To what extent do you feel ready to become a principal in the following roles?

	Percent of Respondents						
	Not at all ready	Not very ready	Somewhat ready	Quite ready	Completely ready	Mean	n
Overall	0	4	32	52	12	3.7	25
Personal leadership	0	0	20	52	28	4.1	25
Adult/team leadership	0	0	12	68	20	4.1	25
Instructional leadership	0	0	24	56	20	4.0	25
Cultural leadership	0	0	28	56	16	3.9	25
Operational leadership	0	8	40	32	20	3.6	25

Source: Emerging Leaders follow-up survey, winter 2019.

Scale: 1 = Not at all ready, 2 = Not very ready, 3 = Somewhat ready, 4 = Quite ready, 5 = Completely ready.

Note. Excludes those already serving as principals and those who did not report they were "interested" or "extremely interested" in becoming princpals.

Exhibit A-39. Skills Needed for Desired Role, 2018–19

Q: To what extent do you feel you need to develop the following skills in your desired role?

	Percent of Respondents						
	Not at all	Not very	Some	Quite a bit	A great deal	Mean	*n*
Creating a vision and goals	10	31	36	14	10	2.8	42
Hiring and staffing	9	19	42	16	14	3.1	43
How to lead other adults	9	28	40	16	7	2.8	43
How to plan agendas and lead meetings	40	16	16	16	12	2.4	43
Setting student achievement goals	16	21	37	14	12	2.8	43
Analyzing data to identify misconceptions	14	40	21	16	9	2.7	43
Differentiating instruction	12	35	23	14	16	2.9	43
Aligning instruction to state standards	14	42	21	12	12	2.7	43
Planning corrective instruction	14	16	37	16	16	3.1	43
Implementation of corrective instruction	7	23	40	19	12	3.1	43
Rigor analysis	5	28	37	23	7	3.0	43
School budgeting	9	23	14	23	30	3.4	43

Source: Emerging Leaders follow-up survey, winter 2019.

Scale: 1 = Not at all , 2 = Not very, 3 = Some, 4 = Quite a bit, 5 = A great deal.

Exhibit A-40. Colleagues Participating in Emerging Leaders Program, 2018–19

Q: Are any colleagues in your school currently participating in the Emerging Leaders program?

	No (%)	Yes (%)	Mean	*n*
Colleagues currently participating in Emerging Leaders	74	26	0.3	43

Source: Emerging Leaders follow-up survey, winter 2019.

Scale: 1 = Not at all , 2 = Not very, 3 = Some, 4 = Quite a bit, 5 = A great deal.

Exhibit A-41. Number of Colleagues Who Are New Leaders Alumni, 2018–19

Q: How many colleagues at your school have already completed New Leaders programs (either the Emerging Leaders program or the Aspiring Principals program), including you?

	Percent of Respondents								
	1	2	3	4	5	7	15	Mean	*n*
Number of colleagues who are alumni	42	33	7	2	5	2	2	2.1	43

Exhibit A-42. Importance of Emerging Leaders Training for Various Concepts, 2018-19

Q: How important has your Emerging Leaders training been for...

	Percent of Respondents						
	Not important	Of minor importance	Of some importance	Very important	Critical	Mean	*n*
Informing collaboration w/ colleagues	0	7	26	42	26	3.9	43
Providing common vocabulary and set of tools	2	5	35	40	19	3.7	43
Having productive conversations	2	2	26	40	30	3.9	43

Source: Emerging Leaders follow-up survey, winter 2019.

Scale: 1 = Not important , 2 = Of minor importance, 3 = Of some importance, 4 = Very important, 5 = Critical.

Exhibit A-43. Instructional Teams' Use of Emerging Leaders Tools or Similar Tools, 2018-19

Q: To the best of your knowledge, are instructional teams in your school currently using any of the following Emerging Leaders tools or tools similar to Emerging Leaders tools? (Select all that apply)

		No (%)	Yes (%)	*n*
Rigor frameworks	Same team I led as part of ELP	79	21	43
	A different team to which I belong	67	33	43
	Other teams in my school	67	33	43
	Any team in my school	37	63	43
	Don't know/not sure	67	33	43
Data Analysis Guide	Same team I led as part of ELP	79	21	43
	A different team to which I belong	67	33	43
	Other teams in my school	65	35	43
	Any team in my school	33	67	43
	Don't know/not sure	70	30	43

		No (%)	Yes (%)	*n*
Corrective Instruction Action Plan (CIAP)	Same team I led as part of ELP	79	21	43
	A different team to which I belong	65	35	43
	Other teams in my school	67	33	43
	Any team in my school	33	67	43
	Don't know/not sure	72	28	43
Analyzing Student Work Template	Same team I led as part of ELP	77	23	43
	A different team to which I belong	60	40	43
	Other teams in my school	63	37	43
	Any team in my school	30	70	43
	Don't know/not sure	70	30	43
Instructional Practice Guide (IPG)	Same team I led as part of ELP	81	19	43
	A different team to which I belong	67	33	43
	Other teams in my school	67	33	43
	Any team in my school	40	60	43
	Don't know/not sure	65	35	43
FAST form	Same team I led as part of ELP	95	5	43
	A different team to which I belong	74	26	43
	Other teams in my school	79	21	43
	Any team in my school	58	42	43
	Don't know/not sure	44	56	43
Other New Leaders tool	Same team I led as part of ELP	100	0	43
	A different team to which I belong	91	9	43
	Other teams in my school	95	5	43
	Any team in my school	91	9	43
	Don't know/not sure	98	2	43

Source: Emerging Leaders follow-up survey, winter 2019.

Exhibit A-44. Number of Connections from Emerging Leaders Cohort, 2018-19

Q: With how many people from your Emerging Leaders cohort are you currently in touch professionally?

	Percent of Respondents					
	None	1–5	6–10	More than 10, but not all	All or almost all	*n*
Number of alumni connections	9	67	9	7	7	43

Source: Emerging Leaders follow-up survey, winter 2019.

Exhibit A-45. Ways Current Emerging Leaders Alumni Network Facilitates Leadership and Growth, 2018-19

Q: In what ways does your current network of Emerging Leaders alumni continue to facilitate your leadership and growth?

	Percent of Respondents						
	Not at all	Not very much	Somewhat	Quite a bit	A great deal	Mean	*n*
Leading a team of teachers	15	13	39	23	10	3.0	39
Preparing for next steps in career	15	10	28	26	21	3.3	39
Improving knowledge of DDI	21	18	31	21	10	2.8	39
Analyzing student data	21	26	28	15	10	2.7	39
Supporting corrective instruction	21	18	33	18	10	2.8	39
Leading a school	10	18	39	18	15	3.1	39
Coaching others to lead DDI	15	15	33	26	10	3.0	39
Using tools from ELP	23	18	28	21	10	2.8	39
Engaging in difficult conversations	15	15	33	28	8	3.0	39
Exercising adaptive leadership	13	10	33	36	8	3.2	39

Source: Emerging Leaders follow-up survey, winter 2019.
Scale: 1 = Not at all, 2 = Not very much, 3 = Somewhat, 4 = Quite a bit, 5 = A great deal.

Impacts on Instructional Leadership as Measured by Emerging Leaders' Outcomes on the New Leaders' Data-Driven Instruction Assessment

The New Leaders staff developed the DDI instrument to assess candidates' knowledge of key skills for leading a team of instructors through a DDI cycle. The assessment is given as part of the application process and again as part of the program assignments, enabling New Leaders to assess participant readiness for the program and to provide summative information to district partners on their readiness to assume more formal leadership positions. The DDI assessment used in the Emerging Leaders program had two components: a written assessment of the candidate's knowledge of and ability to execute a cycle of data-driven inquiry, and video documentation of a team meeting to provide evidence of a candidate's ability to lead a team. The video-based portion of the assessment was not used in this study.

As part of the evaluation, Emerging Leaders candidates agreed to complete the written section of the DDI in early 2018 even if randomized into Group II, providing impact data on three key facets of candidates' DDI leadership knowledge:

- New Leaders Leadership Competency 2.2a: Utilize multiple forms of student-level data to drive sustained gains in student achievement.

- New Leaders Leadership Competency 2.2b: Demonstrate knowledge of and the ability to lead a team through the data- driven instruction cycle and adjust instructional practice to meet the needs of all students.
- New Leaders Leadership Competency 3.1c: Demonstrate deep understanding of efficacy concepts

Members of a trained national assessor corps scored the assessments for consistent measures across all program sites.

DDI Leadership Knowledge Assessment

The DDI leadership knowledge assessment measures participants' leadership knowledge through scenario-based planning. New Leaders gave candidates for Emerging Leaders a set of target standards, a sample assessment, and corresponding assessment data. The candidate had to use these artifacts to complete a structured data analysis guide and devise a corrective instruction plan. The data analysis section of the DDI instrument asked candidates to analyze the data provided by standard (i.e., highest and lowest performing), item, and student. The item-level analysis identified two of the DDI assessment's questions and asked respondents to

- describe the rigor of the question (using, e.g., Bloom's Taxonomy, Webb's Depth of Knowledge, or Hess's Cognitive Rigor Matrix);
- identify the correct answer and the skill students require to answer it; and
- explain the misconception represented by each of the distractor items.

The assessment next asked the respondents to analyze three students' data, explaining each student's strengths and target areas for corrective instruction.

The corrective instruction section of the DDI asked respondents to identify aligned high-impact strategies for the whole class, a small group, and an individual, including supports for student efficacy and a plan for re-assessment. Throughout the DDI, respondents were regularly prompted to support and justify their choices with rationales and assessment data where appropriate.

Administration and Scoring

Emerging Leaders in Group I and Group II completed the assessment at two time points. All candidates for the Emerging Leaders program completed the DDI assessment at baseline in spring 2017 as part of the application process. Group I Emerging Leaders participants completed the DDI again as part of their third assignment, knowing the responses would be scored and those scores provided to them as formative and summative evaluations of their performance in the program. The research team asked Group II respondents to complete the DDI at the same time in exchange for a $25 gift card; they were told that New Leaders staff would use their scores as formative assessment on beginning the program in fall 2018. The research team collected the DDI responses of Group II while Group I was completing their third assignment to ensure timing of the assessments was equivalent between the two groups; each group had approximately 5 weeks in early 2018 to complete the activity.

A member of the national assessor corps read the written responses and assigned a score of 1–4 for each competency. For scoring, three tools were used: a rubric with four scale points, a "cheat sheet" with acceptable answers, and sample responses to provide examples of score points 1, 2, 3, and 4. A score of 1 was "Unproductive," a 2 was "Approaching Proficiency," 3 was "Proficient," and 4 was "Advanced." Half points were allowed, creating a 7-point scale. Responses received higher scores when they were not just accurate and complete, but also provided a through level of detail to support the answers.

New Leaders' staff scored the DDI for to provide up-to-date results for formative feedback and provided the responses to the research team. To ensure an unbiased scoring, we de-identified the responses from Group I and Group II and assigned New Leaders assessors to re-score responses, ensuring that no assessor would score the same response twice and that they would be blind to treatment condition. We assigned two New Leaders assessors to each response to allow for an assessment of the interrater reliability of the DDI assessment (Exhibit A-46).

Exhibit A-46. Measures, Detailed Description, and Interrater Reliability

Competency	Full Description of Item	Interrater Reliability
2.2a1	Utilize multiple forms of student-level data to drive sustained gains in student achievement	0.79
2.2b2	Demonstrate knowledge of and the ability to lead a team through the data-driven instruction cycle and adjust instructional practice to meet the needs of all students	0.85
3.1c1	Demonstrate deep understanding of efficacy concepts	0.90

Note. Interrater reliability is calculated by percentage of agreement within half a point scale.

Descriptive Statistics

Exhibit A-47 presents the unadjusted baseline and outcome data for the DDI assessment.

Exhibit A-47. Leadership Outcome Descriptive Statistics

		Baseline			Outcome		
Leadership Competency		Mean	*SD*	*n*	Mean	*SD*	*n*
2.2a1	Group I	2.01	0.50	48	2.39	0.51	48
	Group II	1.90	0.56	36	1.94	0.41	36
2.2b2	Group I	2.03	0.40	48	2.15	0.55	48
	Group II	2.07	0.45	36	1.72	0.37	36
3.1c1	Group I	NA	NA	NA	2.27	0.66	48
	Group II	NA	NA	NA	1.73	0.45	36

Analysis Methods

The baseline Emerging Leaders sample comprised 58 in Group I and 54 in Group II, for a total of 112 participants (Exhibit A-48). At follow up, 84 of the 112 Emerging Leaders took the DDI assessment again, representing an overall attrition of 25% and differential attrition of 16%. These attrition levels exceeded WWC attrition standards under liberal assumptions, introducing the possibility that the two groups were no longer equivalent in expectation, as they had been at the time of randomization.

Exhibit A-48. DDI Attrition

	Emerging Leaders		
	Group I	Group II	Total
Baseline sample	58	54	112
Analytic sample	48	36	84
Attrition (%)	*17*	*33*	*25*
Differential attrition (%)			*16*

To check the equivalence of the analytic samples at baseline, we divided the difference between the mean of the intervention group and the mean of the comparison group by the pooled within-group standard deviation (SD) on that baseline measure (Exhibit A-49). Baseline difference for 2.2a1 is 0.21. Baseline difference for 2.2b2 is -0.09.

Exhibit A-49. Baseline Equivalence of DDI Analytic Samples

Leadership Competency	Description	Baseline Equivalence
2.2a1	Using multiple forms of data to drive student achievement	0.21
2.2b2	Leading a team through a DDI cycle	-0.09
3.1c1	Building understanding of efficacy concepts	N/A, not measured at baseline

Note. Group I *n* = 48, Group II *n* = 36.

Multiple regression models were conducted to estimate the differences between Group I and Group II Emerging Leaders on leadership development (as measured by the DDI assessment). Because the DDI assessment was given only to Emerging Leaders, this analysis has only one level.

Outcome *Y* for Emerging Leader *i* is given as

$$Y_i = \beta_0 + \beta_1 Treatment + \beta_2 EL + \beta_3 S + \beta_4 D + \beta_5 A + \varepsilon_i$$

where:

Treatment = initial random assignment with 1 for intervention and 0 for control.

EL = Emerging Leader participant baseline scores, subject, and grade level.

S = school characteristics including percentage of students who received free or reduced-price lunch, percentage of English learners, percentage of Latinx students, percentage of African American students, percentage of white students, percentage of Asian students, and percentage of Native American students.

D = district dummy variables.

A = assessor dummy variables.

ε_i = residual terms of the model.

The coefficient β_1 associated with *Treatment* in the above hierarchical linear model (HLM) indicates the average treatment effect in improving Emerging Leaders' leadership practice controlling for Emerging Leaders-level covariates. We used Stata 14's *regress* command for continuous outcomes.

Sensitivity analyses were conducted to check the robustness of the impact of Emerging Leaders across different specifications of the models. We compared the treatment coefficients and associated standard errors from nine models including different sets of district and school covariates and block indicators. The results were consistent across different models.

HLM was conducted on the data with and without imputed independent variables, which left three sets of models for each outcome. The first set used listwise deletion models[17] where an entire record is excluded from the analysis if any independent or dependent variable is missing. The second set of HLM used a dummy variable adjustment imputation approach, which sets the missing pretest scores to zero and adds a dummy variable to indicate the missingness of these scores in the impact model (Puma, Robert, Stephen, & Cristofer, 2009). The last set of the HLM used multiple imputation, where the distribution of the observed data is used to estimate multiple values that reflect the uncertainty around the true value. We imputed for missing data on covariates using the EM (expectation-maximization) algorithm for multiple imputation using Stata 14's *mi* command. Multiple imputation inference proceeds in three distinct phases:

1. The missing data are filled in 10 times to generate 10 complete data sets.
2. The 10 complete data sets are analyzed by using HLM procedure (both Emerging Leaders and teacher team members nested in Emerging Leaders teams).
3. The results from the 10 complete data sets are combined for the inference using the *mi* estimate.

[17] Complete cases analysis might have substantial weaknesses. Listwise deletion limits the statistical power of the tests conducted because it uses a reduced sample size with complete cases (Allison, 2001; Olinsky, Chen, & Harlow, 2003; Roth, 1994). Further, if there is systematic difference between the complete cases and incomplete cases, the statistical inference from complete cases analysis may not be applicable to the population of all cases.

Effect sizes are reported as Hedges's *g* and calculated by dividing the intervention indicator coefficient by the pooled standard deviation of Group I and Group II.

Analysis Results

Exhibit A-50 presents the estimated impact of Emerging Leaders on participants' DDI leadership knowledge. Discussion and interpretation of these results can be found in the narrative text in Chapter 3.

Exhibit A-50. Multiple Regression Results on Leadership Outcomes

Leadership Competency	Group I	Group II	Coeff	*SE*	*t*	*p*		Effect Size	*n*
2.2a1	2.34	2.02	0.32	0.11	2.91	0.006	**	0.95	84
2.2b2	2.12	1.75	0.38	0.13	2.94	0.005	**	0.90	84
3.1c1	2.23	1.78	0.45	0.16	2.84	0.006	**	0.94	84

Note: Columns "Group I" and "Group II" present regression-adjusted outcome scores for Group I and Group II, respectively.
~$p < 0.10$, $*p < .05$, $**p < .01$.

To test whether program impacts were stronger for some participants than others, the research team ran a series of subgroup analyses. We estimated the impact of the Emerging Leaders program on each subgroup separately, and then compared the difference in those impacts, testing whether the difference was significantly different from zero. Where the difference was statistically significant, we can conclude that the program had a greater impact on one subgroup than on another. For example, we estimated the impacts of Emerging Leaders leading math teams separately from those of Emerging Leaders leading ELA instructional teams. We next compared these estimated impacts to each other to examine whether Emerging Leaders learned the constructs measured by DDI assessments more thoroughly when applying them to math or to ELA. Exhibits A-51 and A-52 present the results of these comparisons.

Exhibit A-51. HLM Results for Emerging Leaders Participant Subgroups, Math vs. ELA

Leadership Competency	Emerging Leaders Leading Math Instructional Teams				Emerging Leaders Leading ELA Instructional Teams				Difference	
	Group I	Group II	Effect	Sig	Group I	Group II	Effect	Sig	Effect	Sig
2.2a1	2.20	1.98	0.22		2.24	1.82	0.41	**	-0.19	
2.2b2	2.03	1.74	0.29		2.06	1.61	0.46	**	-0.17	
3.1c1	2.36	2.07	0.02		2.17	1.58	0.07	**	-0.05	
Emerging Leader *n*	*18*	*11*			*27*	*21*				

Source: New Leaders data-driven instructional knowledge assessment, 2017–18
~$p < 0.10$, $*p < .05$, $**p < .01$

Exhibit A-52. HLM Results for Emerging Leaders Participant Subgroups, Teachers vs. Non-Teachers

Leadership Competency	Emerging Leaders Who Were Teachers				Emerging Leaders Who Were Not Teachers				Difference	
	Group I	Group II	Effect	Sig	Group I	Group II	Effect	Sig	Effect	Sig
2.2a1	2.51	2.08	0.42	*	2.29	2.03	0.26		0.16	
2.2b2	2.18	1.69	0.49	*	2.02	1.71	0.31	~	0.18	
3.1c1	2.38	1.75	0.06	*	2.15	1.78	0.04		0.02	
Emerging Leader *n*	*25*	*17*			*23*	*19*				

Source: New Leaders data-driven instructional knowledge assessment, 2017–18
~$p < 0.10$, $*p < .05$, $**p < .01$

Impacts on Instruction

The research team designed an instructional log to capture the Emerging Leaders-aligned instructional behaviors teachers engaged in both inside and outside the classroom. The Emerging Leaders program trains and supports participants in leading instructional teams through DDI cycles that identify gaps in student understanding and skills and which students have them so that past content and skills can be revisited and retaught. The instructional process is intended to improve student learning and increase student achievement. Thus, this evaluation required assessing the extent to which participating in an Emerging Leaders instructional team changed teachers' practices in instructional planning and use of student data.

We emailed teachers a link to an online form to document their daily instructional planning behaviors each day over two separate weeks in the spring semester. This allowed teachers to quickly document their behaviors on the same days they did their instructional planning, maximizing the likelihood they would accurately document what they did each day.

The instructional log was aligned with the Emerging Leaders program logic model and a set of related instructional planning behaviors and in-classroom activities. The four areas of instructional practice were instructional planning behaviors, review of assessment data and student work, corrective instruction, and teachers' and students' self-efficacy.

The analysis of instruction addressed 52 final measures. Of these measures, 21 of which were classified as *primary outcomes*—the foci of the analysis. A log measure was classified as a primary outcome if the instructional practice reflected in it should be reflected in the Emerging Leaders program logic model and practices. For example, the Emerging Leaders program is grounded in collaborative DDI cycles, so the log measure "Planned with others" is strongly reflective of the Emerging Leaders program practices. Concentrating on the 21 primary outcomes provided the greatest insight into the instructional practices most relevant to the Emerging Leaders program.

Teacher Outcome Measures

INSTRUCTIONAL PLANNING BEHAVIORS

The first goal of the instructional log was to document the types of instructional planning behaviors teachers engaged in. If the Emerging Leaders program was impactful, we would expect to see three shifts in teachers' instructional planning practices: (1) more teachers would review student work on a weekly basis, (2) more teachers would spend some amount of time collaboratively planning with colleagues on a

weekly basis, and (3) teachers would spend more minutes each week engaged in instructional planning as a result of reviewing more student data and planning with others.

For the purposes of the log, the research team defined "instructional planning" broadly to increase opportunities for teachers to share their practices with us. Virtually any activity related to lesson planning, reviewing standards or curricula, consulting pacing guides, reviewing formative or summative assessment data, designing assessments, reviewing classwork or homework, setting learning targets, working on instructional materials, or grouping students intentionally was considered to be an instructional planning activity for this project. The only activity we explicitly excluded was grading assignments, as the grading itself was not directly informing future instruction. This was relevant mostly because one of the log measurements was asking teachers to report their daily instructional planning minutes, and we did not want teachers to include time spent grading as instructional planning minutes. Finally, we asked teachers whether they instructionally planned alone or if they spent some time planning with others.

Three primary outcomes were specifically associated with this construct.

- Planning: Reviewed student work (any)
- Planned with others (any)
- Total weekly planning time (minutes)

A secondary outcome was "Planning: Reviewed assessments." This specific log item was incorporated into the combined primary outcome of "Planning: Reviewed student work (any)." This instructional log item affected which teachers received questions on the instructional log related to how they engaged with assessment data during their instructional planning.

REVIEW OF ASSESSMENT DATA AND STUDENT WORK

The second goal of the instructional log was to document how teachers reviewed assessment data and student work. The Emerging Leaders program theory of action posits three shifts in teachers' use of data: (1) more teachers would review specific questions/items and incorrect answers when examining data on a weekly basis, (2) more teachers would disaggregate data in some manner on a weekly basis, and (3) more teachers would identify specific student errors and misunderstandings as they reviewed data.

Teachers' use of assessment data and student work could cover a broad range of practices and behaviors. The instructional log focused on three specific areas of data use: which data teachers examined, which students teachers examined data about, and how teachers reviewed these data.

One instructional log question explored which data teachers examined as they reviewed assessments and student work. Teachers can engage in several different levels of review as they examine assessment data and student work. They can examine summary scores and percentages for the entire assignment, review specific question-/item-level scores for specific problems, and review not just how students performed on certain questions, but also what incorrect answers students provided.

There are different ways of reviewing assessment data, and teachers have several options for whom they review data about. They can focus on reviewing all their students' data using averages or other aggregate measures, or they can examine the data for specific classes, periods, or grade levels separately. Other approaches include disaggregating student data based on students' prior performance, other meaningful subgroups (e.g., English learners), or even by individual students.

Finally, the instructional log explored how teachers reviewed these student data. Teachers can review student data to form a general impressions of student mastery, identify specific student errors and

misunderstandings through the examination of assessment data and student work, or use student data to measure progress made toward specific learning goals.

Five primary outcomes were specifically associated with this construct.

- Disaggregated assessment data (any)
- Assessment data: item-level scores
- Assessment data: incorrect answers
- Assessment data: student progress against goals
- Review student work for errors or misunderstandings

CORRECTIVE INSTRUCTION ACTION PLANNING

The third goal of the instructional log was to document how teachers revisited and retaught past content and skills using corrective instruction. If the Emerging Leaders program was impactful, we would expect to see three shifts in how teachers revisited content: (1) more teachers would decide to revisit and reteach past content and skills, (2) more teachers would select instructional strategies specifically designed to address student errors or to convey the material in a new manner, and (3) more teachers would plan to reassess students' understanding using a scorable instrument (i.e., a formal assessment) after checking whether the revisiting was successful.

A core component of the Emerging Leaders program is supporting teachers in identifying student misconceptions and misunderstandings and using that knowledge to take action through revisiting and reteaching the material needed to correct these misconceptions and misunderstandings. A data-driven instructional cycle is incomplete if it does not result in meaningful action to improve student learning. Additionally, revisiting course content is unlikely to be successful if teachers revisit the material in the same way as students have already been exposed to that learning opportunity. Teachers often need to revisit content in new ways, potentially using a strategy specifically selected to unpack a particular student misunderstanding. Finally, it is important for teachers to intentionally assess what students learned. This applies equally to both teaching and reteaching, and the instructional log asked teachers whether they plan to reassess students in addition to reteaching them.

Five primary outcomes were associated with this construct.

- Decided to revisit content
- Corrective instruction: Instructional strategies (any)
- Corrective Instruction: Grouping based on data
- Corrective instruction: Plan to reassess
- Corrective instruction: Skill gaps

TEACHER SELF-EFFICACY AND STRATEGIES TO DEVELOP STUDENT EFFICACY

The instructional log documented teachers' self-efficacy and their support of students in developing self-efficacy. If the Emerging Leaders program were impactful, we would expect that (1) teachers would have higher average reports of their self-efficacy in teaching using various instructional strategies and promoting student engagement and (2) more teachers would use strategies in the classroom to support students in developing their self-efficacy.

The prior constructs of the instructional log concerned instruction and understanding behaviors, which are largely mental or intellectual components of learning. The self-efficacy goal is focused on important affective components of learning: the extent to which teachers believe they are effective in their capacity to use different instructional strategies and engage students in the classroom and the extent to which teachers support students in believing they can be more successful in the classroom. The instructional log asked teachers to assess their own self-efficacy in these areas and also document whether they used strategies in their classrooms that supported students in the development of their self-efficacy. If the teachers reported using strategies in the classroom, the instructional log asked which specific strategies they used to support student self-efficacy, such as having students make action plans to complete their learning goals or conferencing with students about their progress.

In addition to the measures of instructional practice developed for this evaluation, we included two teacher efficacy scales replicated from earlier studies of the relationship between instructional leadership and teacher self-efficacy (Bellibas et al., 2017, Klassen et al, 2009). These studies found that principals' perceived instructional leadership significantly impacts teacher self-efficacy, both efficacy in instruction and efficacy for promoting student engagement. Exhibit A-52 describes these survey scales. Each daily instructional log included one item from one of these two scales, with Friday's log including the remaining items so that teachers received all 8 items over the course of one 6-day week. Responses to a Likert scale ranging from 1 to 5 were combined across items into a mean scale score.

Eight primary outcomes were specifically associated with this construct.

- Today's lesson: Correct, revise, or improve work
- Today's lesson: Review assessment
- Today's lesson: Goals for next assessment
- Today's lesson: Action plan
- Today's lesson: Teacher conference
- Today's lesson: Post-assessment reflection
- Teacher efficacy: Instruction
- Teacher efficacy: Student engagement

Instructional Log Administration

Instructional logs were administered to Groups I and II for two rounds at each site through Qualtrics, an online survey platform. In each round, teachers were to complete the instructional log for 6 days, starting on Sunday and continuing through Friday. The administration dates were as follows.

- AISD: April 1–6 and April 29–May 4
- SAISD: April 1–6 and April 29–May 4
- SCS: March 18–23 and April 8–13

At 3:00 p.m. Central time on each instructional log administration day, the teachers received an email with a link to the instructional log and a reminder to complete it that day. We sent follow-up emails to teachers who did not complete their instructional logs each day to encourage high response rates. Teachers received a final reminder email the Monday after each administration of the instructional logs.

Emerging Leaders participants who were themselves teachers were included in the instructional log administration. Emerging Leaders participants who had another position (instructional coach, assistant principal) were not included because they did not directly engage in daily instruction.

Participants were provided with modest financial incentives for completing full weeks of instructional logs to encourage high response rates.

Reliability Testing

The research team conducted thorough reliability testing to review and revise the instructional log to maximize its value in measuring instructional practices. The results of the reliability testing indicated "moderate agreement." The aggregate teacher-researcher interrater reliability result was 0.51, which meets the WWC reliability standards of outcome measures having 0.50 or greater reliability.

The research team developed the teacher instructional log iteratively. The first draft was based on existing questionnaire items (particularly for the teacher self-efficacy questions) and new items the research team drafted using language aligned with the Emerging Leaders program content. We then tested the reliability of the instructional log by conducting a series of interviews with volunteer teachers at the three sites. We interviewed the teachers on their most recent day of instructional planning, asking them to describe their instructional planning behaviors and thought processes in as much detail as possible. After each interview, both the interviewer and the teacher completed the instructional log. The teacher used his/her own experience to answer the questions in the instructional log, and the interviewer used the interview notes on the teacher's description of his/her instructional planning behaviors. Then the interviewer reviewed both sets of answers and identified log items where they disagreed. The interviewer followed up with the teacher on these items to identify why the teacher had responded as he/she had. This process was repeated many times during February and March, and several instructional log items were revised or combined into aggregate measures to improve the agreement between the research team members and teachers. The final version of the instructional log underwent a round of reliability testing to generate interrater reliability results. The final reliability testing consisted of a more structured series of interviews in which a research team interviewer led and recorded an interview with a teacher and four other researchers and then watched the recorded interview. All six people (teacher, interviewer, four other research team members) completed instructional logs to finalize the reliability of the instructional log.

Interrater reliability results were calculated by two methods. We first calculated agreement among the five research team members and then calculated agreement between the research team members and the teachers. To compare researchers with one another, we used Fleiss' Kappa (Fleiss, 1971). Fleiss' Kappa was an appropriate measure for interrater reliability because there were more than two raters assigning categorical ratings ("yes/no"). The overall Fleiss' Kappa calculation for the instructional log was 0.68, which according to some indicates "substantial agreement," the second highest of six possible agreement ratings (the only higher rating being "almost perfect agreement") (Landis & Koch, 1977). This also meets the WWC reliability standard of having interrater reliability above 0.50.

We used Cohen's Kappa (Cohen, 1960, 1968) to compare the research team members' responses with the teachers' responses. We calculated one kappa result for each research staff-teacher pair and then averaged them together (Light, 1971) to measure the overall research team-teacher interrater reliability. The overall average Cohen's Kappa calculation for the instructional logs was 0.51, which according to some indicates "moderate agreement," the third highest of six possible agreement ratings (the higher ratings being "substantive agreement" and "almost perfect agreement") (Landis & Koch, 1977). This also meets the WWC reliability standard of having interrater reliability above 0.50.

Exhibits A-53 and A-54 above provide per-item interrater reliability results for both the 21 primary outcomes and 31 secondary outcomes of instructional log. Exhibit A-55 provides reliability statistics for the two survey scales included on the logs.

Exhibit A-53. Primary Outcomes, Log Item Text, and Interrater Reliability

Primary Outcome[a]	Log Item Text (Spring 2018)	Research Team Internal Interrater Reliability	Research Team-Teacher Interrater Reliability
Planning: Reviewed student work or assessments (any)	What kinds of planning did you do for ${subject} today? Reviewed assessments (e.g., exit tickets, quizzes, tests, formative assessments, benchmark assessments, pre-assessments) | Reviewed samples of student work (other than exit tickets or assessments)	0.68	0.51
Collaboration: Planned with others (any)	Did you collaborate with other teachers today, either in person or virtually, to plan instruction for ${subject}? A teacher team that meets regularly (e.g., grade level, department, course, PLC) | A group of teachers that collaborates informally | An instructional coach, mentor teacher, department/grade-level chair or assistant principal | A specialist (e.g., special education specialist, English learner specialist, occupational therapist) | A paraprofessional, student teacher, or teaching assistant | Other	1	1
Total weekly planning time (minutes)	How much time did you spend planning today for ${subject}? Enter the number of minutes you spent planning in the box below. Please include all minutes you spent planning with others and by yourself. DO NOT include any minutes you spent grading or prepping instructional materials (e.g., photocopying).	N/A	N/A
Assessment data: Disaggregated data (any)	When reviewing assessment data today for ${subject}, how did you analyze student performance? By class or class period | By student sub-group (e.g., English learners, special education students) | By group according to prior performance (e.g., the lowest-performing quartile, students who scored basic at the beginning of the year) | Scores disaggregated by standard, skill, or topic	0.60	0.50
Assessment data: Item-level scores	When reviewing assessment data today for ${subject}, what types of scores did you consider? Question/item-level scores (e.g., percentage of students responding to a particular item correctly or with proficiency)	0.08	0.00
Assessment data: Incorrect answers	When reviewing assessment data today for ${subject}, what types of scores did you consider? Incorrect answers selected (e.g., in a multiple choice item, the percent of students choosing an incorrect response)	0.08	0.20
Assessment data: Student progress against goals	When reviewing assessment data today for ${subject}, did you do any of the following? Reviewed students' progress against goals for performance on this assessment or similar assessments	0.58	0.18

Primary Outcome[a]	Log Item Text (Spring 2018)	Research Team Internal Interrater Reliability	Research Team-Teacher Interrater Reliability
Assessment data or student work (any): Identified errors or misunderstandings	When you reviewed assessment data today for ${subject}, did you do any of the following? Identified specific student errors or misunderstandings | When you reviewed samples of student work today for ${subject}, did you do any of the following? Identified specific student errors or misunderstandings	0.43	0.33
Corrective instruction: Decided to revisit content	In any of the instructional planning you completed today for ${subject}, did you decide to have students revisit content or skills covered in a previous lesson?	0.27	0.23
Corrective instruction: Selected new instructional strategies (any)	When planning to have student revisit content or skills covered previously in ${subject}, did you do any of the following today? Selected instructional strategies to address specific student errors | Selected new instructional strategies that are different from the way that you previously taught the skills or content to these same students	0.38	0.52
Corrective instruction: Grouping based on data	When planning to have student revisit content or skills covered previously in ${subject}, did you do any of the following today? Grouped students for instruction on the basis of assessment data	0.38	0.61
Corrective instruction: Plan to re-assess	When planning to have student revisit content or skills covered previously in ${subject}, did you do any of the following today? Planned to re-assess students using an instrument that can be scored (e.g., a quiz, test, or a writing sample scored with a rubric)	-0.04	0.07
Corrective instruction: Skill gaps	When planning to have student revisit content or skills covered previously in ${subject}, did you do any of the following today? Planned to address gaps in skills that you had not previously taught (e.g., skills addressed in previous grades' standards)	Too few rating categories	Too few rating categories
Student efficacy: Correct, revise, or improve work (any)	In today's class(es), did students do any of the following? Correct their own or a classmate's work | Revise and improve work in response to feedback (not simply making corrections)	1	0.37
Student efficacy: Review assessment	In today's class(es), did students do any of the following? Review results from a recent assessment	Too few rating categories	Too few rating categories
Student efficacy: Goals for next assessment	In today's class(es), did students do any of the following? Set performance goals for the next assessment	Too few rating categories	Too few rating categories
Student efficacy: Action plan	In today's class(es), did students do any of the following? Make an action plan for achieving learning goals	Too few rating categories	Too few rating categories
Student efficacy: Teacher conference	In today's class(es), did students do any of the following? Conference with you about their progress or learning goals	0.78	0.83

Primary Outcome[a]	Log Item Text (Spring 2018)	Research Team Internal Interrater Reliability	Research Team-Teacher Interrater Reliability
Student efficacy: Post-assessment reflection	In today's class(es), did students do any of the following? Complete a post-assessment reflection sheet	Too few rating categories	Too few rating categories

[a] Some log items were combined to create composites. Composite items are noted through the use of "(any)" in the outcome name.

Exhibit A-54. Secondary Outcomes, Detailed Description, and Interrater Reliability

Secondary Outcome[a]	Log Item Text (Spring 2018)	Research Team Interrater Reliability	Research Team-Teacher Agreement Reliability
Planning: Planned today	Teachers do not always engage in planning instruction (as defined above) every day. At any point today, did you plan instruction for ${subject}?	N/A	N/A
Planning: Standards, curricula, or pacing guides	What kinds of planning did you do for ${subject} today? Reviewed standards, curricula, or pacing guides	0.25	0.33
Planning: Assessments	What kinds of planning did you do for ${subject} today? Reviewed assessments (e.g., exit tickets, quizzes, tests, formative assessments, benchmark assessments, pre-assessments)		
Planning: Samples of student work	What kinds of planning did you do for ${subject} today? Reviewed samples of student work (other than exit tickets or assessments)	0.13	0
Planning: Learning targets	What kinds of planning did you do for ${subject} today? Set learning target(s) (i.e., statement(s) for students about what they should know and be able to do at the end of the lesson)	1	0.5
Planning: In-class work or homework	What kinds of planning did you do for ${subject} today? Designed task(s) for in-class work or homework	-0.04	-0.06
Planning: Assigned students to groups	What kinds of planning did you do for ${subject} today? Assigned students to groups	0.73	0.83
Planning: Adapted tasks	What kinds of planning did you do for ${subject} today? Adapted tasks for groups of students, based on students' prior knowledge or skills	0.15	0.25
Collaboration: Teacher team	Did you collaborate with other teachers today, either in person or virtually, to plan instruction for ${subject}? A teacher team that meets regularly (e.g., grade level, department, course, PLC)	N/A	N/A
Collaboration: Informal group of teachers	Did you collaborate with other teachers today, either in person or virtually, to plan instruction for ${subject}? A group of teachers that collaborates informally	N/A	N/A

Secondary Outcome[a]	Log Item Text (Spring 2018)	Research Team Interrater Reliability	Research Team-Teacher Agreement Reliability
Collaboration: Instructional coach, mentor teacher, or school leader	Did you collaborate with other teachers today, either in person or virtually, to plan instruction for ${subject}? An instructional coach, mentor teacher, department/grade-level chair or assistant principal	N/A	N/A
Collaboration: Resource teacher	Did you collaborate with other teachers today, either in person or virtually, to plan instruction for ${subject}? A specialist (e.g., special education specialist, English learner specialist, occupational therapist)	N/A	N/A
Collaboration: Para or aide	Did you collaborate with other teachers today, either in person or virtually, to plan instruction for ${subject}? A paraprofessional, student teacher, or teaching assistant	N/A	N/A
Assessment data: Benchmark or interim	What assessment data for ${subject} did you review today? Benchmark or interim assessments (e.g., ${benchmark})	Too few rating categories	Too few rating categories
Assessment data: Common assessments	What assessment data for ${subject} did you review today? Common assessments (districtwide, grade-level, or department-wide) not included above	-0.04	0.65
Assessment data: Pre-assessments	What assessment data for ${subject} did you review today? Pre-assessments (e.g., pre-tests, diagnostic assessments)	0.13	0.33
Assessment data: Previous day	Were the assessment data you reviewed today for ${subject} collected: On the most recent instructional day	0.17	0
Assessment data: Past week	Were the assessment data you reviewed today for ${subject} collected: Within the past week	Too few rating categories	Too few rating categories
Assessment data: Past six weeks	Were the assessment data you reviewed today for ${subject} collected: Within the past six weeks	Too few rating categories	Too few rating categories
Assessment data: More than six weeks	Were the assessment data you reviewed today for ${subject} collected: More than six weeks ago	Too few rating categories	Too few rating categories
Assessment data: Grade level	When reviewing assessment data today for ${subject}, how did you analyze student performance? By grade level	Too few rating categories	Too few rating categories
Assessment data: Class period	When reviewing assessment data today for ${subject}, how did you analyze student performance? By class or by class period	Too few rating categories	Too few rating categories
Assessment data: Student sub-group	When reviewing assessment data today for ${subject}, how did you analyze student performance? By student sub-group (e.g., English Learners, special education students)	Too few rating categories	Too few rating categories

Secondary Outcome[a]	Log Item Text (Spring 2018)	Research Team Interrater Reliability	Research Team-Teacher Agreement Reliability
Assessment data: Prior performance	When reviewing assessment data today for ${subject}, how did you analyze student performance? By group according to prior performance (e.g., the lowest-performing quartile, students who scored basic at the beginning of the year)	Too few rating categories	Too few rating categories
Assessment data: Individual student	When reviewing assessment data today for ${subject}, how did you analyze student performance? By scores for each individual student	0.17	0
Assessment data: By standard, skill, or topic	When reviewing assessment data today for ${subject}, what types of scores did you consider? Scores disaggregated by standard, skill, or topic	-0.14	0.08
Assessment data: Identified errors or misunderstandings	When reviewing assessment data today for ${subject}, did you do any of the following? Identified specific student errors or misunderstandings	0.10	0
Student work: Writing samples	What samples of student work (other than exit tickets or assessments) for ${subject} did you review today? Student writing (e.g., essays, journal entries, responses to a writing prompt)	0.40	0.67
Student work: In-class assignments	What samples of student work (other than exit tickets or assessments) for ${subject} did you review today? In-class assignments (including notes, worksheets, annotation)	0.47	0.42
Student work: Homework assignments	What samples of student work (other than exit tickets or assessments) for ${subject} did you review today? Homework assignments	Too few rating categories	Too few rating categories
Student work: General impression	When you reviewed samples of student work today for ${subject}, did you do any of the following? Formed a general impression of students' mastery of content or skills	N/A	N/A
Student work: Identified errors or misunderstandings	When you reviewed samples of student work today for ${subject}, did you do any of the following? Identified specific student errors or misunderstandings	N/A	N/A
Corrective instruction: Strategies to address specific errors	When planning to have student revisit content or skills covered previously in ${subject}, did you do any of the following today? Selected instructional strategies to address specific student errors	0.39	0.42
Corrective instruction: New instructional strategies	When planning to have student revisit content or skills covered previously in ${subject}, did you do any of the following today? Selected new instructional strategies that are different from the way that you previously taught the skills or content to these same students	0.39	0.42

Secondary Outcome[a]	Log Item Text (Spring 2018)	Research Team Interrater Reliability	Research Team-Teacher Agreement Reliability
Student efficacy: Correct own or a classmate's work	In today's class(es), did students do any of the following? Correct their own work or a classmate's work	1	0.20
Student efficacy: Revise and improve work	In today's class(es), did students do any of the following? Revise and improve work in response to feedback (not simply making corrections)	1	0.20

[a] Some log items were combined to create composites. Composite items are noted through the use of "(any)" in the outcome name.

Exhibit A-55 Primary Outcomes, Teacher Efficacy Scales, Item Text, and Reliability

Primary Outcome	Log Item (Spring 2018)	Response Scale	Reliability (Cronbach's alpha)
Teacher efficacy: Instruction	How much can you do to… a) craft good questions for students? b) implement a variety of assessment strategies? c) provide an alternate explanation when students are confused? d) implement alternative strategies in your classroom?	0 = Not much at all 1 = Very little 2 = Somewhat 3 = Quite a bit 4 = A great deal	.80
Teacher efficacy: Student engagement	How much can you do to… a) assist families in helping their children do well in school? b) motivate students who show low interest in school work? c) get students to believe they can do well in school work? d) help students value learning?	0 = Not much at all 1 = Very little 2 = Somewhat 3 = Quite a bit 4 = A great deal	.85

Attrition for Instructional Logs

The assigned sample for instructional logs includes both (1) instructional team members who were verified at baseline and (2) Emerging Leaders who were themselves teachers. We follow WWC cluster-randomized trial standards to calculate clustered attrition. Following this protocol, Emerging Leaders were attrited from the analytic sample if we collected no outcome data from any member of their instructional team. Teachers from the instructional teams of attrited Emerging Leaders were removed from the attrition denominator if their entire cluster's attrition was already accounted for in the Emerging Leaders attrition calculation. We administered instructional logs to 398 teachers at baseline (350 instructional team members and 48 Emerging Leaders who were also classroom teachers). Of these, 20 teachers were members of instructional teams for which no teachers completed a log. Exhibit A-56 therefore shows a total baseline sample size of 375 teachers, after cluster-level attrition is accounted for.

Our analytic sample represents an overall attrition of 13% for Emerging Leaders, with differential attrition of 7%, and an overall teacher-level attrition of 17%, with differential attrition of 3%. The cluster-level attrition levels meets WWC attrition standards under conservative assumptions.

Exhibit A-56. Teacher Instructional Log Analytic Sample Sizes and Attrition

	Cluster-Level Attrition Emerging Leaders Instructional Teams		
	Group I	Group II	Total
Baseline sample size	58	54	112
Analytic sample size	52	45	97
Attrition %	*10*	*17*	*13*
Differential attrition %			*7*
	Teacher-Level Attrition All Teachers[a]		
	Group I	Group II	Total
Baseline sample size, after cluster-level attrition	201	174	375
Analytic sample size	170	143	313
Attrition %	*15*	*18*	*17*
Differential attrition %			*3*

[a] Includes both Emerging Leaders who were themselves teachers and their instructional team members in 2017–18. Teachers whose instructional team attrited are not included.

Teacher Characteristics

Exhibit A-57 presents the number of Emerging Leaders participants and teacher team members available for the final instructional log analytic sample by subject matter. About 29% of instructional team members were on a team whose stated focus did not match their own instructional responsibility.

Exhibit A-57. Teacher Characteristics by Condition, Final Analytic Sample

Characteristics	Emerging Leaders Teams[a]			Instructional Team Members			All Teachers[b]		
	Group I	Group II	Total	Group I	Group II	Total	Group I	Group II	Total
ELA	28	28	56	76	69	145	84	76	160
Math	21	15	36	59	53	112	70	62	132
Other	3	2	5	8	3	11	16	5	21
Total	**42**	**45**	**97**	**143**	**125**	**268**	**170**	**143**	**313**
Emerging Leaders							27	18	45
Non Emerging Leaders							143	125	268
Total							**170**	**143**	**313**
Team teacher subject mismatch	12	13	25	9	20	29	17	27	44

[a] Some Emerging Leaders teams included both ELA and math teachers. To avoid double counting, this column counts the subject of the Emerging Leaders team focus, not the subject teachers considered for the logs. Teacher subject mismatch includes Emerging Leaders who had at least one teacher team member whose subject did not match the team focus.
[b] This column includes both Emerging Leaders and teacher team members.

Statistical Analysis – Intent to Treat

Intent to treat (ITT) is the average effect of the treatment based on the initial treatment assignment regardless of how many participants actually received the treatment. The ITT impact estimate is the expected effect of the Emerging Leaders program when it was implemented in the real world, with less than perfect implementation and dosage. To estimate the differences in instructional planning practices between Group I participants and Group II Emerging Leaders and teacher team members (measured by instructional logs), HLM was performed to take into account that teacher participants were nested in Emerging Leaders teams. Independent variables included constant, teacher-level covariates (such as pretest scores, grade, subject), team-level average pretest score and treatment indicator, and district fixed effect.

Outcome *Y* for teacher *i* in Emerging Leaders team *j* is given as

$$Y_{ij} = \gamma_{00} + \gamma_{01}Treatment_j + \gamma_{10}P_{ij} + \gamma_{02}T_j + \gamma_{03}S_j + \gamma_{04}D_j + \mu_{0j} + e_{ij}$$

where:

$Treatment_j$ = initial random assignment with 1 for intervention and 0 for control.

P_{ij} = teacher participant baseline scores, subject, and grade level.

T_j = team-level mean baseline score.

S_j = school characteristics including percentage of students receiving free or reduced-price lunch, percentage of students designated as limited English proficiency (LEP), percentage of Latinx students, percentage of African American students, percentage of white students, percentage of Asian students, and percentage of Native American students.

D_j = district dummy variables.

μ_{0j} = team random effect.

e_{ij} = teacher random effect.

The coefficient γ_{01} associated with $Treatment_j$in the above HLM indicates the average treatment effect in promoting improved teacher participant outcomes controlling for all teacher-level and team-level covariates. Because all the covariates were grand-mean centered, our estimates predicted differences for an "average" teacher in the sample. We used Stata 14's *mixed* command for continuous outcomes and *meqrlogit* for the binary outcomes. For the logistic HLM, we transformed the coefficients to estimated probability to present in the report but provide the untransformed coefficients in the exhibits here.

Sensitivity analyses were conducted to check the robustness of the impact of Emerging Leaders across different specifications of the models. We compared the treatment coefficients and associated standard errors from nine models including different sets of district and school covariates and block indicators. The results were consistent across different models.

HLM was conducted on the data with and without imputed independent variables, which left three sets of models for each outcome. The first set used listwise deletion models[18] where an entire record is excluded from the analysis if any independent or dependent variable is missing. The second set of HLM used a dummy variable adjustment imputation approach, which sets the missing pretest scores to zero and adds a dummy variable to indicate missing in the impact model (Puma, Robert, Stephen, & Cristofer, 2009). The last set of the HLM used multiple imputation, where the distribution of the observed data is used to estimate multiple values that reflect the uncertainty around the true value. We imputed for missing data on covariates using the EM (expectation maximization) algorithm for multiple imputation using Stata 14's *mi* command. Multiple imputation inference proceeds in three distinct phases.

1. The missing data are filled in 10 times to generate 10 complete data sets.
2. The 10 complete data sets are analyzed by using HLM (both Emerging Leaders and teacher team members nested in Emerging Leaders teams).
3. The results from the 10 complete data sets are combined for the inference using *mi* estimate.

Effect sizes are reported as Hedges's *g* and calculated by dividing the intervention indicator coefficient by the pooled standard deviation of Group I and Group II.

Exhibits A-58 and A-59 describe instructional log primary outcomes and secondary outcomes, respectively, for the teachers in Group I and Group II.

[18] Complete cases analysis might have substantial weaknesses. Listwise deletion limits the statistical power of the tests conducted because it uses a reduced sample size with complete cases (Allison, 2001; Olinsky, Chen, & Harlow, 2003; Roth, 1994). Second, if there is systematic difference between the complete cases and incomplete cases, the statistical inference from complete cases analysis may not be applicable to the population of all cases.

Exhibit A-58. Primary Outcomes Descriptive Statistics, Group I and Group II

Primary Outcome	Group I			Group II		
	Mean[a]	*SD*	*n*[b]	Mean[a]	*SD*	*n*[b]
Planning: Reviewed student work or assessments (any)	0.88	0.33	321	0.82	0.39	269
Collaboration: Planned with others (any)	0.95	0.21	318	0.94	0.24	266
Total weekly planning time (minutes)	255.88	189.46	319	218.61	168.30	265
Assessment data: Disaggregated assessment data (any)	0.68	0.47	317	0.59	0.49	269
Assessment data: Item-level scores	0.46	0.50	315	0.35	0.48	267
Assessment data: Incorrect answers	0.44	0.50	315	0.33	0.47	267
Assessment data: Student progress against goals	0.55	0.50	315	0.44	0.50	268
Assessment data or student work (any): Identified errors or misunderstandings	0.69	0.46	316	0.62	0.49	268
Corrective instruction: Decided to revisit content	0.92	0.28	321	0.88	0.32	268
Corrective instruction: Selected new instructional strategies (any)	0.76	0.43	321	0.70	0.46	268
Corrective Instruction: Grouping based on data	0.42	0.49	321	0.34	0.47	268
Corrective instruction: Plan to re-assess	0.52	0.50	321	0.42	0.49	268
Corrective instruction: Skill gaps	0.38	0.49	321	0.32	0.47	268
Student efficacy: Correct, revise, or improve work (any)	0.51	0.50	316	0.55	0.50	267
Student efficacy: Review assessment	0.40	0.49	316	0.34	0.47	267
Student efficacy: Goals for next assessment	0.09	0.29	316	0.11	0.31	267
Student efficacy: Action plan	0.82	0.38	316	0.80	0.40	267
Student efficacy: Teacher conference	0.81	0.39	316	0.79	0.41	267
Student efficacy: Post-assessment reflection	0.61	0.49	316	0.65	0.48	267
Teacher efficacy: Instruction	3.23	0.65	321	3.15	0.65	269
Teacher efficacy: Student engagement	3.47	0.52	321	3.40	0.52	269

[a] Group estimates are reported as proportions.

[b] The analytic sample size is reported in the number of instructional log weeks available for analysis. We administered 2 weeks of logs to all teachers. Accordingly, some teachers are represented twice in these analyses as these completed two full weeks of logs. Different items have different numbers of weeks as some teachers omitted answers to some questions.

Exhibit A-59. Secondary Outcomes Descriptive Statistics, Group I and Group II

Secondary Outcome	Group I			Group II		
	Mean[a]	*SD*	*n*[b]	Mean[a]	*SD*	*n*[b]
Planning: Planned today	0.98	0.16	321	0.95	0.22	269
Planning: Reviewed standards, curricula, or pacing guides	0.77	0.42	321	0.74	0.44	269
Planning: Assessments	0.78	0.42	321	0.68	0.47	269
Planning: Samples of student work	0.64	0.48	321	0.60	0.49	269

Secondary Outcome	Group I			Group II		
	Mean[a]	*SD*	*n*[b]	Mean[a]	*SD*	*n*[b]
Planning: Learning targets	0.69	0.47	321	0.65	0.48	269
Planning: In-class work or homework	0.77	0.42	321	0.67	0.47	269
Planning: Assigned students to groups	0.45	0.50	321	0.41	0.49	269
Planning: Adapted tasks	0.62	0.97	321	0.59	0.93	269
Collaboration: Teacher team	0.65	0.48	318	0.56	0.50	266
Collaboration: Informal group of teachers	0.35	0.48	318	0.27	0.45	266
Collaboration: Instructional coach, mentor teacher, or school leader	0.27	0.45	318	0.24	0.43	266
Collaboration: Resource teacher	0.08	0.28	318	0.08	0.27	266
Collaboration: Para or aide	0.03	0.17	318	0.03	0.16	266
Assessment data: Benchmark or interim	0.49	0.50	317	0.44	0.50	269
Assessment data: Common assessments	0.36	0.48	317	0.29	0.46	269
Assessment data: Pre-assessments	0.45	0.50	317	0.33	0.47	269
Assessment data: Previous day	0.46	0.50	317	0.41	0.49	267
Assessment data: Past week	0.50	0.50	317	0.45	0.50	267
Assessment data: Past six weeks	0.22	0.42	317	0.22	0.42	267
Assessment data: More than six weeks	0.10	0.37	321	0.11	0.47	269
Assessment data: Grade level	0.29	0.45	317	0.20	0.40	269
Assessment data: Class period	0.46	0.50	317	0.39	0.49	269
Assessment data: Student sub-group	0.20	0.40	317	0.14	0.35	269
Assessment data: Prior performance	0.32	0.47	317	0.27	0.44	269
Assessment data: Individual student	0.58	0.49	317	0.49	0.50	269
Assessment data: By standard, skill, or topic	0.50	0.50	315	0.41	0.49	267
Assessment data: Identified errors or misunderstandings	0.56	0.50	315	0.44	0.50	268
Student work: Writing samples	0.30	0.46	314	0.27	0.44	267
Student work: In-class assignments	0.54	0.50	314	0.49	0.50	267
Student work: Homework assignments	0.20	0.40	314	0.17	0.38	267
Student work: General impression	0.55	0.50	321	0.48	0.50	269
Student work: Identified errors or misunderstandings	0.44	0.50	321	0.38	0.49	269
Corrective instruction: Strategies to address specific errors	0.50	0.50	321	0.50	0.50	268
Corrective instruction: New instructional strategies	0.68	0.47	321	0.59	0.49	268
Student efficacy: Correct own or a classmate's work	0.20	0.40	316	0.23	0.42	267
Student efficacy: Revise own work	0.36	0.48	316	0.39	0.49	267

[a] Group estimates are reported as proportions.

[b] The analytic sample size is reported in the number of instructional log weeks available for analysis. We administered 2 weeks of logs to all teachers. Accordingly, some teachers are represented twice in these analyses as these completed two full weeks of logs. Different items have different numbers of weeks as some teachers omitted answers to some questions.

Primary estimates of the Emerging Leaders intervention impacts were derived from the ITT analyses. Regardless of the level of implementation, these analyses compared all Emerging Leaders and their team

members who were randomly assigned to Group I (intended to receive the treatment) with those who were randomly assigned to Group II.

Exhibits A-60 and A-61 describe the impact estimate on instructional log primary outcomes and secondary outcomes, respectively, from models employing multiple imputation for missing covariates.

Exhibit A-60. Primary Outcomes HLM Results

Primary Outcome	Cluster *n*	Weeks *n*[a]	Coeff	*SE*	GI Est[b]	GII Est[b]	*p*	Sig
Planning: Reviewed student work or assessments (any)	97	590	0.28	0.30	0.89	0.86	0.35	
Collaboration: Planned with others (any)	97	584	0.51	0.48	0.97	0.95	0.29	
Total weekly planning time (minutes)	97	584	33.47	18.85	252.01	218.55	0.08	~
Assessment data: Disaggregated assessment data (any)	97	586	0.10	0.22	0.67	0.65	0.66	
Assessment data: Item-level scores	97	582	0.37	0.21	0.43	0.34	0.08	~
Assessment data: Incorrect answers	97	582	0.53	0.24	0.42	0.30	0.03	*
Assessment data: Student progress against goals	97	583	0.30	0.20	0.53	0.46	0.15	
Assessment data or student work (any): Identified errors or misunderstandings	97	584	0.18	0.21	0.68	0.64	0.40	
Corrective instruction: Decided to revisit content	97	589	0.48	0.47	0.96	0.93	0.30	
Corrective instruction: Selected new instructional strategies (any)	97	589	0.30	0.27	0.79	0.74	0.27	
Corrective Instruction: Grouping based on data	97	589	0.18	0.24	0.38	0.34	0.44	
Corrective instruction: Plan to re-assess	97	589	0.32	0.28	0.51	0.43	0.25	
Corrective instruction: Skill gaps	97	589	0.28	0.21	0.37	0.31	0.18	
Student efficacy: Correct, revise, or improve work (any)	97	583	-0.17	0.20	0.51	0.55	0.39	
Student efficacy: Review assessment	97	583	0.22	0.23	0.38	0.33	0.34	
Student efficacy: Goals for next assessment	97	583	-0.16	0.38	0.08	0.09	0.69	
Student efficacy: Action plan	97	583	0.17	0.31	0.87	0.85	0.58	
Student efficacy: Teacher conference	97	583	0.15	0.25	0.82	0.80	0.54	
Student efficacy: Post-assessment reflection	97	583	-0.34	0.21	0.60	0.68	0.11	
Teacher efficacy: Instruction	97	590	-0.06	0.06	3.17	3.23	0.33	
Teacher efficacy: Student engagement	97	590	-0.01	0.04	3.43	3.45	0.80	

[a] The analytic sample size is reported in the number of instructional logs available for analysis. We administered 2 weeks of logs to all teachers. Accordingly, some teachers are represented twice in these analyses as these completed two full weeks of logs. Different items have different numbers of weeks as some teachers omitted answers to some questions.

[b] Group estimates are usually reported as proportions. For example, 89% of Group I teachers were estimated to review student work on a weekly basis controlling for other variables. Total weekly planning time is reported in instructional minutes. The personal efficacy questions are reported on a Likert scale from 0–4.
~$p < 0.10$, $*p < .05$. $**p < .01$. $***p < .001$.

Exhibit A-61. Secondary Outcomes HLM Results

Secondary Outcome	Cluster *n*	Weeks *n*[a]	Coeff	*SE*	GI Est[b]	GII Est[b]	*p*	Sig
Planning: Planned today			Did not converge					
Planning: Reviewed standards, curricula, or pacing guides			Did not converge					
Planning: Assessments	97	590	0.20	0.23	0.78	0.74	0.38	
Planning: Samples of student work	97	590	0.03	0.21	0.64	0.63	0.90	
Planning: Learning targets	97	590	0.06	0.21	0.69	0.67	0.77	
Planning: In-class work or homework	97	590	0.48	0.23	0.78	0.69	0.04	*
Planning: Assigned students to groups	97	590	0.18	0.23	0.44	0.40	0.44	
Planning: Adapted tasks	97	590	-0.05	0.20	0.39	0.40	0.82	
Collaboration: Teacher team	97	584	0.34	0.25	0.66	0.58	0.22	
Collaboration: Informal group of teachers	97	584	0.36	0.25	0.31	0.24	0.14	
Collaboration: Instructional coach, mentor teacher, or school leader	97	584	0.24	0.24	0.24	0.20	0.32	
Collaboration: Resource teacher	97	584	-0.31	0.43	0.04	0.06	0.46	
Collaboration: Para or aide	97		Did not converge					
Assessment data: Benchmark or interim	97	586	0.01	0.20	0.47	0.46	0.97	
Assessment data: Common assessments	97	586	0.33	0.26	0.33	0.26	0.20	
Assessment data: Pre-assessments	97	586	0.38	0.21	0.43	0.34	0.07	~
Assessment data: Previous day	97	584	-0.06	0.20	0.42	0.44	0.77	
Assessment data: Past week	97	584	0.03	0.23	0.47	0.46	0.89	
Assessment data: Past six weeks	97	584	-0.003	0.27	0.20	0.20	0.99	
Assessment data: More than six weeks			Did not converge					
Assessment data: Grade level	97	586	0.45	0.25	0.25	0.17	0.07	~
Assessment data: Class period	97	586	0.28	0.21	0.45	0.38	0.18	
Assessment data: Student sub-group	97	586	0.71	0.28	0.18	0.10	0.01	*
Assessment data: Prior performance	97	586	0.01	0.23	0.28	0.28	0.97	
Assessment data: Individual student	97	586	0.34	0.23	0.58	0.49	0.14	
Assessment data: By standard, skill, or topic	97	582	0.01	0.25	0.45	0.45	0.96	
Assessment data: Identified errors or misunderstandings	97	583	0.36	0.20	0.55	0.46	0.07	~
Student work: Writing samples	97	581	-0.02	0.26	0.24	0.24	0.93	
Student work: In-class assignments	97	581	-0.0005	0.20	0.52	0.52	1.00	
Student work: Homework assignments	97	581	0.11	0.31	0.15	0.14	0.73	
Student work: General impression	97	590	0.05	0.21	0.53	0.51	0.81	
Student work: Identified errors or misunderstandings	97	590	0.11	0.20	0.42	0.39	0.58	
Corrective instruction: Strategies to address specific errors	97	589	-0.002	0.22	0.50	0.50	0.99	

Secondary Outcome	Cluster *n*	Weeks *n*[a]	Coeff	*SE*	GI Est[b]	GII Est[b]	*p*	Sig
Corrective instruction: New instructional strategies	97	589	0.22	0.24	0.68	0.63	0.37	
Student efficacy: Correct own or a classmate's work	97	583	-0.27	0.25	0.18	0.22	0.28	
Student efficacy: Revise own work		583	-0.10	0.22	0.34	0.37	0.65	

[a] The analytic sample size is reported in the number of instructional log weeks available for analysis. We administered 2 weeks of logs to all teachers. Accordingly, some teachers are represented twice in these analyses as these completed two full weeks of logs. Different items have different numbers of weeks as some teachers omitted answers to some questions.

[b] Group estimates are usually reported as proportions. For example, 89% of Group I teachers were estimated to review student work on a weekly basis controlling for other variables. Total weekly planning time is reported in instructional minutes. The personal efficacy questions are reported on a Likert scale from 0–4.

~$p < 0.10$, $*p < .05$, $**p < .01$. $***p < .001$.

Subgroup Analyses

Moderation analysis provided information on whether the Emerging Leaders had a differential effect for certain subgroups of teacher participants in the study. We tested reasonably sized subgroups defined by teacher subject and team subject matched or not matched with teacher subject to determine whether such subgroups benefited from the intervention. Exhibits A-62 to A-64 present descriptives and sample sizes by subgroup for each primary teacher outcome. We only included primary outcomes in the subgroup analyses.

The hierarchical linear models described above were modified by adding the moderators as covariates and as grand-mean centered interactions with the treatment indicators. The coefficients of the interaction term were tested using Wald's test to determine whether moderation effects were present. Exhibits A-65–A-67 present the results.

Exhibit A-62. Teacher Subgroup Descriptive Statistics: ELA Teachers vs. Math Teachers

Primary Outcome	Group I				Group II			
	Mean	SD	n_{teams}	$n_{log\ weeks}$	Mean	SD	n_{teams}	$n_{log\ weeks}$
ELA Teachers								
Planning: Reviewed student work or assessments (any)	0.89	0.32	29	160	0.8	0.4	30	142
Collaboration: Planned with others (any)	0.97	0.18	29	159	0.96	0.19	30	139
Total weekly planning time (minutes)	226.32	172.44	29	160	218.55	169.47	30	142
Assessment data: Disaggregated assessment data (any)	0.7	0.46	29	158	0.55	0.5	30	142
Assessment data: Item-level scores	0.37	0.48	29	157	0.3	0.46	30	141
Assessment data: Incorrect answers	0.37	0.48	29	157	0.33	0.47	30	141
Assessment data: Student progress against goals	0.55	0.5	29	157	0.38	0.49	30	142
Assessment data or student work (any): Identified errors or misunderstandings	0.66	0.48	29	157	0.58	0.49	30	142
Corrective instruction: Decided to revisit content	0.9	0.3	29	160	0.91	0.29	30	141
Corrective instruction: Selected new instructional strategies (any)	0.74	0.44	29	160	0.75	0.43	30	141
Corrective Instruction: Grouping based on data	0.43	0.5	29	160	0.34	0.48	30	141
Corrective instruction: Plan to re-assess	0.44	0.5	29	160	0.35	0.48	30	141
Corrective instruction: Skill gaps	0.35	0.48	29	160	0.27	0.45	30	141
Student efficacy: Correct, revise, or improve work (any)	0.51	0.5	29	157	0.65	0.48	30	142
Student efficacy: Review assessment	0.5	0.5	29	157	0.35	0.48	30	142
Student efficacy: Goals for next assessment	0.07	0.26	29	157	0.1	0.3	30	142
Student efficacy: Action plan	0.76	0.43	29	157	0.77	0.42	30	142
Student efficacy: Teacher conference	0.78	0.41	29	157	0.77	0.42	30	142
Student efficacy: Post-assessment reflection	0.58	0.5	29	157	0.56	0.5	30	142
Teacher efficacy: Instruction	3.2	0.66	29	160	3.26	0.61	30	142
Teacher efficacy: Student engagement	3.42	0.51	29	160	3.49	0.49	30	142

Primary Outcome	Group I				Group II			
	Mean	SD	n_{teams}	$n_{log\ weeks}$	Mean	SD	n_{teams}	$n_{log\ weeks}$
Math Teachers								
Planning: Reviewed student work or assessments (any)	0.89	0.31	24	132	0.83	0.38	21	117
Collaboration: Planned with others (any)	0.95	0.21	24	130	0.91	0.29	21	117
Total weekly planning time (minutes)	310.37	207.51	24	130	212.14	160.12	21	113
Assessment data: Disaggregated assessment data (any)	0.72	0.45	24	130	0.65	0.48	21	117
Assessment data: Item-level scores	0.58	0.5	24	129	0.41	0.49	21	116
Assessment data: Incorrect answers	0.56	0.5	24	129	0.32	0.47	21	116
Assessment data: Student progress against goals	0.58	0.5	24	129	0.5	0.5	21	116
Assessment data or student work (any): Identified errors or misunderstandings	0.76	0.43	24	130	0.64	0.48	21	116
Corrective instruction: Decided to revisit content	0.95	0.21	24	132	0.85	0.35	21	117
Corrective instruction: Selected new instructional strategies (any)	0.82	0.39	24	132	0.63	0.48	21	117
Corrective Instruction: Grouping based on data	0.43	0.5	24	132	0.32	0.47	21	117
Corrective instruction: Plan to re-assess	0.65	0.48	24	132	0.51	0.5	21	117
Corrective instruction: Skill gaps	0.45	0.5	24	132	0.37	0.48	21	117
Student efficacy: Correct, revise, or improve work (any)	0.5	0.5	24	130	0.43	0.5	21	115
Student efficacy: Review assessment	0.28	0.45	24	130	0.35	0.48	21	115
Student efficacy: Goals for next assessment	0.1	0.3	24	130	0.12	0.33	21	115
Student efficacy: Action plan	0.91	0.29	24	130	0.83	0.37	21	115
Student efficacy: Teacher conference	0.83	0.38	24	130	0.84	0.36	21	115
Student efficacy: Post-assessment reflection	0.67	0.47	24	130	0.77	0.43	21	115
Teacher efficacy: Instruction	3.24	0.66	24	132	3	0.69	21	117
Teacher efficacy: Student engagement	3.5	0.54	24	132	3.29	0.53	21	117
	0.89	0.31	24	132	0.83	0.38	21	117

Exhibit A-63. Teacher Subgroup Descriptive Statistics: Emerging Leaders with Domain Match vs. Emerging Leaders with Domain Mismatch

Primary Outcome	Group I				Group II			
	Mean	SD	n_{teams}	$n_{log\ weeks}$	Mean	SD	n_{teams}	$n_{log\ weeks}$
Emerging Leaders with Domain Match								
Planning: Reviewed student work or assessments (any)	0.88	0.33	40	290	0.83	0.38	32	218
Collaboration: Planned with others (any)	0.95	0.21	40	287	0.95	0.22	32	215
Total weekly planning time (minutes)	263.31	195.59	40	288	211.66	163.07	32	215
Assessment data: Disaggregated assessment data (any)	0.7	0.46	40	286	0.6	0.49	32	218
Assessment data: Item-level scores	0.46	0.5	40	284	0.34	0.47	32	216
Assessment data: Incorrect answers	0.45	0.5	40	284	0.32	0.47	32	216
Assessment data: Student progress against goals	0.56	0.5	40	284	0.44	0.5	32	217
Assessment data or student work (any): Identified errors or misunderstandings	0.69	0.46	40	285	0.63	0.48	32	217
Corrective instruction: Decided to revisit content	0.91	0.28	40	290	0.9	0.3	32	217
Corrective instruction: Selected new instructional strategies (any)	0.77	0.42	40	290	0.71	0.45	32	217
Corrective Instruction: Grouping based on data	0.43	0.5	40	290	0.33	0.47	32	217
Corrective instruction: Plan to re-assess	0.54	0.5	40	290	0.41	0.49	32	217
Corrective instruction: Skill gaps	0.39	0.49	40	290	0.32	0.47	32	217
Student efficacy: Correct, revise, or improve work (any)	0.5	0.5	40	285	0.56	0.5	32	216
Student efficacy: Review assessment	0.41	0.49	40	285	0.34	0.47	32	216
Student efficacy: Goals for next assessment	0.09	0.29	40	285	0.11	0.31	32	216
Student efficacy: Action plan	0.83	0.38	40	285	0.8	0.4	32	216
Student efficacy: Teacher conference	0.81	0.39	40	285	0.8	0.4	32	216
Student efficacy: Post-assessment reflection	0.62	0.49	40	285	0.65	0.48	32	216
Teacher efficacy: Instruction	3.23	0.66	40	290	3.14	0.67	32	218
Teacher efficacy: Student engagement	3.46	0.52	40	290	3.4	0.52	32	218

Primary Outcome	Group I				Group II			
	Mean	SD	n_{teams}	$n_{log\ weeks}$	Mean	SD	n_{teams}	$n_{log\ weeks}$
Emerging Leaders with Domain Mismatch								
Planning: Reviewed student work or assessments (any)	0.87	0.34	12	31	0.76	0.43	13	51
Collaboration: Planned with others (any)	0.94	0.25	12	31	0.9	0.3	13	51
Total weekly planning time (minutes)	186.94	95.12	12	31	248.5	188.01	13	50
Assessment data: Disaggregated assessment data (any)	0.48	0.51	12	31	0.59	0.5	13	51
Assessment data: Item-level scores	0.48	0.51	12	31	0.41	0.5	13	51
Assessment data: Incorrect answers	0.35	0.49	12	31	0.33	0.48	13	51
Assessment data: Student progress against goals	0.42	0.5	12	31	0.45	0.5	13	51
Assessment data or student work (any): Identified errors or misunderstandings	0.65	0.49	12	31	0.55	0.5	13	51
Corrective instruction: Decided to revisit content	0.94	0.25	12	31	0.82	0.39	13	51
Corrective instruction: Selected new instructional strategies (any)	0.74	0.44	12	31	0.65	0.48	13	51
Corrective Instruction: Grouping based on data	0.39	0.5	12	31	0.37	0.49	13	51
Corrective instruction: Plan to re-assess	0.35	0.49	12	31	0.45	0.5	13	51
Corrective instruction: Skill gaps	0.32	0.48	12	31	0.35	0.48	13	51
Student efficacy: Correct, revise, or improve work (any)	0.61	0.5	12	31	0.51	0.5	13	51
Student efficacy: Review assessment	0.29	0.46	12	31	0.35	0.48	13	51
Student efficacy: Goals for next assessment	0.06	0.25	12	31	0.12	0.33	13	51
Student efficacy: Action plan	0.77	0.43	12	31	0.8	0.4	13	51
Student efficacy: Teacher conference	0.81	0.4	12	31	0.75	0.44	13	51
Student efficacy: Post-assessment reflection	0.52	0.51	12	31	0.63	0.49	13	51
Teacher efficacy: Instruction	3.23	0.61	12	31	3.18	0.58	13	51
Teacher efficacy: Student engagement	3.51	0.48	12	31	3.42	0.49	13	51

Exhibit A-64. Teacher Subgroup Descriptive Statistics: Emerging Leaders Participants vs. Teachers on Instructional Teams

Primary Outcome	Group I				Group II			
	Mean	SD	n_{teams}	$n_{log\ weeks}$	Mean	SD	n_{teams}	$n_{log\ weeks}$
Emerging Leaders Participants								
Planning: Reviewed student work or assessments (any)	0.79	0.41	27	53	0.76	0.44	18	33
Collaboration: Planned with others (any)	0.96	0.2	27	51	0.91	0.29	18	33
Total weekly planning time (minutes)	277.02	188.02	27	53	167.22	120.44	17	32
Assessment data: Disaggregated assessment data (any)	0.7	0.46	27	53	0.52	0.51	18	33
Assessment data: Item-level scores	0.6	0.5	27	52	0.3	0.47	18	33
Assessment data: Incorrect answers	0.6	0.5	27	52	0.18	0.39	18	33
Assessment data: Student progress against goals	0.6	0.49	27	53	0.27	0.45	18	33
Assessment data or student work (any): Identified errors or misunderstandings	0.64	0.48	27	53	0.55	0.51	18	33
Corrective instruction: Decided to revisit content	0.94	0.23	27	53	0.79	0.42	18	33
Corrective instruction: Selected new instructional strategies (any)	0.74	0.45	27	53	0.64	0.49	18	33
Corrective Instruction: Grouping based on data	0.58	0.5	27	53	0.27	0.45	18	33
Corrective instruction: Plan to re-assess	0.51	0.5	27	53	0.36	0.49	18	33
Corrective instruction: Skill gaps	0.42	0.5	27	53	0.3	0.47	18	33
Student efficacy: Correct, revise, or improve work (any)	0.7	0.46	27	53	0.73	0.45	18	33
Student efficacy: Review assessment	0.38	0.49	27	53	0.3	0.47	18	33
Student efficacy: Goals for next assessment	0.11	0.32	27	53	0.09	0.29	18	33
Student efficacy: Action plan	0.79	0.41	27	53	0.79	0.42	18	33
Student efficacy: Teacher conference	0.75	0.43	27	53	0.73	0.45	18	33
Student efficacy: Post-assessment reflection	0.62	0.49	27	53	0.73	0.45	18	33
Teacher efficacy: Instruction	3.54	0.44	27	53	3.35	0.59	18	33
Teacher efficacy: Student engagement	3.77	0.36	27	53	3.54	0.41	18	33

Primary Outcome	Group I				Group II			
	Mean	SD	n_{teams}	$n_{log\ weeks}$	Mean	SD	n_{teams}	$n_{log\ weeks}$
Teachers on Instructional Teams								
Planning: Reviewed student work or assessments (any)	0.89	0.31	25	268	0.83	0.38	27	236
Collaboration: Planned with others (any)	0.95	0.22	25	267	0.94	0.23	27	233
Total weekly planning time (minutes)	251.67	189.81	25	266	225.67	172.86	28	233
Assessment data: Disaggregated assessment data (any)	0.68	0.47	25	264	0.61	0.49	27	236
Assessment data: Item-level scores	0.44	0.5	25	263	0.36	0.48	27	234
Assessment data: Incorrect answers	0.41	0.49	25	263	0.35	0.48	27	234
Assessment data: Student progress against goals	0.54	0.5	25	262	0.46	0.5	27	235
Assessment data or student work (any): Identified errors or misunderstandings	0.7	0.46	25	263	0.63	0.49	27	235
Corrective instruction: Decided to revisit content	0.91	0.29	25	268	0.9	0.3	27	235
Corrective instruction: Selected new instructional strategies (any)	0.77	0.42	25	268	0.71	0.45	27	235
Corrective Instruction: Grouping based on data	0.39	0.49	25	268	0.34	0.48	27	235
Corrective instruction: Plan to re-assess	0.52	0.5	25	268	0.43	0.5	27	235
Corrective instruction: Skill gaps	0.38	0.49	25	268	0.33	0.47	27	235
Student efficacy: Correct, revise, or improve work (any)	0.48	0.5	25	263	0.53	0.5	27	234
Student efficacy: Review assessment	0.4	0.49	25	263	0.35	0.48	27	234
Student efficacy: Goals for next assessment	0.09	0.28	25	263	0.11	0.31	27	234
Student efficacy: Action plan	0.83	0.38	25	263	0.8	0.4	27	234
Student efficacy: Teacher conference	0.82	0.38	25	263	0.79	0.4	27	234
Student efficacy: Post-assessment reflection	0.6	0.49	25	263	0.64	0.48	27	234
Teacher efficacy: Instruction	3.17	0.67	25	268	3.12	0.66	27	236
Teacher efficacy: Student engagement	3.41	0.53	25	268	3.38	0.53	27	236

Exhibit A-65. HLM Results for Teacher Subgroup Analysis: ELA Teachers vs. Math Teachers

Primary Outcome	Math Teachers				ELA Teachers				Difference	
	Group I	Group II	Effect	Sig	Group I	Group II	Effect	Sig	Effect	Sig
Planning: Reviewed student work or assessments (any)	did not converge									
Collaboration: Planned with others (any)	0.97	0.93	0.04		0.98	0.98	0.01		-0.03	
Total weekly planning time (minutes)	268.73	210.98	57.75	~	247.54	234.95	12.59		-45.16	
Assessment data: Disaggregated assessment data (any)	0.51	0.62	-0.11		0.79	0.71	0.08		0.19	*
Assessment data: Item-level scores	0.44	0.32	0.13	~	0.40	0.38	0.02		-0.11	
Assessment data: Incorrect answers	0.52	0.24	0.28	***	0.35	0.35	0.01		-0.28	**
Assessment data: Student progress against goals	0.47	0.47	0.00		0.58	0.46	0.12	~	0.13	
Assessment data or student work (any): Identified errors or misunderstandings	0.73	0.71	0.02		0.65	0.59	0.06		0.04	
Corrective instruction: Decided to revisit content	0.97	0.92	0.05		0.95	0.95	0.00		-0.05	
Corrective instruction: Selected new instructional strategies (any)	0.84	0.65	0.19	*	0.78	0.79	-0.01		-0.20	*
Corrective Instruction: Grouping based on data	0.42	0.36	0.06		0.31	0.33	-0.01		-0.08	
Corrective instruction: Plan to re-assess	0.65	0.56	0.09		0.43	0.33	0.10		0.01	
Corrective instruction: Skill gaps	0.41	0.34	0.07		0.36	0.28	0.08		0.01	
Student efficacy: Correct, revise, or improve work (any)	0.50	0.42	0.08		0.52	0.66	-0.14	*	-0.22	*
Student efficacy: Review assessment	0.25	0.37	-0.12		0.49	0.32	0.17	*	0.29	**
Student efficacy: Goals for next assessment	0.10	0.15	-0.04		0.04	0.06	-0.02		0.02	
Student efficacy: Action plan	0.95	0.89	0.07	*	0.76	0.81	-0.04		-0.11	*
Student efficacy: Teacher conference	0.86	0.86	0.00		0.77	0.78	0.00		-0.01	
Student efficacy: Post-assessment reflection	0.60	0.76	-0.16	*	0.63	0.63	0.01		0.17	~
Teacher efficacy: Instruction	3.13	3.11	0.02		3.20	3.30	-0.10		-0.12	
Teacher efficacy: Student engagement	3.50	3.40	0.10		3.37	3.45	-0.08		-0.18	*

~$p < 0.10$, *$p < .05$, **$p < .01$. ***$p < .001$.

Exhibit A-66. HLM Results for Teacher Subgroup Analysis: Emerging Leaders with Domain Match vs. Emerging Leaders with Domain Mismatch

Primary Outcome	Domain Match				Domain Mismatch				Difference	
	Group I	Group II	Effect	Sig	Group I	Group II	Effect	Sig	Effect	Sig
Planning: Reviewed student work or assessments (any)	0.89	0.85	0.04		0.91	0.91	0.00		-0.03	
Collaboration: Planned with others (any)	0.97	0.95	0.02		0.99	0.97	0.02		0.00	
Total weekly planning time (minutes)	251.48	212.61	38.87	~	245.95	269.64	-23.69		-62.56	
Assessment data: Disaggregated assessment data (any)	0.69	0.64	0.05		0.53	0.69	-0.16		-0.21	
Assessment data: Item-level scores	0.42	0.34	0.08		0.55	0.37	0.18		0.10	
Assessment data: Incorrect answers	0.42	0.30	0.12	*	0.40	0.30	0.11		-0.02	
Assessment data: Student progress against goals	0.54	0.46	0.09		0.42	0.46	-0.04		-0.12	
Assessment data or student work (any): Identified errors or misunderstandings	0.69	0.63	0.06		0.68	0.67	0.01		-0.04	
Corrective instruction: Decided to revisit content	0.96	0.93	0.02		0.98	0.92	0.06		0.04	
Corrective instruction: Selected new instructional strategies (any)	0.79	0.74	0.05		0.80	0.72	0.08		0.03	
Corrective Instruction: Grouping based on data	0.37	0.32	0.05		0.48	0.43	0.05		0.00	
Corrective instruction: Plan to re-assess	0.52	0.42	0.11		0.38	0.48	-0.11		-0.21	
Corrective instruction: Skill gaps	0.38	0.30	0.07		0.36	0.34	0.02		-0.06	
Student efficacy: Correct, revise, or improve work (any)	0.50	0.56	-0.06		0.59	0.53	0.07		0.12	
Student efficacy: Review assessment	0.40	0.33	0.07		0.24	0.33	-0.09		-0.16	
Student efficacy: Goals for next assessment	0.08	0.09	-0.01		0.06	0.10	-0.04		-0.03	
Student efficacy: Action plan	0.87	0.85	0.03		0.84	0.87	-0.02		-0.05	
Student efficacy: Teacher conference	0.82	0.82	0.00		0.83	0.70	0.13		0.13	
Student efficacy: Post-assessment reflection	0.61	0.69	-0.07		0.52	0.65	-0.13		-0.05	
Teacher efficacy: Instruction	3.17	3.20	-0.03		3.13	3.36	-0.23		-0.20	
Teacher efficacy: Student engagement	3.43	3.44	-0.01		3.46	3.46	0.00		0.01	

~$p < 0.10$, $*p < .05$, $**p < .01$. $***p < .001$.

Exhibit A-67. HLM Results for Teacher Subgroup Analysis: Emerging Leaders Participants vs. Teachers on Instructional Teams

Primary Outcome	Instructional Team Members				Emerging Leaders				Difference	
	Group I	Group II	Effect	Sig	Group I	Group II	Effect	Sig	Effect	Sig
Planning: Reviewed student work or assessments (any)	0.90	0.87	0.03		0.81	0.79	0.02		-0.02	
Collaboration: Planned with others (any)	0.97	0.96	0.01		0.98	0.95	0.04		0.03	
Total weekly planning time (minutes)	245.66	227.31	18.35		279.04	187.86	91.18	*	72.83	~
Assessment data: Disaggregated assessment data (any)	0.67	0.66	0.01		0.66	0.52	0.14		0.13	
Assessment data: Item-level scores	0.41	0.35	0.06		0.53	0.26	0.27	*	0.21	
Assessment data: Incorrect answers	0.39	0.32	0.07		0.55	0.11	0.44	***	0.38	**
Assessment data: Student progress against goals	0.52	0.49	0.03		0.58	0.23	0.35	**	0.32	*
Assessment data or student work (any): Identified errors or misunderstandings	0.70	0.65	0.05		0.63	0.54	0.09		0.05	
Corrective instruction: Decided to revisit content	0.95	0.95	0.01		0.98	0.82	0.16	*	0.15	*
Corrective instruction: Selected new instructional strategies (any)	0.80	0.76	0.04		0.77	0.62	0.15		0.11	
Corrective Instruction: Grouping based on data	0.34	0.35	-0.01		0.56	0.22	0.34	**	0.34	**
Corrective instruction: Plan to re-assess	0.52	0.45	0.08		0.42	0.30	0.12		0.05	
Corrective instruction: Skill gaps	0.37	0.32	0.05		0.38	0.26	0.13		0.07	
Student efficacy: Correct, revise, or improve work (any)	0.47	0.52	-0.04		0.72	0.78	-0.06		-0.01	
Student efficacy: Review assessment	0.38	0.33	0.05		0.37	0.31	0.06		0.01	
Student efficacy: Goals for next assessment	0.08	0.09	-0.02		0.10	0.09	0.02		0.03	
Student efficacy: Action plan	0.87	0.86	0.01		0.85	0.77	0.08		0.07	
Student efficacy: Teacher conference	0.83	0.81	0.02		0.77	0.73	0.04		0.02	
Student efficacy: Post-assessment reflection	0.61	0.67	-0.07		0.59	0.74	-0.15		-0.08	
Teacher efficacy: Instruction	3.15	3.22	-0.07		3.30	3.31	-0.01		0.06	
Teacher efficacy: Student engagement	3.41	3.45	-0.03		3.56	3.42	0.14		0.18	~

~$p < 0.10$, *$p < .05$. **$p < .01$. ***$p < .001$.

Treatment-on-the-Treated Analysis

Although ITT analysis suggests the average effect of an intervention, it does not yield the effect of the intervention on the teachers who receive it. In this study, we used a simple approach to estimate the effect of treatment on the treated. We created subsets of the data for the teachers who received full implementation of Emerging Leaders and reran the HLM.

We hypothesized that the impact of intervention on instructional log outcomes would be stronger if we restricted the analyses to teachers who experienced the full dosage of the program. For Group I, this sample was created by first removing any Emerging Leaders (and their teacher team members) who left the program before the end of the year and then removing any individual teacher team members who did not attend all or most of the team meetings. For Group II, Emerging Leaders (and their teacher team members) were removed if they left the district before the end of the year, and individual teachers were removed if they would not have been able to attend team meetings. Teacher attendance at team meetings was based on the Emerging Leaders' responses to the research team's contrast survey. Exhibit A-68 demonstrates the impact of the Emerging Leaders program on this restricted sample.

Exhibit A-68. Treatment-on-Treated Effect on Primary Outcomes

Primary Outcomes	Cluster *n*	Weeks *n*[a]	Coeff	*SE*	GI Est[b]	GII Est[b]	*p*	Sig
Planning: Reviewed student work or assessments (any)	76	411	0.70	0.42	0.91	0.83	0.09	~
Collaboration: Planned with others (any)	76	407	0.86	0.69	0.98	0.95	0.21	
Total weekly planning time (minutes)	76	406	18.11	24.31	243.65	225.53	0.46	
Assessment data: Disaggregated assessment data (any)	76	408	0.09	0.29	0.66	0.64	0.75	
Assessment data: Item-level scores	76	407	0.21	0.28	0.39	0.34	0.45	
Assessment data: Incorrect answers	76	407	0.13	0.32	0.38	0.35	0.70	
Assessment data: Student progress against goals	76	406	0.39	0.27	0.54	0.44	0.16	
Assessment data or student work (any): Identified errors or misunderstandings	76	407	0.45	0.29	0.71	0.61	0.12	
Corrective instruction: Decided to revisit content	76	410	0.67	0.54	0.95	0.90	0.22	
Corrective instruction: Selected new instructional strategies (any)	76	410	0.64	0.39	0.82	0.70	0.10	~
Corrective Instruction: Grouping based on data	76	410	0.06	0.28	0.33	0.32	0.83	
Corrective instruction: Plan to re-assess	76	410	0.24	0.33	0.51	0.45	0.45	
Corrective instruction: Skill gaps	76	410	0.49	0.28	0.37	0.27	0.08	~

Primary Outcomes	Cluster *n*	Weeks *n*[a]	Coeff	*SE*	GI Est[b]	GII Est[b]	*p*	Sig
Student efficacy: Correct, revise, or improve work (any)	76	404	-0.30	0.27	0.47	0.54	0.26	
Student efficacy: Review assessment	76	404	0.55	0.32	0.40	0.28	0.09	~
Student efficacy: Goals for next assessment	76	404	-0.27	0.43	0.09	0.11	0.53	
Student efficacy: Action plan	76	404	0.11	0.47	0.88	0.87	0.82	
Student efficacy: Teacher conference	76	404	-0.05	0.33	0.82	0.83	0.89	
Student efficacy: Post-assessment reflection	76	404	-0.24	0.28	0.59	0.65	0.39	
Teacher efficacy: Instruction	76	411	-0.06	0.08	3.14	3.21	0.41	
Teacher efficacy: Student engagement	76	411	-0.01	0.06	3.41	3.42	0.92	

Note: The treatment-on-treated sample is defined as all Emerging Leaders (and their teacher team members) who did not leave the program or district before the end of the 2017–18 school year and teacher team members who attended all or most team meetings.

[a] The analytic sample size is reported in the number of instructional log weeks available for analysis. We administered 2 weeks of logs to all teachers. Some teachers completed instructional logs twice. Different outcomes have different numbers of weeks as some teachers omitted answers to some questions.

[b] Group estimates are reported as a proportions.

~$p < 0.10$, $*p < .05$. $**p < .01$. $***p < .001$.

Impacts on Student Achievement

The three districts in the study provided the evaluation team with student achievement scores in ELA and math on state standardized assessment in 2016–17 and 2017–18. The data from 2017–18 were the primary outcomes of this study. Students who took alternate assessments were not included in the study.

Two of the three districts provided us NWEA MAP assessment scores for 2017–18. We treat the analysis of the NWEA assessment score as exploratory. Below, in Exhibits A-69 and A-70, we describe the measures used, reliability of the measures, and grade levels included in the analysis.

Exhibit A-69. Primary Outcomes Drawn from Three Districts, Detailed Description, and Reliability

Districts	Assessment	Detailed Description	Reliability	Grades Included in Our Analysis
AISD & SAISD	STAAR State Test	State of Texas Assessments of Academic Readiness (STAAR) are required of all public school students in grades 3–12. The STAAR is based on state curriculum standards in core subject areas for different grade levels: • Reading and math, grades 3–8 • Writing, grades 4 & 7	STAAR reading, 0.89 STAAR math, 0.88 (https://tea.texas.gov/Student_Testing_and_Accountability/Testing/Student_Assessment_Overview/Tech	ELA analysis: 3rd- through 8th-graders who took STAAR ELA test and 9th- and 10th-graders who took English I or English II EOC test Math analysis: 3rd- through 8th-graders who took STAAR

Districts	Assessment	Detailed Description	Reliability	Grades Included in Our Analysis
		• Science, grades 5 & 8 • Social studies, grade 8 • End-of-course (EOC) assessments for English I, English II, Algebra I, Biology, and U.S. history, high school grades	nical_Digest_2017-2018/)	math test and 9th-graders who took Algebra I EOC test
SCS	TNReady State Test	TNReady is a part of the Tennessee Comprehensive Assessment Program (TCAP) and is designed to assess true student learning and understanding in core subject areas including ELA, math, science, and social studies for students in grades 3–8 and high school students.	TNReady ELA and math, 0.86–0.93 across different test forms (Questar & Educational Testing Service, 2018)	ELA analysis: 3rd-through 8th-graders who took ELA test and 9th-and 10th-graders who took English I or English II EOC test Math analysis: 3rd-through 8th-graders who took math test and 9th-graders who took Algebra I EOC test

Note: Students who took alternate assessments were not included in the study. AISD 3rd grade students are excluded from all analyses due ot lack of baseline data.

Exhibit A-70. Additional Outcomes Drawn from Two Districts, Detailed Description, and Reliability

Site	Assessment	Detailed Description	Reliability	Grades Included in Our Analysis
SAISD & SCS	NWEA MAP	Tests are designed to measure what students know and inform teachers what students are ready to learn next so they can adjust teaching accordingly for a subject. NWEA MAP has been used to assess students in grades K–12.	0.84	Reading and math: grades 1–11

Attrition

Although randomization should result in statistically equivalent groups, higher overall attrition and differential attrition between Group I and Group II may jeopardize the initial balance (What Works Clearinghouse, 2008). Our data analysis began with an attrition analysis. This RCT had low overall and differential attrition at all three levels (instructional teams, instructional team members, and student) for state standardized assessments in ELA and math (Exhibit A-71).

- Instructional team. For the state ELA sample, the baseline and analytic sample comprised 19 instructional teams in Group I and 17 instructional teams in Group II. For the state math sample, the baseline and analytic sample included 15 instructional team members in Group I and 13 in Group II. Two instructional teams dropped out of Group I and one dropped out of Group II for the state ELA sample. The attrition rate was low for the state ELA sample. None of the instructional teams dropped out for the state math analysis, nor was there attrition at the instructional team level for the state math.

- Instructional team members. The overall attrition was 12% and differential attrition was 10% at the instructional team member level for state ELA. For state math, one instructional team member left the study, with 54 instructional team members remaining in Group I and all 44 instructional team members in Group I remaining. The overall attrition was 1% and differential attrition was 2% at the instructional team member level for the state math sample.
- Student. We calculated student-level attrition following guidelines on page 26 of WWC handbook 4.0: "[I]ndividual-level non-response is always measured within the sample of non-attriting clusters. Individuals in clusters not represented in the analytic sample do not contribute to the reference sample used in the denominator of the individual-level non-response calculation." Overall student attrition was 9% for the state ELA and 7% for the state math outcomes. Differential student attrition for state ELA outcome was 5% but 2% for state math outcomes.

Exhibit A-71 Student State ELA and Math Outcome Analysis Sample Sizes and Attrition

Outcomes	Groups	Units	Assigned Sample	Analytic Sample	Attrition (%)	Differential Attrition (%)
ELA	Group I	Instructional teams	19	17	11	
		Instructional team members	62	57	8	
		Students	3,615	2,907[a]	6	
	Group II	Instructional teams	17	16	6	
		Instructional team members	51	42	18	
		Students	3,695	2,539[b]	11	
	Total	Instructional teams	36	33	8	5
		Instructional team members	113	99	12	10
		Students	7,310	5,446[c]	9	5
Math	Group I	Instructional teams	15	15	0	
		Instructional team members	55	54	2	
		Students	3,629	3,413[d]	6	
	Group II	Instructional teams	13	13	0	
		Instructional team members	44	44	0	
		Students	3,173	2,904[e]	8	
	Total	Instructional teams	28	28	0	0
		Instructional team members	99	98	1	2
		Students	6,802	6,317[f]	7	2

Note: "Assigned sample" includes instructional teams, instructional team members, and students in the early joiner sample in the grade levels that should have both pretest and posttest scores. "Analytic sample" includes students who were identified in fall 2017 and had spring 2018 test scores and their instructional teams and instructional team members. Missing data on student pretest and demographic characteristics were imputed using multiple imputation (MI).

[a] There were 3,105 students in the non-attriting clusters.
[b] There were 2,852 students in the non-attriting clusters.
[c] There were 5,957 students in the non-attriting clusters.
[d] There were 3,625 students in the non-attriting clusters.
[e] There were 3,173 students in the non-attriting clusters.
[f] There were 6,798 students in the non-attriting clusters.

Descriptive Analysis and Baseline Equivalence

After the attrition analysis, a descriptive analysis was conducted for students in the analytic sample. Exhibit A-67 presents the student background characteristics (gender, race, or disabilities), pretest scores, and posttest scores for participants in the intervention and comparison groups. We calculated baseline differences in standardized effect size unit between the two groups following WWC standards 4.0. The two groups were equivalent on both state ELA and math pretest scores. Exhibits A-72 through A-79 describe state standardized assessment outcomes and district benchmark assessment outcomes for students in Group I compared with students in Group II.

Exhibit A-72. Student Demographics of State Standardized Assessment Analytic Sample – ELA

	Group I			Group II			Total		
	Unimputed		Imputed (*n* = 2,907)	Unimputed		Imputed (*n* = 2,539)	Unimputed		Imputed (*n* = 5,446)
Variable	Mean	*N*	Mean	Mean	*N*	Mean	Mean	*N*	Mean
Male	0.52	2907	0.52	0.51	2539	0.51	0.51	5446	0.51
White	0.04	2907	0.04	0.10	2539	0.10	0.07	5446	0.07
Latinx	0.60	2907	0.60	0.42	2539	0.42	0.52	5446	0.52
African American	0.31	2907	0.31	0.40	2539	0.40	0.35	5446	0.35
Other	0.08	2907	0.08	0.17	2539	0.17	0.12	5446	0.12
Special education	0.08	2907	0.08	0.09	2539	0.09	0.09	5446	0.09
ELL	0.27	2907	0.27	0.20	2539	0.20	0.24	5446	0.24
ECO	0.72	2892	0.72	0.65	2524	0.65	0.69	5416	0.69
Grade 3	0.08	2907	0.08	0.04	2539	0.04	0.06	5446	0.06
Grade 4	0.13	2907	0.13	0.05	2539	0.05	0.09	5446	0.09
Grade 5	0.16	2907	0.16	0.04	2539	0.04	0.10	5446	0.10
Grade 6	0.16	2907	0.16	0.17	2539	0.17	0.16	5446	0.16
Grade 7	0.09	2907	0.09	0.45	2539	0.45	0.25	5446	0.25
Grade 8	0.39	2907	0.39	0.04	2539	0.04	0.23	5446	0.23
Grade 9	0.00	2907	0.00	0.22	2539	0.22	0.10	5446	0.10
Grade 10	0.00	2907	0.00	0.00	2539	0.00	0.00	5446	0.00

Note: ELL = English language learner; ECO = economically disadvantaged.

Exhibit A-73. Student Pretest and Posttest Scores of State Standardized Assessment Analytic Sample – ELA

	Group I					Group II					Total				
	Unimputed			Imputed (n = 2,907)		Unimputed			Imputed (n = 2,539)		Unimputed			Imputed (n = 5,446)	
Variable	Mean	*SD*	*N*	Mean	*SE*	Mean	*SD*	*N*	Mean	*SE*	Mean	*SD*	*N*	Mean	*SE*
ELA pretest	-0.43	0.97	2772	-0.44	0.02	-0.24	1.02	2443	-0.25	0.02	-0.34	1.00	5205	-0.35	0.01
ELA posttest	-0.32	0.95	2907	-0.32	0.02	-0.21	0.96	2539	-0.21	0.02	-0.27	0.96	5446	-0.27	0.01

Note: Test scores were standardized within each grade within each state using means and standard deviations of the test score of that particular grade level from the published state data (https://tea.texas.gov/student.assessment/reports/; https://www.tn.gov/education/assessment/tnready.html). The correlation between pretest and posttest using unimputed data sets is 0.77. The mean for pretest for Group I students who had both pretest and posttest scores is -0.24. The mean for pretest for Group II students who had both pretest and posttest scores is -0.43. The standardized baseline difference in state standardized ELA assessment is 0.19.

Exhibit A-74. Student Demographics of State Standardized Assessment Analytic Sample – Math

	Group I			Group II			Total		
	Unimputed		Imputed (*n* = 3,413)	Unimputed		Imputed (*n* = 2,904)	Unimputed		Imputed (*n* = 6,317)
Variable	Mean	*N*	Mean	Mean	*N*	Mean	Mean	*N*	Mean
Male	0.51	3413	0.51	0.52	2904	0.52	0.52	6317	0.52
White	0.06	3413	0.06	0.10	2904	0.10	0.08	6317	0.08
Latinx	0.37	3413	0.37	0.38	2904	0.38	0.38	6317	0.38
African American	0.52	3413	0.52	0.48	2904	0.48	0.50	6317	0.50
Other	0.09	3413	0.09	0.12	2904	0.12	0.10	6317	0.10
Special education	0.09	3413	0.09	0.09	2904	0.09	0.09	6317	0.09
ELL	0.16	3413	0.16	0.13	2904	0.13	0.15	6317	0.15
ECO	0.60	3403	0.60	0.49	2881	0.49	0.55	6284	0.55
Grade 3	0.07	3413	0.07	0.04	2904	0.04	0.05	6317	0.05
Grade 4	0.05	3413	0.05	0.04	2904	0.04	0.05	6317	0.05
Grade 5	0.06	3413	0.06	0.02	2904	0.02	0.04	6317	0.04
Grade 6	0.10	3413	0.10	0.26	2904	0.26	0.17	6317	0.17
Grade 7	0.29	3413	0.29	0.13	2904	0.13	0.22	6317	0.22
Grade 8	0.14	3413	0.14	0.26	2904	0.26	0.19	6317	0.19
Grade 9	0.29	3413	0.29	0.25	2904	0.25	0.27	6317	0.27

Note: ELL = English language learner; ECO = economically disadvantaged.

Exhibit A-75. Student Pretest and Posttest Scores of State Standardized Assessment Analytic Sample – Math

	Group I					Group II					Total				
	Unimputed			Imputed (n = 3,413)		Unimputed			Imputed (n = 2,904)		Unimputed			Imputed (n = 6,317)	
Variable	Mean	*SD*	*N*	Mean	*SE*	Mean	*SD*	*N*	Mean	*SE*	Mean	*SD*	*N*	Mean	*SE*
Math pretest	-0.36	0.90	3229	-0.38	0.02	-0.42	0.90	2728	-0.43	0.02	-0.39	0.90	5957	-0.40	0.01
Math posttest	-0.26	0.86	3413	-0.26	0.01	-0.43	0.80	2904	-0.43	0.01	-0.34	0.83	6317	-0.34	0.01

Note: Test scores were standardized within each grade within each state using means and standard deviations of the test score of that particular grade level from published state data (https://tea.texas.gov/student.assessment/reports/; https://www.tn.gov/education/assessment/tnready.html). The correlation between pretest and posttest using unimputed data set is 0.72. The mean for pretest for Group I students who had both pretest and posttest scores is -0.36. The mean for pretest for Group II students who had both pretest and posttest scores is -0.42. The standardized baseline difference in state standardized math assessment is 0.05.

Exhibit A-76. Student Demographics of NWEA MAP Analytic Sample – ELA

	Group I			Group II			Total		
	Unimputed		Imputed (*n* = 2,111)	Unimputed		Imputed (*n* = 1,259)	Unimputed		Imputed (*n* = 3,370)
Variable	Mean	*N*	Mean	Mean	*N*	Mean	Mean	*N*	Mean
Male	0.50	2111	0.50	0.51	1259	0.51	0.50	3370	0.50
White	0.03	2111	0.03	0.07	1259	0.07	0.05	3370	0.05
Latinx	0.40	2111	0.40	0.24	1259	0.24	0.34	3370	0.34
African American	0.54	2111	0.54	0.65	1259	0.65	0.58	3370	0.58
Other	0.04	2111	0.04	0.09	1259	0.09	0.06	3370	0.06
Special education	0.09	2111	0.09	0.10	1259	0.10	0.09	3370	0.09
ELL	0.11	2111	0.11	0.08	1259	0.08	0.10	3370	0.10
ECO	0.49	2090	0.49	0.38	1242	0.38	0.45	3332	0.45
Grade 1	0.06	2111	0.06	0.02	1259	0.02	0.05	3370	0.05
Grade 2	0.10	2111	0.10	0.05	1259	0.05	0.08	3370	0.08
Grade 3	0.12	2111	0.12	0.08	1259	0.08	0.10	3370	0.10
Grade 4	0.06	2111	0.06	0.00	1259	0.00	0.04	3370	0.04
Grade 5	0.13	2111	0.13	0.02	1259	0.02	0.09	3370	0.09
Grade 6	0.14	2111	0.14	0.21	1259	0.21	0.17	3370	0.17
Grade 7	0.12	2111	0.12	0.18	1259	0.18	0.14	3370	0.14
Grade 8	0.05	2111	0.05	0.08	1259	0.08	0.06	3370	0.06
Grade 9	0.15	2111	0.15	0.31	1259	0.31	0.21	3370	0.21
Grade 10	0.04	2111	0.04	0.00	1259	0.00	0.03	3370	0.03
Grade 11	0.04	2111	0.04	0.05	1259	0.05	0.04	3370	0.04

Note: ELL = English language learner; ECO = economically disadvantaged.

Exhibit A-77. Student Pretest and Posttest Scores of NWEA MAP Analytic Sample – ELA

	Group I					Group II					Total				
	Unimputed			Imputed (*n* = 2,111)		Unimputed			Imputed (*n* = 1,259)		Unimputed			Imputed (*n* = 3,370)	
Variable	Mean	SD	N	Mean	SE	Mean	SD	N	Mean	SE	Mean	SD	N	Mean	SE
ELA pretest	-0.50	1.20	1939	-0.52	0.03	-0.41	1.06	1127	-0.42	0.03	-0.47	1.15	3066	-0.48	0.02
ELA posttest	0.15	0.97	2111	0.15	0.02	0.11	0.93	1259	0.11	0.03	0.13	0.95	3370	0.13	0.02

Note: Test scores were standardized within each grade within each state using means and standard deviations of the test score of that particular grade level from the national norming sample (https://www.nwea.org/resource-library/research/2015-normative-data-3). The correlation between pretest and posttest using unimputed data set is 0.70. The mean for pretest for Group I students who had both pretest and posttest scores is -0.50. The mean for pretest for Group II students who had both pretest and posttest scores is -0.41.

Exhibit A-78. Student Demographics of NWEA Math Analytic Sample – Propensity Score Weighted

	Group I			Group II			Total		
	Unimputed		Imputed (*n* = 2,437)	Unimputed		Imputed (*n* = 1,951)	Unimputed		Imputed (*n* = 4,388)
Variable	Mean	*N*	Mean	Mean	*N*	Mean	Mean	*N*	Mean
Male	0.50	2437	0.50	0.51	1951	0.51	0.51	4388	0.51
White	0.02	2437	0.02	0.06	1951	0.06	0.04	4388	0.04
Latinx	0.24	2437	0.24	0.31	1951	0.31	0.27	4388	0.27
African American	0.71	2437	0.71	0.60	1951	0.60	0.66	4388	066
Other	0.03	2437	0.03	0.07	1951	0.07	0.05	4388	0.05
Special education	0.07	2437	0.07	0.09	1951	0.09	0.08	4388	0.08
ELL	0.07	2437	0.07	0.09	1951	0.09	0.08	4388	0.08
ECO	0.42	2425	0.42	0.36	1927	0.36	0.40	4352	0.40
Grade 1	0.03	2437	0.03	0.00	1951	0.00	0.02	4388	0.02
Grade 2	0.00	2437	0.00	0.01	1951	0.01	0.00	4388	0.00
Grade 3	0.09	2437	0.09	0.06	1951	0.06	0.08	4388	0.08
Grade 4	0.05	2437	0.05	0.02	1951	0.02	0.04	4388	0.04
Grade 5	0.09	2437	0.09	0.03	1951	0.03	0.06	4388	0.06
Grade 6	0.14	2437	0.14	0.37	1951	0.37	0.24	4388	0.24
Grade 7	0.15	2437	0.15	0.18	1951	0.18	0.16	4388	0.16
Grade 8	0.15	2437	0.15	0.15	1951	0.15	0.15	4388	0.15
Grade 9	0.14	2437	0.14	0.15	1951	0.15	0.15	4388	0.15
Grade 10	0.13	2437	0.13	0.01	1951	0.01	0.08	4388	0.08
Grade 11	0.03	2437	0.03	0.00	1951	0.00	0.01	4388	0.01

Note: ELL = English language learner; ECO = economically disadvantaged.

Exhibit A-79. Student Pretest and Posttest scores of NWEA Math Analytic Sample – Propensity Score Weighted

	Group I					Group II					Total				
	Unimputed			Imputed (*n* = 2,437)		Unimputed			Imputed (*n* = 1,951)		Unimputed			Imputed (*n* = 4,388)	
Variable	Mean	SD	N	Mean	SE	Mean	SD	N	Mean	SE	Mean	SD	N	Mean	SE
Math pretest	-0.37	0.96	2231	-0.39	0.02	-0.64	1.05	1784	-0.66	0.02	-0.49	1.01	4015	-0.51	0.02
Math posttest	-0.39	0.97	2437	-0.39	0.02	-0.80	1.03	1951	-0.80	0.02	-0.57	1.02	4388	-0.57	0.02

Note: Test scores were standardized within each grade within each state using means and standard deviations of the test score of that particular grade level from the national norming sample (https://www.nwea.org/resource-library/research/2015-normative-data-3). The correlation between pretest and posttest using unimputed data set is 0.82 . The mean for pretest for Group I students who had both pretest and posttest scores is -0.37. The mean for pretest for Group II students who had both pretest and posttest scores is -0.64.

Intent-to-Treat Analysis

The student outcome analysis reports the ITT effect of the Emerging Leaders program. A set of two-level HLMs[19] was performed to take into account students who were nested in instructional teams. Dependent variables were the standardized assessment scores in ELA or math listed in Exhibits A-80 and A-81. Independent variables were a constant, a pretest score on the same outcome measure, student demographic characteristics, treatment indicator, school characteristics, and district dummy variables.

Outcome *Y* for student *i* in Emerging Leaders team *j* is given as

$$Y_{ij} = \gamma_{00} + \gamma_{01}Treatment_j + \gamma_{10}P_{ij} + \gamma_{02}D_{ij} + \gamma_{03}S_j + \gamma_{04}D_j + \mu_{0j} + e_{ij}$$

where:

$Treatment_j$ = initial random assignment of Emerging Leaders team *j* with 1 for intervention and 0 for control.

P_{ij} = student participant baseline/pretest scores.

D_{ij} = student demographic characteristics, including gender, race/ethnicity, ELL status, economically disadvantaged status, special education status, and grade level.

S_j = school characteristics including percentage of students receiving free or reduced-price lunch, percentage of LEP, percentage of Latinx students, percentage of African American students, percentage of white students, percentage of Asian students, and percentage of native American students.

D_j = district dummy variables.

μ_{0j} = team random effect.

e_{ij} = teacher random effect.

The coefficient γ_{01} associated with $Treatment_j$in the above HLM indicates the average treatment effect in promoting improved student outcomes. Student assessment scores were converted to cross-state comparable *z* scores by subtracting the state mean and dividing by the state standard deviation for each grade and state (May, Perez-Johnson, Haimson, Sattar, & Gleason, 2009). γ_{01} indicated the effect size of the intervention because the outcome scores were *z* scores[20]. All the covariates were grand-mean centered.

HLM was conducted on the data with imputed independent variables using multiple imputation. The multiple imputation model includes an indicator variable for intervention status, all the covariates that are used for statistical adjustment in the impact estimation model, and the outcome when imputing missing baseline data. Multiple imputation inference proceeds in three distinct phases:

1. The missing data are filled in 10 times to generate 10 complete data sets.

[19] We originally fitted a three-level HLM with students nested in instructional team members and instructional team members nested in instructional teams. However, because of the lack of degrees of freedom to estimate random effect at the instructional team member level, we reduced the model to a two level. According to Kim and Frees (2006), omitting an intermediate level creates little bias in the model parameters.

[20] According to WWC 4.0, "[I]f the outcome is a standardized measure that has been administered to a norming sample (national or state), then the effect size may be calculated using the SD from the norming sample." Hedges's *g* effect sizes are calculated by dividing the intervention indicator coefficient by the pooled standard deviation of Group I and Group II, which is equivalent to intervention indicator coefficient.

2. The 10 complete data sets are analyzed by using HLM procedure (both Emerging Leaders and teacher team members nested in Emerging Leaders teams).

3. The results from the 10 complete data sets are combined for the inference using mi estimate.

Sensitivity analyses were conducted to check the robustness of the impact of Emerging Leaders across different specifications of the models. We compared the treatment coefficients from two-level models, three-level models, listwise deletion models, and models that restricted the sample to grade levels that had both Group I and Group II students. The results were consistent across the different models.

Primary estimates of the Emerging Leaders impacts were derived from the ITT analyses. Regardless of the level of implementation, these analyses compared all students in the treatment condition (who were intended to receive the treatment) with those in the control condition. Exhibit A-80 shows the intervention was successful in improving state math achievement ($p < .10$, effect size = 0.24). We did not find a statistically significant effect on state ELA score. Similarly, we found a significant treatment impact on improving NWEA math achievement ($p < .05$, effect size = 0.38) but not on NWEA reading achievement (Exhibit A-81).

Exhibit A-80. HLM Results on Student State ELA and Math Outcomes Combining Three Districts

Outcome	Coeff	*SE*	GI Est	GII Est	*p*	Sig	n_t	n_s
State assessment ELA	0.04	0.08	0.08	0.04	0.6327		33	5446
State assessment Math	0.20	0.11	0.06	-0.14	0.0749	~	28	6317

Note: GI Est = model-adjusted mean outcome for Group I; GII Est = model-adjusted mean for Group II; n_t = number of instructional teams included in the HLM analysis; n_s = number of students included in the HLM analysis.
The HLM results reported here are based on two-level models (students nested in instructional teams). Students' demographic charaterstics and baseline test scores were imputed if there were any missing. Student outcomes were not imputed if there were any missing data. The HLM controlled for student demographic charateristics, student baseline assessment scores, school characteristics, and district dummy variables.
~$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$.

Exhibit A-81. HLM Results on Student NWEA MAP ELA and Math Outcomes Combining Two Districts

Outcome	Coeff	*SE*	GI Est	GII Est	*p*	Sig	n_t	n_s
NWEA MAP ELA	0.06	0.16	0.43	0.37	0.7015		29	3370
NWEA MAP Math	0.39	0.17	-0.02	-0.41	0.0231	*	22	4388

Note: GI Est = model-adjusted mean outcome for Group I; GII Est = model-adjusted mean for Group II; n_t = number of instructional teams included in the HLM analysis; n_s = number of students included in the HLM analysis.
The HLM results reported here are based on two-level models (students nested in instructional teams). Students' demographic charaterstics and baseline test scores were imputed if there were any missing. Student outcomes were not imputed if there were any missing data. The HLM controlled for student demographic charateristics, student baseline assessment scores, school characteristics, and district dummy variables.
~$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$.

Subgroup Analysis

The HLM analysis provided information on whether the Emerging Leaders program had a differential effect for certain student subgroups. We tested reasonably sized subgroups defined by gender,

race/ethnicity, and achievement level to determine whether such subgroups each benefited from the intervention. The HLM models were modified by restricting the sample of students in the subgroup. The coefficients of the treatment indicate the impact of treatment for that subgroup.

The HLM analysis showed that the Emerging Leaders program resulted in improved math achievement of students in the following subgroups: Latinx, ELL, female, and economically disadvantaged (Exhibit A-82).

Exhibit A-82. HLM Results on State Assessments for Student Subgroups

Outcome	Coeff	*SE*	GI Est	GII Est	*p*	Sig	n_t	n_s
White								
State assessment ELA	-0.12	0.20	0.02	0.14	0.5467		24	387
State assessment Math	0.05	0.36	-0.18	-0.23	0.8785		21	497
African American								
State assessment ELA	0.02	0.09	0.05	0.03	0.8614		31	1907
State assessment Math	0.12	0.10	0.03	-0.08	0.2475		27	3181
Latinx								
State assessment ELA	0.06	0.07	0.07	0.00	0.3884		32	2805
State assessment Math	0.32	0.13	0.11	-0.21	0.0129	*	27	2370
ELL								
State assessment ELA	0.04	0.11	0.02	-0.02	0.7088		28	1295
State assessment Math	0.41	0.15	0.20	-0.21	0.0055	**	25	932
Special Education								
State assessment ELA	0.11	0.13	-0.13	-0.24	0.4042		31	467
State assessment Math	-0.02	0.12	-0.33	-0.31	0.8488		28	584
Female								
State assessment ELA	0.05	0.08	0.10	0.05	0.5417		32	2663
State assessment Math	0.22	0.09	0.09	-0.13	0.0147	*	28	3047
Male								
State assessment ELA	0.03	0.08	0.05	0.02	0.7141		33	2783
State assessment Math	0.18	0.15	-0.14	0.11	0.2456		28	3270
Economically Disadvantaged								
State assessment ELA	0.07	0.08	0.14	0.06	0.3914		33	3711
State assessment Math	0.26	0.12	0.06	-0.20	0.0360	*	28	3440
Low Performing (bottom quartile)								
State assessment ELA	0.06	0.10	-0.29	-0.35	0.5278		33	1248
State assessment Math	0.14	0.11	-0.34	-0.48	0.1929		28	1644
High Performing (top quartile)								
State assessment ELA	-0.05	0.11	0.23	0.28	0.6350		32	1493
State assessment Math	0.16	0.12	0.01	-0.15	0.1743		27	1480

Note: GI Est = model-adjusted mean outcome for Group I; GII Est = model-adjusted mean for Group II; n_t = number of instructional teams included in the HLM analysis; n_s = number of students included in the HLM analysis.
The HLM results are based on two level models (students nested in instructional teams). Students' demographic charaterstics and baseline test scores were imputed if there were any missing. Student outcomes were not imputed if there were any missing data. The two-level HLM controlled for student demographic charateristics, student baseline assessment scores, school characteristics, and district dummy variables.
~p < .1,*p < .05, **p < .01, ***p < .001.

Mediation Analyses

Although the ITT analyses suggested the average effect of the intervention, it did not yield the effect of the Emerging Leaders program for those students who actually received it. This study used the instrumental variable (IV) approach to estimate the effect of treatment on the treated (TOT). The IV approach has been used in a few recently published the RCT studies (Angrist & Imbens, 1995; Rouse & Krueger, 2004). Because random assignment is correlated with the fidelity of implementation measures (since treatment is supposed to increase implementation) but uncorrelated with the error term in the outcome equations, the treatment assignment indicator variable works as an instrument to represent fidelity of implementation (Gennetian, Morris, Bos, & Bloom, 2005). A two-stage least-square model was executed to estimate the TOT. The first-stage regressed fidelity measures (whether or not a teacher received the treatment and DDI score) on the random assignment variable. The second stage regressed student outcome on the predicted value of fidelity from the first stage.

Unlike prior estimations, which used SAS to estimate statistical models, we estimated instrumental variables using the command *ivregress 2sls* command in Stata version 14.2, clustering standard errors within Emerging Leaders. Because Stata does not support imputed data for the *ivregress* command, we lose some sample due to lack of imputed data. We lose additional sample for the DDI mediation analyses due to our response rate on the DDI assessment.

TREATMENT-ON-THE-TREATED ANALYSIS

We defined the treatment-on-treated sample (which we describe in the report as receiving the "full dosage" of the program) using three criteria:

1. Emerging Leaders participants remained employed in the district and (if in Group I) successfully completed the Emerging Leaders program,
2. Teachers on instructional teams attended all or most of the instructional team meetings (Group I) or their Emerging Leader reported on the end-of-year survey that they would have been able to meet on an instructional team had they participated in the program in 2017–18 (Group II)
3. Students completed a full year of instruction in the classroom led by one of these teachers.

We then used the IV approach to examine the effect of receiving the treatment on student achievement. The first stage F statistics were significant at $p < .001$, suggesting that treatment assignment was a valid instrument for the TOT variable (Exhibit A-83). The following exhibit shows the impact of the Emerging Leaders program on student achievement when a treatment flag (representing exposure to the "full dosage" of the program as described above) is added to the models as a mediating variable. The effect sizes shown in Exhibit A-83 represent the impact of the program when students receive the program's full treatment as described above.

Exhibit A-83. Treatment-on-the-Treated Effect of Intervention on Student Outcomes

Outcome	2SLS Coeff	*SE*	*p*	Sig	First-stage *F*	First-stage *p*	R^2	n_t	n_s
State assessment ELA	0.08	0.06	0.203		99.13	0.0000	0.65	30	4986
State assessment Math	0.23	0.10	0.029	*	103.40	0.0000	0.77	28	5875

Note: This was a two-stage least squares analysis to estimate the impact of treatment on student outcomes for those teachers who received treatment. Standard errors are clustered at the Emerging Leaders level.

DDI LEADERSHIP KNOWLEDGE

We also used the IV approach to examine the effect of DDI leadership knowledge on student achievement, hypothesizing that, consistent with the program's theory of action, impacts on Emerging Leaders would drive impacts on student achievement.

The three competency measures assessed on the DDI were combined into a single mediating variable. Factor analysis was done to test the single-factor model (Exhibit A-84). Cronbach's alpha was high (above 0.8), and all three factor loadings were similar and considered to be strong (above 0.5) (Boyce & Bowers, 2018; Dunn et al., 2015). Based on this evidence the three measures were combined into a single factor by calculating a simple mean. Using a simple indicator mean to approximate the factor is appropriate when you want to retain the original scaling of the items, which can aid in easier interpretation, as long as you do not have cross-loaded indicators, heavily correlated factors, or indicators with highly varied weights (DiStefano, Zhu & Mindrila, 2009; Grice, 2001). We do not have concerns with cross-loaded indicators or heavily correlated factors as we are using a single-factor model, and the indicator weights are relatively similar. The factor analysis results were consistent when calculated for the treatment and control groups individually.

Only two of the three DDI competency measures were assessed on the baseline application. These two measures were also combined into a single scale. While these two items had smaller factor loadings, supplemental analyses (available from the authors upon request) confirm that the estimated results change only slightly (small increase in the absolute value of the effect sizes, no change in statistical significance) when using each individual score as its own control. We retained the averaged values as the control variables for parsimony.

Exhibit A-84. Characteristics of New Leaders' Aggregate DDI Leadership Knowledge Score

Scale	Competencies	Cronbach's alpha	Factor loadings
DDI Outcome Scale	Using multiple forms of data to drive student achievement	.89	.87
	Leading a team through a DDI cycle		.85
	Building understanding of efficacy concepts		.81
DDI Baseline Scale	Using multiple forms of data to drive student achievement	.54	.51
	Leading a team through a DDI cycle		.51

In a two-state least squares analysis to estimate the impact of DDI leadership knowledge on student achievement, the first stage F statistics were significant at $p < .001$, suggesting that treatment assignment was a valid instrument for the aggregate DDI leadership knowledge variable (Exhibit A-84). The exhibit shows the impact of the Emerging Leaders program on student achievement when the aggregate DDI leadership knowledge score is added to the models as a mediating variable. The effect sizes shown in Exhibit A-85 represent the impact of the program when program participants achieve a full standard deviation difference in aggregate DDI leadership knowledge, compared with the control group (an effect size roughly equivalent to what the program actually achieved).

Exhibit A-85. Instrumental Variables Estimates of the Effect of Aggregate DDI Leadership Knowledge Score on the Student Outcomes Measured by State Assessment

Outcome	2SLS Coeff	*SE*	*p*	Sig	First-stage *F*	First-stage *p*	R^2	n_t	n_s
State assessment ELA	-0.003	0.02	0.887		38.36	0.0000	0.76	23	3765
State assessment Math	0.24	0.11	0.029	*	13.86	0.0012	0.67	23	4805

Note: This was a two-stage least squares analysis to estimate the impact of DDI leadership knowledge score on student outcomes. Standard errors are clustered at the Emerging Leaders level.

Exhibit A-86. Emerging Leaders Evaluation: Student Impacts Effect Sizes, Improvement Indices, and Percentile Equivalents for Math Analyses Using State Assessment Scores

Exhibit No.	Subject	Sample	Model	Effect Size	*Significance*	Instructional Team N	Student N	Improvement Index	Equivalent to moving from the 50th percentile to the… percentile
14	Math	All Students	ITT	0.20	~	28	6317	+8	
15	Math	African American	ITT	0.12		27	3181	+5	
15	**Math**	**Latinx**	**ITT**	**0.32**	*	**27**	**2370**	**+13**	**63rd**
15	Math	White	ITT	0.05		21	497	+2	
15	**Math**	**Female**	**ITT**	**0.22**	*	**28**	**3047**	**+9**	**59th**
15	Math	Male	ITT	0.18		28	3270	+7	
16	**Math**	**Economically Disadvantaged**	**ITT**	**0.26**	*	**28**	**3440**	**+10**	**60th**
16	**Math**	**English Learner**	**ITT**	**0.41**	**	**25**	**932**	**+16**	**66th**
16	Math	Special Education	ITT	-0.02		28	584	-1	
16	Math	High Performing	ITT	0.16		27	1480	+6	
16	Math	Low Performing	ITT	0.14		28	1644	+6	
18	**Math**	**All Students (excluding those with teachers missing data)**	**Mediation: Full dosage**	**0.23**	*	**28**	**5875**	**+9**	**59th**
19	**Math**	**All Students (excluding those with ELs missing data)**	**Mediation: DDI**	**0.24**	*	**23**	**4805**	**+9**	**59th**

Note: Percentile growth provides the best performing of translations of effect sizes into other terms (Baird & Pane, 2019). In addition to relying on fewer assumptions than other post-hoc translations, percentile points are used by WWC to translate effect sizes into an "improvement index." This translation therefore allows for comparison to other rigorous evaluations within the field. In these tables, we supplement the main report's percentile translations for all student impact estimates with statistically significant ($p < .05$) findings on state assessments, our primary outcome measure, with improvement indexes for all student impact analyses. For statistically significant findings, an improvement index of "+1" indicates that the median control student in our study at the 50th percentile of statewide achievement would have moved, on average, to the 51st percentile of statewide achievement.

~$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$. **Bold text** indicates results are statistically significant at least the $p < .05$ level. ITT = Intent-to-Treat.

Exhibit A-87. Emerging Leaders Evaluation: Student Impacts, Improvement Indices, and Percentile Equivalents for English Language Arts Analyses Using State Assessment Scores

Exhibit No.	Subject	Sample	Model	Effect Size	*Significance*	Instructional Team N	Student N	Improvement Index	Equivalent to moving from the 50th percentile to the… percentile
14	ELA	All Students	ITT	0.04		33	5446	+2	
15	ELA	African American	ITT	0.02		31	1907	+1	
15	ELA	Latinx	ITT	0.06		32	2805	+2	
15	ELA	White	ITT	-0.12		24	387	-5	
15	ELA	Female	ITT	0.05		32	2663	+2	
15	ELA	Male	ITT	0.03		33	2,783	+1	
16	ELA	Economically Disadvantaged	ITT	0.07		33	3711	+3	
16	ELA	English Learner	ITT	0.04		28	1295	+2	
16	ELA	Special Education	ITT	0.11		31	467	+4	
16	ELA	High Performing	ITT	-0.05		32	1493	-2	
16	ELA	Low Performing	ITT	0.06		33	1248	+6	
18	ELA	All Students (excluding those with ELs missing data)	Mediation: Full dosage	0.08		30	4986	+3	
19	ELA	All Students (excluding those with ELs missing data)	Mediation: DDI	0.00		23	3765	0	

Note: Percentile growth provides the best performing of translations of effect sizes into other terms (Baird & Pane, 2019). In addition to relying on fewer assumptions than other post-hoc translations, percentile points are used by WWC to translate effect sizes into an "improvement index." This translation therefore allows for comparison to other rigorous evaluations within the field. In these tables, we supplement the main report's percentile translations for all student impact estimates with statistically significant ($p < .05$) findings on state assessments, our primary outcome measure, with improvement indexes for all student impact analyses. For statistically significant findings, an improvement index of "+1" indicates that the median control student in our study at the 50th percentile of statewide achievement would have moved, on average, to the 51st percentile of statewide achievement.

~$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$. **Bold text** indicates results are statistically significant at least the $p < .05$ level. ITT = Intent-to-Treat.

Exhibit A-88. Emerging Leaders Evaluation: Student Impacts, Improvement Indices, and Percentile Equivalents for Math Analyses Using NWEA MAP Scores, Combining Two Districts

Exhibit No.	Subject	Sample	Model	Effect Size	*Significance*	Instructional Team N	Student N	Improvement Index	Equivalent to moving from the 50th percentile to the… percentile
17	**Math**	**All students in NWEA MAP sample**	**ITT**	**0.39**	*****	**22**	**4388**	**+15**	**65th**
17	ELA	All students in NWEA MAP sample	ITT	0.06		29	3370	+2	

Note. These analyses draw from students in two districts, SAISD and SCS, in grades 2–11 whose teachers administered the NWEA MAP assessment. As we note in the report, we found two issues with these data. First, only two of the three districts administered the assessment. Second, the student data for the NWEA MAP analyses contained significant baseline differences between the treatment students and control students which were not present in the state assessment data. We weighted the NWEA MAP data appropriately to establish baseline equivalence before conducting impact analyses. Finally, although we evened the weighting, it is possible that unobserved factors may have contributed to any measured impact of the Emerging Leaders program. These significant concerns did not exist with the state assessment main findings: baseline equivalence was established, and all three participating sites were represented. We therefore provide these results both in the interest of transparency and as supplemental findings to our main findings, the state assessment analyses.

Percentile growth provides the best performing of translations of effect sizes into other terms (Baird & Pane, 2019). In addition to relying on fewer assumptions than other post-hoc translations, percentile points are used by WWC to translate effect sizes into an "improvement index." This translation therefore allows for comparison to other rigorous evaluations within the field. In these tables, we supplement the main report's percentile translations for all student impact estimates with statistically significant ($p < .05$) findings on state assessments, our primary outcome measure, with improvement indexes for all student impact analyses. For statistically significant findings, an improvement index of "+1" indicates that the median control student in our study at the 50th percentile of statewide achievement would have moved, on average, to the 51st percentile of statewide achievement.

~$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$. **Bold text** indicates results are statistically significant at least the $p < .05$ level.

SRI Education™

SRI Education, a division of SRI International, is tackling the most complex issues in education to identify trends, understand outcomes, and guide policy and practice. We work with federal and state agencies, school districts, foundations, nonprofit organizations, and businesses to provide research-based solutions to challenges posed by rapid social, technological and economic change. SRI International is a nonprofit research institute whose innovations have created new industries, extraordinary marketplace value, and lasting benefits to society.

Silicon Valley
(SRI International headquarters)
333 Ravenswood Avenue
Menlo Park, CA 94025
+1.650.859.2000

education@sri.com

Washington, D.C.
1100 Wilson Boulevard
Suite 2800
Arlington, VA 22209
+1.703.524.2053

www.sri.com/education