Evaluation of TNTP's Turnaround Leadership Teams Strategy, 2018–2022

Final Report

August 30, 2023

Naihobe Gonzalez, Alicia Demers, Marykate Zukiewicz, Francisco Yang, and Dan Welsh

Submitted to:

TNTP
500 7th Avenue, 8th Floor
New York, NY 10018
Attention: Cassandra Coddington, Project Officer

Submitted by:

Mathematica
1100 First Street, NE, 12th Floor
Washington, DC 20002-4221
Phone: (202) 484-9220
Fax: (202) 863-1763
Project Director: Marykate Zukiewicz

This page has been left blank for double-sided copying.

Acknowledgements

This study would not have been possible without the contributions of many individuals and organizations. First and foremost, we thank the TNTP staff who provided essential information about their program over the course of the study. We are also grateful for the cooperation of the school districts, schools, teachers, and students who participated in the study.

The study also benefited from the contributions of many people at Mathematica. Allison McKie provided valuable input on the study design and thoughtful reviews of the final report. Virginia Knechtel and Emilyn Whitesell also provided excellent early leadership on project design. Survey Director Kristina Rall led the study's data collection effort, with assistance from Megan McIntyre, Ryan Callahan and Ae Sengmavong. Ruby Steedle provided excellent research and programming assistance. Tori Rockwell and Karina Edouard supported collection and analysis of implementation data. Sheryl Friedlander and Sharon Clark provided expert production support.

This page has been left blank for double-sided copying.

Contents

Tables

Figures

This page has been left blank for double-sided copying.

Executive Summary: The Turnaround Leadership Teams Strategy, 2018–2022

In fall 2017, TNTP received a $7.8 million Mid-phase Education Innovation and Research (EIR) grant from the U.S. Department of Education to implement its Turnaround Leadership Teams Strategy (TLTS) in two large school districts in the southeast United States. TLTS is a school leadership professional development model that provides two years of training and support to cohorts of schools. Under the EIR grant, TNTP worked with the districts to implement a tailored version of the TLTS program in 15 schools across three cohorts starting in the 2018–2019 school year.

The TLTS program aimed to support professional growth for school leaders and teacher leaders; improve instructional practices, school culture, and family engagement in school activities; and, ultimately, increase student achievement. As a core element of the TLTS program's collective leadership approach, each participating school established a school leadership team (SLT) that engaged in program activities, including trainings and ongoing coaching sessions. SLTs typically included principals, assistant principals, instructional coaches, instructional support specialists and teacher leaders.

Through the TLTS program, TNTP staff provided customized supports to each SLT through two key program components: (1) diagnosing school needs (related to academics and instruction, community engagement, and school culture) and (2) developing and supporting SLTs (through ongoing coaching aimed at developing a school vision and improving leadership competencies). Supports included a three-day leadership training the summer before the first year of implementation and two years of ongoing instructional coaching, which was based on a diagnostic assessment conducted in the fall of the first year and aligned with each school's support plan.

A. Evaluation overview

The evaluation included an impact study and an implementation study. For the impact study, we used a randomized controlled trial design to provide rigorous evidence of TLTS program impacts as implemented under the EIR grant on student achievement, with a program emphasis on Math and ELA, but also considering potential impacts on science and social studies where the data was available. In collaboration with the two participating districts, TNTP and Mathematica used school performance data to identify 28 elementary and middle schools considered most likely to benefit from the program. Mathematica then used matched paired random assignment to establish an intervention group of schools that would participate in TLTS and a comparison group of schools that would proceed as usual without TNTP supports. Finally, we collected student administrative data and teacher survey data to compare outcomes between TLTS and comparison schools.

Schools participated in TLTS for two school years. The impact study's central confirmatory research question therefore focused on student outcomes after two years of implementation.

- **Confirmatory Research Question:** What is the impact of two years of the TLTS program compared to business as usual on students' English language arts (ELA), math, science, and social studies achievement?

The impact study also aimed to answer three exploratory research questions, which could offer insight into the confirmatory analysis findings.

- **Exploratory Research Question 1:** What is the impact of one year of the TLTS program compared to business as usual on students' achievement in ELA, math, science, and social studies?
- **Exploratory Research Question 2:** Did the TLTS program have varying impacts on student achievement across cohorts of participating schools?
- **Exploratory Research Question 3:** What is the impact of two years of the TLTS program on teacher perceptions and practices compared to business as usual?

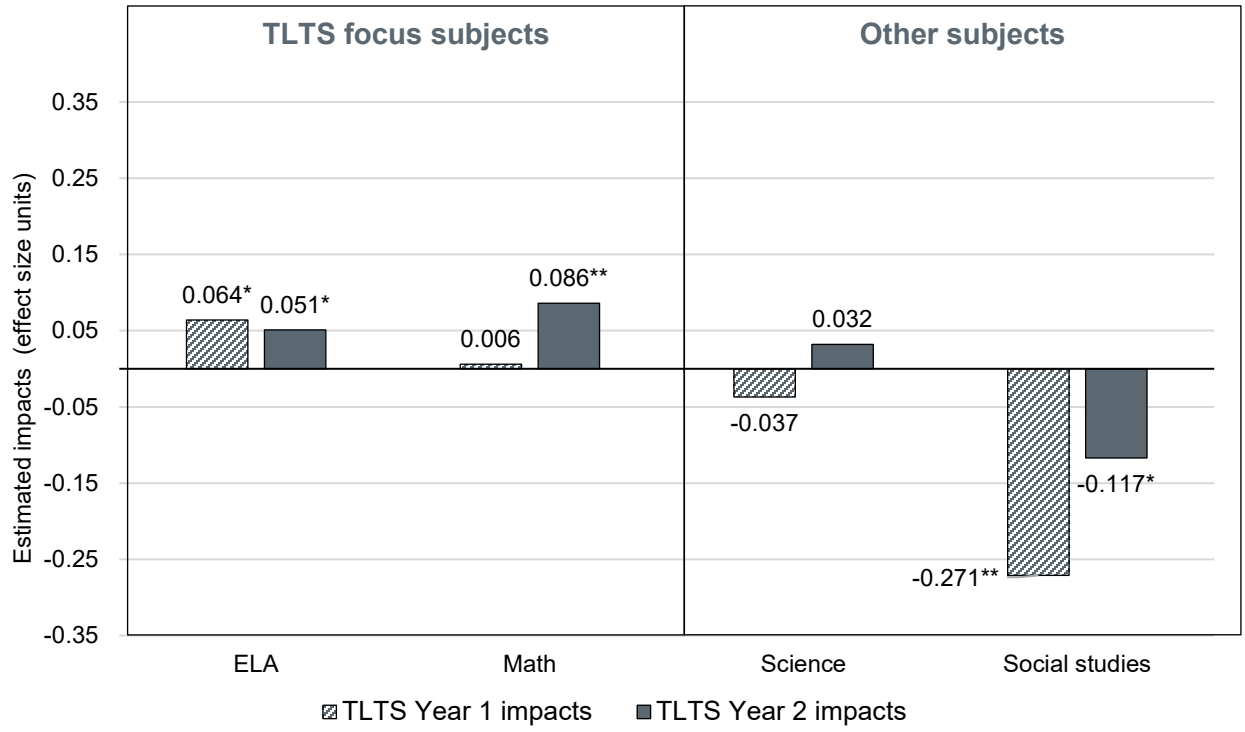
In addition to these impact analyses, the implementation analysis describes the extent to which the program was implemented with fidelity to the program model. We also provide assessments of (a) fidelity to scale-up plans as the program expanded across districts and (b) the program's cost per student.

B. Key findings

After two years of implementation, TLTS had positive effects on student achievement in ELA and math, the two subject areas of focus for TLTS program supports. TLTS schools selected a subject area—either ELA or math—to receive coaching and supports. After two years of TLTS program implementation, students in TLTS schools, on average, performed 0.05 standard deviations higher in ELA than students in comparison schools (p-value = 0.019). Given the growth in ELA that students in the same grades typically demonstrate over an academic year, this represents roughly 1.5 additional months of learning in the subject (Bloom et al. 2008). In math, students in TLTS schools performed, on average, 0.09 standard deviations higher than students in comparison schools (p-value < .001), representing about two additional months of learning.

After two years of implementation, TLTS did not show positive effects on student achievement in other classroom subjects that were not a specific focus of TLTS supports. After two years, TLTS had no detectable effect on student achievement in science and had a negative, statistically significant effect in social studies. The estimated impact on science achievement of 0.03 standard deviations was not statistically significant (p-value = 0.266). On the social studies assessment—administered only to students in grade 8—students in TLTS schools performed, on average, 0.12 standard deviations lower (p-value = 0.030) than students in comparison schools, representing about four months less learning.

Figure ES.1. TLTS impacts on student achievement after one year and two years



Source: District administrative data.

Notes: The study includes students in grades 4–8 who attended study schools in the fall after random assignment. See Table II.2 for summary of outcomes years for each cohort. Estimates of the TLTS impact pool all three school cohorts, control for baseline covariates, and include weights to account for unequal populations across schools. Student analytic sample sizes are as follows: Year 1 ELA = 5,165; Year 2 ELA = 4,322; Year 1 math = 5,113; Year 2 math = 4,288; Year 1 science = 853; Year 2 science = 2,006; Year 1 social studies = 569; and Year 2 social studies = 1,043. Scale scores were converted to z-scores. A z-score of zero indicates a score equal to the mean for all students who took that test (statewide for the state assessment and nationally for the district benchmark assessments) in the same subject, grade, and year. A negative z-score indicates the score was below this mean.

* Impact estimate is statistically significant at the 5 percent level.

** Impact estimate is statistically significant at the 1 percent level.

ELA = English language arts; TLTS = Turnaround Leadership Teams Strategy.

After one year of implementation, TLTS had a positive effect on ELA achievement, whereas the positive effect on math achievement did not emerge until Year 2 (Figure ES.1). The Year 1 exploratory analysis results reflect the emphasis of district and state mandates to improve student achievement in the core subjects of ELA and math as well as the focus of the TLTS program on those two subjects. Most TLTS schools selected ELA as their main focus area for support in Year 1, which may inform Year 1 impacts in that subject. In Year 2, several schools expanded their focus to also include math, which is consistent with impacts in that subject beginning in Year 2. The results also suggest that tradeoffs between core and non-core subjects occurred in both implementation years; however, these tradeoffs may have diminished over time.

TLTS impacts on ELA and math achievement after two years varied by study cohort, generally improving in Cohorts 2 and 3. Schools in Cohort 1—which began implementation in summer 2018—had less favorable results compared to the other cohorts; after two years of implementation, they showed a negative impact on ELA achievement and a null impact on math achievement. Differences in program effectiveness across cohorts could be due to (1) differences between the schools in each cohort, (2) differences in the type and timing of the assessments used in the analyses (Cohort 1 impacts are based on district benchmark tests taken in the winter; Cohort 2 and Cohort 3 impacts are based on the high-stakes state test taken in the spring), or (3) differences in the TLTS program over time. Of note, TNTP provided additional direct supports to schools in 2020 in response to the disruptions caused by the pandemic.

After two years of school participation in TLTS, there was no evidence of effects on self-reported teacher perceptions and practices, although findings should be interpreted with caution. We did not detect a TLTS program impact on teachers' perceptions or practices. However, there were high levels of teacher attrition and survey nonresponse—which was highest among teachers in comparison schools—and we were unable to establish the baseline equivalence of the analytic sample. Therefore, it is possible that the teachers who responded to the survey differed between TLTS and comparison schools, such that these findings may not reflect the true impact of the TLTS program.

TNTP maintained fidelity to its program model during the first year of implementation, which was measured for Cohorts 2 and 3 only. Across schools that began TLTS in 2019–2020 (Cohort 2) and in 2020–2021 (Cohort 3), TNTP achieved fidelity for the program model's two key program components: (1) diagnosing school needs and (2) developing and supporting SLTs.[1] TNTP's fidelity to its model is notable given the challenges the COVID-19 pandemic created for participating schools; the learning environment remained virtual in spring 2020 and was intermittently virtual for specific schools and classrooms in 2020–2021.

Per-pupil costs for the TLTS program averaged $242 per year. For schools engaged in TLTS for the full two years of the program, the average cost per student was $484.40, or $242.20 per student for a single year. These costs reflect expenditures directly related to program provision. Per-pupil costs varied, ranging from a low of $183.88 in 2019–2020, to a high of $446.40 in 2018–2019.

[1] Implementation fidelity was not measured in Cohort 1's first year (2018–2019), as fidelity measures were still in development that year.

I. Introduction

TNTP is a national nonprofit with a mission to help school systems end educational inequity by attracting and training talented teachers and school leaders, ensuring rigorous and engaging classrooms, and creating learning environments that prioritize effective teaching and accelerate student learning.[2] Founded in 1997, TNTP has partnered with more than 10,000 public school districts, charter school networks, and state departments of education in the last four years. TNTP has also recruited or trained more than 41,000 teachers in the last four years; launched the Bridge Fellowship, a program for diverse school leaders; and published a series of policy reports designed to support and develop change in education policy, such as The Opportunity Myth (TNTP 2018a) and The Widget Effect (Weisberg et al. 2009).

In fall 2017, TNTP received a $7.8 million Mid-phase Education Innovation and Research (EIR) grant from the U.S. Department of Education to implement its Turnaround Leadership Teams Strategy (TLTS), a whole-school reform model that provides training and support to cohorts of principals, school leadership teams (SLTs), and teachers at schools in two large school districts in the southeast United States. Under the EIR grant, TNTP worked with these districts to implement a tailored version of its TLTS program.

Under the EIR grant, TNTP added a new cohort of TLTS schools each program year while continuing supports to schools already enrolled in the program. In 2018–2019, TNTP launched the TLTS program in three schools in the first of the two school districts participating in the TLTS program through EIR grant funding. In the 2019–2020 school year, TNTP implemented the program in three additional schools in the first district and three schools in the second district. In 2020–2021, TNTP began working with one final cohort of schools, including one school in the first district and five schools in the second district.

TNTP contracted Mathematica to conduct an independent evaluation of the effectiveness of the EIR-funded program, a requirement for all EIR grantees. This evaluation utilizes a rigorous random assignment design. In collaboration with the two participating districts, TNTP and Mathematica used school performance data to identify the elementary and middle schools in the participating districts most likely to benefit from program participation. From these identified schools, Mathematica randomly selected an intervention group of schools to participate in TLTS and a comparison group of schools to proceed as usual (without TNTP supports). After two years of program implementation, to evaluate the impact of the TLTS program, Mathematica collected and analyzed data from each school on (a) student achievement after one and two years of program implementation and (b) teacher perceptions and practices after two years of program implementation. This report presents the final findings from this impact evaluation. It also includes findings related to analyses of implementation fidelity, scale-up fidelity, and per-student costs.

A. Previous evidence on school leadership programs

School districts across the country face the complex challenge of attracting, training, and retaining leaders with the skillsets needed to transform schools (Béteille et al. 2011). The need for highly qualified leaders creates equity concerns across many districts, as schools serving predominantly low-income students (as measured by student enrollment in the National School Lunch Program) face more challenges in attracting and retaining effective leaders (Branch et al. 2013). Schools have a hard time attracting qualified leaders and keeping them. In 2017, 21 percent of school leaders in high-poverty schools

[2] TNTP is previously known as The New Teacher Project.

nationwide left their positions, compared to 15 percent of school leaders in schools with lower poverty (Goldring and Taie, 2018).

Ample evidence links the quality of school leadership to student outcomes. This relationship is even more pronounced in high-poverty, low-performing schools (Branch et al. 2013; Robinson et al. 2008; Waters et al. 2003; Levin et al. 2020). In fact, the influence of school leaders is second only to teacher influence when it comes to student achievement (Leithwood et al. 2008). Numerous studies have established a relationship between increased principal effectiveness and gains in student achievement and high school graduation rates (Levin et al. 2020). Some interventions focused on improving principal leadership have had promising effects on student achievement. For example, an evaluation of the New Leaders for New Schools (NLNS) program in the Oakland Unified School District found that students in schools with principals who engaged in the NLNS school leadership program for at least three years outperformed students in other district schools by an average of 0.06 standard deviations in English language arts (ELA) and 0.16 standard deviations in math (Booker and Thomas 2014). School leadership teams—which distribute leadership beyond principals to engage teachers in instructional and other decisions—are also a promising leadership model. This approach has been linked to higher student achievement and improved teacher effectiveness, satisfaction, and retention (Booker and Glazerman 2009; Hallinger and Heck 2010; Heck and Hallinger 2009; Leithwood and Mascall 2008; Mascall et al. 2008; Supovitz and Riggan 2012). Evidence suggests that effective school leadership can improve teacher effectiveness, satisfaction, and retention, which, in turn, can lead to improvements in student learning (Hallinger and Heck 2010; Louis et al. 2010; Sebastian and Allensworth 2012).

TNTP's TLTS program aims to build upon proven models for identifying and developing leaders who are (a) attuned to the needs of schools adversely affected by poverty, low expectations, and instability, and (b) able to successfully lead transformation efforts in these schools centered on a strong instructional vision and building coherence at all levels of the school around that vision. Despite growing evidence on the importance of school leadership teams, school turnaround models rarely provide strategic support or professional development to all members of a school leadership team, including principals, vice principals, and instructional coaches (Calkins et al. 2007). However, research suggests that a cohesive school leadership team is essential to address school transformation priorities adequately (Kutash et al. 2010; Seashore Louis et al. 2010). In addition, a unified leadership team structure can help educators get through the difficult first stage of transformation efforts, when new roles are being established and leaders are building foundational relationships and trying to reset school culture (Kutash et al. 2010; Seashore Louis et al. 2010). With these needs in mind, TNTP aimed to create and foster high-functioning leadership teams in low-performing schools, providing them with instructionally coherent supports to ensure they are effective at improving student outcomes.

B. TLTS program description

1. Key program components

This study evaluated the TLTS program implemented under the EIR grant. The TLTS program sought to improve leadership practices among principals, school leadership teams (SLTs), and teachers in the schools randomly selected to implement the program. The program also aimed to support professional growth among school leaders and teachers; improve teacher instructional practices, school culture, and family engagement in school activities; and, ultimately, to increase student achievement.

As a core feature of the TLTS program's collective leadership approach, each participating school established an SLT that engaged in all program activities, including trainings, ongoing coaching sessions, and leadership activities. SLTs typically included principals, assistant principals, instructional coaches, instructional support specialists and teacher leaders who provided leadership within TLTS program schools. Through the TLTS program, TNTP staff provided customized supports to each school's SLTs through two key program components.

1. **Key program component 1: diagnosing school needs.** In the fall of their first year in the TLTS program, schools participated in a highly structured process to diagnose school needs and prioritize focus areas for TLTS support. Early activities to diagnose and plan for school needs included (1) conducting a comprehensive academic diagnostic with each school that focused on the quality and rigor of both instruction and student assignments, (2) conducting a community engagement diagnostic focused on the degree to which families felt welcomed and were engaged in school activities, (3) administering an internally developed Instructional Culture Insight Survey to teachers to learn more about school culture, and (4) using results from the diagnostics and survey to inform the development of the school's plan for ongoing TNTP support.
2. **Key program component 2: developing and supporting SLTs.** After the initial diagnostics and planning, the TNTP team provided ongoing coaching and supports to SLTs to help develop a school vision and improve leadership competencies. These supports included a three-day summer leadership training before the start of their first school year in the program and, following the diagnostic assessments referenced above, ongoing instructional coaching aligned with each school's plan for support. Ongoing coaching also included community engagement supports, further discussed in Appendix A.

Table I.1 summarizes the activities undertaken in each program component and the criteria for assessing implementation fidelity for each component. Chapter IV summarizes how well the TLTS program adhered to the implementation criteria under EIR. More detailed information on the TLTS program, including details on program supports, coaching topics, intensity, duration, and delivery method are provided in Appendix A.

Table I.1. Summary of key program components, activities, and criteria for assessing implementation fidelity

Activities	Description	Implementation criteria
Key program component 1: diagnosing school needs		
Academic diagnostics	In each school, TNTP conducted an academic diagnostic focused on gathering baseline information on the quality and rigor of both instruction and student assignments to help inform the customized school plan.	• Conduct at least one teacher focus group • Conduct at least four classroom observations • Analyze at least 12 student samples
Community engagement diagnostics	Concurrent with the academic diagnostics data collection, TNTP performed a community engagement diagnostic focused on gathering baseline information on the degree to which families are engaged in school activities.	• Conduct interview with each principal • Conduct at least one focus group with parents, community members • Conduct at least one focus group with students • Conduct at least one school walk-through per school

Activities	Description	Implementation criteria
Instructional Culture Insight Survey	TNTP administered its internally developed Insight Survey—a measure of school instructional culture—to eligible school staff members, including all staff who provided direct instruction in classrooms and served in school roles at least 50 percent of the school year.	• At least 40 percent of eligible school staff complete survey
Customized school plan	With input from SLTs, TNTP developed a customized plan for each school to inform ongoing supports. When developing these plans, school leaders identified a content area (math or English language arts), a grade, and specific classrooms in which to focus coaching activities.	• Create customized coaching plan for each school
Key program component 2: developing and supporting SLTs		
Transformational Leadership Summit	This three-day Transformational Leadership Summit with SLTs laid the groundwork for subsequent program activities and supports. The summit included community-building activities, an exploration of collective leadership methods, TNTP's The Opportunity Myth research, and an overview of the TLTS program and planned activities.	• At least 50 percent of all SLT members in each school participate in at least 8 hours of training • All principals participate in at least 12 hours of training
Ongoing coaching activities	Coaching focused on applying tools and methods to ensure that teachers set high classroom expectations and students were highly engaged and had access to grade-appropriate assignments and strong instruction. TNTP coaches worked with SLTs to set goals, practice leadership techniques, and provide feedback on areas for improvement. Schools also participated in district-level meetings that reinforced concepts on strengthening instructional culture and supporting family engagement. In addition, coaches supported the adoption of community engagement strategies within participating schools.	• Align instructional coaching and community engagement plans with school needs • Provide coaching in focus areas specified in coaching plans • Align supports with districts' established priorities

Source: TNTP program data.

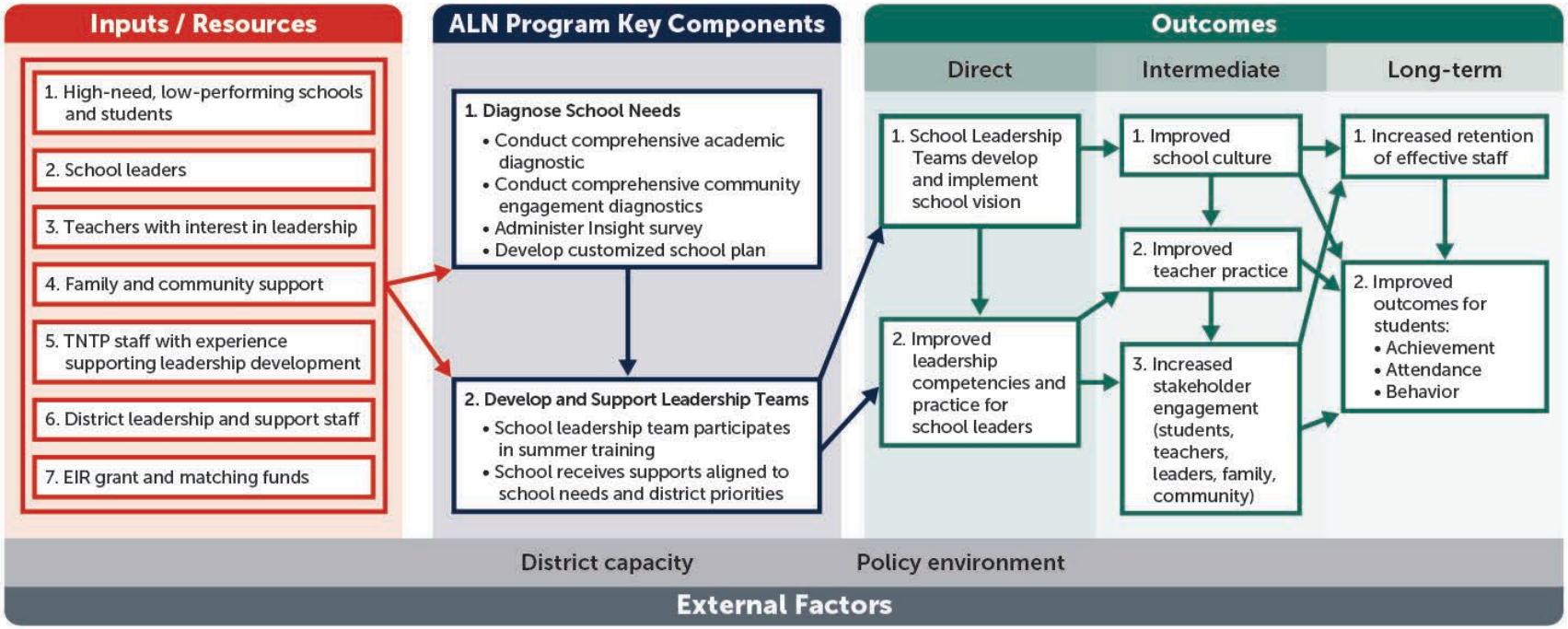
2. TLTS logic model

The logic model (Figure I.1) details how the TLTS program was intended to achieve the target outcomes for the project. The logic model shows external factors; key inputs and resources; key program components of the TLTS program model; and the direct, intermediate, and long-term outcomes the program aimed to achieve. Inputs 1–4 refer to school community members who would be directly involved in program implementation or impacted by it. Inputs 5–7 are the resources at the district and school levels and within TNTP that would help support program implementation.

Through the TLTS program, TNTP staff provided a set of customized supports to SLTs. These supports aimed to diagnose school needs (key program component 1) and develop and support SLTs (key program component 2). School needs diagnostics informed TNTP's support of SLTs, which aimed to improve the SLT's abilities to implement a school vision (direct outcome 1) and improve leadership competencies and practices (direct outcome 2), both directly and through the improved school vision. Two key external factors influence each aspect of the logic model: (1) district capacity and (2) the policy environment.

The strengthened school vision was expected to lead to improved school culture (intermediate outcome 1), which, in conjunction with improved school leadership, was expected to boost teacher practice (intermediate outcome 2). Improved leadership and teacher practices were expected to lead to increased engagement among students, teachers, families, and community members (intermediate outcome 3). In turn, the improved school culture and increased community engagement were expected to help the school retain effective staff (long-term outcome 1). Finally, the improved school culture, teacher practices, and stakeholder engagement were all expected to contribute to improved student outcomes (long-term outcome 2), both directly through increased retention of effective staff.

Figure I.1. TLTS logic model



C. Evaluation goals

The EIR Mid-phase grants seek to expand existing, innovative educational programs that have already been implemented successfully and have demonstrated moderate evidence of effectiveness.[3] To build on this existing evidence, all EIR Mid-phase grantees were required to commission a rigorous, independent evaluation to examine the impact of their programs. This TLTS program evaluation was designed to assess the effectiveness and implementation of the TLTS program as funded by the EIR grant, with the ultimate goal of providing evidence to inform future efforts focused on school leadership development. It includes both an impact and an implementation study. The impact study examines the impact of school participation in the TLTS program on student achievement and teacher perceptions and practices as compared to schools operating with a business-as-usual approach. The implementation study describes how the program model was implemented in practice and the extent to which it was implemented with fidelity to the program model as designed. We also provide an assessment of fidelity to scale-up plans as the program expanded across districts and an examination of per-student costs to understand how program costs relate to program impacts on student achievement.

This evaluation was conducted independently by Mathematica. Mathematica conducted all key aspects of the evaluation, including implementing random assignment and collecting the administrative and teacher survey data used for analysis, estimating impacts, and reporting study findings.[4] To ensure the independence of the impact analysis, TNTP staff reviewed this report for program information accuracy but did not make any modifications to the findings. TNTP staff also facilitated collaboration between Mathematica and the districts to identify schools for the study and provided the information we used to conduct the analyses of implementation fidelity, scale-up fidelity, and costs. However, TNTP staff played no role in randomly assigning schools, testing students, collecting data used in the impact analyses, or analyzing the data.

[3] Previous moderate evidence for a program could be established by citing up to two studies reviewed against the What Works Clearinghouse (WWC).

[4] Prior to collecting any data or performing data analyses, we preregistered the study's choice of primary and confirmatory research questions, outcomes, and analytical methods in the Registry of Efficacy and Effectiveness Studies: https://sreereg.icpsr.umich.edu/sreereg/.

II. Study Description

The Turnaround Leadership Teams Strategy (TLTS) evaluation used a randomized, controlled trial design to provide rigorous evidence on the impacts of the program as implemented under the Education Innovation and Research (EIR) grant. In this chapter, we describe the study's research questions and research design; the participating districts, schools, students, and teachers; and the methods used for the impact analyses.

A. Research questions for the impact evaluation

Schools participated in the TLTS intervention for two school years; thus, the central confirmatory research question the impact evaluation sought to address focused on the program's impact on student outcomes after two years of program implementation.

- **Confirmatory Research Question:** What is the impact of two years of the TLTS program compared to business as usual on students' achievement in English language arts (ELA), math, science, and social studies?

The study also aimed to answer three exploratory research questions that may offer insight into the confirmatory analysis findings.

- **Exploratory Research Question 1:** What is the impact of one year of the TLTS program compared to business as usual on students' achievement in ELA, math, science, and social studies?
- **Exploratory Research Question 2:** Did the TLTS program have varying impacts on student achievement across cohorts of participating schools?
- **Exploratory Research Question 3:** What is the impact of two years of the TLTS program on teacher perceptions and practices compared to business as usual?

B. Research design

To answer these research questions, we randomly assigned schools to either participate in the TLTS program or continue operating as usual. We then compared student and teacher outcomes between schools in the intervention group and schools in the comparison group.

1. Random assignment

Random assignment was performed for the three cohorts of schools participating in the TLTS study: 2018–2019 and 2019–2020 (Cohort 1), 2019–2020 and 2020–2021 (Cohort 2), and 2020–2021 and 2021–2022 (Cohort 3). Starting with the first cohort of schools, Mathematica used district data to pair similar schools into block pairs based on grade level (elementary or middle school) and other shared characteristics. Because student achievement was the evaluation's primary focus, we prioritized pairing schools that were similar in terms of average student achievement (based on the latest year for which state assessment data were available). We also prioritized pairing schools receiving similar district supports. Other factors considered in pairing schools included student demographics, geographic location, and school size. Within each block pair, one school was randomly assigned to the TLTS intervention group, and the other school was assigned to the comparison group. This process was repeated for Cohorts 2 and

3. By using random assignment, the study ensures that any differences in outcomes between the two groups can be attributed to the causal effects of TLTS.

Random assignment occurred on the same schedule across cohorts, beginning in spring 2018 (Table II.1). Specifically, schools were recruited to the study and randomly assigned in the spring before the program began for that cohort. (Schools assigned to the intervention group in the spring began participating in the TLTS program the summer after random assignment.) A total of 28 schools participated in the study, with half of them assigned to the TLTS intervention, and the other half assigned to the comparison group.[5] After random assignment, the study team collected participation forms and formal consent directly from school principals.

Table II.1. Timeline of study activities

Cohort	Number of schools	School recruitment	Random assignment	Intervention start	Year 1 student outcomes	Year 2 teacher outcomes	Year 2 student outcomes
1	6 (all elementary schools)	March 2018	April or May 2018	Summer 2018	Spring 2019	Cancelled due to COVID-19 pandemic	Winter 2020
2	12 (10 elementary, 2 middle schools)	March 2019	April or May 2019	Summer 2019	Winter 2020	Winter 2020–spring 2021	Spring 2021
3	12 (8 elementary, 4 middle schools)	March 2020	April or May 2020	Summer 2020	Spring 2021	Winter 2021–spring 2022	Spring 2022

2. Intervention and comparison conditions

Schools in the intervention group agreed to participate in two sequential years of the TLTS program. Through this program, TNTP program staff offered school-based supports to help school leadership teams transform low-performing schools. For more information on the program model, see the summary in Chapter I and the detailed program description in Appendix A. Comparison schools did not receive any supports from TNTP and continued to proceed with their business-as-usual approach to school operations. Schools in both the intervention and comparison conditions continued to participate in other ongoing programs and receive district supports. Thus, the contrast between the intervention group and the comparison group reflects only the supports TNTP provided through the TLTS program (and excludes the impact of any district supports).

[5] Given the limited number of schools eligible to participate in the study, two comparison group schools from Cohort 1 were re-randomized a second time as part of Cohort 3. That is, they were reassigned after completing the full two years of the study as comparison schools in Cohort 1. These schools were randomly assigned into the third cohort because all eligible schools in the district had already been engaged in the program. In their second random assignment (for Cohort 3,) one of the schools was selected for the TLTS intervention and one as a comparison school. For this reason, the number of schools across each cohort as shown in Table II.1 totals 30, while 28 unique schools participated in the study.

3. Data and outcome measures

As noted in the research questions, our confirmatory analyses compared the ELA, math, science, and social studies achievement of students in TLTS and comparison schools after two years of program implementation. For exploratory analyses, we also examined student achievement after one year of implementation as well as teacher perceptions and practices after two years of implementation. Two main sources of data contributed to the impact study: (1) administrative data, collected to examine TLTS program impacts on student achievement, and (2) a teacher survey, administered to explore program impacts on teacher perceptions and practices.

a. Student data and outcomes

To analyze the impact of TLTS on student achievement, we collected administrative data from districts participating in the study. We obtained three years of data for each cohort; this included one baseline year (before intervention schools started the TLTS program) and two outcome years. Overall, the administrative data spanned the school years 2017–2018 to 2021–2022. Test score data came from three assessments: (1) the state assessment, (2) Renaissance Star Assessment, a district benchmark assessment, and (3) the Northwest Evaluation Association's Measures of Academic Progress (NWEA MAP), another district benchmark assessment. The grade levels and subjects available for each assessment are summarized in Table II.2 by year and cohort. In addition to student test scores, we received a record of every school in which a student enrolled, their enrollment dates, their demographic and educational characteristics, the total number of days they were absent from school, and the number of days they were suspended (either in school or out of school).

We measured student achievement using test scores on the state assessment administered in the spring of each study year, except spring 2020, when the state assessment was cancelled due to the COVID-19 pandemic. To measure achievement that year, we instead used the district benchmark assessments taken by students in winter 2020, before schools transitioned to remote learning. Science and social studies are not tested in the benchmark assessments; we were therefore unable to examine achievement in those subjects in 2020. Although the state exam was administered again in 2021, the state made it optional for students. We discuss how these pandemic-related disruptions affected student attrition later in this chapter.

Because our main analyses include all cohorts, we combined scores from the different assessments into a single outcome measure in each content area (math and ELA). All test scores were first standardized into z-scores by assessment type using year, subject, and grade-level norms.[6] A z-score value of zero can be interpreted as the mean score of all students who took a given assessment in the same year, subject, and grade; negative z-scores therefore indicate a score below this mean. Table II.2 summarizes the assessments used to measure achievement by year and cohort. For example, the Year 2 student outcome measures used in the confirmatory analyses include the winter 2020 district benchmark assessments for Cohort 1, the optional spring 2021 statewide exams for Cohort 2, and spring 2022 statewide exams for Cohort 3.

[6] Statewide assessments were standardized using statewide means and standard deviations (by subject and grade level), which were obtained through the publicly available data on the state's website. District benchmarks were standardized using national means and standard deviations (by subject and grade level), which were obtained from the test developers (Renaissance Learning 2022a, Renaissance Learning 2022b. NWEA 2020).

Table II.2. Assessments used to measure student achievement (by year and cohort)

	Cohort 1 (6 schools)	Cohort 2 (12 schools)	Cohort 3 (12 schools)
Baseline assessment	**Spring 2018 statewide assessments** ELA and math: Grades 3–7	**Spring 2019 statewide assessments** ELA and math: Grades 3–7	**Winter 2020 district benchmarks**[a] Reading and math: Grades 3–7
Year 1 outcomes	**Spring 2019 statewide assessments** ELA and math: Grades 4–8 Science: Grades 5 and 8 Social studies: Grades 5 and 8	**Winter 2020 district benchmarks**[a] Reading and math: Grades 4–8	**Spring 2021 statewide assessments**[b] ELA and math: Grades 4–8 Science: Grades 5 and 8 Social studies: Grade 8[c]
Year 2 outcomes	**Winter 2020 district benchmarks**[a] Reading and math: Grades 5–8	**Spring 2021 statewide assessments**[b] ELA and math: Grades 5–8 Science: Grades 5 and 8 Social studies: Grade 8[c]	**Spring 2022 statewide assessments** ELA and math: Grades 5–8 Science: Grades 5 and 8 Social studies: Grade 8[c]

[a] The study used winter 2020 district benchmark assessments instead of spring 2020 state exams, which were canceled due to the COVID-19 pandemic. One district used the Northwest Evaluation Association's Measures of Academic Progress (MAP) as their benchmark exam, and the other used the Renaissance Star Assessment. District benchmarks assessed math and reading only.

[b] The state made the 2021 assessment optional for students.

[c] The state's social studies exam was discontinued in grade 5 after the 2018–2019 school year.

ELA = English language arts.

b. Teacher data and outcomes

To analyze the impact of TLTS on teacher perceptions and practices, we developed and administered a teacher survey in both TLTS and comparison schools. We did not administer the survey to teachers in Cohort 1 because of school disruptions and closures caused by the start of the COVID-19 pandemic in spring 2020. For Cohorts 2 and 3, we administered the survey in an online format, once per cohort, during the second year of that cohort's program participation (winter 2020-spring 2021 for Cohort 2 and winter 2021-spring 2022 for Cohort 3). To limit the burden on schools and teachers, we randomly selected 10 teachers from each school in Cohort 2 to take the survey. However, following low response rates for the Cohort 2 survey, we surveyed all eligible teachers in Cohort 3 to maximize the sample size for analysis and thus increase our likelihood of detecting small differences between teachers in TLTS and those in comparison schools.

The teacher survey was designed to inform the program implementation analyses and help interpret findings on student achievement. The survey collected teachers' self-assessments of their instructional practices and their perceptions of their school and school leadership, with a focus on core elements of the TLTS model. To learn more about instructional formats during the COVID-19 pandemic, the survey included questions about whether teachers taught in-person, virtually, or a mixture (hybrid) during the survey year and the share of time they spent teaching in each. Finally, we collected teachers' background information, including demographics (gender, race, and ethnicity), highest level of education, number of years teaching, and grades and subjects taught during the survey year.

To measure teacher perceptions and practices, we created several constructs guided by conceptual and statistical methods. First, we created a conceptual framework of outcomes we could measure with the survey, which were informed by the TLTS program's focus on four levers of school transformation: (1) collective leadership, (2) vision and culture, (3) authentic community engagement, and (4) academics (See Appendix A for a description of these levers). We then mapped individual survey items to these

topics within the framework. Because many of the survey questions had varying response options (for instance, Likert scales measuring the degree to which teachers agreed with statements about their schools or Likert scales measuring frequency of school activities), we normalized each survey item on a 0 to 1 scale.[7] This process ensured that larger values (that is, those closer to 1) equated to a more favorable finding (for instance, stronger positive perceptions of school leadership or more frequent practices of a school activity). Next, we measured the internal consistency of the groups of items to validate that responses on different items intended to measure the same construct were consistent with each other. Each group of items achieved adequate internal consistency, with Cronbach's alpha ranging from 0.68 to 0.89.[8] Finally, we averaged the responses for each item to calculate the outcome measures used in the impact analyses. Because the items making up each outcome had been normalized on a 0 to 1 scale, the outcome measures are on the same scale.

Table II.3 provides a description of each survey outcome, the number of survey items used to inform the outcome, and the internal consistency reliability of each measure. Appendix C provides a list of the survey questions used for each outcome as well as additional statistical diagnostics performed to measure the contributions of each survey item to the overall validity of each construct.

Table II.3. Description and internal consistency of outcome measures of teacher instructional practices and perceptions of school leadership

Outcome	Description	Number of items	Reliability Cronbach's Alpha (α)
Collective leadership			
Self-awareness and development	Teachers' perceptions of the degree to which school leadership values different aspects of identity (including the impact of social identity, culture, race, positionality, power, and privilege) and whether the school provides teachers with sufficient training to work with diverse populations of students and address implicit bias in decision making.	3	0.857
Building strong teams	Teachers' perceptions of the degree to which school leadership values diverse opinions, active listening, and equity of voice, and how well leadership understands the unique skills and contributions of individuals.	2	0.678
Vision and culture			
Shared vision for student success	Teachers' perceptions of how effectively school leadership develops and articulates a shared vision for student success in partnership with families, students, and community stakeholders.	3	0.787
Culture and goals	Teachers' perceptions of the degree to which their school leaders and fellow teachers collectively establish and strive to achieve shared goals for student success; whether school leaders have the knowledge and resources to ensure that teachers are providing academically rigorous instruction; and the extent to which teachers develop systems to inform decisions and monitor student progress.	18	0.862

[7] To normalize each survey response, we subtracted the minimum response value from the teacher's reported value, and we then divided this difference by the difference between the maximum and minimum possible values. Before this normalization, we ensured a higher-value response equated to a more favorable response across all survey items. This approach has been used in other analyses of teacher survey data. For an example, see Dolfin et al. (2019).
[8] The What Works Clearinghouse (WWC) recommends a minimum Cronbach's alpha value of 0.50.

Outcome	Description	Number of items	Reliability Cronbach's Alpha (α)
Authentic community engagement			
Empowering others	Teachers' perceptions of the degree to which their schools build trust with families, students, and the community. This outcome also measures whether schools develop systems to collect, respond to, and share community feedback.	19	0.886
Student learning	Teachers' self-assessment of the degree to which they engage and collaborate with parents and guardians about their children's goals for and progress in learning and development.	6	0.699
Academics			
Instructional practices	Teachers' perceptions of the academic expectations the school creates, instructional rigor, and whether instruction is grade appropriate.	3	0.786
Drive teacher improvement	Teachers' self-assessment of the degree to which they participated in or were offered professional development opportunities, coaching, and individualized supports to improve their teaching performance.	9	0.847

Source: Study-developed survey.

C. Recruitment, eligibility, and characteristics of districts and schools

TNTP recruited two school districts to participate in the TLTS program under the EIR grant. These school districts then identified schools eligible for random assignment. Through this recruitment process, a total of 28 elementary and middle schools participated in the study.

1. District recruitment and eligibility

At the launch of the EIR grant, TNTP identified two school districts to participate in the TLTS program under the EIR grant and recruited them for the study based on three criteria: (1) the districts were interested in engaging in the TLTS program, which would provide supports to a subset of their schools, (2) the districts were willing to participate in an impact evaluation of TLTS, including complying with all aspects of the intervention and data collection, and (3) the districts had a sufficient number of schools that served primarily high-need students (75 percent or more of students in the school), defined as students eligible for free or reduced-price lunch under the National School Lunch Program.

2. School recruitment and eligibility

TNTP and Mathematica worked with district liaisons to identify schools with the greatest need for the intervention. Together, we identified cohorts of eligible schools in spring 2018 (Cohort 1), spring 2019 (Cohort 2), and spring 2020 (Cohort 3) (Table II.2). To be eligible for the study, schools had to meet four criteria: (1) be low-performing, as defined by the criteria in the Federal Register notice for Mid-phase grants,[9] (2) serve primarily high-need students (75 percent or more of students in the school), (3) be an elementary or middle school (or both), and (4) could not receive other supports or interventions that

[9] The Federal Register notice for Mid-phase grants defines low-performing schools as "(e)lementary and secondary schools identified, at the time of submission of [grant application], as in need of corrective action or restructuring under the ESEA [Elementary and Secondary Education Act], as authorized and amended by the NCLB [No Child Left Behind Act]; (2) elementary and secondary schools identified, at the time of [grant application], as a priority or focus school by a State under ESEA flexibility; and, (3) secondary (both middle and high schools) in a State that are, at the time of [grant application], equally as low achieving as these Title I schools above and are eligible for, but do not receive, Title I funds."

would interfere with their ability to participate in the TLTS program (as defined by participating districts). For example, participating districts could elect not to include schools in the study if leaders at those schools were receiving extensive supports that would be too burdensome in combination with the TLTS program supports. Districts played a key role in both identifying eligible schools and encouraging their participation in the study. After districts provided the study team with a list of eligible schools, we randomly assigned half of the schools in that cohort to the TLTS program and half to the comparison group, as described above.[10]

3. District and school characteristics

The two districts that participated in the study were large public school districts in the southeastern United States. Based on the National Center for Education Statistics' locale classifications (NCES 2023), one district (District A) was in an urban setting, and the other (District B) was in a suburban setting. Both districts sought to improve student achievement and had a large number of schools with high percentages of low-income students. District A contributed 12 schools to the study, and District B contributed 16 schools. Overall, the study included 22 elementary schools and 6 middle schools, with an average of 647 students enrolled and 43 full-time teachers. On average 96 percent of students in study schools were eligible for free or reduced-price lunch, 76 percent were Black, and 13 percent were Hispanic (Table II.4). All study schools were traditional public schools eligible for Title I. On average, about one in five students in these schools was proficient in ELA or math upon entering the study. Schools assigned to the TLTS intervention or comparison condition had similar characteristics.

Table II.4. Average characteristics of study schools at baseline

Measure	All schools	TLTS schools (N = 15)	Comparison schools (N = 15)
Percent female	47.9	47.7	48.0
Percent Black	75.7	76.9	74.5
Percent Hispanic	13.1	13.1	13.1
Percent eligible for free or reduced-price lunch	95.6	95.1	96.2
Percent proficient in ELA (%)	20.4	19.1	21.8
Percent proficient in math (%)	20.3	20.6	20.0
Average number of students	647	672	622
Average number of FTE teachers	43	48	46
Average student:teacher ratio	14.7	14.9	14.6

Source: National Center for Education Statistics and State Department of Education.

Notes: Table includes 28 distinct schools, including two Cohort 1 schools in the comparison group that were randomly assigned again in Cohort 3. Therefore, the total number of schools across the TLTS and comparison groups sums to 30.

ELA = English language arts; FTE = full-time equivalent; TLTS = Turnaround Leadership Teams Strategy.

D. Eligibility and characteristics of students and teachers

Among the 28 study schools, we identified students and teachers eligible for data collection to form the basis for our analyses. This resulted in 8,934 students and 617 teachers eligible for the study.

[10] To facilitate the paired matching process during random assignment, the districts provided an even number of schools that met the criteria.

1. Student eligibility

To be eligible for the study, students had to be enrolled in a study school at the beginning of the school year after random assignment. Given the changes in enrollment that commonly occur during the summer, we opted to collect data on students enrolled in the fall (after random assignment) rather than in the spring (before random assignment) to limit student attrition. Because students and families were unlikely to be aware of a school's participation in TLTS, we determined there was a low risk that the results of the school-level random assignment would affect whether students left or joined the study schools over the summer. At the same time, we opted to focus on students enrolled in the fall of the first year of program implementation—rather than at the end of program implementation—in case the intervention led to changes in schools that could affect families' enrollment decisions over the course of the following two school years.

Students also had to be enrolled in a tested grade so that we could observe their outcomes one and two years after program implementation began. For example, because math and ELA state assessments are administered to students in grades 3 through 8, our confirmatory analyses focused on students enrolled in grades 4 through 7 in the fall after random assignment so we could obtain (1) baseline math and ELA test scores from the year prior to the intervention and (2) outcome math and ELA test scores from the end of the second year of program implementation. Because science and social studies are only tested in a subset of grades (grades 5 and 8), a smaller number of students were eligible for these analyses (those in grades 4 and 7 in the fall after random assignment, as shown in Table II.8).

2. Student characteristics

In total, we identified 7,274 students eligible for the Year 2 confirmatory analyses. Most of them were students of color and, on average, had lower academic achievement than the average student in the state that took state tests in the year before the study began (Table II.5). Approximately 76 percent of students eligible for the study were Black, 14 percent were Hispanic, 15 percent had a learning disability, and 15 percent had limited English proficiency. Students in study schools had below average achievement in math and ELA compared to other students in the state before the study began, with a mean z-score of -0.62 in math and -0.60 in ELA.[11] In other words, these students, on average, scored about 0.60 standard deviations lower than the average performance of all students who took these tests, which is equivalent to scoring in the 27th percentile.[12]

To help confirm that random assignment resulted in intervention and comparison groups of similar students, we tested for differences between the two groups on key baseline characteristics. Of the 12 baseline characteristics available through administrative data, students in the intervention and comparison groups were statistically significantly different in terms of baseline math and ELA achievement, the percentage of Asian students, and baseline attendance and suspensions.

[11] All eligible students had data available on their demographic and educational characteristics. However, some students did not have baseline achievement or attendance data from the previous year. The sample sizes are noted for each variable reported in Table II.5.

[12] Baseline achievement scores include two assessments, depending on when baseline data were collected for the cohort: (1) scores from the state assessment, which were standardized to the grade-level statewide average for each subject, and (2) benchmark exams standardized to grade-level national averages. Additional information on the transformation of achievement data used in the impact analyses is provided in this chapter.

Table II.5. Characteristics of students eligible for the study

Measure	All schools		TLTS schools		Comparison schools		Difference	
	Sample size	Mean or percent (SD)	Sample size	Mean or percent (SD)	Sample size	Mean or percent (SD)	Standardized difference (Hedges' g)	*p*-value
Prior year achievement								
Mean math z-score	6,609	-0.618 (0.906)	3,433	-0.641 (0.912)	3,176	-0.596 (0.899)	-0.049	0.046*
Mean ELA z-score	6,633	-0.603 (0.978)	3,457	-0.648 (0.964)	3,176	-0.560 (0.991)	-0.090	0.000**
Demographic characteristics								
Percent female	7,274	47.8 (50.0)	3,779	48.4 (50.0)	3,495	47.2 (49.9)	0.022	0.338
Percent Black	7,274	76.0 (42.7)	3,779	76.7 (42.3)	3,495	75.5 (43.0)	0.028	0.232
Percent Hispanic	7,274	14.2 (34.9)	3,779	14.4 (35.2)	3,495	14.0 (34.7)	0.014	0.560
Percent White	7,274	2.64 (16.0)	3,779	2.29 (15.0)	3,495	2.98 (17.0)	-0.043	0.064
Percent Asian	7,274	5.66 (23.1)	3,779	4.94 (21.7)	3,495	6.39 (24.5)	-0.063	0.007**
Percent multiple or other races	7,274	1.20 (10.9)	3,779	1.43 (11.9)	3,495	0.97 (9.81)	0.042	0.074
Educational characteristics								
Percent with learning disability (IEP)	7,274	15.0 (35.7)	3,779	15.2 (35.9)	3,495	14.8 (35.5)	0.011	0.640
Percent with limited English proficiency	7,274	14.9 (35.6)	3,779	15.3 (36.0)	3,495	14.5 (35.3)	0.021	0.370
Prior year attendance and suspensions								
Mean attendance rate	6,945	95.5 (5.2)	3,625	95.5 (5.2)	3,320	95.5 (5.1)	0.005	0.826
Mean suspension rate	7,274	2.31 (15.0)	3,779	1.92 (13.7)	3,495	2.70 (16.2)	-0.052	0.026*

Source: District administrative data.

Notes: Table includes students (a) enrolled in the study schools in the fall in which schools from that cohort began the TLTS program and (b) who had observed data for the relevant baseline characteristic. Means are weighted to account for unequal populations across schools. Standardized differences in means are computed using Hedges' g effect size metric by taking the mean difference between the TLTS and comparison groups and dividing by the pooled standard deviation. Sample sizes vary due to missing data.

* Difference is statistically significant at the 5 percent level.

** Difference is statistically significant at the 1 percent level.

ELA = English language arts; IEP = individualized learning plan; SD= standard deviation; TLTS = Turnaround Leadership Teams Strategy.

In all cases, however, students in the intervention and comparison groups differed by less than 0.10 standard deviations, well below the What Works Clearinghouse's (WWC) threshold of 0.25 standard

deviations to establish baseline equivalence.[13] For example, students in TLTS schools, on average, had somewhat lower baseline achievement than students in comparison schools (Hedges' g = -0.05 for math and -0.091 for ELA). Appendix B includes additional details on the baseline achievement of each analytic sample, which are comprised of students with test score data for each outcome in our analyses (see Table B.2). As discussed in Appendix B, because we control for differences in baseline achievement in our regression analyses, every analytic sample meets the WWC equivalence requirements.

3. Teacher eligibility

Teachers were eligible to take the teacher survey if they were employed full time and were responsible for direct instruction of a complete classroom. Both lead teachers and co-teachers were eligible, as were teachers responsible for special education classrooms and bilingual instruction. Specialists or teachers primarily in support roles (for instance, resource teachers, pull-out reading teachers, paraprofessionals, and instructional coaches) as well as teachers of enrichment classes, such as music, art, and physical education, were not eligible for the survey. We identified eligible teachers from among those employed in study schools in the winter of the second year of program implementation. As discussed earlier, due to high rates of teacher attrition in study schools, we did not survey teachers employed in study schools at the time of random assignment (that is, in the spring before program implementation began). The high levels of teacher attrition compounded by low response rates would have resulted in very small sample sizes for this analysis. Therefore, we opted to survey teachers who were employed shortly before the survey was to be administered (in early spring of the second year of program implementation). As we discuss later, to the extent the intervention may have affected teacher hiring and retention, surveying teachers in the second year of program implementation introduces a risk of bias into this analysis, and the results should be interpreted with caution.

4. Teacher characteristics

In total, we invited 617 eligible teachers to participate in the survey. Among those teachers, 223 teachers (36 percent) took the survey and provided demographic information.[14] Table II.6 shows the characteristics of the invited teachers who took the survey and provided demographic information. To limit the burden on participating districts, we did not collect administrative data on teachers eligible for the survey; we therefore do not have information on the characteristics of all teachers eligible for the survey.

Among the teachers who completed the survey, the majority were female (85 percent) and Black (82 percent). Most commonly, the highest degree earned was a master's degree (52 percent), and 17 percent were certified under the National Board for Professional Teaching Standards. On average, teachers who responded to the survey had 15 years of teaching experience, with six years at their current school. Most taught at an elementary school, and one quarter taught middle school. Over half were general education teachers; the remainder taught specific subjects (most commonly ELA, reading, or math). At the time of the survey (winter/spring 2021 for Cohort 2 and winter/spring 2022 for Cohort 3), about half of the surveyed teachers taught completely or partially through a virtual format.

[13] According to the WWC, differences smaller than 0.25 standard deviations can satisfy the baseline equivalence requirement if the analyses include statistical adjustments to control for these differences. In addition, studies examining achievement outcomes can establish baseline equivalence based on students' prior achievement.
[14] Response rates varied by item. Table II.6 shows the number of teachers who responded to each item.

Table II.6. Characteristics of teachers who responded to the teacher survey

Measure	All schools		TLTS schools		Comparison schools		Differences	
	Number of teachers	Mean or percent (SD)	Number of teachers	Mean or percent (SD)	Number of teachers	Mean or percent (SD)	Standardized difference (Hedges' g)	*p*-value
Demographic characteristics								
Percent female	223	84.7 (36.1)	136	85.4 (35.5)	87	83.9 (36.9)	0.039	0.753
Percent Black	221	81.6 (38.9)	136	77.4 (42.0)	85	85.7 (35.2)	-0.211	0.235
Percent Hispanic or Latino	222	0.9 (9.5)	136	1.4 (11.7)	86	0.4 (6.6)	0.094	0.380
Percent White	221	16.6 (37.3)	136	19.9 (40.1)	85	13.2 (34.1)	0.177	0.350
Percent Asian	221	2.9 (16.8)	136	4.7 (21.3)	85	1.0 (10.2)	0.207	0.065
Percent multiple races	221	1.0 (10.2)	136	2.1 (14.3)	85	0.0 (0.0)	--	0.173
Highest degree earned								
Percent bachelor's degree	223	29.7 (45.8)	136	35.0 (47.9)	87	24.3 (43.1)	0.232	0.132
Percent master's degree	223	52.4 (50.1)	136	52.2 (50.1)	87	52.6 (50.2)	-0.008	0.950
Percent professional degree	223	10.6 (30.9)	136	10.8 (31.1)	87	10.5 (30.9)	0.008	0.959
Percent doctoral degree	223	7.3 (26.1)	136	2.0 (14.2)	87	12.6 (33.4)	-0.447	0.000**
National board-certified teacher status								
Percent Board-certified	223	16.5 (37.2)	136	20.7 (40.6)	87	12.4 (33.1)	0.218	0.173
Years of teaching experience								
Mean years (total)	223	15.1 (9.4)	136	14.1 (9.4)	87	16.2 (9.2)	-0.224	0.053
Mean years at current school	223	6.3 (5.9)	136	6.3 (6.0)	87	6.2 (5.8)	0.024	0.857
School level taught								
Percent elementary school	223	75.0 (43.4)	136	75.0 (43.5)	87	75.0 (43.6)	0.000	1.000
Percent middle school	223	25.0 (43.4)	136	25.0 (43.5)	87	25.0 (43.6)	0.000	1.000

Measure	All schools		TLTS schools		Comparison schools		Differences	
	Number of teachers	Mean or percent (SD)	Number of teachers	Mean or percent (SD)	Number of teachers	Mean or percent (SD)	Standardized difference (Hedges' g)	*p*-value
Subject taught								
Percent general education	223	57.1 (49.6)	136	50.1 (50.2)	87	64.0 (48.3)	-0.279	0.009**
Percent ELA or reading	223	16.5 (37.2)	136	21.9 (41.5)	87	11.2 (31.7)	0.281	0.005**
Percent math	223	23.2 (42.3)	136	26.1 (44.1)	87	20.3 (40.4)	0.135	0.235
Percent social studies or history	223	11.6 (32.1)	136	16.7 (37.4)	87	6.4 (24.7)	0.309	0.002**
Percent science	223	11.3 (31.7)	136	12.7 (33.5)	87	9.8 (29.9)	0.092	0.353
Percent ESOL or ESL	223	4.1 (20.0)	136	3.1 (17.5)	87	5.2 (22.2)	-0.104	0.458
Percent special education	223	3.4 (18.2)	136	3.4 (18.1)	87	3.5 (18.4)	-0.005	0.965
Percent other subject	223	1.3 (11.3)	136	1.2 (10.9)	87	1.4 (11.8)	-0.018	0.867
Percentage of teachers who used instructional format								
Percent in-person	223	37.3 (48.5)	136	34.4 (47.7)	87	40.2 (49.3)	-0.118	0.458
Percent virtual	223	48.0 (50.1)	136	46.4 (50.1)	87	49.6 (50.3)	-0.065	0.58
Percent hybrid (both in-person and virtual)	223	47.5 (50.1)	136	42.0 (49.5)	87	53.0 (50.2)	-0.219	0.251
Percentage of time spent using instructional format								
Mean percentage of time in-person	214	37.9 (42.6)	131	40.1 (43.0)	83	35.7 (42.4)	0.103	0.079
Mean percentage of time virtual	211	47.6 (42.6)	129	46.3 (42.0)	82	48.9 (43.4)	-0.061	0.536
Mean percentage of time hybrid (both in-person and virtual)	205	17.4 (30.2)	123	16.6 (28.8)	82	18.1 (31.7)	-0.051	0.693

Source: Study-developed survey.

Notes: Table includes all teachers who responded to a given survey item. Means were weighted to account for unequal populations across schools. Standardized differences in means are computed using Hedges' g effect size metric by taking the mean difference between the TLTS and comparison groups and dividing by the pooled standard deviation. Sample sizes vary due to missing data.

* Difference is statistically significant at the 5 percent level.

** Difference is statistically significant at the 1 percent level.

ELA = English language arts; ESL = English as a second language; ESOL = English speaker of other languages; TLTS = Turnaround Leadership Teams Strategy.

Although most of these characteristics differed by less than 0.25 standard deviations between surveyed teachers in TLTS schools and those in comparison schools, there were some notable differences. For

example, a much lower share of surveyed teachers in TLTS schools had a doctorate degree than in comparison schools (2.0 percent versus 12.6 percent). In addition, surveyed teachers in TLTS schools were much less likely to teach general education than surveyed teachers in comparison schools (50.1 percent versus 64.0 percent). The surveyed teachers in TLTS schools were more likely to teach ELA or reading (21.9 percent versus 11.2 percent) or social studies or history (16.7 percent versus 6.4 percent) than surveyed teachers in comparison schools. No other differences were statistically significant. As with the analysis of student outcomes, we control for all available teacher characteristics in our regression analyses. However, because we did not collect baseline data on teachers' perceptions and practices before the study began, we cannot establish the equivalence of surveyed teachers in the TLTS and comparison groups.

E. School, student, and teacher attrition

When interpreting the impact analyses findings (presented in the next chapter), it is important to keep in mind attrition rates in the analytic samples—that is, the share of schools, students, and teachers entering each analysis relative to the number eligible for that analysis. High attrition, or nonresponse, can introduce bias in estimates, particularly when one group (TLTS or comparison) is included in the analytic sample at much higher rates than the other group. Below, we discuss (a) school- and student-level attrition for our confirmatory analyses of Year 2 student achievement outcomes and (b) teacher-level attrition for our exploratory analyses of teacher perceptions and practices. Information on the sample sizes and attrition for the exploratory analyses of Year 1 student achievement outcomes appear in Appendix B.

1. School attrition

Because of differences in the grade levels tested for different subjects, the number of eligible schools varied for some analyses. While math and ELA assessments were administered to students in all grades examined in this analysis (grades 3–8), the science assessment was administered to students in grades 5 and 8 and the social studies assessment was administered only to students in grade 8. As a result, only the six middle schools in Cohorts 2 and 3 were eligible for the Year 2 social studies analysis. This analysis does not count those untested grades towards attrition for a given outcome. When schools and students were not eligible for a given analysis due to exclusion criteria that were based on characteristics determined prior to the introduction of the intervention and applied consistently across conditions (such as grade levels served or student's grade level), those schools and students were not considered as attriting. All eligible schools were therefore included in the analyses, resulting in no school-level attrition for any outcomes.

In addition, schools in Cohort 1 were not eligible for the Year 2 science analysis because the state exams that were to be used in the analysis for that outcome year (2018-2019) were cancelled due to the pandemic and the district benchmark exams that replaced them did not assess science or social studies. According to the WWC Procedures and Standards Handbook (version 5.0), the loss of sample after random assignment due to acts of nature, such as pandemics, that affect both the intervention and comparison groups equally, does not constitute attrition. The number of schools eligible for each confirmatory Year 2 outcome analysis is reported in Table II.7.

Table II.7. School sample sizes and attrition rates for the Year 2 analyses on student achievement

Year 2 outcome	TLTS schools			Comparison schools			Overall school attrition rate (percent)	Differential school attrition (percentage points)
	Number of eligible schools at random assignment	Number of schools in the analytic sample	School attrition rate (percent)	Number of eligible schools at random assignment	Number of schools in the analytic sample	School attrition rate (percent)		
ELA	15	15	0.0	15	15	0.0	0%	0.0
Math	15	15	0.0	15	15	0.0	0%	0.0
Science	12	12	0.0	12	12	0.0	0%	0.0
Social studies	3	3	0.0	3	3	0.0	0%	0.0

Source: District administrative data.

Notes: The number of eligible schools includes schools that served grades tested in the indicated subject during the outcome year. The state's social studies exam was discontinued in grade 5 after the 2018–2019 school year; therefore, only the six middle schools in Cohorts 2 and 3 were eligible for the Year 2 social studies analysis. The state exam was cancelled in the 2019–2020 school year and therefore no data were available for science or social studies outcomes that year. This resulted in the six elementary schools in Cohort 1 (whose Year 2 outcomes were measured in 2019–2020) not being eligible for the Year 2 science analysis.

ELA = English language arts; TLTS = Turnaround Leadership Teams Strategy.

2. Student attrition

To be included in the student achievement analyses, eligible students had to have outcome test score data. Students without outcome data (for example, those who were absent from school the day the test was administered or who left the district before the assessment took place) were considered to have left the study. Table II.8 shows the number of students eligible for each Year 2 outcome analysis (that is, those enrolled in study schools in the fall after random assignment) and the number of students for whom we have Year 2 outcome test scores (referred to as the analytic sample). As with the school samples, the number of students eligible for each analysis varied by subject because of differences in which grades and assessments tested each subject. For example, only grades 5 and 8 were tested in science; thus, fewer students were eligible for this analysis than for the analyses of math and ELA achievement. Further, as noted earlier, after the 2018–2019 school year, only students in grade 8 were tested in social studies; therefore, this subject had the smallest sample size.

As shown in Table II.8, the analytic samples for our confirmatory Year 2 outcome analyses experienced considerable attrition overall. This was largely due to lower test participation rates during the pandemic, when the state assessment was optional. However, attrition rates were comparable between the TLTS and comparison groups. According to WWC Standards (version 5.0), the overall and differential attrition in each analytic sample used in the confirmatory analyses meets the guidance for low student-level attrition under both the cautious and optimistic boundaries. However, the optimistic boundary likely applies to the TLTS program because parents and students were unlikely to be aware of the program, and, therefore, we do not expect that school participation in the program affected the decision to leave study schools. As mentioned earlier, we establish the baseline equivalence of each analytic sample in Appendix B.

Table II.8. Student sample sizes and attrition rates for the Year 2 analyses on student achievement

Year 2 outcome	TLTS schools			Comparison schools			Overall student attrition rate (percent)	Differential student attrition (percentage points)
	Number of eligible students at random assignment	Number of students in the analytic sample	Student attrition rate (percent)	Number of eligible students at random assignment	Number of students in the analytic sample	Student attrition rate (percent)		
ELA	3,779	2,261	40.2	3,495	2,061	41.0	40.6	0.8
Math	3,779	2,247	40.5	3,495	2,041	41.6	41.1	0.6
Science	1,731	1,069	38.2	1,579	937	40.7	39.4	1.2
Social studies	923	544	41.1	866	499	42.4	41.7	0.6

Source: District administrative data.

Notes: The number of schools includes schools participating in the study that served grades tested in the indicated subject during the outcome year. The number of students at random assignment includes the reference sample of students who were (a) enrolled in the study schools in the fall after the schools were randomly assigned and (b) eligible to take the exam in Year 2 of the study based on their grade level. For example, science exams are only taken in grades 5 and 8; therefore, the reference sample includes students who were in grades 4 and 7 at the start of the study.

ELA = English language arts; TLTS = Turnaround Leadership Teams Strategy.

3. Teacher attrition

To be included in the analyses of teacher perceptions and practices, eligible teachers had to respond to the teacher survey. We invited all eligible teachers in Cohort 2 and Cohort 3 schools in the second year of implementation – a total of 617 teachers – to participate in the survey (319 teachers from TLTS schools and 298 teachers from comparison schools). Among those surveyed, a maximum of 38 percent (232 teachers) were included in the impact analysis because they responded to the questions relating to a given outcome.[15] Response rates were higher among teachers in TLTS schools. For example, 43 percent (138 teachers) responded to the "driving teacher improvement" questions compared to 32 percent (94 teachers) in comparison schools. The sample sizes for each outcome were similar, but they varied slightly due to differing levels of item nonresponse. Table II.9 shows the number of teachers surveyed and the number of teachers for whom we had sufficient response data to compute outcomes (the analytic sample). Response rates are shown for each measure overall and by study group. The low overall response rates, as well as the large difference in response rates between teachers in TLTS and comparison schools, suggest that the available data may not be representative of all eligible teachers in study schools. As mentioned earlier, we are unable to establish the baseline equivalence of the teacher analytic samples at the time of random assignment.

[15] As noted earlier, response rates varied by item. To measure an outcome, we required the respondent to provide a valid response to each item used to construct the outcome. Table II.9 shows the number of teachers with data for each outcome. For example, 232 teachers provided complete data for the "Driving teacher improvement" outcome.

Table II.9. Sample sizes and response rates for the Year 2 analyses of teacher perceptions and practices

Year 2 outcome	All schools (24 schools, 617 eligible teachers)		TLTS schools (12 schools, 319 eligible teachers)		Comparison schools (12 schools, 298 eligible teachers)	
	Number of teachers in analytic sample	Response rate (percent)	Number of teachers in analytic sample	Response rate (percent)	Number of teachers in analytic sample	Response rate (percent)
Vision and culture						
Culture and goals	218	35.3	133	41.7	85	28.5
Shared vision for student success	224	36.3	137	42.9	87	29.2
Collective leadership						
Building strong teams	226	36.6	138	43.3	88	29.5
Self-awareness and development	225	36.5	137	42.9	88	29.5
Authentic community engagement						
Empowering others	217	35.2	132	41.4	85	28.5
Student learning	226	36.6	138	43.3	88	29.5
Academics						
Driving teacher improvement	232	37.6	138	43.3	94	31.5
Instructional practice	230	37.3	141	44.2	89	29.9

Source: Study-developed survey.

Notes: Response rates are based on the number of teachers who provided a valid response to each survey item that went into each construct. A list of items used for each construct is provided in Appendix C.

TLTS = Turnaround Leadership Teams Strategy.

F. Overview of analytic approach

As summarized below, we used regression analyses to estimate the impact of the TLTS program on student and teacher outcomes. Additional details about the analytic approach, including computation of analytic weights and approaches to handling missing covariates, are provided in Appendix B.

1. Overview of the analytic approach for student outcomes

To estimate the impact of the TLTS program on student achievement, we used an ordinary least squares (OLS) regression model, controlling for student covariates, including grade, race, ethnicity, gender, special education status, English learner status, and prior math and ELA test scores.[16] The model also accounted for the nature of the school's random assignment—which was conducted cohort by cohort based on matched pairs of schools—by including indicators for each school pair and the school's cohort. In addition, the regression weighted students such that each school contributed equally to the analysis regardless of the number of students enrolled.

Thus, the model estimated the TLTS program's impact in the average participating school on math, ELA, science, and social studies achievement among students enrolled in study schools in the fall after random

[16] Data on students' eligibility for free or reduced-price lunch were not available. Study schools participate in the Community Eligibility Provision program and therefore do not directly certify the eligibility of individual students.

assignment. Because students did not necessarily remain in study schools for the full duration of the program, this is referred to as an "intent-to-treat" (ITT) impact estimate. ITT impact models use a conservative approach that includes students in the intervention group even if they left the treatment school at any point after the program started. Therefore, the ITT estimates capture the impact of having been enrolled in a TLTS school at the time the school began participating in the program.

2. Overview of the analytic approach for teacher outcomes

Similar to the analysis of student outcomes, our analysis of teacher outcomes used an OLS regression model to compare outcomes among teachers in TLTS and comparison schools. The model estimated the impact of TLTS on teachers' perceptions and practice outcomes, controlling for teachers' race, ethnicity, gender, highest degree completed, teacher certification status, years of teaching (in total and at the study school), grades and subject taught, and the share of time the teacher spent teaching in virtual, in-person, and hybrid settings. Similar to the student analysis, this model also controlled for the school cohort and for the matched pair of schools to account for the nature of random assignment, and the model weighted teachers such that each school contributed equally to the analysis.

The findings on teacher practice and perceptions should be interpreted with caution for three main reasons:

1. Due to high rates of teacher attrition in study schools, we surveyed teachers who were employed in study schools at the time of the survey administration (Year 2 of program implementation) rather than at the time of random assignment. To the extent that the TLTS program may have affected teacher hiring and retention, changes in the characteristics of teachers in the study schools could bias results.
2. The survey had low response rates, and teachers in TLTS schools were significantly more likely to take the survey than those in comparison schools. This can introduce response bias where, for instance, teachers with strong opinions about their school (favorable or not) may be more inclined to respond. Low response rates also limit the generalizability of our findings and result in smaller sample sizes, which make it more difficult to detect small differences between groups.
3. The study did not administer a baseline survey. We are therefore unable to assess or account for baseline differences in teachers' perceptions and practices at the start of the study when estimating differences at the end of the study.

III. TLTS Impacts on Student Achievement and Teacher Outcomes

This chapter presents findings on the impacts of the Turnaround Leadership Teams Strategy (TLTS) program on student achievement and teacher perceptions and practices of their school's culture. We begin by discussing the findings of the confirmatory analysis, which examines the impacts of TLTS on student achievement two years after the start of the program. We then present the results of the exploratory analyses, which provide additional context for the confirmatory findings. We conclude with a discussion of the overall findings.

A. Confirmatory analysis: TLTS impacts on Year 2 student achievement outcomes

The TLTS program provided supports to participating schools over the course of two school years. To assess whether the full two-year program was effective at improving student achievement, the study's confirmatory analysis examined program impacts on student achievement at the end of the second year of implementation. Because TLTS focused its instructional supports on English language arts (ELA) and math, we first present findings on those two subjects. Given that the program model focused on whole-school reform, we also present findings on the impact of TLTS on student achievement in science and social studies.

TLTS had positive effects on Year 2 student achievement outcomes in ELA and math, the two subject areas of focus for TLTS program supports. Schools that participated in TLTS selected a subject area—either ELA or math—to receive coaching and supports. After two years of TLTS program implementation, students in TLTS schools performed, on average, 0.05 standard deviations higher in ELA than students in comparison schools (p-value = 0.019). Given the amount of growth in ELA that students in the same grades typically demonstrate over an academic year, this represents roughly 1.5 additional months of learning in the subject (Bloom et al. 2008). In math, students in TLTS schools performed, on average, 0.09 standard deviations higher than students in comparison schools (p-value < .001), representing about two additional months of learning. (See Table III.1.)

Another way to interpret effect sizes is using the benchmarks developed by Kraft (2020), which are based on a review of almost 750 randomized controlled trials of educational interventions. That analysis suggests effect sizes between 0.05 to 0.20—which correspond to the 37th to 69th percentiles of all effect sizes reviewed—should be considered medium in magnitude. For example, the median effect size on ELA achievement among the studies reviewed was 0.12 standard deviations, and the median effect size on math achievement was 0.07 standard deviations.

TLTS did not show positive effects on Year 2 student achievement outcomes in other classroom subjects that were not a specific focus of TLTS supports. After two years, TLTS had no detectable effects on student achievement in science and had a negative, statistically significant effect in social studies. The estimated impact on science achievement of 0.03 standard deviations was not statistically significant (p-value = 0.266). On the social studies assessment—administered only to students in grade 8—students in TLTS schools performed 0.12 standard deviations lower on average (p-value = 0.030) than students in comparison schools. This represents about four months less learning. (See Table III.1.)

Table III.1. Impacts of TLTS on Year 2 student achievement outcomes

Outcome	TLTS schools		Comparison schools		Impact estimate (SE)	*p*-value
	Number of students	Mean (SD)	Number of students	Mean (SD)		
Focal subject areas for TLTS supports						
ELA z-score	2,261	-0.547 (0.911)	2,061	-0.598 (0.953)	0.051* (0.020)	0.019
Math z-score	2,247	-0.643 (0.816)	2,041	-0.729 (0.807)	0.086** (0.017)	0.000
Other subject areas						
Science z-score	1,069	-0.564 (0.777)	937	-0.596 (0.812)	0.032 (0.028)	0.266
Social studies z-score	544	-0.626 (0.682)	499	-0.508 (0.783)	-0.117* (0.039)	0.030

Source: District administrative data.

Note: Standard deviations of the unadjusted outcome means are reported in parentheses under each mean z-score. Standard errors are reported in parentheses under each impact estimate. The study includes outcomes for students in grades 4 through 8 who attended study schools in the fall after random assignment. Estimates of the TLTS impact pool outcomes from students across all three school cohorts, control for baseline covariates, and include weights to account for unequal populations across schools. The comparison group mean is unadjusted, and the treatment group mean is the sum of the comparison group mean and the regression-adjusted difference between groups. Scale scores were converted to z-scores. A z-score of zero indicates a score equal to the mean for all students who took that test (statewide for the state assessment and nationally for the district benchmark assessments) in the same subject, grade, and year. A negative z-score indicates the score was below the mean student score.

* Impact estimate is statistically significant at the 5 percent level.

** Impact estimate is statistically significant at the 1 percent level.

ELA = English language arts; SD = standard deviation; SE = standard error; TLTS = Turnaround Leadership Teams Strategy.

B. Exploratory analyses of student and teacher outcomes

To support interpretation of the Year 2 outcome findings, we conducted additional exploratory analyses of student and teacher outcomes. These analyses include an exploration of (a) the impact of TLTS on student achievement after one year of program implementation (midway through the TLTS intervention for participating schools), (b) whether the impacts of TLTS on student achievement varied across cohorts, and (c) whether the program led to impacts on teacher perceptions and practices as measured by the teacher survey administered in the second year of program implementation.

1. TLTS impacts on Year 1 student achievement outcomes

TLTS had positive effects on Year 1 student achievement outcomes in English language arts. One year after the TLTS program began, students in TLTS schools performed, on average, 0.06 standard deviations higher in ELA than students in comparison schools (*p*-value = .013) (Table III.2). This positive impact represents roughly two additional months of learning in the subject. Although the Year 1 impact of 0.06 standard deviations is larger than the Year 2 impact of 0.05 standard deviations, the two estimates are not statistically different. In other words, we cannot rule out the possibility that the program had the same impact on ELA performance in both years (*p*-value of the difference in impacts = 0.676).

TLTS did not show positive effects on Year 1 student achievement outcomes in math, science, or social studies. After one year of program participation, TLTS had no statistically significant impact on student achievement in math. However, by Year 2, the estimated impact increased to 0.09 standard deviations, which is statistically different from the Year 1 impact (p-value of the difference in impacts = 0.007), suggesting the positive impact of the TLTS program on math achievement did not emerge until Year 2.

Although we see a change in the direction of the estimated impact on science achievement between Years 1 and 2 (from -0.04 to 0.03), neither estimate is statistically distinguishable from zero. In addition, the difference between the two impact estimates is not statistically significant (*p*-value of the difference in impacts = 0.193). In other words, TLTS had no detectable impact on science achievement in either program year, and there was little evidence of improvement from one year to the next.

TLTS had a negative impact on social studies achievement after Year 1, with student in TLTS schools performing, on average, 0.27 standard deviations lower (*p*-value = 0.008) than students in comparison schools. This negative impact decreased between Years 1 and 2, from -0.27 to -0.12 standard deviations—the difference between eight fewer months of learning to four fewer less of learning. The difference between the impacts on social studies achievement in Years 1 and 2 is marginally significant (p-value of difference = 0.086), suggesting there may have been a reduction in the negative effect over time.

Table III.2. TLTS impacts on Year 1 student achievement outcomes

Outcome	TLTS schools		Comparison schools		Impact estimate (SE)	*p*-value
	Number of students	Mean (SD)	Number of students	Mean (SD)		
Focal subject areas for TLTS supports						
ELA z-score	2,544	-0.626 (0.918)	2,621	-0.690 (0.975)	0.064* (0.024)	0.013
Math z-score	2,511	-0.693 (0.877)	2,602	-0.700 (0.901)	0.006 (0.024)	0.792
Other subject areas						
Science z-score	432	-0.724 (0.768)	421	-0.687 (0.832)	-0.037 (0.045)	0.420
Social studies z-score	251	-0.967 (0.842)	318	-0.696 (0.868)	-0.271** (0.080)	0.008

Source: District administrative data.

Note: Standard deviations of the unadjusted outcome means are reported in parentheses under each mean z-score. Standard errors are reported in parentheses under each impact estimate. The study includes students in grades 4 through 8 who attended study schools in the fall after random assignment. Estimates of the TLTS impact pool all three school cohorts, control for baseline covariates, and include weights to account for unequal populations across schools. The comparison group mean is unadjusted, and the treatment group mean is the sum of the comparison group mean and the regression-adjusted difference between groups. Scale scores were converted to z-scores. A z-score of zero indicates a score equal to the mean for all students who took that test (statewide for the state assessment and nationally for the district benchmark assessments) in the same subject, grade, and year. A negative z-score indicates the score was below this mean.

* Impact estimate is statistically significant at the 5 percent level.

** Impact estimate is statistically significant at the 1 percent level.

ELA = English language arts; SD = standard deviation; SE = standard error; TLTS = Turnaround Leadership Teams Strategy.

2. TLTS impacts on student achievement by cohort

TLTS impacts on ELA and math achievement after two years varied by study cohort, generally improving in Cohorts 2 and 3. When examining impacts by cohort, we find evidence that later cohorts generally experienced greater impacts. After two years of implementation, the TLTS program had the following impacts on ELA achievement:

- A statistically significant, negative impact in Cohort 1 (-0.06 standard deviations)
- A statistically significant, positive impact in Cohort 2 (0.10 standard deviations)
- No detectable impact in Cohort 3

In math, TLTS schools in Cohorts 1 and 2 experienced no detectable impact, and Cohort 3 showed a statistically significant, positive impact (0.11 standard deviations). When we test whether differences between cohorts are statistically significant, we find that Cohorts 2 and 3 experienced consistently higher impacts than Cohort 1 in both math and ELA. (See Table III.3.) See Appendix B for additional details on these cohort-level analyses.

Table III.3. TLTS impacts on Year 2 student achievement outcomes (by cohort)

Outcome	TLTS schools		Comparison schools		Impact estimate (SE)	*p*-value
	Number of students	Mean (SD)	Number of students	Mean (SD)		
Cohort 1 (district benchmark exam)						
ELA z-score	278	-0.758 (0.998)	245	-0.697 (0.915)	-0.061** (0.011)	0.002
Math z-score	274	-0.812 (1.080)	238	-0.784 (1.069)	-0.028 (0.013)	0.077
Cohort 2 (optional state assessment)						
ELA z-score	439	-0.704 (0.850)	501	-0.808 (0.948)	0.104** (0.026)	0.002
Math z-score	432	-0.804 (0.610)	491	-0.876 (0.697)	0.072 (0.039)	0.094
Science z-score	235	-0.717 (0.675)	254	-0.789 (0.746)	0.072 (0.044)	0.132
Social studies z-score	35	-0.550 (0.500)	53	-0.621 (0.922)	0.071 (0.091)	0.579
Cohort 3 (state assessment)						
ELA z-score	1,544	-0.405 (0.869)	1,315	-0.438 (0.949)	0.033 (0.024)	0.205
Math z-score	1,541	-0.520 (0.720)	1,312	-0.627 (0.679)	0.107** (0.025)	0.001
Science z-score	834	-0.499 (0.801)	683	-0.499 (0.826)	0.000 (0.035)	0.991
Social studies z-score	509	-0.623 (0.680)	446	-0.495 (0.766)	-0.128 (0.045)	0.067

Source: District administrative data.

Note: Standard deviations of the unadjusted outcome means are reported in parentheses under each mean z-score. Standard errors are reported in parentheses under each impact estimate. The study includes students in grades 4 through 8 who attended study schools in the fall after random assignment. Estimates of the TLTS impact control for baseline covariates and include weights to account for unequal populations across schools. The comparison group mean is unadjusted, and the treatment group mean is the sum of the comparison group mean and the regression-adjusted difference between groups. Scale scores were converted to z-scores. A z-score of zero indicates a score equal to the mean for all students who took that test (statewide for the state assessment and nationally for the district benchmark assessments) in the same subject, grade, and year. A negative z-score indicates the score was below this mean.

* Impact is statistically significant at the 5 percent level.

** Impact is statistically significant at the 1 percent level.

ELA = English language arts; SD = standard deviation; SE = standard error; TLTS = Turnaround Leadership Teams Strategy.

3. TLTS impacts on teacher practices and perceptions

After two years of school participation in TLTS, there was no evidence of effects on self-reported teacher perceptions and practices, although findings should be interpreted with caution. We did not detect an impact of the TLTS program on teachers' perceptions or practices in any of the eight outcomes assessed. Each outcome was measured on a scale ranging from 0 to 1, with higher values indicating stronger agreement with a given construct. (See Appendix C for additional information on each outcome.) Table III.4 also shows the impacts expressed in standardized effect size units. As shown in Table III.4, differences between TLTS and comparison schools were not statistically significant, and generally small (below 0.05 standard deviations). However, these teacher survey findings should be interpreted with caution due to the limitations described in Chapter II. Notably, there were high levels of teacher attrition and survey nonresponse—which was highest among teachers in comparison schools—and we were unable to establish the baseline equivalence of the analytic sample. Therefore, it is possible that the teachers who responded to the survey differed between TLTS and comparison schools, such that these findings may not reflect the true impact of the TLTS program.

With these limitations in mind, some trends can be noted among the teachers who responded to the survey. Teachers in both groups (TLTS and comparison) reported most favorably on the Instructional Practices outcome, which indicates whether schools create high academic expectations and provide rigorous and grade-appropriate instruction to all students. Similarly, teachers in both groups tended to report that their school develops a shared vision for student success in partnership with families, students, and community (Shared Vision for Student Success) and that they have the tools and resources needed to achieve these shared goals (Culture and Goals). However, teachers gave the lowest ratings to their schools' drive for teacher improvement, suggesting greater supports may be needed to help improve teacher performance. Teachers also gave relatively lower ratings on Self-Awareness and Development, which refers to how much their school leadership values different aspects of identity, addresses bias in decision-making, and provides sufficient training to work with diverse populations of students.

Table III.4. TLTS impacts on teacher practices and perceptions after two years

Outcome	TLTS schools		Comparison schools		Impact estimate (SE)	Standardized difference (Hedges' g)	*p*-value
	Number of teachers	Mean (SD)	Number of teachers	Mean (SD)			
Vision and culture							
Culture and goals	133	0.809 (0.128)	85	0.818 (0.117)	-0.009 (0.018)	-0.074	0.620
Shared vision for student success	137	0.812 (0.210)	87	0.809 (0.218)	0.00 (0.037)	0.016	0.926
Collective leadership							
Building strong teams	138	0.714 (0.281)	88	0.721 (0.285)	-0.01 (0.052)	-0.026	0.890
Self-awareness and development	137	0.576 (0.300)	88	0.574 (0.299)	0.00 (0.047)	0.007	0.965
Authentic community engagement							
Empowering others	132	0.700 (0.156)	85	0.716 (0.166)	-0.02 (0.026)	-0.101	0.533
Student learning	138	0.625 (0.165)	88	0.634 (0.162)	-0.01 (0.023)	-0.059	0.670
Academics							
Drive teacher improvement	138	0.509 (0.181)	94	0.511 (0.151)	0.00 (0.021)	-0.016	0.895
Instructional practices	141	0.852 (0.188)	89	0.855 (0.168)	0.00 (0.028)	-0.013	0.931

Source: Study-developed teacher survey.

Note: Standard errors are reported in parentheses under each impact estimate. Unadjusted standard deviations are reported under each outcome average. Study includes outcomes for teachers employed at schools randomly assigned to TLTS compared with teachers in comparison group schools. Estimates of the TLTS impact pool Cohorts 2 and 3, control for baseline covariates, and include weights to account for unequal populations across schools. The control mean is unadjusted, and the treatment mean is the sum of the control mean and the regression-adjusted difference between groups. All scales range from 0 to 1, where zero indicates no agreement with the outcome construct and 1 indicates total agreement. Standardized differences in means are computed using Hedges' g effect size metric by taking the mean difference between the TLTS and comparison groups and dividing by the pooled standard deviation.

* Impact is statistically significant at the 5 percent level.

** Impact is statistically significant at the 1 percent level.

C. Discussion

The results of the confirmatory analyses demonstrate that, after two years of implementation, the TLTS program led to positive, medium-sized effects on students' ELA and math achievement that were roughly equivalent to students receiving an additional 1.5–2 months of instruction in those subjects. As described in Appendix A, in the first year of the program, leaders of TLTS schools selected a subject (ELA or math) to focus on during the school year. The positive effects on these two subjects are therefore consistent with the areas where TLTS focused its instructional supports.

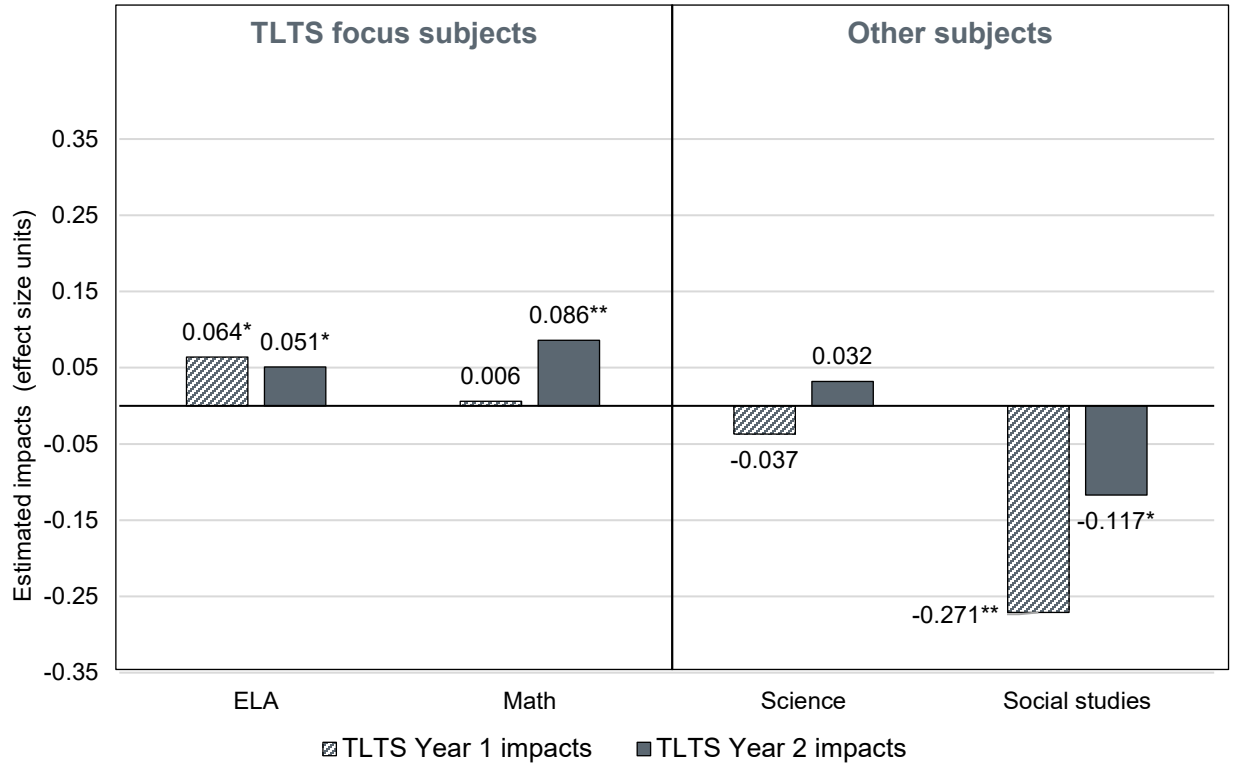
The null impacts on science achievement and the negative impacts on students' social studies achievement suggest there may have been a tradeoff between core subjects (ELA and math) and non-core subjects (science and social studies). As with many other schools nationally, the two districts engaged in the study were focused on math and ELA learning acceleration given the disruptions students experienced due to the pandemic. However, those recovery efforts that emphasized math and ELA in school years 2020–2021 and 2021–2022 would have affected both TLTS and comparison schools. Therefore, the results suggest that TLTS schools may have opted to leverage additional resources and supports provided through TLTS to further double down on these two subjects, which may have had an unintended impact on science and social studies achievement.

Some studies of school improvement efforts have generated similar findings on these tradeoffs. For example, Mathematica's evaluation of the Atlanta Public Schools' school turnaround initiatives in 2017–2019 found some tradeoffs between improvements in core subjects and non-core subjects as instructional time and resources shifted toward the core subjects, which are typically the focus of school transformation and accountability efforts (Hallgren et al. 2019). In that study, implementing a school partnership model in high-need schools had positive impacts on students' math and ELA achievement, with students gaining about one to four months of learning in these subjects; however, there were negative impacts on science and social studies, equivalent to losing roughly seven to nine months of learning.

There are some important caveats to also keep in mind when interpreting science and social studies findings in this study. Although similar findings have appeared in other contexts, a smaller set of schools and grades were included in the analyses of science and social studies achievement than in the analyses of math and ELA achievement because science and social studies are not tested in all grades. For example, whereas the Year 2 impacts on math and ELA achievement include students in grades 4–8 in 28 elementary and middle schools (that is, all study schools), the Year 2 impact on social studies achievement includes only grade 8 students in the six middle schools that participated in the study. It is possible that the tradeoffs observed may have been different if we had been able to examine science and social studies achievement in all grades and schools.

The results of the Year 1 exploratory analyses suggest that TLTS had a positive effect on ELA achievement after just one year of program implementation, whereas the positive effect on math achievement did not emerge until Year 2 (Figure III.1). Although it is possible that schools may have needed the full two years of support to see a positive impact on math achievement, this finding could also reflect the fact that more schools shifted their focus to math instruction in Year 2. Most SLTs chose to focus on literacy instruction and supports as the academic priority in the program's first year because of its relevance across all subjects and the district's priorities. If a school showed improvement in its initial priority subject in Year 1, they often chose to use that momentum to improve instructional supports in other subjects (typically math) the following year. The results of the Year 1 exploratory analyses also suggest that any tradeoffs between core and non-core subjects were present in both years of implementation, although there was some evidence that these lessened over time (Figure III.1).

Figure III.1. TLTS impacts on student achievement after one year and two years



Source: District administrative data.

Notes: The study includes students in grades 4–8 who attended study schools in the fall after random assignment. Estimates of the TLTS impact pool all three school cohorts, control for baseline covariates, and include weights to account for unequal populations across schools. Student analytic sample sizes are as follows: Year 1 ELA = 5,165; Year 2 ELA = 4,322; Year 1 math = 5,113; Year 2 math = 4,288; Year 1 science = 853; Year 2 science = 2,006; Year 1 social studies = 569; and Year 2 social studies = 1,043. Scale scores were converted to z-scores. A z-score of zero indicates a score equal to the mean for all students who took that test (statewide for the state assessment and nationally for the district benchmark assessments) in the same subject, grade, and year. A negative z-score indicates the score was below this mean.

* Impact estimate is statistically significant at the 5 percent level.

** Impact estimate is statistically significant at the 1 percent level.

ELA = English language arts; TLTS = Turnaround Leadership Teams Strategy.

While we see positive impacts on ELA and math achievement in our confirmatory analyses—which combine all three cohorts of schools—the exploratory analyses of cohort-level impacts suggest there were differences in how effective TLTS was for each cohort of schools. In particular, schools in Cohort 1, which began implementation in summer 2018, had less favorable results compared to the other cohorts, including a negative impact on ELA achievement and a null impact on math achievement in the second year of implementation. Improvements across cohorts can often be attributed to program start-up effects, which suggest a program is implemented with higher fidelity in later years compared to the first year. However, TNTP reported that the program was implemented consistently and with high fidelity in each year. A more plausible explanation could relate to the added direct supports TNTP began providing schools in 2020 in response to the disruptions caused by the pandemic. Students in Cohort 1 TLTS schools, which were nearing the end of the study at that time, did not experience these added services, which may have helped boost achievement, to the same extent as the latter two cohorts.

The differences in program impacts across cohorts could also be due in part to differences in the timing and type of assessments used in the analyses. Namely, the impacts for Cohort 1 are based on district benchmark assessments taken in the winter, whereas the impacts for Cohorts 2 and 3 are based on the high-stakes state assessments taken in the spring (although for Cohort 2, the state assessment was optional in Year 2). To the extent that program impacts can take longer to emerge or are more likely to be perceived on high-stakes assessments, these differences could also help explain why the results for Cohort 1 are less favorable. It is also important to note that Cohort I was potentially impacted by the project team scaling to a second district during the first year of impact analysis work and program supports may have been less robust given the roll out in district two. Finally, it is also worth noting that Cohort 1 schools had lower baseline achievement levels than schools in the later cohorts. For example, at baseline, students in Cohort 1 schools were half as likely to be proficient in math on the state assessment as students in Cohorts 2 and 3 (11 percent in Cohort 1 versus 23 percent in both Cohorts 2 and 3). Thus, Cohort 1 schools could have differed from schools in later cohorts in other ways that affected program effectiveness.

Our final exploratory analysis examined whether TLTS had an effect on teacher perceptions and practices as measured by a study-administered teacher survey in the second year of program implementation. There were limitations of the survey data described above and in Chapter II which do not show evidence of significant changes in school leadership, school culture, or teacher instruction in TLTS schools. This remains an area for future research.

This page has been left blank for double-sided copying.

IV. Fidelity of Program Implementation

This chapter summarizes implementation fidelity of TNTP's Turnaround Leadership Teams Strategy (TLTS) program in two urban school districts in the southeast United States during two years of the Education Innovation and Research (EIR) grant (2019–2020 and 2020–2021). For each year of the analysis, we examine fidelity across the sample of schools that began TLTS in that school year: Cohort 2 in 2019–2020, and Cohort 3 in 2020–2021. We did not assess implementation fidelity for the first cohort that began the program in 2018–2019 as fidelity measures were still being developed in that year. The analysis focuses on the first year of implementation in each school because, while some coaching supports continued in year 2, core intervention supports were concentrated in Year 1.

Relying on TNTP program data and feedback from district staff and principals, the analysis assesses whether the two program components of TLTS described in the program logic model (Figure I.1) — namely, diagnosing school needs and developing and supporting school leadership teams (SLTs)—were adequately implemented within these two cohorts. TNTP provided input on indicators and scoring to determine whether these key program components were implemented with fidelity.

For each of the two key program components, the fidelity analysis examines both implementation and the specific indicators used for assessment separately. Indicators describe the specific activities that were to be implemented under each component. For each key program component, the fidelity analysis assigns a single point for each indicator successfully met, sums the indicator scores, and then determines whether the total score meets the threshold for adequate implementation at the component level.

A. Key program component 1: diagnosing school needs

1. Measuring key program component 1

The first key program component of TLTS—diagnosing school needs—describes activities that TNTP undertook with each new participating school in the early months of the program to diagnose each school's needs for support and develop a plan for the coaching TNTP staff provide. Diagnosing school needs involved nine school-level indicators categorized into four types of diagnostic and planning activities for each school: (1) academic diagnostics, (2) community engagement diagnostics, (3) TNTP's Instructional Culture Insight Survey, and (4) customized instructional coaching and community engagement plans. Table IV.1 describes the nine indicators that measured component-level implementation, their operational definition, and scoring at the school and sample levels.

Table IV.1. Measuring implementation fidelity for key program component 1 (diagnosing school needs)

Indicator	Unit of measurement	Operational definition	Indicator scoring at school level	Adequate fidelity at sample level
Academic diagnostics				
Conduct teacher focus groups	School	At least one focus group conducted	0: Not completed 1: Completed	At least 80% of schools receive a score of 1
Conduct class observations	School	≥ 4 observations conducted	0: < 4 observations 1: ≥ 4 observations	At least 80% of schools receive a score of 1
Conduct student work analysis	School	≥ 12 work samples analyzed	0: < 12 samples 1: ≥ 12 samples	At least 80% of schools receive a score of 1
Community engagement diagnostics				
Interview with school principal	School	Principal participated in interview	0: Not completed 1: Completed	At least 80% of schools receive a score of 1
Conduct parent focus groups	School	At least one focus group conducted	0: Not completed 1: Completed	At least 80% of schools receive a score of 1
Conduct student focus groups	School	At least one focus group conducted	0: Not completed 1: Completed	At least 80% of schools receive a score of 1
Conduct school walk-through	School	School walk-through conducted	0: Not completed 1: Completed	At least 80% of schools receive a score of 1
TNTP Instructional Culture Insight Survey				
Conduct survey	School	At least 40% of eligible staff complete survey[a]	0: < 40% complete 1: ≥ 40% complete	At least 80% of schools receive a score of 1
Customized instructional coaching and community engagement plans				
Create customized coaching plan	School	Coaching plan created	0: No plan created 1: Plan created	At least 80% of schools receive a score of 1
Key component 1 total score:				**Number of indicators that met the threshold for adequate fidelity at the sample-level** **(Range = 0–9)** **Adequate = 7+ indicators met threshold**

[a] Eligible staff were defined as staff responsible for direct instruction for a roster of students and who were employed by the school at least 50 percent of the school year.

2. Findings on implementation fidelity for key program component 1

Table IV.2 presents the implementation fidelity analysis related to TNTP's work to diagnose school needs for the schools beginning TLTS in 2019–2020 (Cohort 2), the schools beginning in 2020–2021 (Cohort 3), and the full implementation fidelity sample (Cohorts 2 and 3).

Academic diagnostics. As part of initial diagnostics, TNTP collected baseline data for newly participating schools focused on the quality and rigor of the school's academics. These diagnostics involved three indicators for each school: (1) conducting a teacher focus group, (2) conducting at least four classroom observations, and (3) analyzing the quality of at least 12 samples of student work. All

schools beginning the program in 2019–2020 (Cohort 2) and 2020–2021 (Cohort 3) met each indicator related to academic diagnostics.

Community engagement diagnostics. Initial diagnostics for new schools also assessed the degree to which students' family members feel welcomed and engaged in school activities. This assessment involved four indicators: (1) completing a principal interview to understand community context, (2) conducting at least one focus group with families, (3) conducting at least one student focus group, and (4) hosting a school walk-through with families. In 2019–2020, schools met all indicators. In 2020–2021, schools completed the first three indicators; however, due to the lack of on-campus activities during the COVID-19 pandemic, TNTP was unable to complete school walk-throughs that academic year.

TNTP's Instructional Culture Insight Survey. TNTP also conducted the internally developed TNTP Instructional Culture Insight Survey with school staff as a part of early diagnostics to gather their perspective on the school environment. TNTP's standard for successful completion of the survey was to collect responses from at least 40 percent of eligible staff.[17] Both cohorts met this standard; more than half of eligible staff at each school completed the fall survey.

Table IV.2. Implementation fidelity for key program component 1 (diagnosing school needs)

Indicator	Sample-level implementation					
	Cohort 2 (2019–2020) 6 schools		Cohort 3 (2020–2021) 6 schools		Full sample (2019–2021) 12 schools	
	Schools meeting threshold (percent)	Adequate implementation (≥ 80% of schools)	Schools meeting threshold (percent)	Adequate implementation (≥ 80% of schools)	Schools meeting threshold (percent)	Adequate implementation (≥ 80% of schools)
Academic diagnostics						
Conduct teacher focus groups	100	Yes	100	Yes	100	Yes
Conduct class observations	100	Yes	100	Yes	100	Yes
Conduct student work analysis	100	Yes	100	Yes	100	Yes
Community engagement diagnostics						
Interview with school principal	100	Yes	100	Yes	100	Yes
Conduct parent and community focus groups	100	Yes	100	Yes	100	Yes
Conduct student focus groups	100	Yes	100	Yes	100	Yes
Conduct school walk-through[a]	100	Yes	0	No	50	No

[17] Eligible staff were defined as staff responsible for direct instruction of a full classroom (that is, staff responsible for a roster of students) and who were employed by the school at least 50 percent of the school year.

Indicator	Sample-level implementation					
	Cohort 2 (2019–2020) 6 schools		Cohort 3 (2020–2021) 6 schools		Full sample (2019–2021) 12 schools	
	Schools meeting threshold (percent)	Adequate implementation (≥ 80% of schools)	Schools meeting threshold (percent)	Adequate implementation (≥ 80% of schools)	Schools meeting threshold (percent)	Adequate implementation (≥ 80% of schools)
TNTP Instructional Culture Insight Survey						
Survey conducted	100	Yes	100	Yes	100	Yes
Customized instructional coaching and community engagement plans						
Customized plan created for school	100	Yes	100	Yes	100	Yes
Total indicators adequately implemented		**9 indicators**		**8 indicators**		**8 indicators**
Cohort met implementation fidelity for component?		**Yes, ≥ 7+ indicators adequately implemented**		**Yes, ≥ 7+ indicators adequately implemented**		**Yes, ≥ 7+ indicators adequately implemented**

Source: TNTP program data.

[a] In 2020–2021, TNTP was unable to conduct a school walk-through at any of the six schools in Cohort 3 due to the impact of the COVID-19 pandemic on schools.

Customized instructional coaching and community engagement plans. Under the TLTS model, each school was expected to use diagnostic data to develop a customized plan for instructional support, including identifying a specific grade, classrooms, and content area for subsequent coaching supports. All new schools joining TLTS in 2019–2020 (Cohort 2) and 2020–2021 (Cohort 3) developed an instructional support plan with TNTP during the early months of engagement.

3. Component-level assessment for key program component 1

TNTP defined adequate component implementation as the completion of at least seven (of the nine) indicators. Schools that began the program in 2019–2020 (Cohort 2) met all nine indicators; schools that began the program in 2020–2021 (Cohort 3) completed eight of the nine indicators. Thus, TNTP achieved fidelity in implementing this component for all schools and for the sample level overall.

B. Key program component 2: developing and supporting SLTs

1. Measuring key program component 2

TNTP provided ongoing professional development and coaching throughout the school year. For the second key program component of TLTS—developing and supporting SLTs—evaluation involved five indicators grouped into three activities: (1) a Transformational Leadership Summit the summer before program launch, (2) supports aligned with school needs, and (3) supports aligned with district priorities. Table IV.3. describes the five indicators that measured component-level implementation, their operational definition, and scoring at the school and sample levels.

Table IV.3. Measuring implementation fidelity for key program component 2 (developing and supporting SLTs)

Indicator	Unit of measurement	Operational definition	Indicator scoring at school level	Adequate fidelity at sample level
Transformational Leadership Summit				
SLT participation	School	50% of SLT participates in ≥ 8 hours of training[a]	0: <50% participated 1: ≥ 50% participated	More than 80% of schools receive a score of 1
Principal participation	School	Principal participates in ≥12 hours of training[a]	0: Principal did not participate 1: Principal participated	At least 80% of schools receive a score of 1
Supports aligned with school needs				
Coaching plan is aligned with school needs	School	Principal reports that plan is aligned with school needs	0: Plan is not aligned 1: Plan is aligned	At least 80% of schools receive a score of 1
Coaching is in school focus area (based on plan)	School	Principal reports that their SLT received coaching in focus areas[b]	0: Principal does not receive coaching 1: Principal receives coaching	At least 80% of schools receive a score of 1
Supports aligned with district priorities				
Supports align with each district's priorities	District (score at school level)[c]	District leader reports that plan is aligned with district priorities[c]	0: Support is not aligned 1: Support is aligned	At least 80% of schools receive a score of 1
Key program component 2 total score:				**Number of indicators that met the threshold for adequate fidelity at the sample-level** **(Range = 0–5)** **Adequate = 4+ indicators met threshold**

[a] Includes training and prework.

[b] Plans identified three focus areas in which coaches would provide supports: (1) a specific grade, (2) a content area (either English language arts or math), and (3) select classrooms in the focal grade.

[c] If a district leader reported that supports were aligned with district priorities, all schools in that district received an indicator of 1.

SLT = school leadership team.

2. Findings on fidelity of implementation for key program component 1

Table IV.4. presents the fidelity analysis related to TNTP's work to develop and support SLTs in schools that began TLTS in 2019–2020 (Cohort 2) and 2020–2021 (Cohort 3).

Transformational Leadership Summit. The Transformational Leadership Summit was a three-day summer intensive training for SLTs. Held before their first school year in TLTS, it was a core element of the support the program provides to schools. The summit laid the groundwork for subsequent program activities and supports. As a part of its fidelity plan, TNTP specified that the summer summit should satisfy two key indicators, both focused on attendee participation: (1) at least 50 percent of all SLT members participate in at least 8 hours of training, and (2) all principals participate in at least 12 hours of training. All schools in both cohorts met both indicators.

Supports aligned with school needs. TNTP also provided intensive, individualized coaching to principals and SLTs throughout the school year. As part of its fidelity plan, TNTP identified two critical indicators that inform this coaching: (1) plans for TNTP's instructional coaching and community engagement support should be aligned with school needs, including the grade, classrooms, and content area designated to receive coaching supports, and (2) schools should receive coaching in the focus areas specified in their coaching plans. Schools reported on these supports through written feedback. All principals reported that school plans aligned with their school needs and that subsequent support was aligned with those plans.

Supports aligned with district priorities. Based on written feedback forms, leaders of both districts reported that TNTP supports were aligned with their district's priorities for the 2019–2020 and 2020–2021 school years. Thus, TNTP completed this activity with fidelity for all cohort schools and for the program overall.

Component-level assessment for developing and supporting SLTs. Adequate implementation of this component was defined as the completion of at least four of the five key indicators described in Table IV.3 above. All schools in both cohorts satisfied all five indicators for developing and supporting SLTs.

Table IV.4. Implementation fidelity for key program component 2 (developing and supporting SLTs)

Indicator	Sample-level implementation					
	Cohort 2 (2019–2020) 6 schools		Cohort 3 (2020–2021) 6 schools		Full sample (2019–2021) 12 schools	
	Schools meeting threshold (percent)	Adequate implementation (≥ 80% of schools)	Schools meeting threshold (percent)	Adequate implementation (≥ 80% of schools)	Schools meeting threshold (percent)	Adequate implementation (≥ 80% of schools)
Transformational Leadership Summit						
SLT participation	83	Yes	100	Yes	92.5	Yes
Principal participation	100	Yes	100	Yes	100	Yes
Supports aligned with school needs						
Coaching plan is aligned with school needs	100	Yes	100	Yes	100	Yes
Coaching is in school focus area (based on plan)	100	Yes	100	Yes	100	Yes
Supports aligned with district priorities						
Supports align with each district's priorities	100	Yes	100	Yes	100	Yes
Total indicators adequately implemented		**5 indicators**		**5 indicators**		**5 indicators**
Cohort met implementation fidelity for component?		**Yes, ≥ 4 indicators adequately implemented**		**Yes, ≥ 4 indicators adequately implemented**		**Yes, ≥ 4 indicators adequately implemented**

Source: TNTP program data and district leader and principal interviews.

SLT = school leadership team.

C. Conclusion

Across schools that began TLTS in 2019–2020 (Cohort 2) and schools that began the program in 2020–2021 (Cohort 3), TNTP achieved fidelity for the program model's two key program components: (1) diagnosing school needs and (2) developing and supporting SLTs. TNTP's fidelity to its model is notable given the extraordinary challenges the pandemic created for participating schools, as the learning environment remained virtual for spring 2020 and then intermittently virtual for specific schools and classrooms in 2020-2021. Despite the many barriers the pandemic posed for schools across the country, TNTP continued most of its planned activities to support participating districts and schools (See Appendix A for further discussion of supports during the pandemic). This included transitioning to a virtual environment across both cohorts (including holding the Transformational Leadership Summit training virtually), documenting most aspects of school diagnostics virtually, and providing ongoing coaching while navigating remote learning instruction. With the exception of the school walk-throughs with students' families, which were not conducted in 2020–2021 due to the cancellation of in-person school activities, TNTP completed all planned activities with participating schools in these cohorts.

This page has been left blank for double-sided copying.

V. Fidelity to EIR Scale-up Goals and Cost Analysis

As an Education Innovation and Research (EIR) Mid-phase grantee, TNTP established goals for the Turnaround Leadership Teams Strategy (TLTS) scale-up grant that it aimed to achieve by the third year of implementation (2020–2021). TNTP also designed proactive strategies for overcoming potential barriers to reaching these goals. As a separate grant requirement, TNTP also conducted a cost analysis that summarized per student costs of TLTS. This chapter presents an analysis of TNTP's fidelity to its planned scale-up goals as well as an analysis of per student costs under the grant.

A. Fidelity to EIR scale-up goals

We conducted a quantitative analysis to measure and document the extent to which TNTP achieved its stated goals for the EIR grant and the degree to which it implemented planned strategies for accomplishing these goals. To describe implementation of the scale-up, we synthesized data gathered from TNTP's written plans for the scale-up, program documentation and files, correspondence with TNTP, and correspondence with participating districts and schools.

1. TNTP's scale-up goals

Under the EIR scale-up, TNTP identified and successfully achieved two scale-up goals: (1) expanding TLTS to two local education agencies by the third year of implementation (2020–2021) and (2) expanding the TLTS program to support 14 schools across these two districts by 2020–2021. TNTP accomplished the first goal by expanding TLTS to two local education agencies (the two districts in the southeast United States engaged in this evaluation). One of these districts launched the program in its schools in 2018–2019, and the other district launched the program in 2019–2020. TNTP achieved the second goal by successfully serving 15 schools through TLTS by the start of the 2020–2021 school year (Table V.1).

Table V.1. TNTP attainment of EIR scale-up goals

Scale-up goals	Successful implementation threshold	Data collection or measure	Results under EIR grant	Minimum threshold met?
Goal 1: Expand TLTS to serve two local education agencies by grant end	At least two local education agencies served by third year of implementation (2020–2021)	Records indicating that two local education agencies are engaged in TLTS	Expanded to two local education agencies	Yes
Goal 2: Expand TLTS to serve 15 schools by grant end	At least 14 schools served by third year of implementation (2020–2021)	Records documenting school participation	15 program schools engaged	Yes

Source: TNTP program data.

TLTS = Turnaround Leadership Teams Strategy.

2. Strategies to address barriers to scale-up

Before launching TLTS under the EIR grant, TNTP also identified two key barriers the program would need to address to successfully scale the TLTS model: (1) intervention costs limit the number of schools TNTP can serve, and (2) school and district leaders may be less likely to fully incorporate and sustain TLTS practices over time given that TNTP is an organization external to district and school administration (Table V.2).

To address the first barrier—cost—TNTP implemented a **cost reduction strategy** with three key mechanisms:

1. Increasing the average number of schools each coach served (to at least three schools per coach) in order to reduce the intervention's human capital requirements per school.
2. Developing district staff capacity and capabilities to implement TLTS strategies in order to increase the number of schools that could benefit from TLTS strategies beyond the grant period. Specifically, TNTP provided coaching on core aspects of school supports to key administrative staff members in each district. (This mechanism also addressed the second barrier.)
3. Codifying the recommendations on practice and sustainable approaches for districts on a website created during the grant to support district capacity-building.

To address the second barrier—TNTP's position as a district outsider—TNTP implemented an **engagement strategy** with two key mechanisms:

1. Hiring staff from the metro areas surrounding each district. All non-temporary TNTP coaching staff resided in or near the relevant district at their time of hire. With their knowledge of the local context, these staff members could forge lasting relationships with local partners.
2. Training district staff as program coaches (as discussed above).

Table V.2. TNTP EIR scale-up strategies and mechanisms

Scale-up strategies and mechanisms	Successful implementation threshold	Data source	Results under EIR grant	Minimum threshold met?
Strategy 1: Cost reduction				
Mechanism 1: TNTP will increase the number of schools each coach serves	Coaches work with an average of at least three schools per year	Coaching logs	Coaches served an average of 3.7 schools per year	Yes
Mechanism 2: TNTP will coach at least one staff person in each district[a]	District staff identified by TNTP confirm they received training	Documentation by staff member that received coaching	Six staff trained in District 1; five trained in District 2	Yes
Mechanism 3: TNTP will codify recommended practices for each district	TNTP creates website that codifies best practices by grant end	Review of website codifying best practices	TNTP generated website codifying best practices	Yes
Strategy 2: Program engagement				
Mechanism 1: TNTP will hire regional staff in each district to provide direct coaching to schools	All non-temporary TNTP coaching staff reside in the district's metro area at their time of hire	TNTP submits records listing all non-temporary staff and their addresses	All coaching staff (six coaches) reside in their respective district's metro area	Yes
Mechanism 2: TNTP will coach at least one staff person in each district on TLTS program model[a]	District staff identified by TNTP confirm they received training	Documentation by staff member that received coaching	Six staff trained in District 1; five trained in District 2	Yes

Source: TNTP program data.

[a] TNTP indicated that training district staff served as a mechanism to both reduce program costs overtime (by building district-level capacity to support ongoing efforts) and foster program engagement.

TLTS = Turnaround Leadership Teams Strategy.

B. Cost analysis for the TLTS program

As part of performance reporting, EIR projects must report per-pupil expenditures annually, including information on how expenditures were calculated. The cost analysis in this section provides information on per-pupil costs for each school year. Per-pupil cost estimates could help TNTP replicate the program in the future in the future; they could also provide useful information for school districts and members of the larger education community that may want to implement the program in the future.

1. Costs per student and per school

For schools engaged in the TLTS program for the full two years of the TLTS program, the average cost per student was $484.40, or $242.20 per student for a single year (Table V.3). TNTP provided program costs to the study team for all program years (2018–2022). These costs reflected the following expenditures directly related to program provision: (1) staff time of core project team and national staff who supported diagnostic activities and surveys, (2) program materials and meeting costs, (3) program-related staff travel, and (4) stipends paid to districts to help support implementation at participating schools. These program costs exclude expenses that were not associated with program provision, such as program evaluation costs and grant costs, including grant management.

To calculate per-pupil costs for a single year of program participation, the study team divided total program costs by the total number of students served in all 15 participating program schools during both years of each district's program participation (based on student enrollment data from the Common Core of Data), resulting in 16,579 students. The study team calculated the average total cost of two program years by multiplying a single year's cost by two to reflect two years of engagement.[18]

Table V.3 also presents the average program cost per school. Because the TLTS program provided supports at the school level, and not at the student level, school-level costs may also be informative for district leaders and the larger education community. For two years of TLTS program engagement, the average cost per school was $267,701.00, or $133,850.50 for a single year (Table V.3).

In practice, program costs varied across the school years in which TLTS was implemented under EIR. This was largely due to the fixed costs required to initiate the program at the district and school levels and the number of students TLTS participating schools served each year. In Table V.3, we also summarize the average per-pupil and per-school costs for each school year in which TLTS was implemented under the EIR grant. These costs were calculated using the same program costs TNTP provided to the study team, the total number of schools implementing TLTS in a given school year, and the number of students enrolled in each TLTS school, regardless of whether the schools were in their first or second year of TLTS implementation. Per-pupil costs varied, ranging from a low of $183.88 in 2019–2020, to a high of $446.40 in 2018–2019. Similarly, per-school costs varied from $103,237.44 in 2019–2020, to $200,566.67 in 2018–2019.

[18] Although TLTS program activities were more heavily concentrated in a school's first year of participation, program-level costs could not be isolated from a specific school's costs each year. Thus, costs are presented as an annual average across all schools.

Table V.3. TLTS per-pupil and per-school costs (by years of engagement and school year)

Engagement period	Per-pupil cost (average, US$)	Per-school cost (average, US$)
Years of school engagement in the TLTS program		
One year of school engagement in TLTS program	$242.20	$133,850.50
Two years of school engagement in TLTS program	$484.40	$267,701.00
School year		
2018–2019	$446.40	$200,566.67
2019–2020	$183.88	$103,237.44
2020–2021	$247.46	$116,141.75
2021–2022	$239.93	$181,829.50

Source: TNTP program cost data and Common Core of Data student enrollment figures.

TLTS = Turnaround Leadership Teams Strategy.

2. Alignment of student impacts with per-pupil costs

To support interpretation of the relative costs of achieving program impacts at the district level, Table V.4 summarizes impact estimates for each two-year student achievement outcome (as described in Chapter III) aligned with the average per pupil costs for the program.

Table V.4. TLTS impacts on primary outcomes and per-pupil costs

Primary outcome measure	Impact estimate (standardized effect size)	Average per-pupil cost
Math: two-year impacts	0.085**	$484.40
ELA: two-year impacts	0.050*	$484.40
Science: two-year impacts	0.035	$484.40
Social Studies: two-year impacts	-0.120*	$484.40

Source: TNTP program cost data and findings from impact analysis (See Chapter III).

* Impact estimate is statistically significant at the 5 percent level.

** Impact estimate is statistically significant at the 1 percent level.

TLTS = Turnaround Leadership Teams Strategy.

References

Achieve Inc. "EQuIP Student Work Protocol." 2016. https://www.achieve.org/our-initiatives/equip/all-equip-resources/student-work-protocol/student-work-protocol

Béteille, T., D. Kalogrides, and S. Loeb. "Stepping Stones: Principal Career Paths and School Outcomes." NBER Working Paper No. 17243. Cambridge, MA: National Bureau of Economic Research, July 2011. https://ssrn.com/abstract=1892667.

Bloom, H.S., C.J. Hill, A.R. Black, and M.W. Lipsey. "Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions." *Journal of Research on Educational Effectiveness*, vol. 1, no. 4, October 2008, pp. 289–328. https://doi.org/10.1080/19345740802400072.

Booker, K., and J. Thomas. "Impacts of New Leaders on Student Achievement in Oakland." Washington, DC: Mathematica Policy Research, May 2014. https://mathematica.org/publications/impacts-of-new-leaders-on-student-achievement.

Booker, K., and S. Glazerman. "Effects of the Missouri Career Ladder Program on Teacher Mobility." Washington, DC: Mathematica Policy Research, May 2009. https://files.eric.ed.gov/fulltext/ED507470.pdf.

Branch, G.F., E.A. Hanushek, and S. Rivkin. "School Leaders Matter: Measuring the Impact of Effective Principals." *Education Next*, vol. 13, no. 1, Winter 2013, pp. 62–69. https://www.educationnext.org/school-leaders-matter/.

Calkins, A., W. Guenther, G. Belfiore, and D. Lash. "The Turnaround Challenge: Why America's Best Opportunity to Dramatically Improve Student Achievement Lies in our Worst-performing Schools." Boston, MA: Mass Insight Education and Research Institute, 2007. https://files.eric.ed.gov/fulltext/ED538298.pdf.

Dolfin, S., S. Richman, and J. Choi. "Evaluation of the Teacher Potential Project." Washington, DC: Mathematica, June 2019. https://www.mathematica.org/publications/evaluation-of-the-teacher-potential-project.

Goldring, R., and S. Taie. "Principal Attrition and Mobility: Results From the 2016–17 Principal Follow-up Survey." First Look (NCES 2018-066). National Center for Education Statistics, U.S. Department of Education, July 2018. https://nces.ed.gov/pubs2018/2018066.pdf.

Hallgren, K., N. Gonzalez, and K. Kelly. "Year 2 Report of the Atlanta Public Schools Turnaround Strategy." Princeton, NJ: Mathematica, February 2019. https://mathematica.org/publications/year-2-report-of-the-atlanta-public-schools-turnaround-strategy.

Hallinger, P., and R.H. Heck. "Collaborative Leadership and School Improvement: Understanding the Impact on School Capacity and Student Learning." *School Leadership & Management*, vol. 30, no. 2, April 2010, pp. 95–110. https://doi.org/10.1080/13632431003663214.

Heck, R. H., and P. Hallinger. "Assessing the Contribution of Distributed Leadership to School Improvement and Growth in Math Achievement." *American Educational Research Journal*, vol. 46, no. 3, September 2009, pp. 659–689. https://doi.org/10.3102/0002831209340042.

Henderson, A.T., K.L. Mapp, V.R. Johnson, and D. Davies. *Beyond the Bake Sale: The Essential Guide to Family-School Partnerships*. The New Press, 2017. https://thenewpress.com/books/beyond-bake-sale.

Institute of Education Sciences, and the Regional Educational Laboratory Southeast at Florida State University. "Improving Instruction Through Professional Learning Communities." Washington, DC: Institute of Education Sciences, U.S. Department of Education, n.d. https://ies.ed.gov/ncee/edlabs/infographics/pdf/REL_SE_Improving_Instruction_through_Professional_Learning_Communities.pdf.

Kraft, M.A. "Interpreting Effect Sizes of Education Interventions." *Educational Researcher*, vol. 49, no. 4, April 2020, pp. 241–253. https://doi.org/10.3102/0013189X20912798.

Kutash, J., E. Nico, E. Gorin, S. Rahmatullah, and K. Tallant. "The School Turnaround Field Guide." Boston, MA: FSG Social Impact Advisors, September 2010. https://www.wallacefoundation.org/knowledge-center/documents/the-school-turnaround-field-guide.pdf.

Leithwood, K., A. Harris, and D. Hopkins. "Seven Strong Claims About Successful School Leadership." *School Leadership & Management*, vol. 28, no. 1, January 2008, pp. 27–42. https://doi.org/10.1080/13632430701800060.

Leithwood, K., and B. Mascall. "Collective Leadership Effects on Student Achievement." *Educational Administration Quarterly*, vol. 44, no. 4, August 2008, pp. 529–561. https://doi.org/10.1177/0013161X08321221.

Levin, S., C. Scott, M. Yang, M. Leung-Gagné, and K. Bradley. "Supporting a Strong, Stable Principal Workforce: What Matters and What Can Be Done." Research Brief. Palo Alto, CA: Learning Policy Institute, May 2020. https://learningpolicyinstitute.org/product/supporting-strong-stable-principal-workforce-brief.

Levin, S., M. Leung-Gagné, A.K. Edgerton, and C. Scott. "Elementary School Principals' Professional Learning: Current Status and Future Needs." Palo Alto, CA: Learning Policy Institute, October 2020. https://learningpolicyinstitute.org/product/elementary-school-principals-professional-learning-report.

Louis, K.S., K. Leithwood, K.L. Wahlstrom, and S.E. Anderson. "Investigating the Links to Improved Student Learning: Final Report of Research Findings." Minneapolis, MN: University of Minnesota, 2010. http://www.wallacefoundation.org/knowledge-center/Documents/Investigating-the-Links-to-Improved-Student-Learning.pdf.

Mascall, B., K. Leithwood, T. Straus, and R. Sacks. "The Relationship Between Distributed Leadership and Teachers' Academic Optimism." *Journal of Educational Administration*, vol. 46, no. 2, 2008, pp. 214–228. https://www.hsredesign.org/wp-content/uploads/2018/06/09578230810863271.pdf.

National Center for Education Statistics. "Locale Definitions." 2023. https://nces.ed.gov/programs/edge/docs/LOCALE_CLASSIFICATIONS.pdf

NWEA. "2020 NWEA MAP Growth normative data overview." 2020. https://teach.mapnwea.org/impl/MAPGrowthNormativeDataOverview.pdf.

Renaissance Learning. "Star Assessments™ for Math Technical Manual." 2022a. https://docs.renaissance.com/R54119.

Renaissance Learning. "Star Assessments™ for Reading Technical Manual." 2022b. https://docs.renaissance.com/R43843.

Robinson, V.M.J., C.A. Lloyd, and K.J. Rowe. "The Impact of Leadership on Student Outcomes: An Analysis of the Differential Effects of Leadership Types." *Educational Administration Quarterly*, vol. 44, no. 5, December 2008, pp. 635–674. https://doi.org/10.1177/0013161X08321509.

Seashore Louis, K., B. Dretzke, and K. Wahlstrom. "How Does Leadership Affect Student Achievement? Results from a National U.S. Survey." *School Effectiveness and School Improvement*, vol. 21, no. 3, July 2010, pp. 315–336. https://doi.org/10.1080/09243453.2010.486586.

Sebastian, J., and E. Allensworth. "The Influence of Principal Leadership on Classroom Instruction and Student Learning: A Study of Mediated Pathways to Learning." *Educational Administration Quarterly*, vol. 48, no. 4, February 2012, pp. 626–663. https://doi.org/10.1177/0013161X11436273.

Steele, J.L., E.D. Steiner, and L.S. Hamilton. "Priming the Leadership Pipeline: School Performance and Climate Under an Urban School Leadership Residency Program." *Educational Administration Quarterly*, vol. 57, no. 2, 2021, pp. 221–256. https://files.eric.ed.gov/fulltext/EJ1291815.pdf.

Supovitz, J.A., and M. Riggan. "Building a Foundation for School Leadership: An Evaluation of the Annenberg Distributed Leadership Project, 2006-2010." Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania, 2012. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1013&context=cpre_researchreports.

TNTP. "The Opportunity Myth: What Students Can Show Us About How School Is Letting Them Down—and How to Fix It." New York, NY: TNTP, September 2018a. https://tntp.org/publications/view/student-experiences/the-opportunity-myth. Accessed July 27, 2020.

TNTP. "Student Experience Toolkit: Assignment Review Protocols." 2018b. https://tntp.org/student-experience-toolkit/view/assignment-review-protocols.

TNTP. "COVID-19 School Response Toolkit: Partnering with Families Virtually." 2020a. https://tntp.org/covid-19-school-response-toolkit/view/partnering-with-families-virtually.

TNTP. "About Insight." 2020b. https://tntp.org/teacher-talent-toolbox/insight-survey.

Ujifusa, A. "Teachers Have Trust Issues." *Education Week*, Tuesday, January 8, 2019, Federal section. https://www.edweek.org/policy-politics/teachers-have-trust-issues/2019/01.

Waters, T., R.J. Marzano, and B. McNulty. "Balanced Leadership: What 30 Years of Research Tells Us about the Effect of Leadership on Student Achievement." Working Paper. Aurora, CO: Mid-Continent Research for Education and Learning, 2003. https://files.eric.ed.gov/fulltext/ED481972.pdf.

Weisberg, D., S. Sexton, J. Mulhern, and D. Keeling. "The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness." New York, NY: TNTP, 2009. https://tntp.org/publications/view/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness.

Yeh, S.S. "The Cost Effectiveness of 22 Approaches for Raising Student Achievement." *Journal of Education Finance*, vol. 36, no. 1, Summer 2010, pp. 38–75. http://doi.org/10.1353/jef.0.0029.

This page has been left blank for double-sided copying.

Appendix A. Description of Program Supports Provided to Districts and Schools through TLTS

This appendix describes program supports provided to schools in two districts in the southeast United States that participated in TNTP's Turnaround Leadership Teams Strategy (TLTS) and summarizes how TNTP staff delivered those supports. TNTP supported three cohorts of schools, each of which began the program in a different school year: 2018-2019, 2019-2020, and 2020-2021. The description of supports draws from TNTP program data and interviews with TNTP staff and district and school leaders.

A. Overview

TNTP's TLTS is a leadership development model designed to prepare school leaders for the challenge of turning around low-performing schools. TLTS strives to improve leadership practices among (1) principals, (2) school leadership teams (SLTs), which, in addition to principals, typically include assistant principals, instructional coaches, and instructional support specialists, and (3) teachers in select grades that have expressed an interest in improving teaching practices. The program aims to support professional growth among school leaders; improve teacher instructional practices, school culture, and family engagement in school activities; and, ultimately, to increase student achievement at TLTS program schools (Figure I.1). During the Education Innovation and Research (EIR) grant, program supports provided to schools through TLTS focused on four levers for school transformation:

1. **Collective leadership.** Building trusting relationships to work collectively toward a shared vision.
2. **Vision and culture.** Co-creating an inclusive culture and a vision for student success.
3. **Authentic community engagement.** Engaging students and families as equal partners in education.
4. **Academics.** Developing strong instructional practices to ensure all students have access to key resources to improve student learning.

As a core feature of TLTS's collective leadership approach, each participating school established an SLT that would engage in all program activities, including trainings, ongoing coaching sessions, and leadership activities. Through TLTS, TNTP staff then provided customized supports to each school's SLTs through two key program components. Throughout TLTS, TNTP would also connect with district-level partners to set and refine priorities for collaboration with schools.

1. Key program component 1: diagnosing school needs

In the fall of their first year in TLTS, TLTS schools and districts worked closely with TNTP staff to establish focus areas for coaching and training support and to conduct preliminary diagnostics with schools during the first semester of the school year. District-level liaisons to the TNTP program provided guidance and priorities for school-level supports. Schools new to the program then undertook a structured process to diagnose school needs and prioritize focus areas for support during TLTS. Early activities to diagnose and plan for school needs included (1) conducting a comprehensive academic diagnostic with each school that focused on the quality and rigor of instruction and student assignments, (2) conducting a community engagement diagnostic that focused on the degree to which families feel welcomed and are engaged in school activities; (3) fielding an internally-developed Instructional Culture Insight Survey among teachers to learn more about school culture; and (4) utilizing results from the diagnostics and survey to directly inform the development of the school's plan for ongoing TNTP support.

2. Key program component 2: developing and supporting SLTs

Following initial diagnostics and planning, a TNTP coach assigned to the TLTS school provided ongoing coaching and supports to SLTs to develop a school vision and improve leadership competencies. These supports included a three-day summer leadership training before the beginning of their first school year in the program and, following the diagnostic assessments referenced above, ongoing instructional coaching provided by the assigned coach that aligned with each school's plan for support.

TLTS supports under the EIR grant spanned two years, though most program activities were concentrated in the first year of program participation. An assessment of the degree to which TNTP maintained fidelity to these two planned program components is summarized in Chapter 5.

3. Background and experience of program staff

TLTS was supported by a total of 12 program staff who directly engaged districts and schools across the program span. Most academic coaches and program directors were former classroom teachers or administrators (principals or assistant principals), and several staff brought expertise in community engagement to the program, having formerly served as community advisors, advocates, or program directors. All coaches and the program director were certified as TNTP diagnostic reviewers, which requires annual recertification and a passing score on qualifying diagnostic assessments.

B. Component 1: Diagnosing school needs

1. Setting priorities with district staff

In the first year of participation, TLTS schools and districts worked closely with TNTP staff to establish focus areas for coaching and training support and to conduct preliminary diagnostics with schools during the first semester of the school year. Prior to engaging with participating schools, TNTP staff met with key district staff to identify their core priorities for school-level support and ensure that the focus of TNTP coaching for each school aligned with those priorities. District liaisons served in leadership roles focused on academics and instruction. Throughout the implementation of TLTS in schools, TNTP would continue to connect with the district partners on an ongoing basis to provide updates on school-level work and progress and to refine priorities for collaboration with schools.

2. Conducting diagnostics with TLTS schools

To launch school engagement, TNTP conducted diagnostics with every school participating in TLTS to gather baseline information. During the fall months of the first program year, TNTP conducted a series of diagnostic activities with each school that began TLTS to gather baseline information on school academics and community-school partnerships. TNTP staff used a systematic, integrated diagnostic approach that included (1) an academic diagnostic, (2) a community engagement diagnostic, and (3) an Instructional Culture Insight Survey developed by TNTP and administered to instructional staff (TNTP 2020b).

Academic diagnostics. TNTP conducted school site visits that included several activities to assess academic quality in each participating school at baseline (Table A.1). TNTP collected baseline data on two key academic topics: (1) teachers' instructional practices and (2) the quality and rigor of student assignments. During these academic site visits, school principals, other SLT members, and district staff accompanied TNTP staff. Academic site visits took place over the course of several days. They were

guided by highly structured protocols and included school principal interviews, teacher focus groups, classroom observations, student assignment reviews, and debriefings with school and district leaders.

Table A.1. Academic diagnostic activities conducted in TLTS schools

Diagnostic activity	Topics	Minimum quantity	Minimum duration	Format	Modality
Informal interviews with school principals	Conducted during initial meetings to capture perceptions on leadership, central office support, vision, culture, instructional vision, curriculum and resources, community engagement, and expectations for students and teacher support	One per school	Varied	Individual meetings	In-person
Teacher focus groups	Gathered teacher input on strengths and challenges in their school community, perceptions of what excellent instruction looks like, curriculum materials available to teachers, professional development opportunities, teachers' approaches to family engagement, and principal expectations	One per school	60 minutes	Group meetings	In-person
Classroom observations	Examined evidence of student engagement, instruction alignment with standards, opportunities for student practice, connection to real-world contexts, use of high-quality materials, and quality of student tasks	Two classrooms per grade level	25–30 minutes	Observations	In-person
Student assignment reviews	Examined the academic rigor and quality of student assignments, including the degree to which assignments aligned with standards, offered meaningful practice opportunities, and connected to real-world issues or context. Reviews also examined student performance on the assignment if students had completed the task	12 work samples per school	NA	Work review	Work samples collected during site visits and examined by TNTP staff following review
Debriefing with school and district leaders	Conducted with school and district partners between observations and at the end of the school day to discuss trends and key findings	One per school	30-minute debriefing at end of school day as well as short debriefs between observations	Individual or group meetings	In-person

Source: TNTP program data.

TLTS = Turnaround Leadership Teams Strategy.

Community engagement diagnostics. TNTP also conducted additional on-site data collection during the first semester of the first program year to assess the school's level of community engagement at baseline (Table A.2.). Data collection examined the degree to which students' family members and community partners felt welcomed in each school and engaged in school activities. Community engagement diagnostics examined several key aspects of family engagement, including (1) each school's established connections to families, (2) the school's promotion of extracurricular learning opportunities for students, and (3) efforts to invite families to inform each school's instruction, planning, and decision-making

processes.[19] TNTP used structured protocols to gather community engagement data. The community engagement site visits gathered data from family focus groups, student focus groups, school walkthroughs with family members, and interviews with school principals.

Table A.2. Community engagement diagnostic activities conducted in TLTS schools

Diagnostic activity	Topics	Quantity	Duration	Format	Modality
Family focus groups	Collected information on parents' current and desired levels of engagement with schools	One per school	Approximately 90 minutes	Group meetings	In-person
Student focus groups[a]	Focused on the degree to which students felt welcomed and supported in their school and challenged by school assignments, and teacher expectations of students	One per school	Approximately 60 minutes	Group meetings	In-person
School walkthroughs with family members[b]	Examined the degree to which the school environment is welcoming to students and their families	One per school	Approximately 60 minutes or less	Group meetings	In-person
Informal interviews with school principals	Conducted during initial meetings that gathered information on both academic culture and community engagement	One per school	Varied	Individual meetings	In-person

Source: TNTP program data.

[a] Activities in 2018-2019 and 2019-2020 also included student shadowing. TNTP discontinued student shadowing activities in 2020-2021 due to COVID-19 restrictions on building access.

[b] TNTP did not conduct school walkthrough activities in 2020-2021 due to COVID-19 restrictions on building access.

TLTS = Turnaround Leadership Teams Strategy.

Instructional Culture Insight Survey. Developed by the national TNTP program office, this survey gathered teacher input across 15 domains pertaining to staff, school leadership, and family engagement (TNTP 2020b).[20] TNTP administered this survey to TLTS schools in the fall of their first program year as part of initial school diagnostics and again in the spring. TNTP staff aimed to administer a survey to all staff in TLTS schools who provide direct instruction to students and work in the school at least 50 percent of the school year. The survey was also administered to schools during their second year of engagement in the TLTS program.

3. Creating instructional coaching and community engagement plans

Following diagnostic activities, TNTP presented each SLT with a tailored report on diagnostic findings for its school. A single coach would collaborate with a school on planning activities, and this coach would continue to provide individualized supports to the school on an ongoing basis. Together with the TNTP coach, each SLT co-developed a customized plan for ongoing support based on the diagnostic findings. The plan specified three key areas where TNTP and SLTs would focus activities and supports during the second semester of the year:

[19] TNTP assessed community engagement based on the Family-School Partnership Rubric developed by Henderson et al. (2017).

[20] Domains addressed in the survey included instructional planning for student growth; professional development; peer culture; evaluation; learning environment; observation and feedback; academic opportunity; leadership; diversity, equity, and inclusion; family and community engagement; school operations; career progression; workload; teacher compensation; and the school's hiring process.

1. A single grade level that would benefit most from additional academic support based on diagnostic input, student assessment results, or both.
2. Two or three focus classrooms within the targeted grade in which teachers had demonstrated openness to adopting new instructional practices and had a minimum of three years of experience.
3. A focus content area—English language arts (ELA) or math—that would benefit most from additional academic supports based on diagnostics results, assessment results from the previous year, or both.

TNTP worked with SLTs using structured worksheets to analyze diagnostic findings, prioritize areas for support, and create school plans. TNTP also gathered additional information to support ongoing coaching, including logistics (such as identifying shared planning times for teachers) and upcoming curriculum modules that these classrooms would cover in the spring.

C. Component 2: Developing and supporting SLTs

In addition to diagnostic activities, TNTP provided ongoing coaching and supports to schools throughout their participation in the program. Coaching supports began with a summer summit, which launched prior to the first year of program participation, and continued through ongoing, individualized coaching for each school.

1. Transformational Leadership Summit

In the first year of program participation—prior to the launch of school-level diagnostic activities described above—TNTP conducted a summer summit with newly joining schools to kick off program activities (Table A.3). TNTP's Transformational Leadership Summit, a three-day summer intensive workshop for SLTs, included principals, assistant principals, instructional coaches, and select teachers from the new cohort of schools. District staff that served as liaisons to the TNTP program also attended. During the Transformational Leadership Summit, TNTP staff led discussions with SLTs on topics that would ground their TLTS work, including school leadership and vision, building community among participating schools, and exploring collective leadership methods.

Attendance. As indicated in the summary of implementation fidelity (Chapter IV), attendance was high among schools in both districts. Additionally, district-level liaisons most involved in implementing TLTS in both districts participated in half of the first day of the summit.

Duration. The full training encompassed approximately 18 hours of programming. TNTP required SLT members who were unable to attend the summit to attend a make-up session, which consisted of eight hours of programming over the course of one day and featured many of the same, albeit condensed, activities as the full session.

Focus and structure. The summit's structure included full-group presentations, large-group and small-group team-building exercises, and individual reflection sessions. SLTs also participated in large-group team-building exercises. The summit focused on three overarching learning objectives for SLTs during the summit: (1) understanding their identities as transformational leaders, (2) exploring models of leadership to enhance their current leadership practice, and (3) building community among participating schools.

During the first day of the training, TNTP facilitators introduced a core concept that would guide summit activities: the student academic opportunity gap. The opportunity gap refers to challenges that many

students—especially students of color, those from low-income families, those with disabilities, and English language learners—face in accessing four important educational resources that are critical for success: (1) grade-appropriate assignments, (2) strong classroom instruction, (3) deep engagement with classroom assignments, and (4) teachers with high expectations (TNTP 2018a). TNTP facilitators summarized findings from a 2018 TNTP report, *The Opportunity Myth*, which presents evidence on the opportunity gap and the ways the education system has systematically reproduced social and economic inequalities and led a discussion on the ways school transformation can address these inequalities. Summit sessions also focused on leadership style, collective leadership, leadership development, and personal reflection. The last day of the training concluded with an overview of TLTS's summer and fall activities, including school-level diagnostics and planning.

Table A.3. Components of TLTS's Transformational Leadership Summit

Component	Summary
Timing	Summer months prior to start of TLTS
Duration	3 days of training • Summer 2018 and summer 2019: minimum 18 hours • Summer 2020: 11-15 hours [a]
Format	Group format (SLT and district leader participants)
Topics	• Collective commitments to engage in training activities • Understanding student "opportunity gaps" and structural oppression in education • Leadership style and its impact on school communities • Collective leadership and community building • Asset-mapping activities to identify team strengths • School team commitments on how they will implement lessons learned from the summer training
Structure	Full-group overview and presentations, large-group and small-group team-building exercises, and individual reflection sessions
Modality	In-person (summer 2018, summer 2019) Virtual (summer 2020)

Source: TNTP program data.

[a] In summer 2020, training was virtual, given disruptions to in-person activities during the pandemic. It included 9 hours of online group training and 2-3 hours of independent work. Principals also participated in an additional three hours of summer learning through one-on-one coaching supports. TNTP held make-up sessions for attendees that could not participate during the originally scheduled sessions.

TLTS = Turnaround Leadership Teams Strategy; SLT = school leadership teams.

2. Ongoing coaching and other supports

Following diagnostics and planning, TNTP staff provided ongoing coaching support to schools throughout the second semester of the first TLTS year for each school. TNTP coaches individually supported each school through instructional coaching cycles (ICCs) and held bimonthly, district-specific, all-cohort meetings with TLTS schools to build community and discuss pertinent themes. In 2020 due to the Covid-19 pandemic, TNTP transitioned all-cohort meetings from in-person to virtual. As a result, schools across both districts participated in joint meetings and eventually TNTP shifted to a cross-district elementary school meetings and separate middle school meetings. Participants responded favorably to this shift and collaboration across schools and districts were enhanced. TNTP continued this format throughout the duration of the program.

Instructional coaching cycles. TNTP coaches assigned to each school supported SLTs at least every other week through individualized coaching. TNTP provided coaching through a series of ICCs that focused on applying tools and methods to ensure that each school's students had access to grade-appropriate assignments, strong instruction, deep student engagement, and high teacher expectations. ICCs focused on the specific content area, grade level, and classrooms identified in each school's support plan. As part of ICCs, SLT members participated in learning sessions, completed student work analysis, and collaborated with teachers in selected classrooms to examine and improve the rigor of classroom assignments. The teachers in the selected classrooms analyzed student work, submitted student work to TNTP for analysis, and incorporated input from the SLT and the TNTP coach into their lesson planning and instructional activities. To the extent possible, TNTP aimed to incorporate ICCs into existing professional learning communities and common planning time structures in each school.

During the second semester (winter and spring) of each cohort's first year of participation, TNTP conducted two consecutive six-week ICCs. The first ICC introduced SLTs to tools and processes for examining assignments and student work. The second ICC built on the work conducted during the first ICC by observing lessons and further assessing the quality of student performance on work assignments. In 2019-2020, due to the impact of COVID-19 and the new priorities created among schools and districts, some TLTS schools did not complete the second ICC.

A suite of TNTP tools to assess the quality of student work, referred to as Assignment Review Protocols (ARPs), served as the foundation for many ICC activities (TNTP 2018b). Informed by Achieve Inc.'s Educators Evaluating Quality Instructional Products (EQuIP) Student Work Protocol, , the ARPs provide teachers and administrators with a structured approach for assessing assignment quality, including the degree to which assignments (1) align with grade-level standards, (2) require students to read the text and apply what they learned, (3) offer meaningful practice opportunities, and (4) are relevant to students' lives and connect to a real-world context (Achieve Inc. 2016). Coaches also guided a review of student performance on the assignment in cases where students completed the task. The ARPs included content-specific versions of the protocol in ELA, math, science, and social studies. Although the protocols were not grade-specific, they instructed reviewers to examine the questions in tandem with grade-specific standards. TNTP coaches also supported school leaders in efforts to establish systems for regular collection and analysis of student work and continued to collect staff feedback and surveys.

Community engagement support. In addition to academic coaching, TNTP provided schools with ongoing support related to community engagement. Initially, SLTs included community representatives, such as parent liaisons, parent volunteers or members of the PTA. In 2020, TLTS coaches expanded support to work directly with each school's larger community engagement team, such as PTA members, staff and/or parent volunteers supporting community engagement, in aligning community activation strategies with each school's improvement plan. Supports to community engagement teams included training using the TLTS framework; onboarding support; and succession planning, given that many were either elected or volunteered to serve on these teams. TNTP also collaborated with district-level community advisory groups that supported parents and staff working to engage parents in each school.

All-cohort meetings. TNTP also held all-cohort meetings with TLTS schools every other month throughout the school year. Districts held separate meetings and included all schools within the district in their first and second year of the program. Attendees included principals, SLTs and community engagement teams from each TLTS school, and leaders from each district office.

All-cohort meetings aimed to supplement ongoing coaching supports to strengthen academic and instructional culture and foster connections among participating TLTS schools. These meetings addressed topics such as the importance of gathering classroom data to improve instruction; fostering community engagement; and understanding ways in which the identity of teachers, administrators, students, and families impact the school community dynamic. TNTP collected feedback surveys from participants after each session.

3. Ongoing supports provided in the second year of program participation

While TLTS intended for the first year of school supports to be the most intensive, TLTS schools received continued supports in the second year of program engagement through ongoing, individualized coaching supports and engagement in all-cohort meetings. Coaches tailored the intensity of second year supports to the interests and availability of school leaders, and many schools continued to receive supports every other week. Coaching scope and sequence continued to align with the four levers of the TLTS program, with a focus on student work analysis and instructional practices in either ELA or math. Most schools continued to focus on the priority subject identified in year one of engagement (ELA or math), although some schools switched their focus between math and ELA from year one to year two.

D. TNTP adaptation and supports in the context of COVID-19

1. Impact of COVID-19 on TLTS

When the COVID-19 pandemic led to the closure of school buildings in mid-March 2020, the transition to remote learning and virtual instruction took priority. This transition was challenging for district administrators and principals, as teachers and administrators found that remote learning required a different infrastructure (such as virtual classes). As a result, the transition to remote learning consumed most of the attention from school and district leadership, shifting their priorities during the spring months. Educators also found that remote learning required new skill sets to facilitate classes virtually, use new remote learning tools, and proactively engage with families online.

Facing these new pressures and competing priorities, schools and districts had a reduced capacity to engage in TLTS. When interviewed, several principals indicated that the priority of establishing virtual learning communities and supporting students without technology resources sometimes hindered their ability to engage fully in TLTS. Principals often felt overwhelmed with the responsibility of transitioning schools to remote learning as planning and emergency meetings often dominated their schedules.

2. TLTS school supports during COVID-19

Before the COVID-19 disruption in spring 2020, TNTP staff focused on implementing the second ICC of the year, which centered on the core topics of grade-appropriate instruction and aligning assignments with standards. Following school closures, TNTP staff connected with principals to revise individual support plans as they transitioned to remote learning. When TNTP staff asked school principals how they could better support them during COVID-19, many requested more support implementing remote learning. In response, TNTP refocused its efforts to provide strategies for effective virtual instruction and communication with families. TNTP offered optional teacher trainings on using online platforms, structuring virtual instruction sessions, and collecting student feedback through online tools. TNTP also engaged families through feedback surveys, held online community meetings, and helped schools develop

student instructional schedules with parents. TNTP also held a school leadership roundtable where school leaders shared best practices for supporting remote learning.

In one district, TNTP provided coaching virtually from March 2020 to late fall 2021, and from March 2020 to winter 2022 in the other district. Given restricted access to buildings, TNTP cancelled in-person observations during that time that schools were closed, limiting in-person support to implementation and monitoring of instructional practices. In early months of the COVID-19 pandemic, the analysis of student work was also limited as schools transitioned to online assignments.

TNTP staff noted the importance of being mindful of district and school capacity immediately at the start of the pandemic, including the frequency of communications, shifts in priorities, and the need to be flexible in collaborating with partners. Each principal determined the level of communication that would suit their needs. In some cases, TNTP would connect with a principal sporadically; other principals requested more support and had multiple touchpoints per week. One district requested that TNTP suspend supports to schools for the period immediately after the pandemic's outset so that schools could adjust to remote learning. TNTP also sent all principals more frequent email updates, increasing from monthly to weekly, to update them on new resources for remote learning and other emerging topics.

Appendix B. Additional Technical Details for the Analyses of Student Achievement Outcomes

This appendix presents additional information about the analytic approach used to estimate the impacts of the Turnaround Leadership Teams Strategy (TLTS) program on student achievement, the attrition and baseline equivalence of the analytic samples used to estimate those impacts, and detailed findings from our exploratory and sensitivity analyses on student achievement.

A. Analytic approach to estimate TLTS impact on student achievement

Our confirmatory analysis assessed the impact of TLTS on student assessment scores in the second year of program implementation for all three cohorts of schools combined. In addition, we explored the impacts of TLTS after one year of implementation and, separately, by cohort. To estimate these impacts, we compared students who attended a school that participated in the TLTS program to students who attended a comparison school, accounting for any differences in students' baseline characteristics and achievement. Specifically, we estimated the regression model in Equation (B1):

$$\text{(B1)}\quad y_{isc} = \alpha + \beta X_i + \theta T_s + s + c + p_s + e_{isc}$$

where y represents the outcome (math, ELA, science, or social studies score) for student i at school s in cohort c; X is a vector of baseline student covariates, including grade, race and ethnicity, gender, special education status, English learner status, and prior math and ELA test scores;[21] T is an indicator for being enrolled at a TLTS school the fall after random assignment;[22] s is a set of indicators for the study school where the student was initially enrolled; c is a set of indicators for the cohort in which the school participated; p is a set of indicators for each matched pair of schools to account for the nature of random assignment; e is a student-level error term; and α, β, and θ are parameters to be estimated. In this framework, the parameter θ represents the impact of being enrolled in a TLTS school. We estimated equation (1) with ordinary least squares using robust standard errors that accounted for the clustering of students at the school level.[23]

These "intent-to-treat" (ITT) impact estimates use a conservative approach that includes students in the treatment group even if they left the treatment school at any point after the start of the program. Therefore, the ITT estimates capture the impact of having been enrolled in a TLTS school at the time the school began participating in the program.

Weights. The model weights each student observation so that each school in the sample contributes equally to the analysis, ensuring the relative weight of the treatment and control school within each matched pair is the same. Specifically, we computed analytic weights for the student analyses using the following formula:

[21] To account for instances where the baseline exam was different than the outcome exam, the model also includes interactions of the baseline test score and the type of test (either district benchmark or statewide assessment) and whether the baseline assessment was different than the outcome assessment.
[22] Students were included in the study if they were enrolled in a study school in grades 4 through 7 before September of the first implementation year for each cohort. Students enrolled after September 1 were excluded from the study.
[23] Cluster-robust standard errors were computed using Huber-White sandwich estimation approach in STATA (version 17).

$$\text{(B2)}\quad w_{isc} = \left(\frac{1}{I_{sc}}\right) x \left(\frac{1}{S}\right) \sum_{s=1}^{S} I_{sc}$$

where w represents the weight for student i at school s in cohort c; I is the total number of students in school s; S is the total number of schools; and $\sum_{s=1}^{S} I_s$ is the total number of students across all schools.

Missing data. Missing covariates, including baseline test scores, were imputed using a dummy imputation approach by setting all missing covariates to a constant value (zero) and included missing indicators in the regression.[24–25] The study did not impute any outcome data.

Sensitivity analyses. We conducted two analyses to test the sensitivity of our confirmatory findings to the data and methods used. First, we conducted a separate analysis that excluded test scores from spring 2021, when state testing was optional, to assess whether the impacts were affected by this atypical year. Second, to assess the extent to which our results were sensitive to our approach to imputing missing baseline data, we conducted a sensitivity analysis that excluded students with any missing baseline data (complete-case analysis).

B. Attrition and baseline equivalence

Below we present information on the attrition of schools and students in each analytic sample for both the confirmatory analyses of Year 2 impacts and the exploratory analyses of Year 1 impacts, along with information on the baseline achievement of each analytic sample.

1. Sample for confirmatory analyses of Year 2 impacts on student achievement

a. Attrition of schools in the analyses of Year 2 impacts

As mentioned in Chapter II, all eligible schools that participated in the study contributed to the student impact analyses. That is, there was no attrition at the school level. However, some outcomes included fewer eligible schools depending on which grades and subjects were tested each year (Table II.7).

b. Attrition of students in the analyses of Year 2 impacts

As with the school samples, the number of students eligible for each analysis varied by outcome year and subject because not all grades and subjects were always tested. As discussed in Chapter II and shown in Table II.8, the analytic samples for our confirmatory Year 2 outcome analyses experienced considerable attrition overall. This was largely due to lower test participation rates during the pandemic, when the state assessment was optional. However, attrition rates were comparable between the TLTS and comparison groups. According to What Works Clearinghouse (WWC) Standards (version 5.0), the overall and differential attrition in each analytic sample used in the confirmatory analyses meet the guidance for low student-level attrition under the optimistic boundary, which, as discussed in Chapter II, likely applies to this study given the nature of the TLTS intervention.

[24] While we use these imputed baseline covariates in our main analysis, none of the imputed values were included in the tests of baseline equivalence of the analytic samples discussed in this appendix or in the baseline characteristics of eligible students presented in Chapter II. For the analysis of baseline equivalence, students missing data on a given variable were excluded from the sample.

[25] According to the WWC Standards (version 5.0), mean and dummy imputation are acceptable approaches to handling missing baseline data in low attrition randomized controlled trials and should therefore not affect the rating of this study.

c. Baseline equivalence of students in the analyses of Year 2 impacts

Considering the sizeable attrition of students in the confirmatory analyses, we conducted tests of equivalence to confirm that students in the intervention and comparison schools entering the analyses remained similar with respect to their baseline achievement, as required by the WWC. Table B.1 presents baseline achievement data among students in each of the confirmatory analytic samples with observed baseline and outcome test scores. Although several of the differences in achievement are statistically significantly different in each analytic sample, the magnitude of the standardized differences in baseline achievement (computed using Hedges' *g*) is smaller than the WWC's threshold of 0.25 standard deviations, ranging from 0.002 standard deviations in math achievement among students in the social studies analysis, to 0.182 standard deviations in ELA achievement among students in the science analysis.

Because the baseline assessment sometimes differs from the outcome assessment, we calculated pairwise correlations of baseline and outcome scores to determine whether the assessments measured similar competencies. As with the standardized differences, these correlations were calculated using the sample of students in each analysis with observed baseline data. The correlations range from 0.60 and 0.77 and, as expected, are typically highest when the baseline and outcome tests are in the same subject. For example, the correlation between baseline and Year 2 ELA scores is 0.77, and the correlation between baseline and Year 2 math scores is 0.70. The WWC recommends that the baseline measures used to establish baseline equivalence have a correlation of 0.60 or higher with the outcome. Each outcome meets this requirement with both baseline measures.

As noted earlier, we control for baseline achievement in the impact regressions. This satisfies the WWC's baseline equivalence requirement given the magnitude of the observed differences and of the correlations between the baseline and outcome measures. Because the analytic samples used in each of the confirmatory analyses have low attrition and establish baseline equivalence, all Year 2 outcomes are eligible to meet the highest WWC rating of Meets Group Design Standards Without Reservations.

Table B.1. Baseline achievement for students in the Year 2 analyses on student achievement

Measure	TLTS group			Comparison group			Standardized difference (Hedges' *g*)	*p*-value of difference	Correlation between pre and post-test
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation			
Year 2 ELA sample									
Math z-scores	2,132	-0.640	0.957	1,953	-0.592	0.918	-0.051	0.205	0.707
ELA z-scores	2,150	-0.656	1.003	1,951	-0.579	1.009	-0.077	0.052	0.775
Year 2 math sample									
Math z-scores	2,119	-0.634	0.956	1,934	-0.587	0.919	-0.049	0.245	0.696
ELA z-scores	2,136	-0.652	1.004	1,932	-0.575	1.009	-0.077	0.052	0.598
Year 2 science sample									
Math z-scores	1,004	-0.671	0.953	887	-0.559	0.948	-0.118	0.028*	0.647

Measure	TLTS group			Comparison group			Standardized difference (Hedges' *g*)	*p*-value of difference	Correlation between pre and post-test
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation			
ELA z-scores	1,008	-0.723	0.962	885	-0.544	1.000	-0.182	0.001**	0.628
Year 2 social studies sample									
Math z-scores	507	-0.537	0.819	473	-0.538	0.924	0.002	0.984	0.676
ELA z-scores	508	-0.551	1.001	470	-0.643	1.052	0.089	0.209	0.664

Source: District administrative data.

Notes: Baseline scores included the statewide math and ELA assessments for Cohorts 1 and 2 and the district benchmark math and reading assessments for Cohort 3. Pretest scores were standardized using statewide norms for the state assessment and national norms for the benchmark assessments at the subject, year, and grade levels. Missing data were not imputed in this table; sample sizes differ by row due to variations in student attrition. Effect sizes were computed using means and standard deviations that weight each school equally, using the same weights used in the impact analysis. Both math and ELA scores are included as covariates in the study's impact model for all outcomes.

ELA = English Language Arts.

2. Sample of exploratory analyses of Year 1 impacts on student achievement

a. Attrition of schools in the analyses of Year 1 impacts

Similar to the confirmatory analyses of Year 2 outcomes, all schools eligible for the Year 1 analyses were included in the analyses as long as outcome data were available (Table B.2). All study schools contributed data to the ELA and math analyses for both Years 1 and 2, meaning there was no attrition at the school level. For the science and social studies analyses, which were only assessed in grade 5 and/or 8, the number of schools eligible for these analyses differs between Year 1 and Year 2. In Year 2, all six Cohort 1 schools were excluded from the Year 2 science and social studies assessments because the district benchmark assessments we used in place of the cancelled state assessment that year do not test these subjects. In Year 1, the 12 schools in Cohort 2 used 2020 district benchmark scores as outcomes and were therefore excluded from the science and social studies analyses. Moreover, because the state discontinued the grade 5 social studies exams after spring 2019, our Year 2 social studies analysis included only students in grade 8 and, therefore, only the six middle schools in the study (two in Cohort 2 and four in Cohort 3). The Year 1 social studies analysis included both grade 5 and grade 8 students; this analysis therefore included all six elementary schools in Cohort 1 as well as the four middle schools in Cohort 3.

Table B.2. School sample sizes and attrition rates for the Year 1 analyses on student achievement

Year 2 outcome	TLTS schools			Comparison schools			Overall school attrition rate (percent)	Differential school attrition (percentage points)
	Number of eligible schools at random assignment	Number of schools in the analytic sample	School attrition rate (percent)	Number of eligible schools at random assignment	Number of schools in the analytic sample	School attrition rate (percent)		
ELA	15	15	0.0	15	15	0.0	0%	0.0
Math	15	15	0.0	15	15	0.0	0%	0.0

Year 2 outcome	TLTS schools			Comparison schools			Overall school attrition rate (percent)	Differential school attrition (percentage points)
	Number of eligible schools at random assignment	Number of schools in the analytic sample	School attrition rate (percent)	Number of eligible schools at random assignment	Number of schools in the analytic sample	School attrition rate (percent)		
Science	9	9	0.0	9	9	0.0	0%	0.0
Social studies	5	5	0.0	5	5	0.0	0%	0.0

Source: District administrative data.

Notes: The number of eligible schools includes schools that served grades tested in the indicated subject during the outcome year. The state's social studies exam was discontinued in grade 5 after the 2018–2019 school year; therefore, starting with Cohort 2, only middle schools were eligible for the Year 1 social studies analysis. The state exam was cancelled in the 2019–2020 school year and therefore no data were available for science or social studies outcomes that year. This resulted in Cohort 2 schools (whose Year 1 outcomes were measured in 2019–2020) not being eligible for the Year 1 science or social studies analyses.

ELA = English language arts; TLTS = Turnaround Leadership Teams Strategy.

b. Attrition of students in the analyses of Year 1 impacts

Table B.3 shows the number of students eligible for each Year 1 analysis as well as the number of students included in each Year 1 analytic sample (those with test score outcome data). Under the WWC Standards (version 5.0) optimistic boundary, all of our Year 1 outcomes have high student attrition; they are therefore not eligible to meet the highest WWC rating of Meets Group Design Standards Without Reservation. The relatively high rates of attrition at the student level mainly stem from low test-taking rates among Cohort 3 students. As noted earlier, for Cohort 3 (the largest cohort), Year 1 outcomes were measured using the optional 2021 state assessment, which many students opted not to take.

Table B.3. Student sample sizes and attrition rates for the Year 1 analyses of student achievement

Year 2 outcome	TLTS schools			Comparison schools			Overall student attrition rate (percent)	Differential student attrition (percentage points)
	Number of eligible students at random assignment	Number of students in the analytic sample	Student attrition rate (percent)	Number of eligible students at random assignment	Number of students in the analytic sample	Student attrition rate (percent)		
ELA	4,672	2,544	45.5	3,495	4,262	2,621	38.5	7.0
Math	4,672	2,511	46.3	3,495	4,262	2,602	38.9	7.4
Science	1,093	432	60.5	1,579	968	421	56.5	4.0
Social studies	781	251	67.9	866	682	318	53.4	14.5

Source: District administrative data.

Notes: The number of students at random assignment includes the reference sample of students who were (a) enrolled in the study schools in the fall after schools were randomly assigned and (b) eligible to take the exam in Year 1 of the study based on their grade level. For example, science exams are only taken in grades 5 and 8; therefore, the reference sample includes students in grades 5 and 8 at the start of the study.

ELA = English Language Arts; TLTS = Turnaround Leadership Teams Strategy.

c. Baseline equivalence of student sample for the analyses of one-year impacts

Despite the high prevalence of missing Year 1 outcome scores, the students included in each analysis of Year 1 outcomes were similar between TLTS and comparison schools were similar in terms of their baseline achievement (Table B.4). Although some of the differences are statistically significant, all standardized differences fall within the WWC's adjustment range of 0.05 to 0.25 standard deviations. All Year 1 outcomes are therefore eligible to receive a WWC rating of Meets Group Design Standards With Reservations. Because we control for baseline ELA and math achievement in our regression analyses, and these baseline measures all have a correlation of 0.66 or higher with the outcomes, we establish baseline equivalence for each analytic sample.

Table B.4. Baseline achievement for students in the Year 1 analyses of student achievement

Measure	TLTS group			Comparison group			Standardized difference (Hedges' *g*)	*p*-value of difference	Correlation between pre and post-test
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation			
Year 1 ELA sample									
Math z-scores	2,380	-0.631	0.900	2,446	-0.587	0.877	-0.049	0.414	0.706
ELA z-scores	2,385	-0.652	0.938	2,449	-0.573	0.976	-0.083	0.141	0.792
Year 1 math sample									
Math z-scores	2,349	-0.628	0.902	2,425	-0.582	0.877	-0.052	0.378	0.761
ELA z-scores	2,354	-0.642	0.936	2,427	-0.573	0.978	-0.072	0.198	0.660
Year 1 science sample									
Math z-scores	402	-0.634	0.988	400	-0.595	0.830	-0.042	0.677	0.665
ELA z-scores	406	-0.603	1.031	402	-0.621	1.016	0.018	0.833	0.704
Year 1 social studies sample									
Math z-scores	234	-0.669	0.854	298	-0.792	0.768	0.152	0.378	0.668
ELA z-scores	234	-0.641	0.979	300	-0.841	0.964	0.206	0.075	0.746

Source: Administrative data.

Note: Baseline scores included the statewide math and ELA assessments for Cohorts 1 and 2 and the district benchmark math and reading assessments for Cohort 3. Pretest scores were standardized using statewide norms for the state assessment and national norms for the benchmark assessments at the subject, year, and grade levels. Missing data were not imputed; sample sizes differ by row due to variation in student attrition. Effect sizes were computed using means and standard deviations that weight each school equally. Both math and ELA scores are included as covariates in the study's impact model for all outcomes.

* Difference is statistically significant at the 5 percent level.

** Difference is statistically significant at the 1 percent level.

ELA = English Language Arts; TLTS = Turnaround Leadership Teams Strategy.

C. Exploratory analysis by cohort

To explore whether there were differences in impacts across cohorts, we repeated our analyses of Year 1 and Year 2 achievement outcomes separately by study cohort. Cohort-level impact estimates of the Year 1 analyses are presented in Table B.5 below (the results of the Year 2 analyses are presented in Table III.3 of the main text). We also tested whether the differences in impacts between cohorts were statistically significant, comparing the impact of Cohort 1 to Cohort 2, Cohort 1 to Cohort 3, and Cohort 2 to Cohort 3. Table B.6 reports the *p*-values of the differences in impacts across cohorts for the Year 1 and Year 2 analyses. As discussed in Chapter III, these exploratory results suggest that the TLTS program was least effective among Cohort 1 schools and most effective among Cohort 2 schools. However, because of differences in the implementation context across cohorts—including pandemic-related disruptions, fluctuations in test-taking rates, and changes in assessments used in the analyses—care should be taken when comparing results across cohorts. Detailed results for the other exploratory analyses we conducted appear in Chapter III.

Table B.5. Impacts of TLTS on student achievement after 1 year (by cohort)

Outcome	TLTS schools		Comparison schools		Impact estimate (SE)	*p*-value
	Number of students	Mean (SD)	Number of students	Mean (SD)		
Cohort 1 (state assessment)						
ELA z-score	358	-0.757 (1.015)	310	-0.756 (0.945)	-0.001 (0.022)	0.974
Math z-score	358	-0.768 (0.887)	310	-0.701 (0.807)	-0.067 (0.033)	0.098
Science z-score	177	-0.787 (0.795)	147	-0.711 (0.813)	-0.077 (0.043)	0.134
Social studies z-score	177	-0.954 (0.841)	147	-0.661 (0.873)	-0.293* (0.082)	0.016
Cohort 2 (district benchmark exam)						
ELA z-score	1,492	-0.592 (0.890)	1,541	-0.684 (1.004)	0.092* (0.040)	0.044
Math z-score	1,467	-0.593 (0.948)	1,519	-0.642 (0.999)	0.049 (0.031)	0.141
Cohort 3 (optional state assessment)						
ELA z-score	694	-0.595 (0.858)	770	-0.613 (0.932)	0.018 (0.035)	0.619
Math z-score	686	-0.811 (0.655)	773	-0.855 (0.703)	0.044 (0.039)	0.276
Science z-score	255	-0.628 (0.727)	274	-0.652 (0.860)	0.024 (0.069)	0.733
Social studies z-score	74	-0.916 (0.877)	171	-0.929 (0.812)	0.013 (0.014)	0.406

Source: District administrative data.

Note: Standard deviations of the unadjusted outcome means are reported in parentheses under each mean z-score. Standard errors are reported in parentheses under each impact estimate. The study includes students in grades 4 through 8 who attended study

schools in the fall after random assignment. Estimates of the impact of TLTS control for baseline covariates and include weights to account for unequal populations across schools. The comparison group mean is unadjusted, and the treatment group mean is the sum of the comparison group mean and the regression-adjusted difference between groups. Scale scores were converted to z-scores. A z-score of zero indicates a score equal to the mean for all students who took that test (statewide for the state assessment and nationally for the district benchmark assessments) in the same subject, grade, and year. A negative z-score indicates the score was below this mean.

* Difference is statistically significant at the 5 percent level.

** Difference is statistically significant at the 1 percent level.

ELA = English Language Arts; SD = standard deviation; SE = standard error; TLTS = Turnaround Leadership Teams Strategy.

Table B.6. Statistical significance of the differences in impacts of TLTS on student achievement across cohorts

Outcome	Cohort 1 impact estimate (SE)	Cohort 2 impact estimate (SE)	Cohort 3 impact estimate (SE)	*p*-value of difference in impacts between Cohorts 1 and 2	*p*-value of difference in impacts between Cohorts 1 and 3	*p*-value of difference in impacts between Cohorts 2 and 3
Year 1 outcomes						
ELA z-score	-0.001 (0.022)	0.092* (0.040)	0.018 (0.035)	0.044†	0.652	0.165
Math z-score	-0.067 (0.033)	0.049 (0.031)	0.044 (0.039)	0.010‡	0.029†	0.928
Science z-score	-0.077 (0.043)	--	0.024 (0.069)	--	0.214	--
Social studies z-score	-0.293* (0.082)	--	0.013 (0.014)	--	0.000‡	--
Year 2 outcomes						
ELA z-score	-0.061** (0.011)	0.104** (0.026)	0.033 (0.024)	0.000‡	0.000‡	0.045†
Math z-score	-0.028 (0.013)	0.072 (0.039)	0.107** (0.025)	0.016†	0.000‡	0.455
Science z-score		0.072 (0.044)	0.000 (0.035)	--	--	0.198
Social studies z-score		0.071 (0.091)	-0.128 (0.045)	--	--	0.051

Source: District administrative data.

Note: Standard errors are reported in parentheses under each impact estimate. The study includes students in grades 4 through 8 who attended study schools in the fall after random assignment. Estimates of the impact of TLTS pool all three school cohorts, control for baseline covariates, and include weights to account for unequal populations across schools.

* Difference is statistically significant at the 5 percent level.

** Difference is statistically significant at the 1 percent level.

† Difference in impacts between cohorts is statistically significant at the 5 percent level.

‡ Difference in impacts between cohorts is statistically significant at the 1 percent level.

ELA = English Language Arts; SE = standard error; TLTS = Turnaround Leadership Teams Strategy.

D. Sensitivity analyses

We conducted two analyses to test the sensitivity of our confirmatory Year 2 findings to the data and methods used. The first sensitivity analysis excluded data from the optional 2021 state assessment, and the second excluded students with missing baseline data.

1. Sensitivity analysis excluding the optional 2021 state assessments

The state assessment used in our study must be administered at the school and in person. However, as mentioned earlier, the state made the exam optional in spring 2021 due to differences in access to in-person learning across the state caused by the COVID-19 pandemic. As a result, even if students did return to in-person learning, they could opt out of taking the state exam in 2021. Only 22 to 31 percent of study students took state tests across subjects (math, ELA, science, and social studies) in 2021, compared to 92 percent in 2019, before the pandemic.

Therefore, test-taking patterns in spring 2021 could have introduced selection bias in our analytic samples by only including students who both resumed in-person learning in time to take the exams and opted to take the exams. The decision to return to in-person learning sooner or to take the state exam may have differed between students in TLTS and comparison schools; thus, differences in the outcomes of the two groups could be due to reasons other than the TLTS program.

We therefore replicated our Year 2 impacts while excluding 2021 outcome scores (which meant excluding Cohort 2 from this analysis). The findings, presented in Table B.7, are generally consistent with the main results (which are based on all cohorts and years) in terms of their sign and magnitude. For instance, the TLTS program's impact on math achievement after two years remained positive and statistically significant: 0.09 standard deviations (p-value < 0.001) for the full sample compared to 0.07 standard deviations (p-value = 0.009) for the sample excluding 2021 scores. The two estimates are not statistically different from each other.

Table B.7. Impacts of TLTS on student achievement after 2 years (excluding 2021 optional state assessments)

Outcome	TLTS schools		Comparison schools		Impact estimate (SE)	*p*-value
	Number of students	Mean (SD)	Number of students	Mean (SD)		
ELA z-score	1,822	-0.508 (0.916)	1,560	-0.527 (0.945)	0.019 (0.020)	0.363
Math z-score	1,815	-0.615 (0.860)	1,550	-0.681 (0.835)	0.065** (0.022)	0.009
Science z-score	834	-0.499 (0.801)	683	-0.499 (0.826)	0.000 (0.035)	0.991
Social studies z-score	509	-0.623 (0.680)	446	-0.495 (0.766)	-0.138 (0.045)	0.067

Source: District administrative data.

Note: Standard deviations of the unadjusted outcome means are reported in parentheses under each mean z-score. Standard errors are reported in parentheses under each impact estimate. The study includes students in grades 4 through 8 who attended study schools in the fall after random assignment. Estimates of the impact of TLTS pool all three school cohorts, control for baseline covariates, and include weights to account for unequal populations across schools. The comparison group mean is unadjusted,

and the treatment group mean is the sum of the comparison group mean and the regression-adjusted difference between groups. Scale scores were converted to z-scores. A z-score of zero indicates a score equal to the mean for all students who took that test (statewide for the state assessment and nationally for the district benchmark assessments) in the same subject, grade, and year. A negative z-score indicates the score was below this mean.

* Impact estimate is statistically significant at the 5 percent level.

** Impact estimate is statistically significant at the 1 percent level.

ELA = English Language Arts; SD = standard deviation; SE = standard error; TLTS = Turnaround Leadership Teams Strategy.

2. Sensitivity analysis excluding students with missing baseline covariates

To test whether our approach to imputing missing baseline data affected the study's confirmatory findings, we estimated impacts on Year 2 outcomes using the same model, but only among students with observed values for all baseline covariates. In other words, we did not impute any missing baseline data. The results of this sensitivity analysis based only on students with complete baseline data are consistent with our primary findings (Table B.8), suggesting the results are not sensitive to our imputation approach. Baseline achievement data among students in each of the complete case analytic samples was also similar to our primary findings (Table B.9). All standardized differences in baseline achievement remain smaller than the WWC's threshold of 0.25 standard deviations, ranging from 0.018 standard deviations in math achievement among students in the social studies analysis, to 0.182 standard deviations in ELA achievement among students in the science analysis.

Table B.8. Impacts of TLTS on student achievement after 2 years (complete case sample)

Outcome	TLTS schools		Comparison schools		Impact estimate (SE)	*p*-value
	Number of students	Mean (SD)	Number of students	Mean (SD)		
ELA z-score	2,125	-0.550 (0.911)	1,939	-0.594 (0.954)	0.045 (0.023)	0.059
Math z-score	2,112	-0.644 (0.815)	1,920	-0.733 (0.803)	0.088** (0.017)	0.000
Science z-score	998	-0.563 (0.785)	877	-0.597 (0.810)	0.035 (0.029)	0.241
Social studies z-score	504	-0.633 (0.696)	463	-0.508 (0.792)	-0.125* (0.042)	0.032

Source: District administrative data.

Note: Standard deviations of the unadjusted outcome means are reported in parentheses under each mean z-score. Standard errors are reported in parentheses under each impact estimate. The study includes students in grades 4 through 8 who attended study schools in the fall after random assignment. Estimates of the impact of TLTS pool all three school cohorts, control for baseline covariates, and include weights to account for unequal populations across schools. The comparison group mean is unadjusted, and the treatment group mean is the sum of the comparison group mean and the regression-adjusted difference between groups. Scale scores were converted to z-scores. A z-score of zero indicates a score equal to the mean for all students who took that test (statewide for the state assessment and nationally for the district benchmark assessments) in the same subject, grade, and year. A negative z-score indicates the score was below this mean. Students missing any baseline covariates were excluded from this analysis.

* Impact is statistically significant at the 5 percent level.

** Impact is statistically significant at the 1 percent level.

ELA = English Language Arts; SD = standard deviation; SE = standard error; TLTS = Turnaround Leadership Teams Strategy.

Table B.9. Baseline achievement for students in the Year 2 analyses on student achievement (complete case sample)

Measure	TLTS group			Comparison group			Standardized difference (Hedges' *g*)	*p*-value of difference	Correlation between pre and post-test
	Sample size	Mean	Standard deviation	Sample size	Mean	Standard deviation			
Year 2 ELA sample									
Math z-scores	2,125	-0.638	0.952	1,939	-0.594	0.919	-0.046	0.239	0.708
ELA z-scores	2,125	-0.656	1.004	1,939	-0.579	1.006	-0.077	0.056	0.774
Year 2 math sample									
Math z-scores	2,112	-0.631	0.951	1,920	-0.589	0.920	-0.045	0.283	0.697
ELA z-scores	2,112	-0.652	1.005	1,920	-0.575	1.006	-0.077	0.055	0.599
Year 2 science sample									
Math z-scores	998	-0.666	0.940	877	-0.562	0.948	-0.111	0.036*	0.647
ELA z-scores	998	-0.723	0.960	877	-0.545	1.002	-0.182	0.001**	0.626
Year 2 social studies sample									
Math z-scores	504	-0.538	0.819	463	-0.553	0.924	0.018	0.818	0.675
ELA z-scores	504	-0.548	1.003	463	-0.652	1.054	0.101	0.187	0.666

Source: District administrative data.

Notes: Baseline scores included the statewide math and ELA assessments for Cohorts 1 and 2 and the district benchmark math and reading assessments for Cohort 3. Pretest scores were standardized using statewide norms for the state assessment and national norms for the benchmark assessments at the subject, year, and grade levels. Missing data were not imputed in this table; sample includes students with no missing covariates and with observed outcomes. Effect sizes were computed using means and standard deviations that weight each school equally, using the same weights used in the impact analysis. Both math and ELA scores are included as covariates in the study's impact model for all outcomes.

ELA = English Language Arts.

Appendix C. Additional Technical Details for the Analyses of Teacher Perception and Practice Outcomes

This appendix presents additional information about the analytic approach used to estimate the impacts of the Turnaround Leadership Teams Strategy (TLTS) program on teacher perception and practice outcomes and the construction of those outcomes.

A. Analytic approach to estimate TLTS impacts on teacher outcomes

To estimate the impacts of the TLTS program on teacher perceptions and practices, we compared teachers who taught in a TLTS school at the time the survey was administered in the winter/spring of Year 2 to teachers who taught in comparison schools during the same period.[26] Specifically, we estimated the regression model in Equation (C1):

$$\text{(C1)}\quad y_{isc} = \alpha + \beta X_i + \theta T_s + s + c + p_s + e_{isc}$$

where y represents a survey outcome for teacher i at school s in cohort c; X is a vector of teacher covariates, including the teacher's race, ethnicity, gender, highest degree completed, teacher certification status, years of teaching (in total and at the study school), grades and subject taught, and the share of time the teacher spent teaching in virtual, in-person, or hybrid settings; T is an indicator of employment at a TLTS school at the time of the survey; s is a set of indicators for the study school where the teacher was employed; c is a set of indicators for the cohort in which the school participated; p is a set of indicators for each matched pair of schools to account for the nature of random assignment; e is a teacher-level error term; and $\alpha, \beta,$ and θ are parameters to be estimated. In this framework, the θ term represents the impact of teaching in a TLTS school. We estimated Equation (C1) with ordinary least squares using robust standard errors that accounted for the clustering of students at the school level.

As discussed in Chapter II, the results of this exploratory analysis should be interpreted with caution. Specifically, the analysis results may reflect differences between teachers that are not a direct result of the TLTS program due to (1) high levels of teacher attrition (which led us to survey teachers employed in study schools in Year 2 of program implementation rather than before the start of the program) and (2) low survey response rates, which differed substantially between teachers in TLTS and comparison schools. However, the study did not administer a baseline survey, so we are unable to establish the baseline equivalence of teachers' perceptions and beliefs in the analytic samples or assess whether the teachers who responded to the survey are representative of the teachers eligible to take the survey. For these reasons, the teacher perception and practice outcomes analyzed in this study are not eligible to meet WWC Group Design Standards With or Without Reservations.

Weights. Similar to the student model, the teacher model also weights the sample so that each school contributes equally to the estimates, and the relative weight of the treatment and control school within

[26] Due to pandemic disruptions, only teachers in Cohorts 2 and 3 were surveyed. Teachers were eligible for the survey if they were employed full time in a study school in the winter of the second implementation year and were responsible for direct instruction of a complete classroom; they did not have to be employed at a study school at the time of random assignment. See Chapter II for more information on the survey administration, teacher eligibility, and response rates.

each matched pair is the same. Specifically, we computed analytic weights for the teacher analyses using the following model:

$$(C2)\ w_{tsc} = \left(\frac{1}{T_{sc}}\right) x \left(\frac{1}{S}\sum_{s=1}^{S} T_{sc}\right)$$

where w represents the weight for teacher t at school s in cohort c; T is the total number of teachers in school s; S is the total number of schools; and $\sum_{s=1}^{S} T_s$ is the total number of students across all schools.

Missing data. In this analysis, all missing covariates were imputed using a dummy imputation approach by setting all missing covariates to zero and including imputation indicators in the regression model. The study did not impute any outcome data.

B. Construction of teacher perception and practice outcomes

Table C.1 presents the survey items related to each outcome along with (1) the correlation between each item and the overall score for the outcome (test correlation), (2) the correlation of all other items when excluding each individual survey item (rest correlation), and (3) the internal consistency of scores on each of the eight outcomes when excluding each individual survey item within an outcome (item leave-out alpha). The correlation coefficients, also known as validity coefficients, range from –1 to +1, where coefficients close to 1 in absolute value indicate high predictive validity of the items, and coefficients close to 0 indicate little or no association between the items. Leave-out alpha values closer to 1 indicate that responses on items within that outcome are more similar when excluding the item than if that item were included, suggesting that the item may be dissimilar from the others in the construct. Overall, the statistics in Table C.1 suggest that each outcome area had high levels of internal consistency and was constructed using related items that measured the outcome of interest.[27]

[27] A similar approach has been used to assess the validity of teacher practice and perception outcomes using survey data in other studies. For an example, see Dolfin et al. 2019.

Table C.1. Item-level reliability of the teacher survey outcomes

Survey items	Test correlation	Rest correlation	Item leave-out alpha (α)
Collective leadership			
Teacher self-awareness and development: 3 items, Cronbach's alpha (α) = 0.857			
Thinking about this school year, to what extent do you agree with the following statements about your school?			
• School leadership leads teachers in meaningful discussions about difficult topics (e.g., racism, sexism, or homophobia).	0.866	0.686	0.841
• My school provided me with sufficient training on working with diverse student populations (e.g., income level or race).	0.899	0.773	0.760
• My school provided me with sufficient training on working with homeless families.	0.882	0.731	0.798
Building strong teams: 2 items, Cronbach's alpha (α) = 0.678[a]			
Thinking about this school year, to what extent do you agree with the following statements about the school leadership team (SLT)?			
• Seeks upward feedback from the teachers.	NA	NA	NA
Thinking about this school year, to what extent do you agree with the following statements about your school?			
• My school encourages the sharing of diverse perspectives.	NA	NA	NA
Vision and culture			
Shared vision for student success: 3 items, Cronbach's alpha (α) = 0.787			
Thinking about this school year, to what extent do you agree with the following statements about the SLT?			
• Communicates a clear vision for the school.	0.814	0.590	0.747
Thinking about this school year, to what extent do you agree with the following statements about your school's work in building relationships with parents/guardians?			
• Teachers and administrators in your school work closely with parents/guardians to make sure students succeed.	0.857	0.676	0.656
Thinking about this school year, to what extent does your school do the following?			
• Incorporates feedback from families into school planning and decision making.	0.850	0.616	0.730

Survey items	Test correlation	Rest correlation	Item leave-out alpha (α)
Culture and goals: 18 items, Cronbach's alpha (α) = 0.862			
Thinking about this school year, on average, how often have you used student data (of any type) for each of the following purposes?			
• Monitor student progress toward priority learning goals.	0.344	0.260	0.863
• Plan whole-class instruction.	0.444	0.368	0.859
• Plan small-group instruction and/or change students' small-group assignments.	0.491	0.419	0.857
• Plan individualized instruction for students.	0.568	0.494	0.854
• Identify students in need of pull-out services or other intensive interventions.	0.546	0.442	0.857
• Adjust the amount of time you spend on specific topics (e.g., reteach concepts that students find challenging).	0.382	0.304	0.861
• Identify student needs relative to the standards (e.g., understanding what students should learn).	0.513	0.453	0.857
Thinking about this school year, on average, how often have you discussed or shared student data with the following people?			
• Teachers	0.391	0.308	0.862
• Administrators	0.529	0.449	0.856
Thinking about this school year, to what extent do you agree with the following statements about your school?			
• School leaders have the necessary resources to ensure that teachers are providing academically rigorous instruction.	0.648	0.560	0.851
• School leaders have the necessary knowledge to ensure that teachers are providing academically rigorous instruction.	0.625	0.543	0.851
Thinking about this school year, to what extent do you agree with these statements about teachers in your school?			
• Teachers set high standards for teaching.	0.581	0.516	0.853
• Teachers set high standards for students' learning.	0.538	0.473	0.855
• Teachers work hard to make sure all students are learning.	0.550	0.491	0.855
Thinking about this school year, to what extent do you agree with the following statements about the SLT?			
• Sets clear expectations for meeting instructional goals.	0.598	0.512	0.853
• Knows what's going on in your classroom.	0.697	0.615	0.848
• Follows through on goals and priorities.	0.750	0.683	0.843
• Celebrates the school's successes.	0.610	0.536	0.852

Survey items	Test correlation	Rest correlation	Item leave-out alpha (α)
Authentic community engagement			
Empowering others: 19 items, Cronbach's alpha (α) = 0.886			
Thinking about this school year, on average, how often have you discussed or shared student data with the following people?			
• Parents/guardians	0.441	0.374	0.885
• Students	0.456	0.363	0.885
Thinking about this school year, about how often have you engaged in the following practices with your students?			
• Asked students to develop individual academic goals.	0.533	0.466	0.882
• Asked students to assess their progress towards achieving their goals.	0.516	0.447	0.883
• Asked students about their personal and career goals beyond the classroom.	0.535	0.471	0.882
• Asked students to provide input on how engaging they found classroom activities and topics.	0.542	0.471	0.882
• Asked students to connect academic standards to real-world issues and/or contexts.	0.439	0.363	0.885
• Used student post-secondary goals to inform your lesson planning and classroom instruction.	0.529	0.446	0.883
Thinking about this school year, to what extent do you agree with the following statements about your school?			
• My school is a safe, caring, welcoming, and respectful place that is fair to all students.	0.516	0.444	0.883
• The school community puts sufficient effort into addressing issues of difference, fairness, and inclusion.	0.655	0.595	0.878
• Students experience school as fair to all students, regardless of their social identity (e.g., race and gender).	0.468	0.399	0.884
• My school tries hard to make people from different backgrounds feel included.	0.627	0.572	0.879
• My school conducts community outreach with diverse communities (e.g., African American, Hispanic, or Asian communities).	0.618	0.541	0.880
Thinking about this school year, to what extent do you agree with the following statements about your school's work in building relationships with parents/guardians?			
• My school keeps the line of communication open for parents/guardians.	0.509	0.455	0.883
Thinking about this school year, to what extent does your school do the following?			
• Develops formal networks to link all families with each other (for example, sharing parent/guardian directories, providing a website for parents/guardians to connect with one another, etc.).	0.663	0.589	0.878
• Builds trust with families/cultivates strong relationships.	0.728	0.682	0.875
• Helps families understand how their feedback was used.	0.767	0.720	0.873
• Encourages more involved parents/guardians to reach out to less involved parents/guardians.	0.660	0.589	0.878
• Engages members of the school community (students, teachers, and parents/guardians) in creating a positive learning environment.	0.723	0.675	0.875

Survey items	Test correlation	Rest correlation	Item leave-out alpha (α)
Student learning: 6 items, Cronbach's alpha (α) = 0.699			
Thinking about this school year, on average, how much total time in a week did you spend…			
• Talking or emailing with parents/guardians about their child's performance in your class (e.g., calling home or emailing about missed assignments)?	0.549	0.363	0.679
Thinking about this school year, on average, how often have you participated in the following parent/guardian engagement activities?			
• Solicited information from parents/guardians about their child.	0.651	0.469	0.648
• Communicated to parents/guardians about the work students are doing in their class.	0.721	0.595	0.621
• Held student-led parent-teacher conferences.	0.572	0.316	0.702
Thinking about this school year, please indicate the extent to which…			
• Your teaching has been influenced by what parents/guardians share with you about their child's learning style.	0.657	0.442	0.656
• You collaborated with parents/guardians to set goals for their child's learning and development.	0.686	0.465	0.649
Academics			
Instructional practices: 3 items, Cronbach's alpha (α) = 0.786			
Thinking about this school year, to what extent do you agree with the following statements about your school?			
• My school creates high academic expectations for all students.	0.832	0.565	0.799
• Students are given instruction that asks them to think and engage deeply with challenging material.	0.873	0.720	0.618
• Students are given grade-appropriate classroom assignments.	0.820	0.618	0.721
Drive teacher improvement: 9 items, Cronbach's alpha (α) = 0.847			
Thinking about this school year, including the preceding summer, have you participated in any…			
• Professional development activities?	0.306	NA[b]	0.865
Thinking about this school year, including the preceding summer, on average, how often have you participated in the following professional development activities?			
• Received individualized (one-on-one) coaching	0.734	0.617	0.826
• Received any support in a small group or team setting with other teachers (e.g., professional learning community)	0.707	0.601	0.826
• Received school level professional development.	0.664	0.574	0.831
• Received whole district level professional development.	0.539	0.439	0.843

Survey items	Test correlation	Rest correlation	Item leave-out alpha (α)
Thinking about this school year, including the preceding summer, about how often did a principal, assistant principal, instructional coach, or grade-level chair provide the following coaching or individualized support for your teaching?			
• Modeled instruction for you.	0.806	0.723	0.812
• Facilitated a role-playing exercise.	0.787	0.702	0.815
• Helped with lesson planning.	0.776	0.663	0.821
• Helped with reviewing data.	0.701	0.593	0.828

Source: Study-developed teacher survey.

Note: As with the construction of the outcomes, missing values were handled by pairwise deletion; that is, if a teacher was missing a value on one of the items used for an outcome, this teacher was excluded from the analytic sample for that particular correlation.

[a] Item-level correlations and leave-out alphas cannot be computed for the Building Strong Teams outcomes because only two items are included in this construct.

[b] The rest correlation cannot be computed for binary variables because the left-out item only takes on the values of 0 or 1.

Mathematica Inc.

Our employee-owners work nationwide and around the world.

Find us at **mathematica.org** and **edi-global.com**.