Scaling Up Teacher Induction: Implementation and Impact on Teachers and Students

Evaluation of the New Teacher Center's i3 Scale-up Grant, Revised Final Report

January 2025

Rebecca A. Schmidt
Aliya R. Pilchen
Katrina Laguarda
Haiwen Wang
Deepa Patel

SRI Education™
A DIVISION OF SRI INTERNATIONAL

January 2025

Funding for this study came from the U.S. Department of Education under an Investing in Innovation (i3) scale-up grant to the New Teacher Center (#U411 A 150005-15A). The findings and conclusions in this report do not necessarily represent the official position or policies of the New Teacher Center or the U.S. Department of Education.

Suggested citation:

Schmidt, R. A., Pilchen, A. R., Laguarda, K., Wang, H., & Patel, D. (2025). *Scaling up teacher induction: Implementation and impact on teachers and students.* (Evaluation of the New Teacher Center's i3 Scale-up Grant, Revised Final Report). SRI International.

Abstract

Beginning teachers enter a profession that places particularly challenging demands on novice practitioners. The New Teacher Center's (NTC) induction model provides intensive, instructionally focused coaching to teachers during their first two years in the classroom, in-depth training for induction mentors, a suite of tools to guide coaching cycles, and capacity-building for district leaders to sustain induction mentoring programs after NTC's direct involvement in the district ends.

With funding from a U.S. Department of Education Investing in Innovation (i3) scale-up grant, NTC tested strategies for scaling its validated induction model to 301 schools in five school districts serving high proportions of students of color and students from low-income households. NTC adapted its model to support district adoption at scale, including an option for deploying part-time, school-based mentors, reduced requirements for mentor training, and online training and video-sharing tools.

SRI's evaluation of the implementation and impact of NTC's i3 scale-up grant employed a cluster-randomized controlled trial design with schools as the unit of randomization. All first-year teachers in randomized schools were included in the study. Treatment teachers received induction supports from NTC-trained mentors, while control teachers received the supports provided by their districts under business-as-usual conditions.

The evaluation examined fidelity of implementation to the model as designed, the contrast between the induction supports in the treatment and control conditions, and impacts on three key outcomes: 1) teachers' classroom practice as measured by the Danielson Framework for Teaching, 2) student achievement on state standardized assessments in mathematics and English Language Arts (ELA) in grades 4 through 8, and 3) teacher retention within district. [1]

NTC's induction model was not implemented with adequate fidelity in any of the five sites according to thresholds set by NTC, and the mentoring received by NTC treatment teachers was not substantially different in key respects from the mentoring received by control teachers. There were no statistically significant impacts of the model as implemented on overall teacher practice, student achievement, or teacher retention.

Exploratory findings suggest conditions under which NTC might see a greater impact. There was a positive correlation between students' mathematics achievement and mentoring that met NTC's fidelity thresholds for frequency and duration, as well as between mathematics achievement and mentoring that met NTC's expectations for instructional focus. NTC induction supports also had a positive impact on student ELA achievement in schools with higher proportions of historically underserved students.

These findings indicate the importance of ensuring high-quality implementation of a program. Under the i3 scale-up grant, NTC attempted to adapt its model for scaling, but the partner districts failed to fully implement key components and mediators as intended. There is evidence that the model has promise when fully implemented, particularly in schools with higher proportions of historically underserved students, but without further research this evidence is simply suggestive.

[1] A previous version of this report was published in December 2020. This report provides updated information for review under WWC version 5.0 standards. See "2025 Report Updates" for a list of the changes made.

Acknowledgments

The findings reported here culminated from the efforts of a large team over 5 years. The authors thank Viki Young for her early leadership of the study, and Rebecca Goetz, Mark Sieber, Hannah Cheever, Hannah Kistler, Jenna Rush, Allison Milby, and Veronica Madrigal for their valuable contributions to data collection and analysis. The classroom observation team also included Naa Ammah-Tagoe, Carmen Araoz, Jennifer Bland, Lauren Cassidy, Andrew Ezekoye, Annie Fikes, Bowyee Gong, Irmary Garcia-Maisonet, Sarah Gerard, Erin Harless, Rex Long, Nyema Mitchell, Jaunelle Pratt-Williams, Joseph Shields, Naomi Tyler, and Kaily Yee. Hannah Kelly, Brenda Waller, and Katie Roberts produced the final report. The authors thank Ali Picucci and the program staff at the New Teacher Center for their strong engagement with and feedback on the evaluation. Not least, the authors are grateful to the program leads, lead coaches, induction mentors, principals, and teachers in the five partner sites for their participation in the evaluation.

2025 Report Updates

This revised final report on SRI's evaluation of the New Teacher Center's i3 scale-up grant replaces an earlier version originally published in December 2020. In order to facilitate review by the What Works Clearinghouse (WWC) under version 5.0 standards, the report's authors have provided additional information about the analytic samples included in the impact study and have organized key sections of the report to ensure that all of the data required for WWC review is easily accessible to reviewers. None of the report's main findings or conclusions have changed since the version first released in 2020.

The changes in this 2025 update include:

1) Adding additional subheadings and reordering the sections in Chapter 3.

2) Clarifying the eligible sample for the student achievement analysis in Chapter 3.

3) Clarifying the role of teacher and student joiners in each analysis in Chapter 3.

4) Separating cluster-level attrition from representativeness, clarifying the denominator used for attrition calculations, and providing additional information on representativeness in Chapter 3 and Appendix G.

5) Adding additional subheadings and more explicit information on outcome measures, assignment to intervention and comparison conditions, compositional change, representativeness, and baseline equivalence necessary for WWC review to Appendix G.

Contents

List of Exhibits

Executive Summary

The first two years in the classroom are a critical period in developing beginning teachers' foundational skills. During this time, teachers experiment—and often struggle alone—with developing strong instructional practice to support students' diverse academic and social needs and begin to set the foundation for habits and dispositions that persist throughout their careers (Snyder & Bristol, 2015). If teachers lack adequate support during these critical early years in the classroom, there can be considerable consequences for teacher retention, teaching quality, and student outcomes. Thus, interventions designed to retain beginning teachers in the classroom and enable them to attain proficiency as instructors as quickly as possible are necessary to ensuring equitable access to high-quality instruction for all students.

The New Teacher Center Induction Model

The New Teacher Center (NTC) aims to ensure that beginning teachers have the support they need through intensive and instructionally focused mentoring. NTC's induction model includes training and in-field coaching for induction mentors, a suite of tools to guide instructionally focused coaching cycles with beginning teachers, and activities to build the capacity of district leaders and coaches to sustain induction mentoring programs after NTC's direct involvement in the district ends. For more than 20 years, NTC has developed and refined this comprehensive mentor-based induction model in their work with dozens of districts across the country.

Study Design

In 2016, NTC received an Investing in Innovation (i3) scale-up grant to test strategies for scaling its validated induction model in a broad array of district contexts. The model as designed for the scale-up study includes NTC's three key program components: 1) providing NTC supports, including tools, curricula, and funded staff positions to support induction mentors; 2) assigning mentors with sanctioned time for mentoring and limited caseloads of beginning teachers; and 3) providing mentor training and development activities. It also includes one key mediator: providing high-quality mentoring that meets expectations for duration, intensity, and instructional focus.[2]

Building on lessons learned from prior research on its model, NTC adapted its approach to make it more flexible and less expensive, thus making it more feasible for districts to adopt. These adaptations, or "scaling strategies," included changes to the selection and deployment of mentors. NTC collaborated with three of the five scale-up sites to recruit and train classroom teachers and school-based instructional coaches to serve as mentors part-time with a much smaller caseload of teachers, rather than requiring that districts hire dedicated full-time induction coaches. NTC also reduced the amount of time mentors were required to spend in training (taking mentors away from schools less often), and offered online training, video-sharing, and video-conferencing options for in-field mentor support and

[2] In keeping with guidance from the i3 grant program, this report uses the term "mediator" to refer to the behaviors, processes, and skills targeted by the intervention's key components. Mediators are expected to lead, in turn, to teacher and student outcomes (Boulay et al., 2018). These aspects of the program were not implemented by NTC directly, but rather by mentors trained by NTC, working with assigned beginning teachers. These mediators were not tested in a statistical mediation model.

meetings with beginning teachers. NTC expected that these tools would reduce the time that full-time coaches would spend traveling between schools and offer more flexibility in scheduling classroom observations and debriefing sessions.

For this study, NTC partnered with five school systems that serve large proportions of students from low-income households (receiving free and reduced-price meals) and students of color: Chesterton School District, Taftville Unified School District, Hayes Park Public Schools, Fillmore County Public Schools, and Garfield City Department of Education.[3] SRI conducted the evaluation of NTC's scaled version of its induction model under the i3 scale-up grant. This technical report presents the results of this evaluation.[4]

Research Questions

The study included two implementation research questions, three confirmatory impact research questions, and three exploratory research questions.

Implementation research questions

(1) What is the level of implementation fidelity to the NTC model in the five study sites?

(2) To what extent is there a contrast between the treatment and the control condition in the level and type of mentoring teachers received?

Confirmatory impact research questions

(3) What is the impact of two years of NTC induction supports on teaching practices overall and in the domains of classroom environment and instruction, compared with the control condition?

(4) What is the impact of NTC induction supports on student achievement in English language arts (ELA) and math in grades 4 through 8, after one year of exposure to beginning teachers in their second year of teaching, compared with the control condition?

(5) What is the impact of two years of NTC induction supports on teacher retention into a third year in the same district, compared with the control condition?

Exploratory research questions

(6) Is effective instruction related to student achievement? If NTC induction supports have impacts on teacher practice and student achievement (as examined by research questions 3 and 4), does effective instruction mediate the NTC effect on student outcomes?

(7) Are higher levels of mentoring (e.g., in frequency and instructional focus) related to higher student achievement?

(8) Is the effect of NTC induction supports on student achievement moderated by any school- or student-level variables?

[3] The district names used throughout the report are pseudonyms.

[4] A previous version of this report was published in December 2020, this report provides updated information for review under WWC version 5.0 standards. See "2025 Report Updates" for a list of the changes made.

Methods and Data

To answer the study research questions, SRI used a cluster-randomized controlled trial (RCT) design, with clusters (schools) randomized within blocks. All eligible schools had a 50 percent chance of receiving treatment, and all beginning teachers in the randomized schools were included in the study. NTC provided induction supports to beginning teachers in the treatment schools, while teachers in the control schools received business-as-usual induction supports.

The implementation study relied on data from NTC's online mentoring platform (Learning Zone) and from teacher surveys administered each spring. SRI worked with NTC to create quantitative indicators and thresholds for implementation fidelity for each of the three key components and the one key mediator of its model at the site level and for the study overall. This analysis used descriptive statistics and t-tests to examine the level of fidelity to the NTC induction model on each of these indicators, as well as the difference between the treatment and control (business-as-usual) conditions.

The confirmatory and exploratory impact studies relied on extant data provided by the sites (student demographics and achievement, teacher demographics and employment), publicly available extant data (baseline school characteristics), and classroom observation data gathered by trained SRI observers. SRI used an intent-to-treat approach to measure the impact of NTC induction supports on teacher practice, student achievement, and teacher retention. Thus, data for teachers were analyzed in their original randomized condition (treatment or control) even if they changed their school assignments. Both overall and differential attrition rates were within acceptable bounds set by What Works Clearinghouse (WWC, 2022) for all three confirmatory impact analyses.

Teacher practice

Trained SRI observers rated study teachers on eight components of the Danielson Framework for Teaching (FFT) at baseline (fall of their first year of teaching) and at follow-up (spring of their second year of teaching). Teachers were eligible to be observed if they taught core subjects (mathematics, English language arts, science, or social studies), or if they taught in a multi-subject elementary classroom. Observers were blind to teachers' treatment status. Impact models accounted for teachers' baseline observation scores, teacher and school demographics, and the blocking variables used in random assignment. The analysis used multilevel models to account for the nesting of teachers within schools.

Student achievement

SRI examined the one-year impact of NTC induction supports on students' scores on their state's standardized tests in mathematics and English language arts (ELA) after their teachers had received two years of induction support. Because the study sites were spread across multiple states, SRI standardized the students' scores to their district mean and standard deviation (Hayes Park, Fillmore County, Chesterton, and Taftville) or to the sample mean and standard deviation, where district statistics were not available (Garfield City). Students in grades 4 through 8 were included in the analysis. Impact models accounted for students' prior-year test scores, student, teacher and school demographics, and the blocking variables used in random assignment. The analysis used multilevel models to account for the nesting of students within teachers and teachers within schools.

Teacher retention

Using district employment records, SRI examined the impact of NTC induction supports on the retention of teachers in instructional positions into their third year of teaching. Impact models accounted for teacher and school demographics, baseline teacher retention rates at the school level, baseline student attendance rates at the teacher level, and the blocking variables used in random assignment. The analysis used multilevel models to account for the nesting of teachers within schools.

Exploratory research questions

To address each of the exploratory research questions, SRI fit the same models as were used in the main student achievement analysis. The treatment indicator in the models was replaced with measures of classroom practice to answer research question 6 and with levels of mentoring to answer research question 7. Interactions between student- and school-level variables and the treatment indicator were added to address research question 8.

Results

Results from SRI's implementation and impact studies cover three years of implementation (2016–17 through 2018–19) and combine data across cohorts of schools, mentors, and beginning teachers.

Fidelity to the NTC induction model was low.

NTC's induction model includes three key components (NTC tools and resources, mentor assignment, mentor development and accountability) and one key mediator (provision of high-quality mentoring). None of these key components or the mediator was implemented with fidelity at the study level in any of the three years of the intervention. Although by Year 3, at least one site met the site-level threshold for fidelity of implementation for each component, the study-level threshold was not met in any of the components in any of the three years of the study. Overall, the implementation fidelity results suggest that NTC's intervention fell short of NTC's intended design during the scale-up grant period.

The contrast between the treatment and business-as-usual conditions was not as strong as expected.

Based on their responses to surveys administered each spring during the study, teachers in treatment schools were more likely to be formally assigned a mentor than teachers in control schools, and teachers with NTC mentors were more likely to report receiving instructionally focused supports than teachers in control schools. However, there was no contrast in the frequency or duration of meetings between mentors and teachers in treatment and control schools. Less than half of treatment and control teachers reported meeting with their mentor weekly or more often for one hour or more, which was the level of contact consistent with NTC's expectations for the frequency and duration of mentoring.

This study detected no impact of NTC induction supports on an overall measure of classroom practice, on student achievement in math or ELA, or on teacher retention. However, NTC did have a positive impact on teachers' classroom practice in Communicating with Students.

There was no statistically significant difference between treatment and control teachers' overall classroom practice or their practice in the domains of Classroom Environment or Instruction on the FFT. Similarly, although the coefficients on the impact of NTC induction supports on math and ELA achievement were both positive, the variation around these estimates was large, so the coefficients were not statistically distinguishable from zero. Likewise, after accounting for differences between teachers and schools, the difference between treatment and control teachers in retention rates was not statistically significant.

However, the confirmatory impact study did find a positive impact on one component of classroom practice: NTC had a statistically significant and positive impact on ratings of teachers' practice in Communicating with Students (0.10 points on the FFT scale of 1–4). The effect size of the impact is one quarter of a standard deviation, which is a moderate impact in education (Lipsey, et al., 2012).

FFT measures of classroom practice were strongly related to both math and ELA achievement.

The strong relationship between FFT and student achievement establishes the predictive validity of the FFT. It indicates that the lack of an impact on student achievement may be because any impact on teacher practice (detectable only in the Communicating with Students component) may have been too small to translate into detectable changes in student achievement.

Students' mathematics achievement was associated with the type of mentoring their teachers received.

The SRI evaluation team hypothesized that the lack of a relationship between NTC induction supports and student achievement may have been due, at least in part, to the low level of implementation fidelity and minimal contrast between the treatment and control conditions.

When looking across the treatment and control group, broad measures of the frequency (weekly or more) and duration (meetings of an hour or more) of mentoring were not significantly associated with student achievement. However, when looking only within the treatment group, math teachers who received the full frequency and duration of mentoring from an NTC-trained mentor saw higher math achievement among their students than treatment teachers who received less NTC mentoring. Additionally, when looking both within the treatment group and across treatment and control groups, there was a statistically significant relationship between instructionally focused mentoring activities and student mathematics achievement.

This supports NTC's hypothesis that the simple presence of an induction mentor is not sufficient to affect student achievement. Rather, mentors need training on working with teachers to improve their instructional practice, they need to spend an hour or more weekly with beginning teachers, and their mentoring must include a regular focus on instructional practice.

NTC induction had a statistically significant impact on student ELA achievement in schools with high proportions of historically underserved students.

Finally, the SRI evaluation team hypothesized that the impact of NTC induction supports may be stronger in some types of schools or with some groups of students. If NTC's impact is stronger in some schools than in others, the overall impact estimates above may mask these differences.

Although there was no statistically significant moderation effect of any student-level variables on the relationship between NTC induction supports and student achievement, there was significant moderation of school-level variables on ELA achievement. In particular, in schools with an above-average concentration of students receiving free or reduced-price meals (higher poverty schools), there was a positive impact of NTC induction supports on student achievement, with a small to moderate effect size of 0.12 standard deviation. NTC induction did not have the same impact in the schools with below-average concentrations of students receiving free or reduced-price meals (lower poverty schools), and the difference in impacts between the two types of schools was statistically significant.

Likewise, in schools with an above-average concentration of English learner students there was a positive impact of NTC induction supports on ELA achievement, with a small to moderate effect size of 0.14 standard deviation. NTC induction did not have the same impact in schools with below-average concentrations of English learner students, and the difference in impacts between the two types of schools was statistically significant.

Conclusion and Implications

NTC induction as designed for the i3 scale-up grant was intended to provide intensive mentoring support to beginning teachers during their crucial first two years of teaching, while also adapting the validated model to make it more flexible and less expensive, and thus more feasible for districts to adopt. Despite the attempt to make the model more feasible to adopt, this study found that the level of fidelity to the model was inadequate across all components and years, and the contrast between the treatment and control conditions was less than expected. Additionally, there was no impact of the induction supports as implemented on overall teacher practice, student achievement, or teacher retention.

Nonetheless, there was a correlational relationship between higher levels of mentoring with an instructional focus and student achievement in mathematics. Additionally, NTC induction supports had an impact on student achievement in ELA in higher poverty schools and schools with more English learner students. These exploratory findings are suggestive of a relationship between NTC's model of mentoring and student achievement that was not detected in the overall study. They suggest that NTC may have an impact when implemented with fidelity and when implemented in schools with a higher proportion of historically underserved students.

Thus, this study shows the critical importance of ensuring high-quality implementation of a program and the consequences of failure to do so. There is some evidence that the model has promise when fully implemented, particularly in schools with higher proportions of historically underserved students, but without further research, this evidence is simply suggestive.

Chapter 1. Introduction

The first two years in the classroom are a critical period in beginning teachers' development. Although they have the same responsibilities as veteran teachers, early career teachers are still developing foundational skills such as classroom management and instructional pedagogy. These skills will enable them to create equitable and productive classroom environments and provide all students with access to content that meaningfully engages them in intellectual work. In the first years of their careers, teachers experiment—and often struggle alone—with managing student behavior, mastering the curriculum, engaging students in learning, and attending to students' diverse academic and social needs. During these formative years, beginning teachers can also set the foundation for habits and dispositions that persist throughout their careers (Snyder & Bristol, 2015).

A lack of adequate support for teachers during these critical early years in the classroom can have considerable consequences for both teachers and students. Retention rates are lower for early career teachers than at any other stage (Ingersoll, 2001; Smith & Ingersoll, 2004), and this turnover comes at a high cost for students, especially those who are historically underserved (e.g., non-White students, English learners, and students who experience poverty) (Ronfeldt, Loeb, & Wyckoff, 2013). Complex, adaptive instructional skills that enable teachers to accelerate student learning are developed in the classroom, and beginning teachers make rapid gains in skills in their first years in the classroom that are reflected in stronger student outcomes (Kini & Podolsky, 2016; Kagan, 1992). As the U.S. teacher workforce has grown younger and less experienced (Kini & Podolsky, 2016), students in schools with high rates of teacher turnover spend more years in classrooms with less experienced, less proficient teachers. This is especially true of low-income students and students of color (Goldhaber et al., 2018; Scafidi et al., 2007). Thus, interventions designed to retain beginning teachers in the classroom and enable them to attain proficiency as instructors as quickly as possible are critical to ensuring equitable access to high-quality instruction for all students.

The New Teacher Center (NTC) aims to improve educational equity for historically underserved students by accelerating the effectiveness of beginning teachers during their first two years in the classroom. NTC provides induction mentors with intensive training, in-field coaching, and a suite of tools to guide instructionally focused coaching cycles with beginning teachers. NTC also works to build the capacity of district leaders and coaches to sustain induction mentoring programs after NTC's direct involvement in the district ends. For more than 20 years, NTC has developed and refined this comprehensive mentor-based induction model in their work with dozens of districts across the country.

In 2013, NTC received an Investing in Innovation (i3) validation grant to test the efficacy of its induction model in three diverse contexts (a large urban district, a large countywide system, and a consortium of rural districts). SRI served as the external evaluator. Through the grant, NTC funded, selected, and trained full-time mentors to serve all beginning teachers hired into study schools. NTC induction mentors provided regular, intensive, instructionally focused mentoring to beginning teachers in these schools for two full years. NTC also documented the key components and mediators of its intervention and developed expectations for fidelity of implementation that district staff used to communicate internally about expectations for mentoring and that SRI researchers used to measure program fidelity. SRI's evaluation of NTC's i3 validation grant found that all of the key components and mediators of the

program were implemented with high fidelity in all three sites,[5] and that NTC mentoring represented a significant departure from business-as-usual induction practices. In two sites, SRI measured the impact of induction support on teacher practice, teacher retention, and student achievement via a randomized control trial (RCT). In those sites, NTC mentoring had a positive, statistically significant impact on student achievement in both English language arts (ELA) and mathematics for students in grades 4 through 8. These effects[6] were equivalent to 2 to 4.5 months of additional learning, as measured by state assessments. There were no measurable impacts of the program on either teacher practice or teacher retention.[7] (Young et al., 2017).

In 2016, NTC received an i3 scale-up grant to test strategies for scaling its validated induction model to more schools in a larger array of district contexts. Building on lessons learned from the i3 validation study, NTC adapted its mentoring model to make it more flexible and less expensive and thus more feasible for districts to adopt. Despite these changes, the scale-up grant retained a focus on all of NTC's key program components and a clear set of expectations for the duration, intensity, and instructional focus of mentoring.

SRI conducted the evaluation of NTC's i3 scale-up grant. The evaluation featured a rigorous mixed-methods design to measure implementation fidelity and impacts on teacher and student outcomes in the five participating sites. The evaluation team used multiple measures to capture implementation fidelity and provided timely feedback to NTC and to the sites to inform program management and support. The evaluation team also explored the contribution of implementation fidelity to patterns in student outcomes.

This technical report begins with a discussion of the NTC induction mentoring logic model in which key program components and mediators, scaling strategies, and intended outcomes are specified. Next, the report describes the research design, including recruitment and randomization; site, teacher, and student samples; and data collection and analysis methods. Finally, this report presents the findings on program implementation, teacher practice outcomes, student learning outcomes, and teacher retention. Appendices to the report provide additional information on and supplementary analyses to the main results.

The NTC Induction Support Logic Model

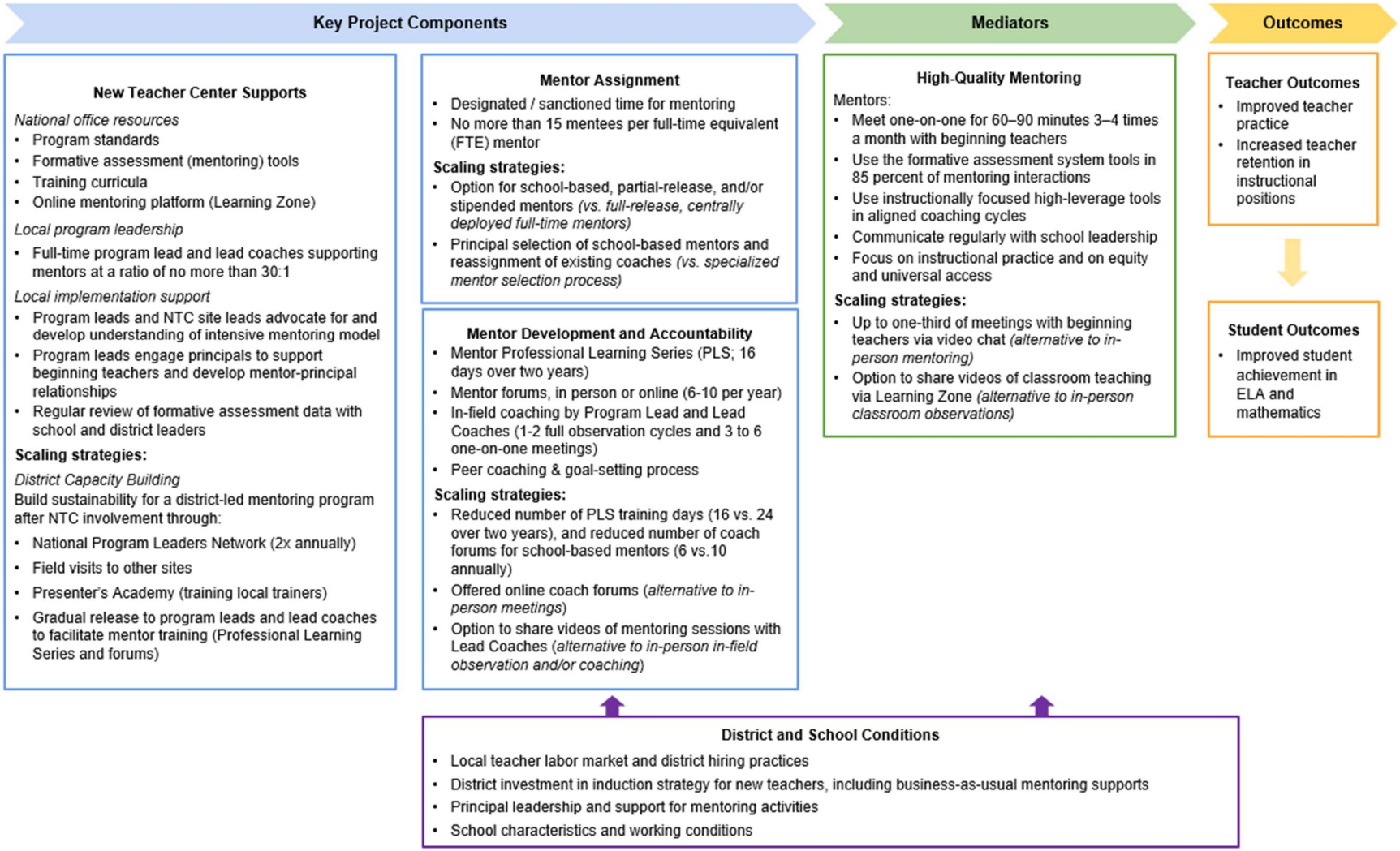
NTC induction supports are designed to help teachers develop foundational teaching skills across all subject areas. The goal of these supports is to improve teachers' instructional practice and retention in instructional positions in their schools and districts, with the ultimate outcome of improving students' academic achievement in ELA and mathematics (Exhibit 1).

[5] Implementation fidelity was lower in the first year of the project but consistently met the threshold for high fidelity in both the second and third years.

[6] Effect sizes were 0.09 ($p < .05$) in English language arts and 0.15 ($p < .01$) in mathematics.

[7] Impact estimates for the rural site, where SRI employed a QED design to estimate the program's effects, were inconclusive because the analysis was limited by extremely small sample sizes.

Exhibit 1. Logic model for the New Teacher Center scale-up grant



Key Program Components

NTC's induction model includes three key components, shown in the blue boxes in Exhibit 1: (1) New Teacher Center supports, including tools, curricula, and funded program leadership positions, (2) mentor assignment and dedicated time, and (3) mentor development and accountability. The model also includes one key mediator (the green box): provision of high-quality mentoring. For each program component and mediator, NTC designed scaling strategies intended to make the program more cost-effective, easier to implement in a diverse range of local contexts, and to make more efficient use of mentors' time.

New Teacher Center Supports

The first key component, New Teacher Center supports, includes tools, program resources, and staff funded by the grant or provided by NTC's national office.

Tools and program resources

The New Teacher Center national program office developed a common set of mentoring tools and other program resources that they provided to each study site, including program standards, formative assessment tools, curricula for mentor training, and an online mentoring platform (Learning Zone). One of the distinguishing characteristics of NTC's model is the suite of formative assessment system (FAS) tools to be used during mentor-teacher coaching meetings. NTC developed and refined these tools over many years in working with dozens of districts. NTC expected mentors to use one or more of these tools to organize and guide their conversations in nearly all of their meetings with teachers. NTC also designated three of the tools as "high-leverage." High-leverage tools focus on key phases of instruction, including planning lessons aligned to standards, designing and delivering engaging instruction that supports equitable access to the content for all students, and formative review of student data to inform future planning (see below). NTC expected that mentors would use these high-leverage tools in combination to support regular coaching cycles with teachers (three cycles per year).

> **High-Leverage Coaching Tools to Support Aligned Coaching Cycles**
> ***Planning Conversation Guide (PCG)***
>
> Mentors use the PCG to support teachers in analyzing or planning a standards-aligned lesson based on knowledge of students and content, with the aim of meeting the needs of all learners. Using prompts in the tool, mentors help beginning teachers to design lessons that will increase student engagement and access to the content.
>
> ***Observation Cycle Tools: Pre-Observation Conversation Guide, Selective Scripting or Seating Chart, and Post-Observation Co-Assessment (POCA)***
>
> Using the Observation Cycle tools, mentors and teachers examine classroom practices that contribute to a learning environment necessary for effective teaching and learning; assess alignment of instructional strategies with content standards and student learning needs; analyze observation data for evidence of effective practices and areas for growth; and determine student learning needs to inform future planning.

> ***Analysis of Student Learning (ASL)***
>
> Mentors use the ASL to guide beginning teachers through the process of analyzing artifacts of student learning, so that teachers can better understand the results of standards-based instruction. Mentors use the ASL as an entry point for a coaching cycle, to follow up after the classroom observation by analyzing the student work produced during that lesson, or to help teachers better understand how students respond to specific types of learning tasks.

Local program leadership

Under the i3 scale-up grant, NTC supported local program leadership by funding several full-time positions in each site. The program lead was employed by each district in the district's teacher development or beginning teacher support office and was responsible for coordinating with NTC on all aspects of local program implementation. This position was fully funded by the grant. In addition, the grant funded two to six lead coach positions in each district, with the number depending on the number of mentors trained and deployed in study schools. Lead coaches were responsible for ongoing mentor development and support via mentor forums and in-field coaching.

Local implementation support

NTC site leads (i.e., the NTC staff responsible for building relationships with the specific site and fulfilling NTC's contract with the site) and local program leads were responsible for supporting local implementation of the NTC's induction model by developing district leaders' understanding of NTC's intensive mentoring model and advocating for this model within the district, holding regular meetings with key stakeholders, and engaging principals to support their beginning teachers. NTC offered two half-day trainings for school leaders to introduce them to the mentoring model and expectations for their role in supporting beginning teacher development. Program leads were expected to meet monthly with school leaders to review formative implementation data and to ensure that beginning teachers had dedicated time to meet with their mentors and other supports needed for their development.

Scaling strategies

As a scaling strategy, NTC site leads built the capacity of program leads and lead coaches to train new cohorts of teachers after the grant period ended. NTC offered a Presenter's Academy to train local program staff, which replicated 16 days of foundational mentor training for new cohorts of mentors. Program leads and lead coaches also co-facilitated mentor forums with NTC staff, gradually assuming full responsibility for planning and implementing this component of mentor development. In the last two years of the grant, NTC convened program leads and district leaders in a National Program Leaders Network to plan for sustaining induction supports in their districts after the grant period ended.

Mentor Assignment

The second key component of the NTC induction model under the scale-up grant was mentor assignment. Under this component, NTC and local district representatives were expected to assign mentors with designated time for mentoring, and each mentor was expected to serve no more than 15 mentees.[8]

[8] Proportional to the 1:15 ratio expected for full-time mentors, NTC expected that half-time mentors would serve no more than eight teachers, .25 full-time equivalent (FTE) members would serve no more than four teachers, and stipended classroom teachers with no release time would serve no more than two teachers.

NTC's scaling strategies for mentor assignment were twofold. First, alternative approaches to staffing mentor positions allowed districts to adapt and enhance existing induction programs. Under the i3 validation grant, all mentors were full-time, served multiple schools in their districts, and did not have any duties at the school or district outside of supporting beginning teachers. As a scaling strategy, NTC allowed districts to use partial-release school-based mentors, as the full-time release model was perceived to be cost prohibitive for some districts.

Second, the scale-up sites used alternate strategies for selecting and assigning mentors, including reassignment of existing instructional coaches. These school-based mentors were classroom teachers or instructional coaches who provided induction mentoring part-time to much smaller caseloads of beginning teachers in their own schools. This was a significant shift from the model tested under the validation grant, in which NTC oversaw a highly selective recruitment, screening, and hiring process for all mentors.

Mentor Development and Accountability

The mentor development and accountability component stipulated that mentors receive intensive training and engage in peer coaching and goal setting. This component included four types of mentor support. First, mentor academies were a series of structured multi-day professional development sessions on instructional mentoring. Second, mentor forums were professional learning communities designed to develop mentors' skills. Third, mentor-to-mentor shadowing, for full-time release mentors only, provided peer support for mentors. Finally, NTC site leads provided ongoing support and feedback to all mentors. In addition to these supports, NTC mentors were expected to participate in peer coaching and goal setting, including completing self-assessment and professional growth tools and collaborative assessment logs.

Under this key component, scaling strategies included: (1) modifying mentor training and mentor forums to reduce the number of days school-based mentors must be out of the classroom; (2) offering online monthly mentor forums to reduce travel time for mentors dispersed across large geographic regions or congested urban areas; and (3) leveraging online collaboration tools for lead coaches to observe mentors during in-field coaching, thus reducing coaches' travel time among schools.

Key Mediator

These three program components produce impacts on teacher and student outcomes via a key mediator, the provision of high-quality mentoring to beginning teachers.[9] NTC staff did not provide mentoring to teachers directly, but rather trained mentors who then worked with their assigned beginning teachers. NTC defined high-quality instructional mentoring by setting expectations for the frequency, duration, and nature of mentoring activities; use of NTC-developed mentoring tools; and advocacy support. Mentors were expected to meet one-on-one with each of their mentees for at least 60 minutes at least three times per month and focus their mentoring on instructional practice and on equity and universal access. Mentors were also expected to use one of NTC's formative assessment (mentoring) tools with each beginning teacher in at least 85 percent of their interactions and use key

[9] In keeping with guidance from the i3 grant program, this report uses the term "mediator" to refer to the behaviors, processes, and skills targeted by the intervention's key components. Mediators are expected to lead, in turn, to teacher and student outcomes (Boulay et al., 2018). These aspects of the program were not implemented by NTC directly, but rather by mentors trained by NTC, working with assigned beginning teachers. These mediators were not tested in a statistical mediation model.

high-leverage tools in regular coaching cycles at least three times a year. Finally, mentors were expected to communicate regularly with school leadership to advocate for their beginning teachers.

As a scaling strategy, centrally deployed mentors who served beginning teachers across multiple schools were offered video and virtual coaching tools to observe teachers and provide coaching feedback virtually, reducing the time spent traveling among schools while maintaining face-to-face interactions for the majority of mentoring activities. School-based mentors who were also classroom teachers had access to these same tools when teaching schedules made observations and in-person meetings difficult to schedule.

Outcomes

NTC's key program components and mediators are hypothesized to lead to both improved teacher and student outcomes, as shown in Exhibit 1. The teacher outcomes—improved teacher practice and teacher retention—are, in turn, expected to mediate NTC's impact on student outcomes, specifically improved student achievement in math and ELA.

District and School Context

Finally, the logic model in Exhibit 1 includes key contextual factors that affect mentors' opportunities to work with beginning teachers and shape beginning teachers' evolving practice. Local district and school conditions at each partner site, such as a district's existing investments in induction for beginning teachers and principals' leadership and support for mentoring activities, can facilitate or hinder the implementation of NTC's model and the outcomes it can achieve.

Implementing Sites

Under its i3 scale-up grant, NTC aimed to provide comprehensive induction services in five sites across diverse regions in the country, including large, urban districts (Garfield City, Fillmore County, Chesterton) and smaller towns serving rural communities (Hayes Park and Taftville).[10] NTC had not provided comprehensive teacher induction in these sites before (with the exception of some schools in Garfield City, which were not served under the scale-up grant). The five school systems serve a large proportion of students of color and students who were eligible for free or reduced-price meals (Exhibit 2).

[10] The district names used throughout the report are pseudonyms.

Exhibit 2. Average site characteristics at baseline (2015–16 school year)

Site	Number of K–8 schools	Percent students of color	Percent eligible for free or reduced-price meals	Percent English learners	Percent receiving special education services
Hayes Park	164	58	79	11	11
Garfield City	> 600	85	71	13	20
Fillmore County	516	93	72	19	10
Chesterton	114	86	56	26	12
Taftville	100	90	86	22	11

Study Design

To investigate the implementation and impact of the NTC model as described above in the scale-up sites, SRI designed and implemented a randomized control trial, with randomization at the school level. All beginning teachers in schools randomized to the treatment condition received NTC induction support, and all beginning teachers in schools randomized to control received the indication supports their district normally provides under business-as-usual conditions.

Research Questions

The study was designed to answer research questions about implementation, confirmatory research questions about impacts on teachers and students, and additional exploratory questions about impacts.

Implementation research questions

(1) What is the level of implementation fidelity to the NTC model in the five study sites?

(2) To what extent is there a contrast between the treatment and the control condition in the level and type of mentoring teachers received?

Confirmatory impact research questions

(3) What is the impact of two years of NTC induction supports on teaching practices overall and in the domains of classroom environment and instruction, compared with the control condition?

(4) What is the impact of NTC induction supports on student achievement in English language arts (ELA) and math in grades 4 through 8, after one year of exposure to beginning teachers in their second year of teaching, compared with the control condition?

(5) What is the impact of two years of NTC induction supports on teacher retention into a third year in the same district, compared with the control condition?

Exploratory research questions

(6) Is effective instruction related to student achievement? If NTC induction supports have impacts on teacher practice and student achievement (as examined by research questions 3 and 4), does effective instruction mediate the NTC effect on student outcomes?

(7) Are higher levels of mentoring (e.g., in frequency and instructional focus) related to higher student achievement?

(8) Is the effect of NTC induction supports on student achievement moderated by any school- or student-level variables?

School Recruitment and Eligibility

Within the five study sites, two groups of schools were recruited for participation in the study: one group in the summer of 2016 and another group in the summer of 2017. SRI collected baseline data from school year 2015–16 on schools and teachers to identify eligible participants before randomly assigning the schools to the treatment or the control (business-as-usual) condition. Within the study schools, the SRI evaluation team tracked mentors and teachers for two years and collected data on their participation in induction activities, level and type of induction support provided and received, classroom practice, student achievement, and teacher retention in the study sites.

In Hayes Park, Chesterton, and Taftville, all schools that hired beginning teachers were eligible for the study. In Garfield City, the sample included only schools within select geographic regions whose principals opted into the study (five regions in 2016, with a sixth added in 2017). Finally, in Fillmore County, the sample excluded schools targeted for additional supports and oversight by the district. These schools were excluded from the study to avoid duplication of services.

In all five sites, the first group of eligible schools was randomly assigned to the treatment or control condition, within blocks, in the summer of 2016. The blocking variables used were site, grades served (elementary or middle), Title I status, achievement level (Fillmore County), and geography (Hayes Park and Garfield City). An additional group of schools was randomly assigned within these blocks in the summer of 2017. Within each block, all eligible schools were randomly assigned to treatment or business-as-usual with a 50 percent chance of receiving treatment.

Characteristics of the sample of randomized schools are shown in Exhibit 3. Overall, the differences between treatment and control study schools on student characteristics were all small due to randomization. On the other hand, while the demographics of the study schools were similar to the sites in which they were located, there were some key differences. First, in Hayes Park and Garfield City, the percent of students receiving free or reduced-price meals and the percent students of color was substantially higher[11] in study schools than in the sites overall. Second, in every site except Garfield City, the average percent of English learners at a school was substantially higher in the study schools than in the site overall. Thus, the schools in this study can be characterized as having higher-than-average proportions of historically underserved students than their sites overall.

[11] "Substantial" differences are those greater than five percentage points.

Exhibit 3. Average site and study sample characteristics at the time of random assignment

Districts	Study condition	Number of schools	Percent students of color	Percent free or reduced-price meals	Percent English learners	Percent special education
Hayes Park	Treatment	39	85	87	22	9
	Control	33	88	93	23	12
	Overall study	**72**	**87**	**90**	**22**	**10**
	Overall site	***164***	***58***	***79***	***11***	***11***
Garfield City	Treatment	39	95	85	12	23
	Control	32	96	89	14	22
	Overall study	**71**	**96**	**87**	**13**	**22**
	Overall site	***> 600***	***85***	***71***	***13***	***20***
Fillmore County	Treatment	29	90	73	25	9
	Control	37	93	77	28	11
	Overall study	**66**	**92**	**75**	**27**	**10**
	Overall site	***516***	***93***	***72***	***19***	***10***
Chesterton	Treatment	23	84	58	32	12
	Control	26	85	62	37	11
	Overall study	**49**	**85**	**60**	**34**	**12**
	Overall site	***114***	***86***	***56***	***26***	***12***
Taftville	Treatment	21	93	91	27	10
	Control	22	92	88	26	12
	Overall study	**43**	**93**	**89**	**27**	**11**
	Overall site	***100***	***90***	***86***	***22***	***11***

Note. Overall site statistics reflect all K–8 schools in the site. Overall study statistics reflect only the K–8 schools included in the study.

Teacher Eligibility

For schools that were randomized in the summer of 2016, the study included eligible beginning teachers who started in fall 2016 (Cohort 1 teachers) as well as those who started in fall 2017 (Cohort 2 teachers). For schools that were randomized in the summer of 2017, the study included only beginning teachers who started in fall 2017 (Cohort 2 teachers).

This study used a relatively narrow definition of a beginning teacher to determine teacher eligibility. This definition is based on assumptions about the traditional path of entry into the profession, in which teachers graduate from college or complete a master of arts in teaching and enter as brand new teachers in the fall of the following year, with no prior full-time teaching experience. The purpose of

these restrictions was twofold: first, the definition aligned with NTC's target demographic, and second, it allowed a consistent definition of the sample between treatment and control groups.

For the purposes of this study, a beginning teacher:

- Entered their first year of full-time teaching as the teacher of record in fall 2016 (Cohort 1 teachers) or fall 2017 (Cohort 2 teachers);
- Was hired between April 1 and October 1, 2016 (Cohort 1 teachers) or between April 1 and October 1, 2017 (Cohort 2 teachers);
- Was staffed as a full-time classroom teacher of record, including teachers of special education,[12] gifted education, bilingual education, or dual language immersion;
- Held a regular preliminary or professional teaching certificate, indicating the teacher was new to the profession and did not hold an emergency or short-term credential;[13] and
- Had no previous teaching experience, defined as fewer than two months of prior teaching experience, either inside or outside of the district. Teachers who taught for more than two consecutive months as a long-term substitute teacher in a single classroom were excluded from the study sample.[14]

The research team collected data from a variety of sources to determine teacher eligibility for the study, including human resources (HR) data (e.g., hire dates, salaries, contract type, and hiring history); teacher self-report of prior teaching experience on the spring 2017 beginning teacher survey; and email and phone conversations during the observation scheduling process in fall 2016 and fall 2017. The demographics of the full teacher sample are shown in Exhibit 4.[15] Overall, teachers in this sample were predominantly female and disproportionately White (non-Hispanic[16]), compared with the demographics of the students in their districts. While teacher demographics were similar between the treatment and control groups, there were some differences in teacher race and gender by study condition. All impact models accounted for these differences, which arose due to chance during the randomization process, by including teacher race and gender as covariates.

[12] Special education teachers in most sites receive NTC supports and are, thus, included in the study. However, Chesterton excluded special education teachers from the NTC program because they receive supports from another organization focused on supporting teachers in high-need positions, i.e., special education. The district and NTC together decided that providing NTC mentoring in addition to the support provided by the other organization would not be practical. Thus, these teachers are in neither the treatment nor the control group.

[13] Teachers working under emergency teaching credentials designed to fill acute staffing shortages were excluded. Teachers receiving significant support from alternative certification programs, including Teach for America, local site teaching residencies, and university or district interns, were also excluded. However, the study sample includes teaching fellows in Garfield City, who comprise a large proportion of the beginning teaching force in that site and receive less support from their alternative certification program than Teach for America corps members or other sites' residency programs.

[14] Prior teaching experience does not include student teaching experience, or teaching experience in higher education, early childhood education, or in an international setting. Thus, teachers with experience teaching in those settings are included in the study.

[15] SRI also attempted to collect data on teacher certification type, but that data was incomplete, and the sites could not confirm that all values were from baseline.

[16] This report uses the term "Hispanic," as opposed to "Latino" or "Latinx," to reflect the term used in the sites' extant data.

Exhibit 4. Average study sample teacher characteristics at the time of random assignment

		n teachers	Percent White, non-Hispanic	Percent Black, non-Hispanic	Percent Hispanic	Percent another race	Percent female
Hayes Park	Treatment	149	74	5	13	7	79
	Control	102	71	11	15	4	82
	Overall	251	73	8	14	6	80
Garfield City	Treatment	133	50	21	19	10	81
	Control	103	50	24	17	9	74
	Overall	236	50	22	18	10	78
Fillmore County	Treatment	60	27	67	5	2	77
	Control	60	32	62	7	0	79
	Overall	120	29	64	6	1	78
Chesterton	Treatment	46	46	15	2	37	83
	Control	56	45	21	5	29	71
	Overall	102	45	19	4	32	76
Taftville	Treatment	43	33	37	0	30	72
	Control	43	37	19	7	37	77
	Overall	86	35	28	3	34	74

Implementation Study

The implementation study provides data on the extent to which sites implemented NTC's induction model with fidelity and on differences between NTC's model and the business-as-usual induction supports (implementation research questions 1 and 2). These data are critical to understand the degree of contrast in mentoring supports provided to teachers in treatment and control schools and may help to shed light on findings from the impact study. Data for the implementation study came from NTC's online mentoring platform, Learning Zone, and from teacher surveys administered each spring.

Impact Study

The impact study used a school-level randomized controlled trial design to estimate the impact of NTC mentoring on teacher practice, student achievement, and teacher retention. The study addresses research questions 3 through 5 about the impacts of NTC induction on teacher practice, student achievement, and teacher retention. The impact study also addresses three exploratory questions (research questions 6 through 8) about the relationship between instructional practice and student achievement, the relationship between the level of mentoring and student achievement, and the moderation effect of student and school characteristics. Data for the impact study came from extant data provided by the sites, as well as SRI's observations of classroom practice.

***

This technical report presents findings from the analyses described above. Chapter 2 presents findings from the implementation study, including implementation fidelity and treatment-control contrast. Chapter 3 presents findings that address the study's confirmatory impact questions, including impacts on teacher practice, student achievement, and teacher retention. Chapter 4 presents exploratory analyses to address additional research questions about the relationship between NTC's induction supports and student outcomes. The report ends with a discussion of implications for future research.

Chapter 2. Implementation Findings

The implementation study addresses two primary research questions:

(1) What is the level of implementation fidelity to the NTC model in the five study sites?

(2) To what extent is there a contrast between the treatment and the control conditions in the level and type of mentoring beginning teachers received?

Answering these research questions is critical in describing the NTC induction model as implemented in this study and may help to shed light on findings from the impact study.

Implementation Fidelity

Fidelity of program implementation is the extent to which key program components and mediators are carried out as intended. Measuring implementation fidelity can help identify areas for program improvement.

Key Components and Mediators

The NTC induction model as designed for the scale-up study included three key components and one key mediator, which align with the key components and mediators in the logic model. As discussed in Chapter 1, these key components and mediators are: New Teacher Center supports, mentor assignment, mentor development and accountability, and provision of high-quality mentoring (Exhibit 5).

Exhibit 5. NTC induction model key components and mediator

Component or mediator	Definition
NTC supports	National NTC office created and distributed standards, tools, training, and the online mentoring platform; advocated for sanctioned mentoring time; engaged principals; and helped build district capacity to continue the induction mentoring program beyond the grant period.
Mentor assignment	Mentors had adequate sanctioned time for mentoring and appropriate caseloads of beginning teachers.
Mentor development and accountability	Sites held trainings and discussion forums; mentors attended those trainings and forums; mentors received one-on-one support from NTC; and mentors engaged in peer coaching.
High-quality mentoring	Mentors met with teachers at the required frequency and duration; used NTC's tools; and met regularly with school leaders.

SRI worked with NTC to create quantitative indicators and thresholds for adequate implementation of each program component at the site level and for the study overall. Each component had between two and eight indicators. At the site level, the threshold for adequate fidelity of implementation for every component was to receive a score of "high" on 60 percent or more of individual indicators for that component and to receive a score of "low" on less than 20 percent of the indicators. Adequate study-

level implementation for each component was defined as four of five sites meeting the threshold for that component. SRI used descriptive statistics to examine the level of fidelity of the NTC induction model on each of these indicators at the site level and then rolled up the results to the study level to determine the overall level of fidelity to the NTC induction model on each of the key components and mediators.

See Appendix A for the full implementation fidelity matrix, including indicator-level fidelity thresholds and results.

Implementation Fidelity Findings

Implementation varied by site, by program component, and over time. However, in all three years of the study, implementation was inadequate overall, and it was lowest in the mediator, provision of high-quality mentoring. See Appendix A for fidelity at the indicator level for each site and study year.

- Two sites failed to meet the threshold for adequate fidelity of the first component (New Teacher Center Supports) because of low levels of principal engagement and because program leads did not meet requirements for building district capacity to sustain the induction mentoring program beyond the grant period.
- Several sites failed to meet the threshold for adequate fidelity of the second component (Mentor Assignment) because mentors did not have adequate sanctioned time for mentoring.
- Sites often failed to meet the threshold for adequate fidelity of the third component (Mentor Development & Accountability) because mentors did not attend the foundational NTC trainings and because mentors did not receive adequate one-on-one support from program leads or meet requirements for peer coaching and goal setting.
- All five sites struggled to meet the threshold for adequate fidelity of the high-quality mentoring practices expected to mediate teacher outcomes. Sites often failed because mentors were not regularly using NTC's formative assessment tools during meetings with beginning teachers and because mentors did not meet regularly with school leaders.

The threshold for program-level fidelity in each component was four of five sites implementing with adequate fidelity. Across the four components, this threshold was not met in any of the three years of the study (Exhibit 6). This indicates that, overall, none of the key components or mediators of the NTC program were implemented as intended.

Exhibit 6. Implementation fidelity for each key component and mediator of NTC's logic model

Implementation of Scaling Strategies

For each key component of its teacher induction model, NTC designed a corresponding scaling strategy, as shown in the initiative's logic model (Exhibit 1) and discussed in Chapter 1. Each of these scaling strategies was intended to make NTC's induction model less expensive for sites to adopt, more flexible to local conditions, and therefore easier to implement. The fidelity indicators designed for this study built in assumptions about the scaling strategies adopted for each key component.

In a district that hires hundreds of teachers each year, funding enough full-time induction mentor positions to serve all beginning teaches in caseloads of 15 per mentor is a significant program cost. One of the key scaling strategies NTC offered sites was to allow them to recruit and train classroom teachers or school-based instructional coaches to serve as part-time induction mentors for beginning teachers in their own schools. In school-based sites, classroom teachers and instructional coaches either accommodated induction mentoring work within their existing responsibilities, received release time to devote to mentoring, or received a modest stipend[17] in recognition of the additional time they were spending with beginning teachers.

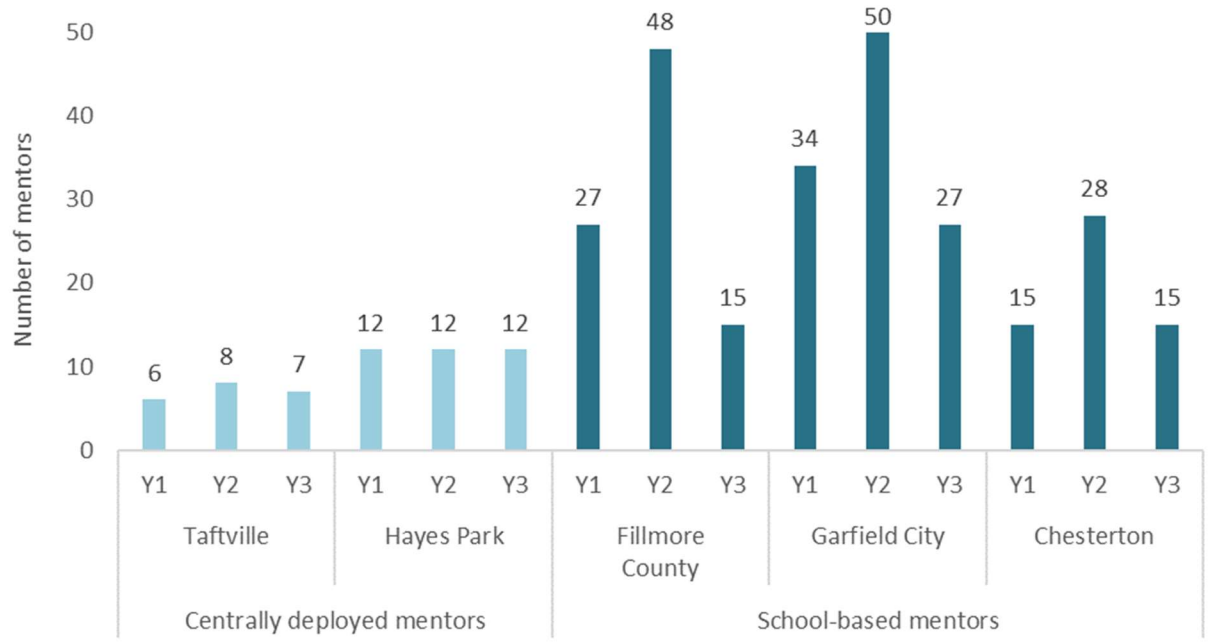
Three of the five study sites implemented this scaling strategy, with two sites using exclusively school-based mentors, one using a combination of centrally deployed and school-based mentors, and two sites using exclusively full-time, centrally deployed mentors. Program developers expected that school-based mentors' relatively limited time and flexibility in scheduling would interfere with their ability to attend some training or engage in the program in the same way as full-time mentors based in district offices. Therefore, NTC adjusted fidelity thresholds set at the individual mentor or teacher level on many of its indicators to reflect this scaling strategy. For example, NTC reduced expectations for ongoing mentor

[17] The stipend consisted of $1,000 per beginning teacher in Fillmore County.

development such that school-based mentors were expected to attend six mentor forums, compared with 10 for centrally deployed mentors. Similarly, school-based mentors were expected to complete two observation cycles per year with each teacher they supported, compared with three observation cycles per year for centrally deployed mentors. See Appendix A for additional detail.

In the three scale-up sites that adopted a school-based mentor deployment strategy, principals recruited and selected mentors from within their schools to work with small caseloads of teachers, depending on the proportion of release time available for mentoring. As a consequence, these sites recruited and trained up to four times as many mentors as the sites that created centrally deployed mentor positions (Exhibit 7). Average caseloads were also much lower in the school-based sites, where mentors served one to three beginning teachers, on average, compared with 13 in the centrally deployed sites.

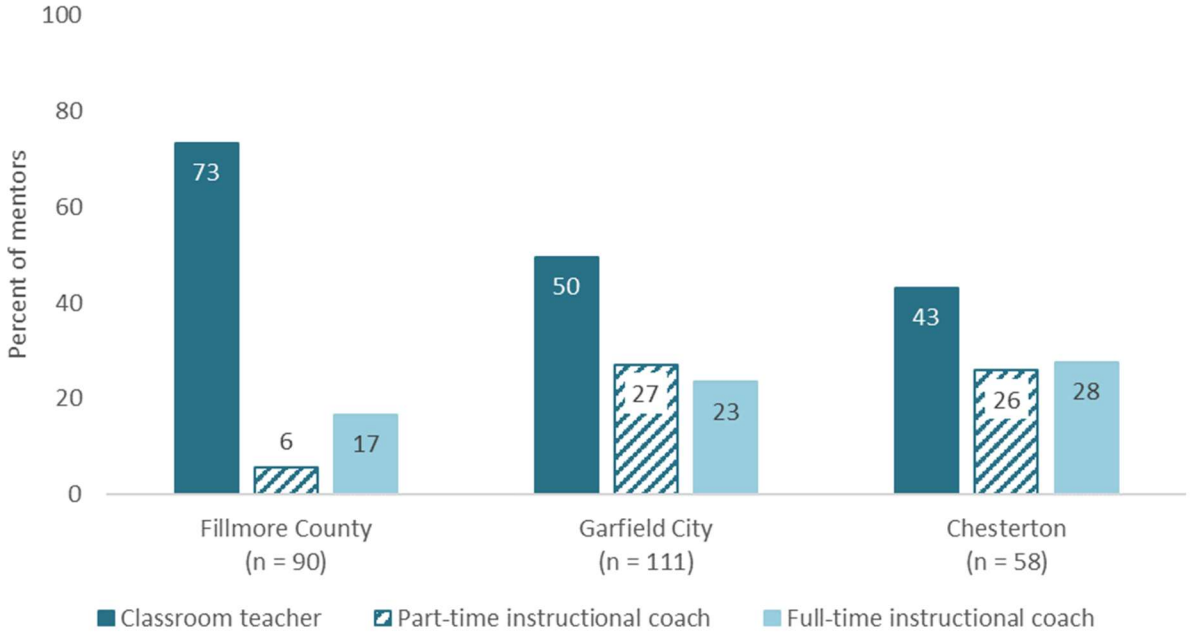
Exhibit 7. Number of mentors matched to study teachers, by site



Note. Mentors served beginning teachers in Cohort 1 only in Y1, in both cohorts in Y2, and in Cohort 2 only in Y3. Therefore, the number of school-based mentors required to serve all treatment teachers in the study was higher in Y2.

Among school-based sites, Fillmore County relied more heavily on classroom teachers to serve as mentors than other districts; nearly three-quarters of mentors in this site were themselves full-time classroom teachers (Exhibit 8). Despite the variation in roles, release time, and proximity to beginning teachers between school-based and centrally deployed mentors, SRI found no systematic differences in implementation fidelity between the sites that employed school-based mentors as a scaling strategy and sites that did not.

Exhibit 8. School-based mentor roles, by site



In addition to offering the option of school-based mentors to sites, NTC developed several other scaling strategies, as described in Chapter 1. They included: (1) district capacity-building to replicate NTC's mentor development strategy via train-the-trainer and co-facilitation activities, (2) reduced requirements for mentor training, and (3) virtual and video-based alternatives to in-person mentor training and to in-person mentoring or classroom observation for beginning teachers. Uptake of these additional scaling strategies varied across sites.

In the case of district capacity-building, NTC's expectations for program fidelity in this key component remained unchanged from past iterations of the intervention, and they were met in three of the five sites. Two sites did not send program leads and other district staff to all train-the-trainer activities as envisioned by NTC.

To support scaling of its approach to mentor training, NTC reduced the number of days required for its foundational training, the Professional Learning Series (PLS), from 24 days to 16 days over the course of two years. This reduction in training time represented a departure from past iterations of the PLS, including the version tested under NTC's i3 validation grant. NTC provided this revised training to all mentors in all five scale-up sites. Despite reduced time demands, mentors in the school-based sites still found it difficult to leave their buildings to attend training. In the final two years of the program, none of the school-based sites met the threshold for fidelity with respect to mentor training (80 percent of

mentors attend at least 80 percent of training days). Mentor attendance at training reached expected levels only in the two sites where mentors were centrally deployed.

Finally, NTC made video-based and virtual tools available for mentors and lead coaches to use at their discretion. NTC held the same expectations for the content, duration, and amount of training provided to mentors and for the mentoring provided to beginning teachers, regardless of whether the support was in-person or virtual. NTC maintained that expectations for the work itself should not change based on the mode of interaction. Mentor-teacher interaction logs recorded in Learning Zone did not capture the mode of interaction (virtual or in-person), and so SRI was not able to systematically assess whether virtual forms of mentoring differed from traditional in-person interactions with respect to implementation fidelity. However, interviews indicated that uptake of these strategies was low. Mentors reported that they did not adopt video sharing as a regular practice, preferring in-person interactions whenever possible.

Contrast Between Treatment and Business-as-Usual

In addition to measuring the fidelity of implementation among treatment schools, mentors, and teachers, SRI also examined the contrast between the supports received by treatment teachers and those received by control teachers in their first and second year of teaching. Finding a strong contrast between the treatment and control conditions would suggest that the NTC model differs from the existing business-as-usual supports for beginning teachers, increasing the likelihood for impact on teachers and students.

Each spring, NTC administers a Program Quality Survey to all teachers nationwide receiving support from their NTC-trained mentors. For the purposes of this study, SRI provided feedback on that survey, suggested additional items to add to measure the contrast between the treatment and control conditions, and ensured that all study teachers (both treatment and control) were sent the survey. Thus, NTC administered the teacher survey to study teachers at the end of their first year of teaching (spring 2016 for Cohort 1 and spring 2017 for Cohort 2) and at the end of their second year of teaching (spring 2017 for Cohort 1 and spring 2018 for Cohort 2). The response rates were above 75 percent for both conditions in both years (Exhibit 9).

Exhibit 9. Survey response rates by treatment condition and years of teaching experience

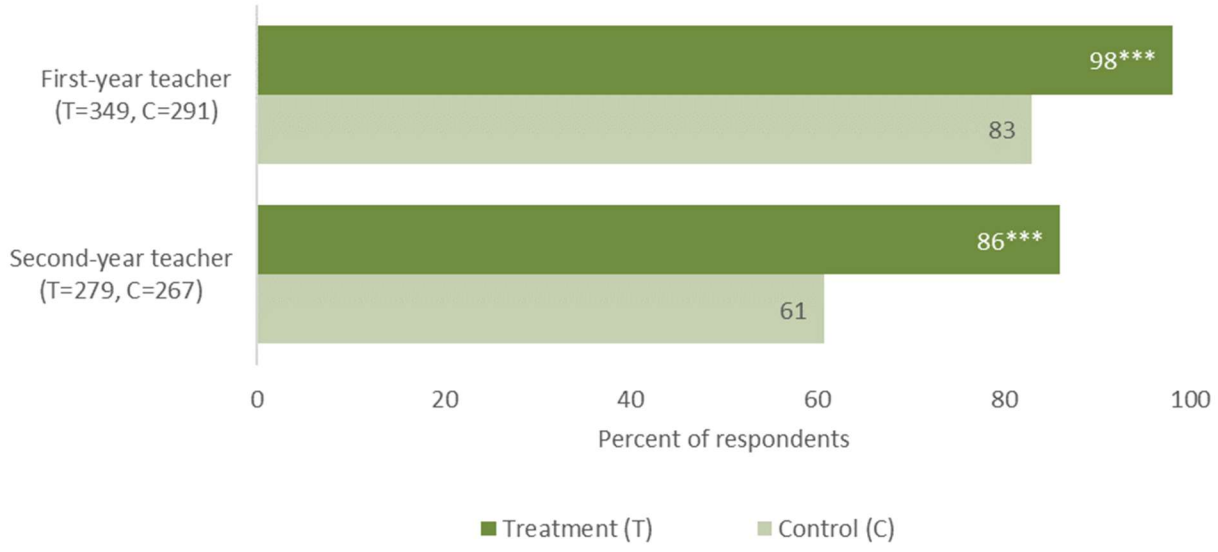
		Treatment	Control	Overall
First-year teacher	Surveyed	344	307	651
	Responded	279	233	512
	Response rate	81%	76%	79%
Second-year teacher	Surveyed	279	267	546
	Responded	254	235	489
	Response rate	91%	88%	90%

In addition to examining item-level responses to questions about the level and type of mentoring teachers received, SRI also created a series of variables based on sets of items that functioned together as scales. The SRI evaluation team used factor analysis to examine the properties of each scale and then created the scales based on an average of the items. See Appendix B for the properties of each scale. SRI used descriptive statistics and t-tests to examine the difference in the item-level percentages and scale means in the treatment and control (business-as-usual) conditions.

Mentor Assignments

Two study sites had formal induction programs for first-year teachers and two sites had formal induction programs for both first- and second-year teachers, before the NTC program was implemented. Therefore, it is not surprising that a large proportion of control teachers reported that their school had formally assigned a mentor to them in their first year (83 percent), and a substantial proportion reported that they had a mentor in their second year (61 percent). However, the proportion of treatment teachers reporting that they were formally assigned a mentor was significantly higher than the proportion of control teachers in both years (Exhibit 10).

Exhibit 10. Beginning teachers formally assigned a mentor, by years of teaching experience and treatment status

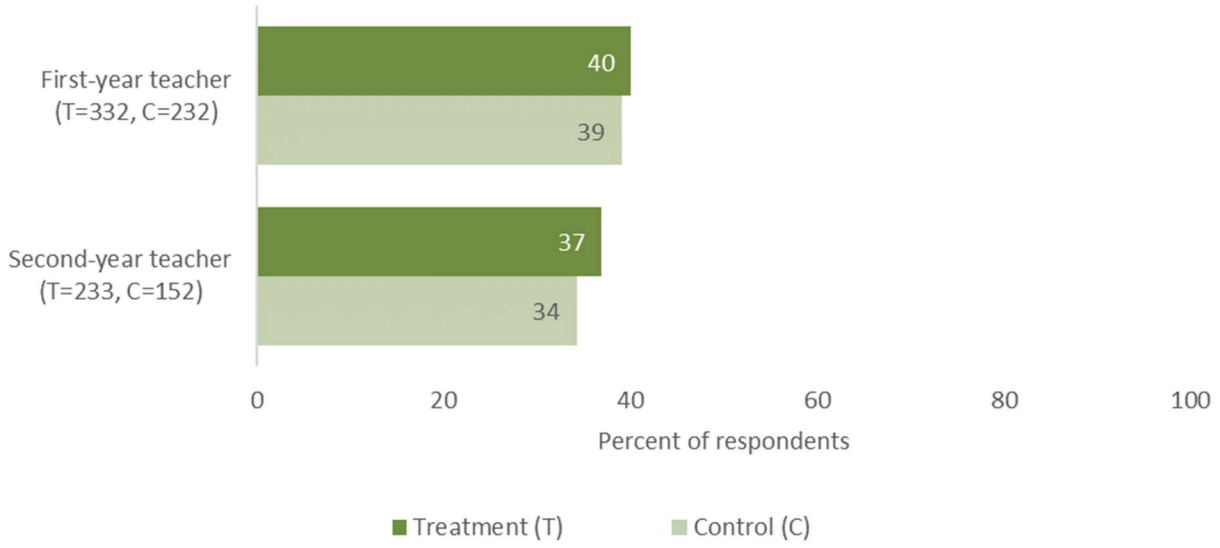


*$p < .05$, **$p < .01$, ***$p < .001$.

Frequency and Duration of Meetings Between Teachers and Mentors

Among those who were assigned a mentor, treatment and control teachers met with their mentors at a similar frequency and for similar durations, in both their first and second years of teaching (Exhibit 11). As a part of NTC's induction model in this study, treatment mentors were expected to meet with their teachers approximately weekly for one hour or more. Only two in five first-year treatment teachers who were assigned a mentor reported meeting this threshold, and a similar percentage of control teachers (39 percent) reported this same frequency and duration. Likewise, only 37 percent of treatment teachers reported meeting this threshold in their second year of teaching, and a similar percentage of control teachers (34 percent) also met with their mentors weekly for an hour or more in their second year of teaching. These findings reinforce the implementation data that suggest NTC mentors did not meet with treatment teachers with the frequency and duration that the NTC program intended.

Exhibit 11. Percent of beginning teachers with a mentor who met with that mentor weekly for an hour or more, by years of teaching experience and treatment status

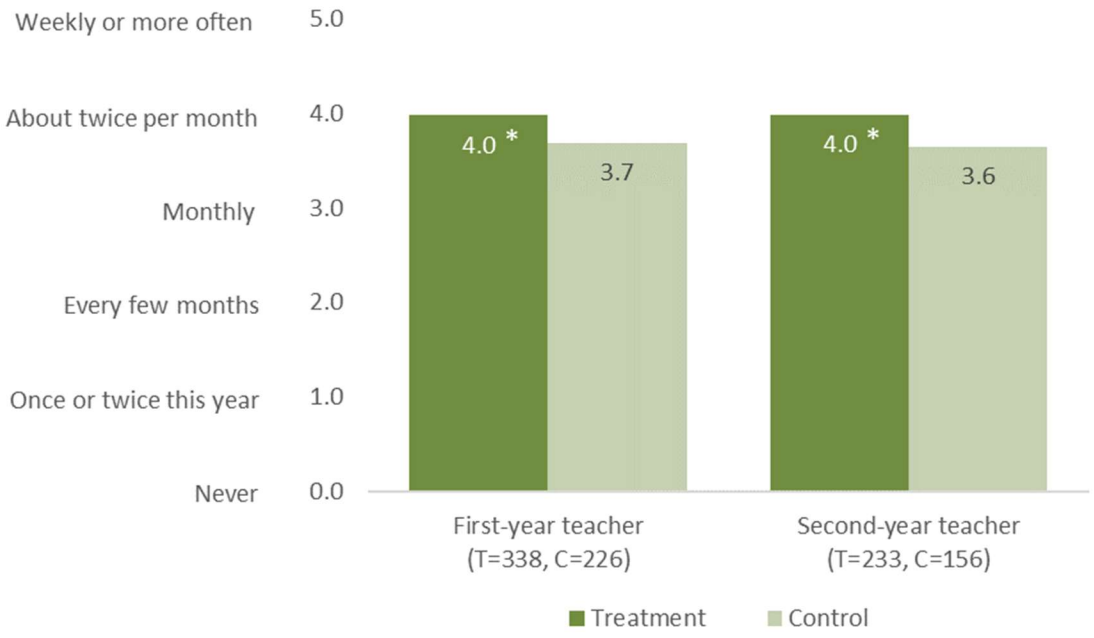


*$p < .05$, **$p < .01$, ***$p < .001$.

Mentors' Use of Instructionally Focused Strategies with Beginning Teachers

NTC identified a list of instructionally focused mentoring activities that they hypothesized were more likely to lead to improvements in instructional practice and student achievement (e.g., observations and feedback, analysis of student work, and discussing assessment data to inform instruction). The NTC tools and training are designed to support mentors in carrying out these activities with their beginning teachers, but control mentors may also conduct similar activities. When asked on the survey, both first- and second-year treatment teachers were more likely than control teachers to report that their mentors carried out these mentoring activities (Exhibit 12). This contrast indicates that NTC's induction program was more instructionally focused than the business-as-usual supports in the scale-up sites. However, it is worth noting that a substantial percentage of control teachers also reported receiving these supports from their mentors. Appendix B includes the list of items that contribute to the Instructionally Focused Mentoring scale and the properties of the scale, as well as graphs of the contrast between treatment and control for each individual item.

Exhibit 12. Average frequency of instructionally focused mentoring activities, by years of teacher experience and treatment status

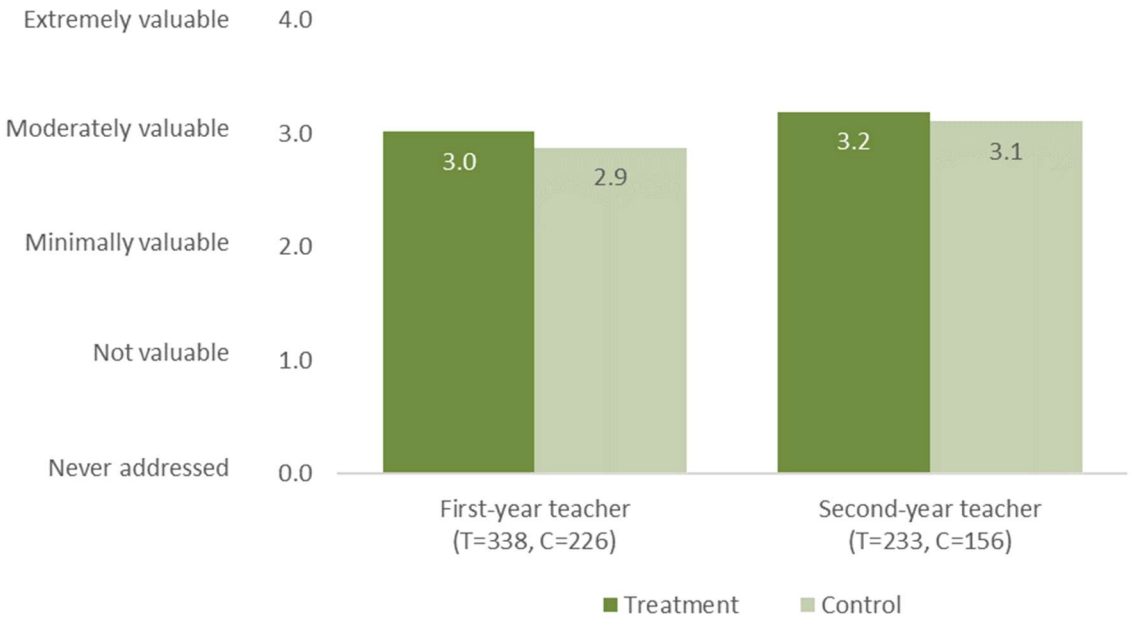


*$p < .05$, **$p < .01$, ***$p < .001$.

Teachers' Perceptions of the Value of the Mentoring They Received

On the survey, teachers also reported how valuable they found their mentors in supporting their instructional practice and data use, differentiating instruction, and supporting students' social and emotional learning. Overall, there was not a statistically significant difference between treatment and control teachers in how valuable they found the mentoring they received (Exhibit 13). Appendix B includes the list of items that contribute to the Value of Mentoring scale, the properties of the scale, as well as graphs of the contrast between treatment and control for each individual item.

Exhibit 13. Teachers' perceptions of the value of the mentoring they received, by treatment status and years of experience



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

Summary of Implementation Findings

Overall, the treatment schools implemented the NTC model with inadequate fidelity across all key components and mediators of the program in all three years of the study. In addition, the contrast between the treatment and control conditions was less than expected. Although teachers in treatment schools were more likely to be formally assigned a mentor than teachers in control schools, and teachers with NTC mentors were more likely to report receiving instructionally focused supports than teachers in control schools, there was not a significant difference in the frequency or duration of mentoring meetings between treatment and control teachers. This inadequate implementation of the intervention, along with the low level of contrast between treatment and control mentoring support, indicates that the study conditions were not ideal for detecting an impact of the NTC induction model as designed.

Chapter 3. Impact Findings

When implemented with fidelity, NTC induction supports are intended to improve teacher practice, student achievement, and teacher retention in the profession. This chapter shares results of the impact of NTC induction supports on teacher and student outcomes after two full years of implementation for both cohorts. This chapter addresses the three confirmatory impact research questions:

(3) What is the impact of two years of NTC induction supports on teaching practices overall and in the domains of classroom environment and instruction, compared with the control condition?

(4) What is the impact of NTC induction supports on student achievement in English language arts (ELA) and math in grades 4 through 8, after one year of exposure to beginning teachers in their second year of teaching, compared with the control condition?

(5) What is the impact of two years of NTC induction supports on teacher retention into a third year in the same district, compared with the control condition?

Impacts on Teacher Practice

As depicted in the logic model in Chapter 1 (Exhibit 1), the first hypothesized proximal outcome of NTC induction supports is improving beginning teachers' classroom practice. This is measured in this study via direct observations of teacher practice using the Danielson Framework for Teaching (FFT) observation rubric. This section presents the results of the analysis examining the impact of NTC on FFT scores after two years of induction support, across both study cohorts.

Measures

Trained observers rated each observed teacher on 11 elements under Domain 2: Classroom Environment and 15 elements under Domain 3: Instruction on the Danielson Framework for Teaching (FFT) (Danielson, 2013), shown in Exhibit 14.**Error! Reference source not found.**

Exhibit 14. Framework for Teaching domains, components, and elements observed in this study

Domain 2: The Classroom Environment	
Component	**Elements**
Creating an Environment of Respect and Rapport	Teacher interactions with students
	Student interactions with other students
Establishing a Culture for Learning	Importance of the content and of learning
	Expectations for learning and achievement
Managing Classroom Procedures	Management of instructional groups
	Management of transitions
	Management of materials and supplies
	Performance of classroom routines

Domain 2: The Classroom Environment	
Component	**Elements**
Managing Student Behavior	Expectations
	Monitoring of student behavior
	Response to student misbehavior
Domain 3: Instruction	
Communicating with Students	Expectations for learning
	Directions for activities
	Explanations of content
	Use of oral and written language
Using Questioning and Discussion Techniques	Quality of questions/prompts
	Discussion techniques
	Student participation
Engaging Students in Learning	Activities and assignments
	Grouping of students
	Instructional materials and resources
	Structure and pacing
Using Assessment in Instruction	Assessment criteria
	Monitoring of student learning
	Feedback to students
	Student self-assessment and monitoring of progress

A total of 356 teachers were observed at both baseline and follow-up, across two cohorts. Of these, 33 (9.3 percent) were observed and rated by more than one observer on the same lesson and examined for inter-rater reliability. The average reliability of ratings between observers within one point was 95 percent. See Appendix G for additional information on the outcome measure, including reliability information.

After observers scored each element, the SRI evaluation team created a set of 11 variables for the main (confirmatory) and exploratory analyses. The main analysis uses an "Overall Classroom Practice" variable that is an average of all the elements listed in Exhibit 14. In addition to this measure of overall classroom practice, SRI examined a "Classroom Environment" variable that is an average of the elements included under the FFT Classroom Environment domain and an "Instruction" variable that is an average of the elements included under the FFT Instruction domain. Finally, SRI also examined component-level scores, which were created by averaging the elements in each component shown in Exhibit 14. Each analytic variable has a range consistent with the FFT's original scale, where a one represents "Unsatisfactory," and a four represents "Distinguished."

Sample

Within the study sample of teachers described in Chapter 1, teachers were eligible to be observed if they taught a core subject (mathematics, reading/English language arts, social studies, or science), or if they taught in a multi-subject elementary classroom.[18] Trained, independent observers conducted observations during instruction in core subjects in both treatment and control schools, blind to treatment status. They conducted baseline observations shortly after beginning teachers entered the classroom (fall 2016 for Cohort 1 and in fall 2017 for Cohort 2 teachers), and follow-up observations near the end of teachers' second full year of teaching (spring 2018 and spring 2019). These teachers had received nearly two full school years of mentoring at the time of follow-up.

Attrition

Attrition from this analysis occurred when teachers eligible for observation were not observed at baseline (e.g., because of principal or teacher refusal), and when those observed at baseline were not observed at follow-up (e.g., because teachers left the study site or did not respond to requests to schedule a follow-up observation). Schools attrited from the sample when all of the teachers who were eligible for observation within the school attrited (i.e., were not observed at both time periods). Exhibit 15 shows the reduction of the school sample from the full set of schools included in the study (299) to the number with eligible teachers (264), and finally the number of schools with teachers included in the observation analysis (201).

Exhibit 15. Overall school sample and attrition for the observation analysis

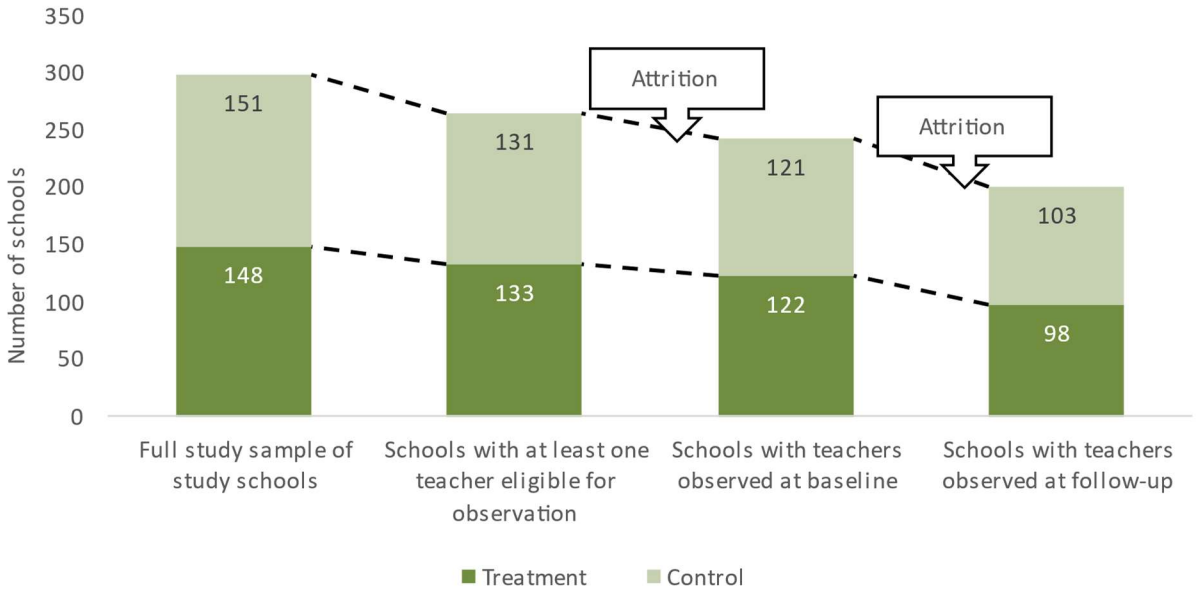


Exhibit 16 displays the school-level attrition rates by condition. WWC standards for attrition consider both overall attrition and the difference in attrition between treatment and control groups (differential attrition). With both cohorts combined, overall school-level attrition was 24 percent, with differential

[18] This excluded special education self-contained and English language development classrooms in Cohort 1 but included some of these classrooms in Cohort 2.

attrition of 4.9 percentage points. This is within the "optimistic boundary," and just at the "cautious boundary" as defined by WWC design standards (WWC, 2022).

Exhibit 16. School-level attrition ratings

	Treatment	Control	Overall	Differential
Schools with at least one eligible teacher	133	131	264	
Schools with eligible teachers observed at both time points	98	103	201	
Percent attrited	26%	21%	24%	4.9%
		Met attrition standards		

Because the unit of random assignment was schools and the unit of analysis was teachers, this study has a high risk of bias due to joiners (Figure 9, p. 48; WWC, 2022). Therefore, SRI tested the sensitivity of the results to the inclusion of these joiner teachers, and results are shown in Appendix C.

Representativeness

As an RCT with low cluster-level attrition, but a high risk of bias due to joiners, the study must demonstrate the representativeness of teachers within non-attrited schools (WWC, 2022). Representativeness is calculated in the same way as attrition, except that the reference sample is the individuals (teachers) in non-attrited clusters (schools) at follow-up. The overall teacher-level attrition with respect to eligible teachers within non-attrited schools at follow-up was 28 percent, with differential attrition of 0.4 percentage points (Exhibit 17). This is within the "cautious boundary." Thus, the teachers included in this analysis are representative of clusters at follow-up (WWC, 2022).

Exhibit 17. Representativeness of clusters of follow-up – Teacher Practice

	Treatment	Control	Overall	Differential
Teachers eligible for observation in non-attrited schools	254	239	493	
Teachers eligible for observation in non-attrited schools who were observed at both time points	184	172	356	
Percent attrited	27.6%	28.0%	27.8%	0.4
		Met attrition standards		

Methods

Using the sample described above, SRI compared the scores of treatment teachers to the scores of control teachers on Overall Teacher Practice, the Classroom Environment and Instruction domains, and component-level measures, using multilevel models with teachers nested within schools. The form of the models was:

$$Obs_{jk} = \gamma_{00} + \gamma_{01}Treatment_k + \gamma_{02}S_k + \gamma_{03}D_k + \gamma_{10}Z_{jk} + \gamma_{04}S_k * D_k + \gamma_{20}Z_{jk} * D_k + r_{jk} + u_{0k}$$

Where Obs_{jk} is the average observation score (overall, for each domain, and for each component) for teacher *j* in school *k*. The treatment indicator is entered at the school level, to reflect the random

assignment approach. S_k is a vector of school control variables, including the blocking variables used in random assignment. D_k represents site fixed effects, and Z_{jk} is a vector of teacher-level covariates, including baseline observation scores. Each school- and teacher-level covariate is also interacted with the site indicators ($S_k * D_k$ and $Z_{jk} * D_k$), to account for differential relationships between these predictors and the outcome across sites. The full set of control variables is shown in Exhibit 18.

Exhibit 18. Control variables included in the model estimating the impact of NTC on teacher practice

	Control variable
School-level variables	Percent of students in poverty
	Percent of students who are English learners
	Percent of students whose race/ethnicity is African American and/or Hispanic
	Randomization block
Teacher-level variables	Teacher race/ethnicity is African American and/or Hispanic
	Teacher gender is female
	Baseline observation score

There was a small amount of missing data in covariates, and SRI used multiple imputation to fill in these missing values,[19] including baseline scores. The imputation model used the same covariates as the outcome model, including the treatment indicator, as well as the outcome measure. Missing outcome scores were not imputed.

Results

There was no detectable effect of NTC induction supports on the measure of overall teacher practice or on the separate Classroom Environment or Instruction domains (Exhibit 19). However, the NTC induction supports had a statistically significant and positive impact on ratings of teachers' practice in the Communicating with Students component (0.10 points on the FFT scale of 1–4). The effect size of the impact is equivalent to about one-quarter of a standard deviation, which is a moderate impact in education.

[19] One teacher was missing several baseline observation scores, three teachers were missing school-level covariates, and 33 teachers were missing race/ethnicity. SRI used the *mi impute* command in Stata to impute for missing values, specifying a chained equations approach and adding five imputations for each teacher-level observation. In analysis, SRI accounted for the imputation approach using the *mi estimate* prefix on all commands.

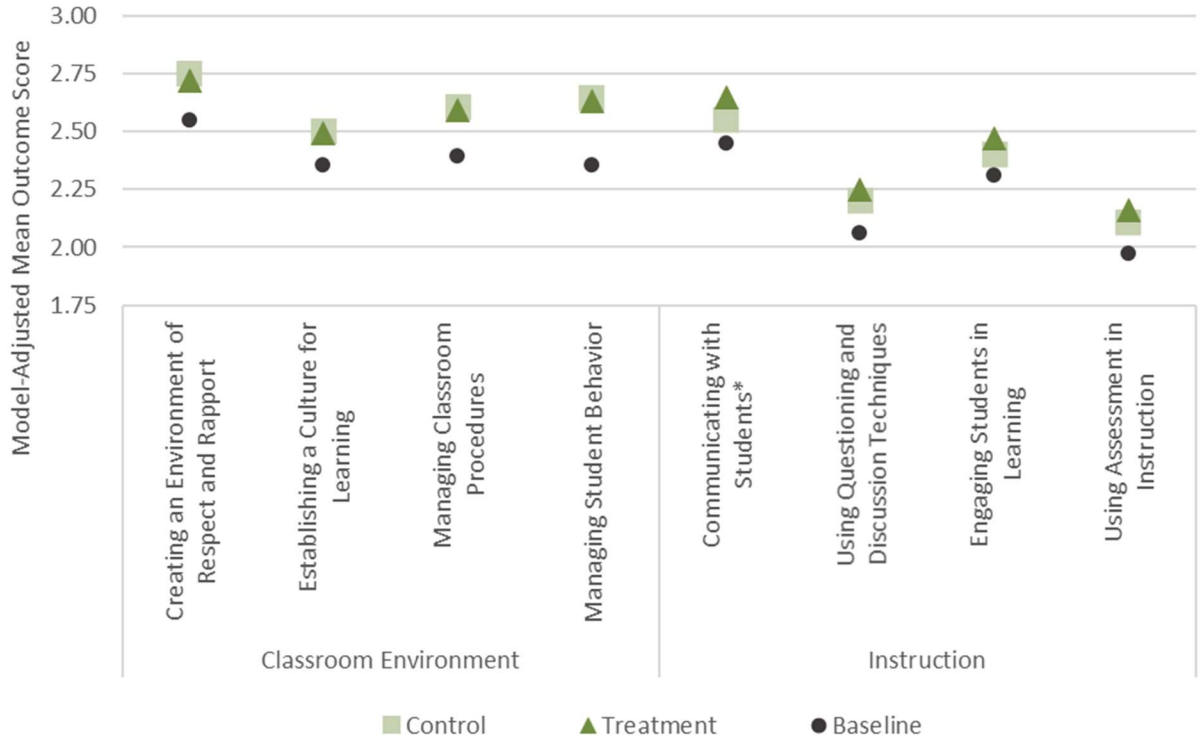
Exhibit 19. Impact of the NTC model on teacher practice outcomes

Overall teacher practice				
	Impact	***p* value**	**Pooled *SD***	**Effect size**
Overall teacher practice	0.02	0.72	0.41	0.04
Domain 2: Classroom Environment				
Classroom Environment Domain	-0.02	0.62	0.47	-0.05
Creating an Environment of Respect and Rapport	-0.03	0.65	0.62	-0.05
Establishing a Culture for Learning	-0.01	0.86	0.48	-0.02
Managing Classroom Procedures	-0.01	0.80	0.48	-0.03
Managing Student Behavior	-0.02	0.83	0.68	-0.02
Domain 3: Instruction				
Instruction Domain	0.07	0.13	0.42	0.16
Communicating with Students	0.10*	0.03	0.40	0.24
Using Questioning and Discussion Techniques	0.05	0.44	0.61	0.08
Engaging Students in Learning	0.07	0.21	0.52	0.13
Using Assessment in Instruction	0.05	0.36	0.48	0.10
Sample size				
n schools	201			
n teachers	356			

*$p < 0.05$.

Exhibit 20 graphically displays the estimated component-level outcome observation scores, accounting for baseline observation scores and school- and teacher-level covariates. These graphs show the variation in average outcome scores by domain, with generally higher average scores on the components in the Classroom Environment domain than in the Instruction domain, for both treatment and control teachers. They also show that the difference between treatment and control is very small on the components in the Classroom Environment domain, and larger in the components in the Instruction domain. Finally, they show that the average growth from baseline to outcome varies by component. The one statistically significant result—Communicating with Students—shows the largest gap between treatment and control at outcome.

Exhibit 20. Graphic display of predicted baseline and outcomes scores on the FFT, by treatment status and FFT component



*$p < 0.05$.

See Appendix C for the findings of additional supplementary analyses on the teacher practice outcomes, including impacts by mentor type (centrally deployed versus school-based mentors) and sensitivity of the results to the inclusion of teacher joiners.

Impacts on Student Achievement

Through an impact on teacher practice, NTC induction support is expected to improve students' achievement on standardized tests in mathematics and reading/English language arts (ELA). This section presents the results of the analysis examining the one-year impact on students' achievement after their teachers had received two full years of induction mentoring.[20]

Sample

Within the randomly assigned sample of schools described in Chapter 1, students were eligible to be included in the impact sample if they took mathematics or ELA classes with eligible teachers (that is, beginning teachers) in grades 4 through 8 (the grades with state standardized testing in those subjects).

[20] Note that this is not a two-year impact on students' achievement because students did not usually have a study teacher in both years of the study, and the models control for students' achievement in the prior year (not two years prior). Thus, this analysis tests the impact of having an NTC-supported teacher for one school year, after that teacher had received two full years of NTC induction supports.

In the study sites, students were tested in mathematics and ELA in grades 3 through 8. As discussed below, students' prior achievement is a covariate in the models. Therefore, grade 3 students are excluded from the analysis because they do not have a prior-year test score. Students identified for special education and English learners in an inclusion setting were included in the analysis as long as they took the traditional state standardized assessment.

Attrition

As in the teacher practice analysis, attrition is calculated at the school level (Exhibit 21). Schools attrited from the achievement analysis if all eligible study students left the study before testing or if none of the eligible students in a school had valid outcome test scores. The overall school-level attrition rate was 3.6 percent in mathematics and 4.1 percent in ELA, with 3.8 percentage points differential attrition in math and 1.8 percentage points in ELA. This is below WWC's cautious boundary for both outcomes (WWC, 2022). Thus, the risk of bias due to attrition in this analysis is low.

Exhibit 21. School-level attrition ratings

Mathematics				
	Treatment	**Control**	**Total**	**Differential**
Schools with students eligible to be included in the analysis	57	54	111	
Included schools	56	51	107	
Percent attrited	1.5%	5.6%	3.6%	3.8
		Met attrition standards		
ELA				
	Treatment	**Control**	**Total**	**Differential**
Schools with students eligible to be included in the analysis	62	60	122	
Included schools	60	57	117	
Percent attrited	3.2%	5.0%	4.1%	1.8
		Met attrition standards		

Because the unit of random assignment was schools and the unit of analysis was students, and not all classrooms in each school were included in the study, this study has a high risk of bias due to joiners. Therefore, SRI tested the sensitivity of the results to the inclusion of joiner teachers, and results are shown in Appendix D. Data provided by the sites was not sufficient to identify student joiners and exclude them from the analysis.

Representativeness

As an RCT with low cluster-level attrition, but a high risk of bias due to joiners, the study must demonstrate the representativeness of students within non-attrited schools. Representativeness is calculated in the same way as attrition, except that the reference sample is the eligible individuals (students) in non-attrited clusters (schools) at follow-up. Students with missing baseline or outcome

data are included as attrition. For that reason, the total number of students included in the numerator (students with baseline and outcome scores) is lower than the number of students in the analysis, where baseline achievement was imputed. The overall student-level attrition with respect to eligible students within non-attrited schools at follow-up was 36.6 percent, with differential attrition of 0.9 percentage points in mathematics. In ELA the overall student-level attrition with respect to eligible students within non-attrited schools at follow-up was 34.4% with differential attrition of 2.7 percentage points (Exhibit 22). These are both within the WWC's "cautious boundary." Thus, the students with complete baseline and outcome data are representative of clusters at follow-up (WWC, 2022). See Appendix G for the representativeness of the sample including cases with imputed baseline data.

Exhibit 22. Representativeness of clusters at follow-up – student achievement

Mathematics				
	Treatment	**Control**	**Total**	**Differential**
Eligible students in non-attrited schools	4,321	3,448	7,769	
Students with baseline and outcome scores	2,758	2,170	4,928	
Percent of students attrited from non-attrited schools	36.2%	37.1%	36.6%	0.9 points
		Met attrition standards		
ELA				
	Treatment	**Control**	**Total**	**Differential**
Eligible students in non-attrited schools	5,163	3,730	8,893	
Students with baseline and outcome scores	3,444	2,387	5,831	
Percent of students attrited from non-attrited schools	33.3%	36.0%	34.4%	2.7 points
		Met attrition standards		

Methods

For the student achievement analysis, the SRI evaluation team collected students' scaled scores on their state's assessment in mathematics and ELA and standardized the students' scores to their district mean and standard deviation (Hayes Park, Fillmore County, Chesterton, and Taftville) or to the sample mean and standard deviation, where district statistics were not available (Garfield City). SRI fit the following three-level models with students nested within teachers, and teachers nested within schools:

$$Achievement_{ijk} = \gamma_{000} + \gamma_{001}\mathrm{Treatment_k} + \gamma_{002}D_\mathrm{k} + \gamma_{003}S_\mathrm{k} + \gamma_{010}Z_{jk} + \gamma_{100}X_{ijk}$$

$$+ \gamma_{003}\mathrm{S_k} * \mathrm{D_k} + \gamma_{020}Z_{jk} * \mathrm{D_k} + \gamma_{200}\mathrm{X_{ijk}} * \mathrm{D_k} + u_{00k} + r_{0jk} + e_{ijk}$$

Where $Achievement_{ijk}$ is the standardized math or ELA score for student *i* with teacher *j* in school *k*. The treatment indicator is entered at the school level, and the model includes a vector of site fixed effects

(D_k), school-level covariates (S_k), teacher-level covariates (Z_{jk}), and student-level covariates (X_{ijk}), including the student's prior-year test score. Each school, teacher, and student-level indicator was also interacted with the site indicators, to account for differential relationships between these predictors and the outcome across sites (S_k* D_k, Z_{jk}* D_k, and X_{ijk}* D_k) The full set of school, teacher and student covariates is shown in Exhibit 23.

Exhibit 23. Covariates included in the model estimating the impact of NTC on student achievement

	Control variable
School-level variables	Percent of students in poverty
	Percent of students who are English learners
	Percent of students whose race/ethnicity African American and/or Hispanic
	Randomization block
Teacher-level variables[21]	Teacher race/ethnicity is African American and/or Hispanic
	Teacher gender is female
Student-level variables	Prior year test score
	Student race is African American, non-Hispanic
	Student ethnicity is Hispanic, any race
	Student gender is female
	Receives special education services
	Designated as an English learner
	Eligible for free or reduced-price meals
	Grade level

After excluding students with no valid outcome scores, approximately 34 percent of math students and 33 percent of ELA students were missing data on one or more control variables included in the models. Multiple imputation filled in these missing values, including baseline scores. The imputation model included the outcome scores and all the covariates included in the impact model, including an indicator for treatment status. Missing outcome scores were not imputed.[22] Students missing baseline data were counted as attrited in the above attrition analysis, but they were included with imputed data in the models.

Results

Overall, there was no statistically significant impact of NTC induction supports on student achievement in either mathematics or ELA (Exhibit 24). Although the coefficients in both cases are positive, they are

[21] SRI also planned to include teacher certification and highest degree earned, but the data on these variables at baseline were unreliable. Both variables may change over time, and most sites did not store historical data on teachers' status as of their hire date. See Appendix D for models that include the teacher certification data that sites did provide.

[22] As a sensitivity test, SRI also ran the models using casewise deletion to eliminate all records missing values on any independent variables. The conclusions were consistent with those reported here. See Appendix D.

small, and the variation around these estimates is large, so they are not statistically distinguishable from zero.

Exhibit 24. Impact of the NTC model on student mathematics and English language arts achievement

	Impact	*p* value	Pooled *SD*	Effect size	*n* schools	*n* teachers	*n* students
Math	0.02	0.74	0.97	0.02	107	142	7,012
ELA	0.03	0.50	0.98	0.05	117	165	8,260

See Chapter 4 for additional exploratory analyses using student achievement test scores as the outcome. Also see Appendix D for the findings of additional sensitivity tests on the main student achievement outcomes.

Impacts on Teacher Retention in Instructional Positions

NTC induction supports are also designed to improve the retention of teachers in instructional positions in their districts into the third year of teaching. This section presents the results of the analysis examining the impact on teacher retention after the two full years of induction support. The outcome is measured using Human Resources records. Teachers are defined as "retained" if they continued in the district for a third year of teaching.

Sample, Attrition, and Representativeness

The sample for the retention analysis is the full sample of study teachers, as described in the introduction. For this analysis, SRI collected data from the sites on whether study teachers were still employed in instructional positions in the site as of the fall of their third year of teaching (October 1, 2018 for Cohort 1 and October 1, 2019 or Cohort 2 teachers). Teachers attrited from this analysis only if the district was unable to provide any information on the teacher's employment status as of those dates. All sites were able to provide retention data on all teachers in the study. Thus, this analysis had no attrition at either the school or teacher level.

As in the above analyses, this analysis includes joiner teachers. Thus, the study must demonstrate the representativeness of teachers within non-attrited schools. As all eligible teachers and schools were included in the analysis, the representativeness criterion is also met.

Methods

To examine the impact of NTC induction supports on teacher retention, SRI fit the following two-level model, with teachers nested within schools:

$$Retained_{jk}^{*} = \gamma_{00} + \gamma_{01}Treatment_k + \gamma_{02}S_k + \gamma_{03}D_k + \gamma_{10}Z_{jk} + \gamma_{04}S_k * D_k + \gamma_{20}Z_{jk} * D_k + r_{jk} + u_{0k}$$

Where $Retained_{jk}*$ represents the latent, or underlying, propensity for a teacher to be retained in an instructional position within the site (a binary outcome). As in other analyses, this model controls for the blocking variables used in randomization, school- and teacher-level demographics, site-level fixed

effects, and interactions between the school and teacher variables and the site indicators. The control variables are the same as shown in Exhibit 18, except teachers' baseline observation scores are replaced with two measures of baseline equivalence recognized by the WWC for teacher retention outcomes: the average attendance rate of the teacher's students at baseline, and the school's baseline teacher retention rate (WWC, 2019).

Results

After controlling for differences between teachers and schools, the estimated retention rate for treatment teachers was 81 percent, compared with 79 percent among control teachers (Exhibit 25). This difference was not statistically significant.

Exhibit 25. Impact of the NTC model on teacher retention

Control	Treatment	Impact estimate in log odds	*p* value	*n* schools	*n* teachers
79%	81%	0.11	0.60	299	795

See Appendix E for the findings of additional supplementary analyses on the teacher retention outcomes.

Summary of Impact Findings

Overall, there were no significant differences between the treatment and control conditions in teachers' overall instructional practice, student achievement in mathematics or ELA, or teacher retention in instructional positions. While the direction of the impact estimates was consistently positive, the estimates were not statistically significant. However, when examining teacher practice on a more detailed level, by looking at FFT component-level scores, there was a positive impact of NTC induction supports on teachers' practice in the Communicating with Students component. The effect size of 0.24 is a moderate impact in education.

Chapter 4. Exploratory Findings

In addition to the confirmatory research questions addressed in Chapter 3, SRI investigated three exploratory research questions, digging deeper into variation in students' achievement by variables other than the binary treatment indicator. The three exploratory research questions addressed in this chapter are:

(6) Is effective instruction related to student achievement? If NTC induction supports have impacts on teacher instructional practices and student achievement (as examined by questions 3 and 4), does effective instruction mediate the NTC effect on student outcomes?

(7) Are higher levels of mentoring (e.g., in terms of frequency and instructional focus) related to higher student achievement?

(8) Is the effect of NTC induction supports on student achievement moderated by any school- or student-level variables?

Relationship Between Effective Instruction and Student Achievement

In the confirmatory analyses, SRI found an impact of NTC induction supports on one component of the FFT: Communicating with Students. While this is an important finding, there was not a link to the more distal outcome hypothesized in the logic model: changes in students' mathematics or ELA achievement. This could be because changing teacher practice in only one area was too weak to mediate student achievement. An alternative explanation might be that the one FFT measure that improved is not associated with improvements in student achievement. To check the plausibility of the alternative explanation, SRI examined whether the FFT measures of effective instruction, including Communicating with Students, were significantly associated with student achievement.

Sample and Methods

The sample for this analysis is teachers who were both observed and had students tested in math or ELA in grades 4 through 8. The SRI evaluation team fitted models of the same form used in the main student achievement analysis, replacing the binary treatment indicator with continuous measures of instruction, to test whether these measures of instruction are associated with student achievement, regardless of whether a teacher is in a treatment or a control school, or has received NTC induction support or not. One model was estimated for each of the eight components and two domains, as well as one for the overall teacher practice measure.

Results

There were positive relationships between nearly all the FFT measures, including Communicating with Students, and student achievement in both ELA and mathematics (Exhibit 26). These results indicate that the FFT has predictive validity—higher scores on the FFT are associated with higher student achievement. However, these analyses cannot be used to infer a causal relationship between FFT practices and student achievement.

Exhibit 26. Association between FFT components and student achievement

	Mathematics	ELA
Environment of Respect and Rapport	0.07	0.13***
Establishing a Culture for Learning	0.09	0.15**
Managing Classroom Procedures	0.25***	0.14**
Managing Student Behavior	0.08*	0.11***
Communicating with Students	0.23**	0.23***
Questioning and Discussion Techniques	0.12**	0.09*
Engaging Students in Learning	0.18***	0.19***
Using Assessment in Instruction	0.13*	0.08
Classroom Environment Domain	0.12*	0.18***
Instruction Domain	0.19**	0.21***
Overall teacher practice	0.16**	0.22***
n schools	75	82
n teachers	87	98
n students	4,652	5,540

Note. The sample size for this analysis is smaller than in the student achievement analysis because not all teachers who taught math and ELA in the second year also taught those subjects (and thus were eligible for observation) in their first year. Additionally, some observed teachers attrited from the study before testing at the end of their second year.
*$p < 0.05$ ** $p < 0.01$, ***$p < 0.001$.

Evidence of the predictive validity of the FFT in its association with student achievement indicates that the mismatch in the findings was not due to a mismatch in the instruments used. Instead, it may be that the student achievement outcome was too distal for the program to impact in the limited time frame of the study. Alternatively, the impact on teacher practice may have been too small or narrow (i.e., improving only one component of teacher practice) to translate into detectable changes in student achievement.

This conclusion is further supported by the evidence from the achievement models, combined with the coefficients reported here. With the observed sample and variances in the achievement model, the smallest impact of NTC induction supports on student achievement that could be detected was 0.09 in math and 0.08 in ELA. To achieve these effect sizes through the mechanism of improving teachers' Communicating with Students scores, NTC induction supports would have to improve those FFT scores

by 0.39 points in math (0.09 divided by 0.23) and 0.35 points in ELA (0.08 divided by 0.23). However, the actual impact on Communicating with Students was only 0.10 points. Thus, NTC induction supports as implemented in this study did not have a strong enough impact on teachers' practice to influence student achievement through this path.[23] This analysis did not investigate a full mediation model, however, because there was no overall impact on student achievement.

Relationship Between Implementation Fidelity and Student Achievement

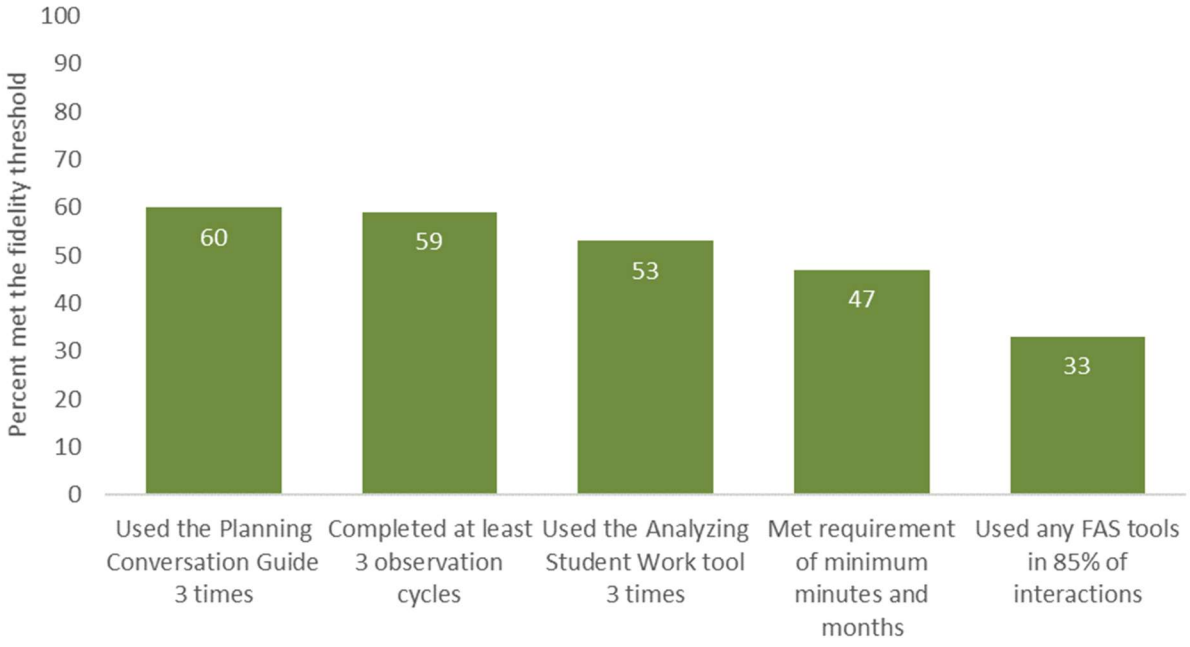
As discussed in Chapter 2, the NTC induction model was not implemented with fidelity in any of the five sites in any year, which may have contributed to the lack of impact discussed in Chapter 3, particularly on the more distal outcome of student achievement. However, although implementation was inadequate overall, there was variation in the mentoring teachers received, in particular in relation to the component that NTC hypothesized would mediate the impact of the program on teacher practice: the provision of high-quality mentoring. Therefore, as a further test of whether elements of the NTC induction model are associated with student achievement outcomes, SRI examined quantitative measures of implementation fidelity in the treatment group. Using variation in the mentoring received at the teacher level, the SRI evaluation team explored whether fidelity to the model as designed was associated with higher student achievement in mathematics or ELA. SRI examined the following indicators, all from the "provision of high-quality mentoring" component of NTC's logic model discussed in Chapter 2:

- Met the threshold for minimum minutes and months of mentoring
- Used formative assessment system (FAS) tools in at least 85 percent of all interactions
- Used NTC's Planning Conversation Guide (PCG) at least three times during the year
- Used NTC's Analyzing Student Learning (ASL) tool at least three times during the year
- Completed an NTC observation cycle at least three times during the year

Exhibit 27 displays the percentages of teachers in the achievement analysis who met the fidelity threshold for each of these indicators. Note that treatment teachers who did not appear in NTC's Learning Zone system were counted as receiving no mentoring. As this graph shows, mentors did not meet the implementation thresholds with all their teachers on any of these indicators, yielding teacher-level variation in these binary variables.

[23] NTC could have an impact on student achievement through other mechanisms besides a change in teacher practice. However, that is beyond the scope of this study as designed.

Exhibit 27. Percent of teachers in the math or ELA analysis who met thresholds for high implementation fidelity on indicators in "provision of high-quality mentoring"



Sample and Methods

To examine the relationship between implementation fidelity and student achievement, SRI used a subset of the sample included in the overall achievement analysis: treatment teachers and their students. SRI fit a series of three-level models, one for each fidelity measure in mathematics and one for each fidelity measure in ELA. The models were of the same form shown in Chapter 3, with the above measures (*FidelityMeasure*$_{jk}$) taking the place of the treatment indicator:

$$Achievement_{ijk} = \gamma_{000} + \gamma_{010}\text{Fidelity Measure}_{\text{jk}} + \gamma_{001}D_{\text{k}} + \gamma_{002}S_{\text{k}} + \gamma_{020}Z_{jk} + \gamma_{100}X_{ijk}$$

$$+ \gamma_{003}\text{S}_{\text{k}} * \text{D}_{\text{k}} + \gamma_{030}Z_{jk} * \text{D}_{\text{k}} + \gamma_{200}\text{X}_{\text{ijk}} * \text{D}_{\text{k}} + u_{00k} + r_{0jk} + e_{ijk}$$

Note that, because the fidelity measures were only available for treatment teachers, these results show the relationship between level of mentoring and student achievement, within the treatment condition only.

Results

There was a strong, statistically significant relationship between measures of implementation fidelity and mathematics achievement (Exhibit 28). Controlling for baseline differences in achievement, mathematics teachers who received at least 180 minutes of mentoring a month for seven months from an NTC-trained mentor saw higher achievement among their students than treatment teachers who received less mentoring. Similarly, mathematics teachers whose NTC-trained mentors completed a Planning Conversation Guide (PCG) tool, Analyzing Student Learning (ASL) tool, and/or an observation cycle with them at least three times during their second year of teaching saw higher achievement

among their students than treatment teachers whose mentors did not complete these tools three times. The broader measure of tool use, whether mentors used any tool (including non-instructionally focused tools) in at least 85 percent of interactions, was not significantly associated with student achievement.

These findings support NTC's hypothesis that these particular mentoring practices are likely to be important for improving student learning. Results show that beginning mathematics teachers whose mentors met NTC's standards for implementing the model with fidelity saw stronger student performance in math than those whose mentors did not. However, these same measures were not significantly related to students' ELA performance.

Exhibit 28. Relationship between dichotomous measures of implementation fidelity and student mathematics achievement

	Mathematics		ELA	
Measure	**Effect size**	***p* value**	**Effect size**	***p* value**
Met minimum minutes and months	0.25***	< 0.001	-0.02	0.74
Used any tool in 85 percent of interactions	0.11	0.06	-0.07	0.20
Used the PCG tool at least three times	0.22**	0.001	-0.06	0.39
Used the ASL tool at least three times	0.31***	< 0.001	-0.11	0.14
Completed an observation cycle at least three times	0.18*	0.01	-0.06	0.31
n schools	56		60	
n teachers	73		89	
n students	3,801		4,754	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

It is important to note that these results are correlational only. Teachers were not randomly assigned to different levels of mentoring. In fact, the reasons for the level and type of mentoring teachers received may have been related to their students' achievement or their own effectiveness as teachers. Therefore, this analysis does not establish that higher levels of mentoring caused improved student achievement.

Relationship Between Survey Measures of Mentoring Practice and Student Achievement

In addition to overall inadequate implementation of the NTC induction model in the treatment condition, there was also a low level of contrast between the mentoring supports treatment and control teachers received. Under business-as-usual conditions, a high percentage of beginning teachers in all five study sites were assigned a mentor in their first year in the classroom (see Exhibit 10 in Chapter 2). Further, there was variation in the dosage and types of mentoring teachers received in both the treatment and control groups. Thus, SRI examined the relationship between student achievement and a set of binary measures of mentoring based on survey responses across both treatment and control teachers. These survey measures reflect that not all treatment teachers received the full intervention,

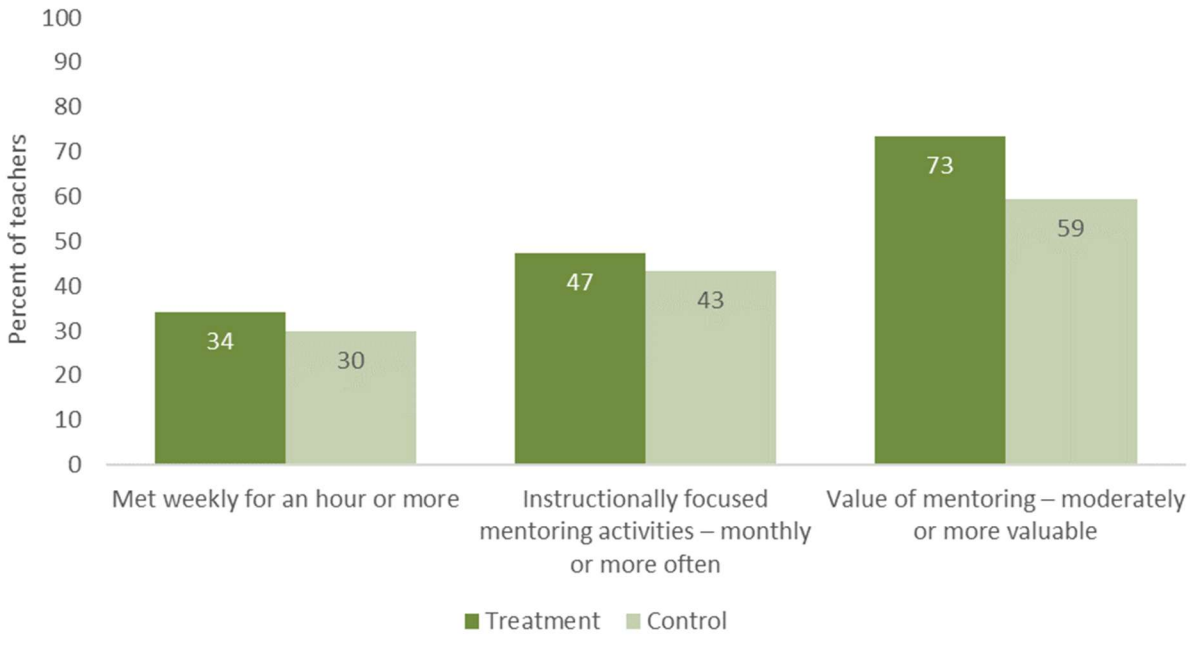
and some control teachers also received induction supports that met NTC's standards for fidelity to their model. For this analysis, SRI examined a set of three measures created from teacher survey items:

- Frequency and duration of mentoring (met with a mentor weekly for an average of an hour or more/did not meet weekly for an average of an hour or more)
- Frequency of instructionally focused mentoring activities (monthly or more often/less than monthly)
- Perceived value of mentoring activities (moderately or extremely valuable/not valuable or minimally valuable)

As discussed in Chapter 2, instructionally focused mentoring activities were those identified by NTC in their logic model as practices that were likely to lead to improvements in teacher practice and student achievement. These included observing teachers' classrooms and providing feedback, analyzing samples of student work, and planning lessons together. See Appendix B for the full list of survey items that contributed to the scales used here.

Exhibit 29 displays the percentages of teachers in the achievement analysis who reported receiving the recommended frequency and duration of mentoring and the level of instructionally focused mentoring prescribed by the NTC model, and that moderately or highly valued the mentoring they received. As this graph shows, teachers varied in the amount, type, and perceived value of mentoring they received, in both the treatment and the control conditions.

Exhibit 29. Percent of teachers in the math or ELA analysis who reported a high level or value of mentoring



Sample and Methods

Because this sample relies on teachers' survey responses, the sample is the subset of teachers in the main achievement analysis who also responded to the survey.

Once again, the analysis used a series of three-level models, one for each survey measure in mathematics and one for each survey measure in ELA. The models were of the same form shown in Chapter 3, with the above measures (*Survey Measure*$_{jk}$) taking the place of the treatment indicator:

$$Achievement_{ijk} = \gamma_{000} + \gamma_{010}\text{Survey Measure}_{\text{jk}} + \gamma_{002}D_{\text{k}} + \gamma_{003}S_{\text{k}} + \gamma_{020}Z_{jk} + \gamma_{100}X_{ijk}$$

$$+ \gamma_{004}\text{S}_{\text{k}} * \text{D}_{\text{k}} + \gamma_{030}Z_{jk} * \text{D}_{\text{k}} + \gamma_{200}\text{X}_{\text{ijk}} * \text{D}_{\text{k}} + u_{00k} + r_{0jk} + e_{ijk}$$

Note that, because the survey measures replace the treatment variable, these results show the relationship between level or value of mentoring and student achievement, across both the treatment and control condition.

Results

There was no statistically significant relationship between the duration and frequency of mentoring or the value of mentoring and student achievement (Exhibit 30). However, there was a significant relationship between instructionally focused mentoring activities and student mathematics achievement. This supports NTC's hypothesis that having an induction mentor, even if they meet often and are highly valued by their beginning teacher, may not be enough to have an impact on student achievement. Rather, a focus on instruction may also be required. This may be why beginning mathematics teachers whose mentors focused on instructional practice at least monthly saw stronger student performance than those whose mentors put less emphasis on instruction.

Exhibit 30. Relationship between dichotomous measures of mentoring practice and student achievement

Mathematics					
Measure	Effect size	*p* value	*n* schools	*n* teachers	*n* students
Met with a mentor weekly for an hour or more	0.02	0.67	90	115	5,746
Instructionally focused mentoring monthly or more	0.12*	0.04	69	85	4,655
Mentoring was moderately or more valuable	0.09	0.14	70	86	4,660
English Language Arts					
Measure	Effect size	*p* value	*n* schools	*n* teachers	*n* students
Met with a mentor weekly for an hour or more	0.00	0.97	96	130	6,550
Instructionally focused mentoring monthly or more	0.02	0.64	69	90	4,668
Mentoring was moderately or more valuable	0.03	0.54	72	94	4,754

Note. Sample sizes vary due to item-level nonresponse on the survey.
*$p < 0.05$.

As in the previous section, these results are correlational only. Teachers were not randomly assigned to different types of mentoring. In fact, the reasons for the level and type of mentoring teachers received may have been related to their students' achievement or their own effectiveness as teachers. Therefore, this analysis does not establish that higher levels of mentoring caused higher student achievement.

Moderation Effects of Student and School Characteristics

Finally, SRI examined whether the impact of NTC induction supports was stronger in some types of schools or with some groups of students. If NTC's impact is stronger with some populations than in others, the overall impact estimates above may mask these differences. In particular, given NTC's focus on improving educational equity for historically underserved students, the SRI evaluation team hypothesized that there may be a stronger impact with these students or in schools with high concentrations of these students.

Sample and Methods

The sample for this analysis is identical to the overall achievement sample described in Chapter 3. This analysis tested moderation effects of eight student-level variables:

- Student grade level (elementary vs. middle grades)
- Student race (White vs. students of color)
- Student gender (female vs. non-female)
- Student receives free or reduced-price meals[24] (yes/no)
- Student receives special education services (yes/no)
- Student designated as an English learner (yes/no)
- Student's prior achievement was below average for the sample (yes/no)
- Student's prior achievement was below the 25th percentile for the sample (yes/no)

The analysis also tested the moderation effects of five school-level variables:

- School above average percent receiving free or reduced-price meals (yes/no)
- School above average percent English learners (yes/no)
- School above average percent Black or Hispanic[25] students (yes/no)
- School above average percent proficient on state assessments at baseline (yes/no)

School averages were based on publicly available data. They were dichotomized to indicate whether the school was above or below the average for study schools in their site. In other words, a school with "above average" percent receiving free or reduced-price meals had a higher percentage of students receiving free or reduced-price meals than the other study schools in their site. For each variable, SRI

[24] SRI did not have access to a more accurate indicator of students experiencing poverty.
[25] This report uses the term "Hispanic," as opposed to "Latino" or "Latinx," to reflect the term used in the sites' extant data.

fitted the same three-level models used for the main impact analysis, with the addition of an interaction term between the variable of interest and treatment status to test the moderation effect of the variable.

Results

There was no significant moderation effect of any student-level variables on the relationship between NTC induction supports and student achievement in mathematics or ELA. In other words, the lack of a relationship between NTC and student achievement held for students in elementary and middle grades, White students and students of color, and all other student groups listed above. Likewise, there was no significant moderation effect of any school-level variables on NTC's impact in mathematics. See Appendix F for the results of these models.

However, in ELA, there was a positive impact of NTC induction supports on student achievement in higher poverty schools (schools with an above-average concentration of students receiving free or reduced-price meals), with a moderate effect size of 0.12 standard deviation (Exhibit 31). The same impact was not present in lower-poverty schools (those with below-average concentrations of students receiving free or reduced-price meals). The difference in impacts between the higher poverty and lower poverty schools was statistically significant.

Likewise, in schools with an above-average concentration of English learner students, there was a positive impact of NTC induction supports on ELA achievement, with a moderate effect size of 0.14 standard deviation (Exhibit 31). The same impact was not present in schools with below-average concentrations of English learner students. The difference in impacts between schools with above average and below average concentrations of English learner students was statistically significant.

There was no significant moderation effect of the concentration of Black or Hispanic students or of school-level baseline achievement on the impact of NTC on either math or ELA achievement. See Appendix D for both these and the student-level moderation results.

Exhibit 31. Moderation effect of student and school characteristics in ELA

School percent free/reduced-price meals	Impact	*p* value	*n* schools	*n* teachers	*n* students
At or below average	-0.08	0.24	41	52	3,420
Above average	0.12*	0.02	76	113	4,840
Difference in impacts	0.20*	0.03	117	165	8,260
School percent English learners	**Impact**	***p* value**	***n* schools**	***n* teachers**	***n* students**
At or below average	-0.05	0.35	63	79	4,264
Above average	0.14*	0.01	54	86	3,996
Difference in impacts	0.19*	0.02	117	165	8,260

*$p < 0.05$.

Summary of Exploratory Findings

Although the main confirmatory analyses investigated in this study did not find any overall causal relationship between NTC induction supports and student achievement, these exploratory findings support key assumptions in that undergird the NTC model. First, a significant association between the FFT and student achievement indicates that NTC may have been able to affect student achievement if there had been a larger impact on the areas of teachers' practice measured by the FFT. Second, significant associations between instructionally focused mentoring and math achievement suggest the critical importance of this focus when mentoring. Finally, the finding that NTC had a significant impact on ELA achievement in higher poverty schools and schools with higher percentages of English learners suggests that NTC is more beneficial in schools with higher proportions of historically underserved students. While SRI hypothesizes that this difference may be due to differences in hiring, staffing, and the supports available to beginning teachers at these schools, thus making NTC support more valuable and necessary when compared with business-as-usual conditions, the data are not available to test this theory..[26]

[26] The only teacher characteristics collected were teacher certification (full or partial/provisional) and hire date. Neither of these variables was correlated with the school characteristics discussed here.

Chapter 5. Conclusion and Implications

In the 2017–18 school year, approximately 318,000 teachers entered their first or second year of teaching in the United States (National Center for Education Statistics, 2019). Beginning teachers embark on a profession that places particularly challenging demands on novice practitioners. In an effort to support these teachers, NTC has provided induction support to over 25,000 beginning teachers in schools across the country over the past 20 years.

After finding positive impacts of the NTC induction model on student outcomes in a prior i3 validation study, NTC hoped to scale up its supports to a broader range of sites, allowing adaptations to its approach in an effort to improve flexibility and reduce costs. This scale-up study was designed to examine the implementation of that revised model and its impact on teachers and students.

In contrast to the prior validation study, SRI found an overall inadequate level of implementation, indicating that the scale-up model was not implemented with fidelity in any of the five sites across the three years of the study. Likewise, SRI found that the mentoring received by NTC treatment teachers was not substantially different, in many ways, from the mentoring received by teachers in the business-as-usual condition. Finally, SRI found no impact of the model as implemented on overall teacher practice, student achievement, or teacher retention.

The low implementation fidelity, minimal contrast between treatment and business-as-usual, and no impact on teachers and students are all disappointing results. SRI cannot say for certain whether NTC's adapted model would have had an impact if it had been implemented with fidelity, or if it had been implemented at the same level in districts with less robust mentoring programs of their own.

However, the exploratory findings discussed here do shed some light on the conditions under which NTC might see a greater impact. First, there was a correlational relationship between duration and frequency of mentoring among treatment teachers and student achievement, but no corresponding relationship when control teachers were included. This indicates that more time with a mentor is not inherently better, but when that time is spent with an NTC-trained mentor, it may be related to student achievement. Second, the correlational relationship between instructionally focused mentoring and math achievement, both among treatment and control teachers, supports NTC's hypothesis that the *content of mentoring* matters. Teachers whose mentors focused on analyzing student work, observing their classrooms and providing feedback, and adjusting instruction for diverse learners saw higher math achievement among their students than teachers whose mentors did not have this instructional focus.

Finally, the results of the moderation analysis indicate that there is a positive impact of NTC induction supports on student ELA achievement in schools with higher proportions of historically underserved students. Given NTC's focus on educational equity, this result is promising.

Thus, the findings of this study indicate the importance of ensuring high-quality implementation of a program and the consequences of failure to do so. Under its i3 validation grant, NTC used substantial grant resources to ensure that the intensive induction model was implemented as intended, including funding full-time induction mentor positions and hiring staff to fill those positions through a highly selective process. Under those conditions, the model was implemented with fidelity and produced moderate, positive impacts on student achievement (Young et al., 2017). Under the i3 scale-up grant,

NTC attempted to adapt that model for scaling, but failed to fully implement key components and mediators as intended. There is some evidence that the model has promise when fully implemented, particularly in schools with higher proportions of historically underserved students, but without further research, this evidence is simply suggestive.

References

Boulay, B., Goodson, B., Olsen, R., McCormick, R., Darrow, C., Frye, M., Gan, K., Harvill, H., & Sarna, M. (2018). *The Investing in Innovation Fund: Summary of 67 evaluations: Final report* (NCEE 2018-4013). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Danielson, C. (2013). *The framework for teaching evaluation instrument: 2013 edition*. The Danielson Group.

Goldhaber, D., Quince, V., & Theobald, R. (2018). Has it always been this way? Tracing the evolution of teacher quality gaps in U.S. public schools. *American Educational Research Journal*, *55*(1), 171–201. https://doi.org/10.3102/0002831217733445

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*(3), 499–534. https://doi.org/10.3102/00028312038003499

Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research*, *62*(2), 129–169. https://doi.org/10.3102/00346543062002129

Kini, T., & Podolsky, A. (2016). *Does teaching experience increase teacher effectiveness? A review of the research*. Learning Policy Institute. https://learningpolicyinstitute.org/our-work/publications-resources/does-teaching-experience-increase-teacher-effectiveness-review-research

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M.D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013–3000). U.S. Department of Education, Institute of Education Sciences, National Center for Special Education Research. http://ies.ed.gov/ncser

National Center for Education Statistics. (2019). *Digest of Education Statistics 2018*, Table 208.20. Public and private elementary and secondary teachers, enrollment, pupil/teacher ratios, and new teacher hires: Selected years, fall 1955 through fall 2028. https://nces.ed.gov/programs/digest/d18/tables/dt18_208.20.asp

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4–36.

Scafidi, B., Sjoquist, D., & Stinebrickner, T. (2007). Race, poverty, and teacher mobility. *Economics of Education Review*, *26*(2), 145–159. https://doi.org/10.1016/j.econedurev.2005.08.006

Smith, T., & Ingersoll, R. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Education Research Journal*, *41*(3), 681–714.

Snyder, J., & Bristol, T. J. (2015). Professional accountability for improving life, college, and career readiness. *Education Policy Analysis Archives, 23*(16). http://files.eric.ed.gov/fulltext/EJ1070474.pdf

What Works Clearinghouse (WWC). (2019). *Review protocol for teacher excellence, version 4.0 (May 2019)*. https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/Wwc_te_protocol_v4.0_508.pdf

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). This report is available on the What Works Clearinghouse website at https://ies.ed.gov/ncee/wwc/Handbooks

Young, V. M., Schmidt, R., Wang, H., Cassidy, L., & Laguarda, K. (2017). *A comprehensive model of teacher induction: Implementation and impact on teachers and students* (Evaluation of the New Teacher Center's i3 Validation Grant, Final Report. Prepared for the New Teacher Center). SRI International.

Appendix A. Implementation Fidelity

Exhibit A-1. Fidelity of implementation matrix and thresholds

Component	Indicator	Site-level threshold for adequate fidelity
1: New Teacher Center supports	1a. Program leads develop and advocate for sanctioned time for mentoring	Present
	1b. Principals attend Role of the School Leader training	80 percent of principals attend
	1c. Program leads engage principals in annual one-on-one meetings	80 percent of principals met with the program lead
	1d. Program leads build sustainability for a district-led induction mentoring program after NTC involvement by attending all trainings	Program lead attended all trainings
	1e. Program leads build sustainability for a district-led induction mentoring program after NTC involvement by co-facilitating all mentor forums	Program lead facilitated all mentor forums
	1f. Development of standards, tools, training materials, and an online mentoring platform	All elements developed
2: Mentor assignment	2a. Mentors released from teaching assignments and/or provided with appropriate sanctioned time for mentoring	80 percent of mentors released full-time from teaching responsibilities or received sanctioned time for mentoring
	2b. Appropriate ratio for mentor caseloads	80 percent of mentors' caseloads meet requirements for their amount of release time

Exhibit A-1, continued. Fidelity of implementation matrix and thresholds

Component	Indicator	Site-level threshold for adequate fidelity
3: Mentor development and accountability	3a. Mentor retention in school-based sites	80 percent of mentors retained in their position of the full year
	3b. Mentors participate in Mentor Professional Learning Series (PLS)	80 percent of mentors present for at least 80 percent of offered PLS days
	3c. Site holds the number of Professional Learning Series days specified by NTC	Site held at least 90 percent of specified PLS days
	3d. Mentors participate in mentor forums	80 percent of mentors present for at least 80 percent of offered mentor forum days
	3e. Site holds the number of mentor forums specified by NTC	Site held at least 90 percent of specified mentor forum days
	3f. Mentors receive support and feedback from program leads	80 percent of mentors met with program lead for required number of observations and one-on-one meetings
	3g. Mentors engage in peer coaching and goal-setting process	80 percent of mentors completed all activities

Exhibit A-1, concluded. Fidelity of implementation matrix and thresholds

Component	Indicator	Site-level threshold for adequate fidelity
4: Provision of high-quality mentoring	4a. Mentors meet regularly with beginning teachers	80 percent of mentors met with each of their beginning teachers for a minimum of 180 minutes per calendar month, over a minimum of seven (7) months between October and May. Two thirds (120 minutes) of meetings must be face-to-face; up to one third of meetings (60 of 180 minutes) may be virtual
	4b. Mentors used any of NTC's formative assessment system tools, excluding the Administrator-Mentor Collaboration log	80 percent of teachers' mentors used a Formative Assessment System tool for at least 85 percent of their interactions
	4c. Mentors meet minimum expectations for use of key tools – Planning Conversation Guide (PCG)	80 percent of teachers' mentors used the PCG at least three times during the school year
	4d. Mentors meet minimum expectations for use of key tools – Analysis of Student Learning (ASL)	80 percent of teachers' mentors used the ASL at least three times during the school year
	4e. Mentors meet minimum expectations for use of key tools – Observation cycle	80 percent of teachers' mentors used the observation cycle tool at least three times during the school year
	4f. Mentors meet regularly with school leadership	80 percent of mentors met with school leadership with the required frequency
	4g. Teachers agree that "The new teacher support program is committed to supporting equitable and inclusive learning environments"	80 percent of teachers indicated that they agree or strongly agree with the survey item
	4h. Beginning teachers find the support that they receive from their NTC mentors to be valuable	80 percent of teachers report that they found the mentoring they received "moderately" or "extremely" valuable

Exhibit A-2. Site-level fidelity by indicator

		Hayes Park			Garfield City			Fillmore County			Chesterton			Taftville		
Com-ponent	Indi-cator	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
1	1a	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	1b	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	~	✗	~	✗	✗
	1c	✓	✓	✓	✓	✓	✓	✓	~	✓	✗	✗	✗	~	✗	✗
	1d	✓	✗	✓	✓	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗
	1e	N/A	✓	✓	N/A	✓	✓	N/A	✓	✓	N/A	✓	✓	N/A	✗	✓
	1f	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fidelity		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No
2	2a	✓	✓	✓	✗	~	~	✗	~	✓	✗	✗	~	✓	✓	✓
	2b	✗	✓	✓	~	✓	✓	✓	✓	✓	~	✓	✓	~	✓	✗
Fidelity		No	Yes	Yes	No	No	No	No	No	Yes	No	No	No	No	Yes	No
3	3a	N/A	N/A	N/A	✓	✓	✓	✓	✓	✓	✓	✓	✓	N/A	N/A	N/A
	3b	✓	✓	✓	✓	~	✗	✓	~	✗	✗	✗	✗	✓	✓	✓
	3c	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓
	3d	✓	✓	✓	✗	✗	✗	~	~	~	✓	✓	✓	✓	✓	✓
	3e	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	3f	✓	✗	✗	✓	✓	✗	✓	✓	✓	✗	~	✓	✗	✓	✗
	3g	✓	~	~	✗	~	✗	✓	✓	~	✗	✗	✓	✗	✗	~
Fidelity		Yes	Yes	Yes	No	No	No	Yes	Yes	No	No	No	No	No	Yes	Yes
4	4a	✓	~	✓	✗	✗	✗	✓	~	~	~	✓	✓	✓	✓	~
	4b	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
	4c	✓	✓	✓	✗	✗	✗	✓	✓	~	✓	✓	✓	✓	✓	~
	4d	✓	~	✓	✗	✗	✗	✓	✓	~	✓	✓	✓	✓	✓	✗
	4e	✓	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗
	4f	~	✗	~	✓	✗	✗	✓	✗	✗	✗	✗	✗	✓	✗	~
	4g	~	✗	✗	~	~	~	✗	✓	✓	~	✗	~	✓	~	~
	4h	~	✓	~	~	✓	~	~	✓	~	~	✓	~	✓	✓	~
Fidelity		No	No	Yes	No	No	No	No	No	No	No	No	No	Yes	No	No

Note. At the indicator level, ✓ indicates "High fidelity," ✗ indicates "Low fidelity," and ~ indicates "Medium fidelity." At the component level, sites must receive a score of "high" on *60 percent or more* of individual indicators for that component AND receive a score of "low" on *less than 20 percent* of the indicators to achieve adequate fidelity.

Appendix B. Treatment-Control Contrast Supplementary Information

This appendix presents supplementary information for the survey data discussed in Chapter 2.

Survey Item Text

The following survey questions were administered to both treatment and control teachers in the spring of 2019. The same or similar items were also administered in spring 2017 and spring 2018.

Were you formally assigned a new teacher mentor/coach for the current school year through your school or district?
Yes
No

How often do you meet with your mentor/coach?
Weekly or more often
About twice per month
Monthly
Every few months
A few times per year
Never

Considering all forms of communication (face-to-face, phone, email, text, video chat, etc.) how many minutes do you spend interacting one-on-one with your mentor/coach per week, on average? Please enter a numeric value: _____

Instructionally Focused Mentoring Activities

How often does your mentor/coach do this?
Scale: Never, Once or twice this school year, Every few months, Monthly, About twice per month, Weekly or more often

Help you to understand rigorous state standards*
Discuss with you your mentor/coach's observation data and plan next steps*
Model instruction in your classroom*
Observe you teaching in your classroom and provide you with feedback*
Watch a video recording of your teaching and provide you with feedback
Discuss instructional challenges, problems, and strategies with you*
Plan lessons with you*

Analyze samples of your students' work with you*
Talk to you about the strengths and/or needs of specific students*
Discuss student assessment data to help you make instructional decisions*
Help with your organization
Help you gather resources and instructional materials
Help you to manage job-related stress (e.g., time management, etc.)
Provide emotional support

* Item contributes to the “instructionally focused mentoring activities” scale

Instructional Practice and Data Use

How valuable has your mentor/coach been to your development in the following areas:
Scale: Not valuable, minimally valuable, moderately valuable, extremely valuable, N/A my mentor/coach has not addressed this

Deepening knowledge of your content area(s)*
Analyzing the alignment of students' work with rigorous standards*
Planning lessons that are focused on grade-level standards*
Using high-level questions and prompts to scaffold learning*
Increasing student discussion of each other's thinking*
Requiring students to provide supporting evidence*
Tracking students' academic progress*
Setting goals for individual student achievement*
Designing lesson plans to address gaps in students' learning*
Designing and using formal and informal assessments*

* Item contributes to the overall “Value of Mentoring” scale

Classroom Environment

How valuable has your mentor/coach been to your development in the following areas:
Scale: Not valuable, minimally valuable, moderately valuable, extremely valuable, N/A my mentor/coach has not addressed this

Employing classroom routines, etc. to create a climate of respect and trust
Responding appropriately to student misbehavior
Managing classroom procedures
Using language/actions to convey confidence that all students can meet standards

Differentiation

How valuable has your mentor/coach been to your development in the following areas:
Scale: Not valuable, minimally valuable, moderately valuable, extremely valuable, N/A my mentor/coach has not addressed this

Engaging students with content in multiple ways according to their needs*
Providing multiple ways for students to demonstrate their learning*
Using evidence of student learning to plan instruction*
Working with English Language Learners*
Working with students with identified learning needs*
Using techniques to meet the needs of students from diverse backgrounds*
Addressing issues of equity and inclusivity in your class*

* Item contributes to the "instructionally focused mentoring activities" scale

Socioemotional Learning

How valuable has your mentor/coach been to your development in the following areas:
Scale: Not valuable, minimally valuable, moderately valuable, extremely valuable, N/A my mentor/coach has not addressed this

Helping improve students' ability to give feedback to peers*
Helping improve students' ability to develop a growth mindset*
Helping improve students' ability to respect diverse perspectives*

* Item contributes to the "instructionally focused mentoring activities" scale

Professional Support

How valuable has your mentor/coach been to your development in the following areas:
Scale: Not valuable, minimally valuable, moderately valuable, extremely valuable, N/A my mentor/coach has not addressed this

Demonstrating professionalism (e.g., ethical conduct, compliance with district/school regulations)
Developing positive and collaborative relationships with colleagues
Collaborating with resource personnel to support student learning
Collaborating with families to support student learning
Managing job-related stress
Providing emotional support

Scale Properties

Exhibit B-1 displays the eigenvalues, alpha reliability coefficients, items, and item loadings from factor analyses run on the survey items above before creating scales using a simple average of the non-missing items.

Exhibit B-1. Scale properties of the survey scales discussed in the report

Scale	Eigen-value	Alpha	Items	Item loading
Instructionally focused mentoring	4.6	0.91	Help you understand rigorous state standards	0.78
			Discuss your mentor/coach's observation data	0.84
			Observe you and provide feedback	0.71
			Watch a video of your teaching	0.52
			Discuss instructional challenges	0.75
			Plan lessons with you	0.77
			Analyze samples of your students' work with you	0.81
			Discuss student assessment data	0.83
Perceived value of mentoring	15.1	0.98	Deepening knowledge of your content area(s)	0.85
			Analyzing the alignment of students' work with standards	0.90
			Planning lessons based on standards	0.87
			Using high level questions	0.86
			Increasing student discussion	0.88
			Requiring students to provide evidence	0.87
			Tracking students' progress	0.87
			Setting student goals	0.91
			Designing lesson plans	0.89
			Designing and using assessments	0.88
			Engaging students in multiple ways	0.89
			Providing multiple ways to demonstrate learning	0.89
			Using evidence to plan instruction	0.90

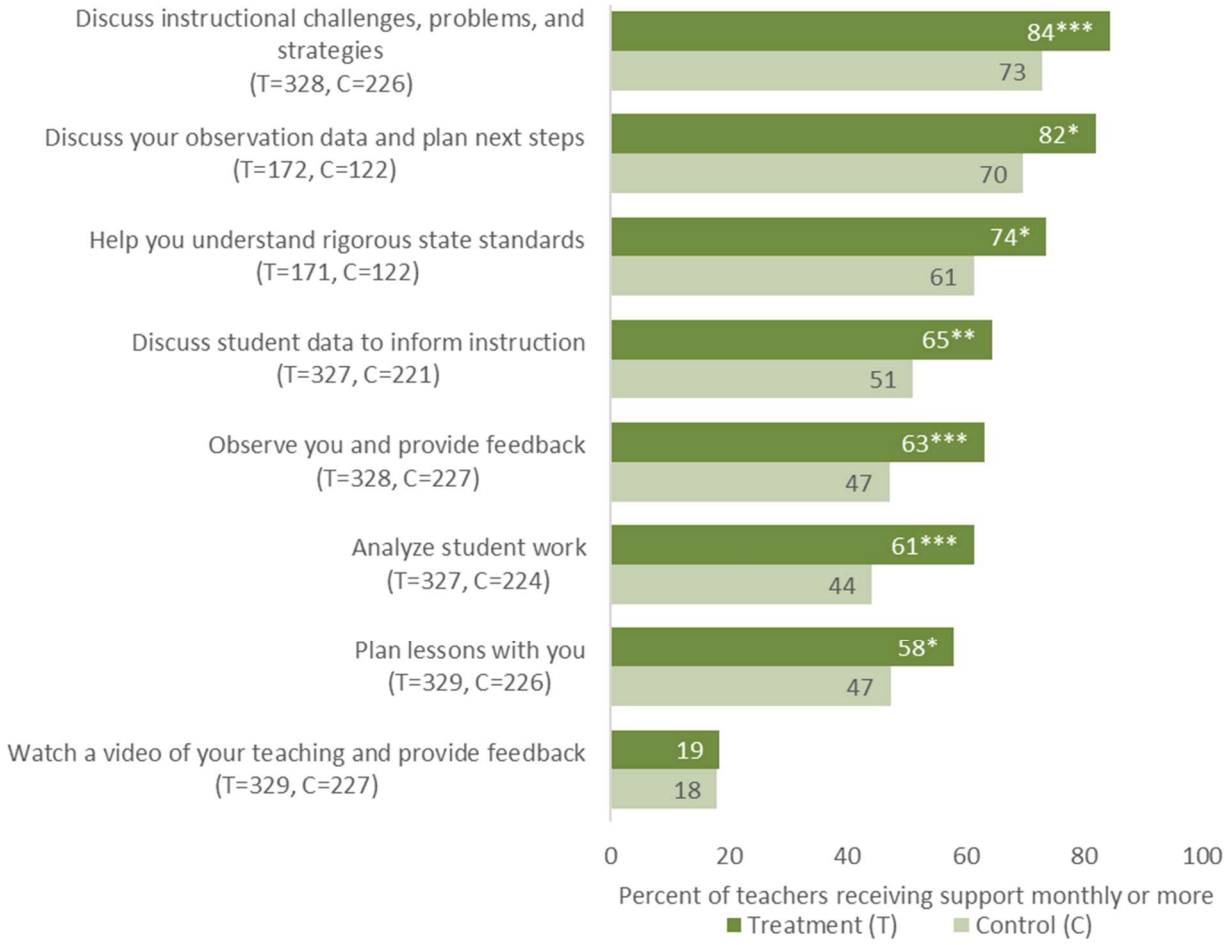
Exhibit B-1, concluded. Scale properties of the survey scales discussed in the report

Scale	Eigen-value	Alpha	Items	Item loading
Perceived value of mentoring (continued)	15.1	0.98	Working with English learners	0.72
			Working with students with identified learning needs	0.82
			Techniques for students from diverse backgrounds	0.87
			Addressing equity	0.85
			Improving active listening	0.87
			Improve students' ability to respect diversity	0.88
			Improve students' growth mindset	0.89

Supplementary Exhibits

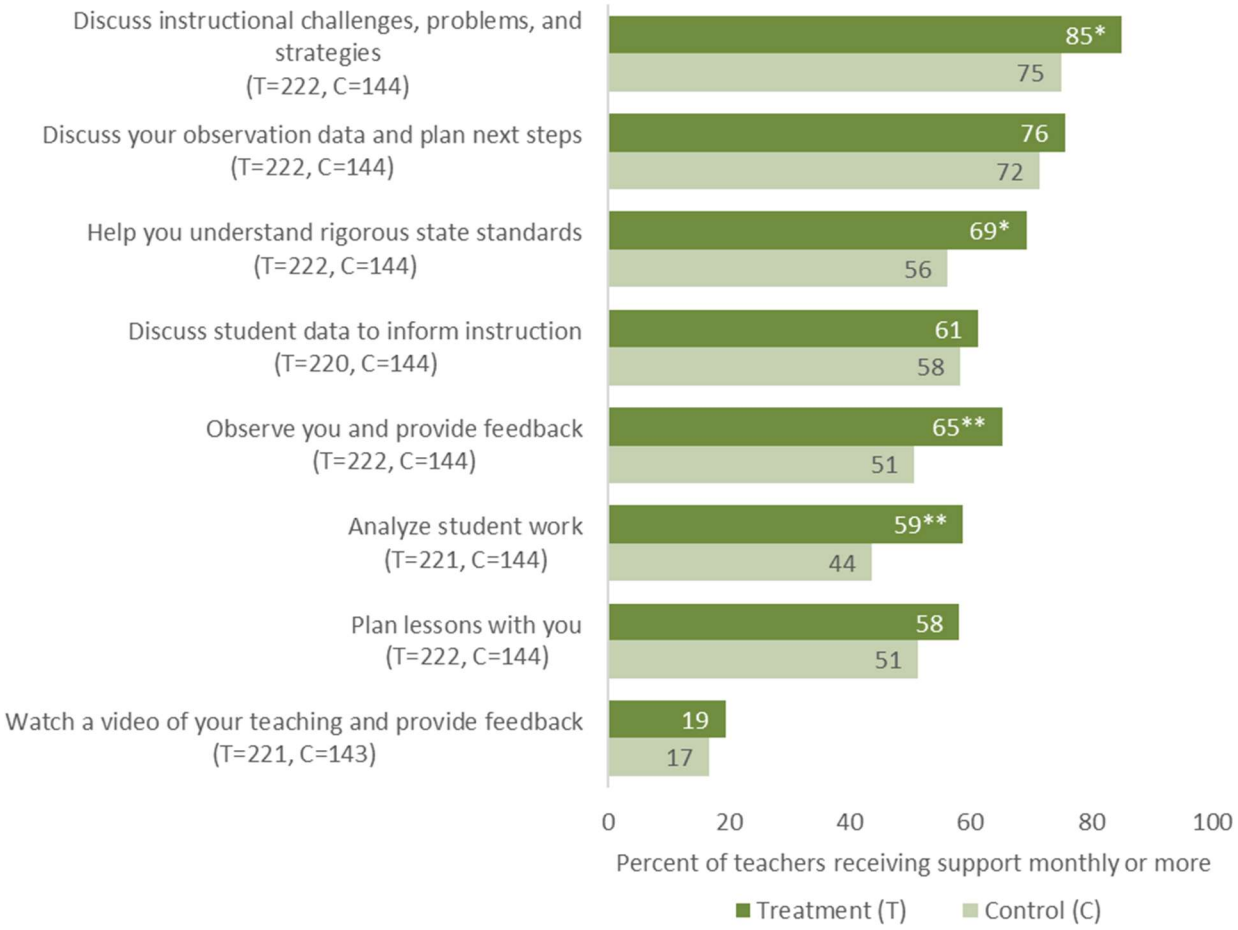
Exhibits B-2 and B-3 provide the item-level differences between treatment and control teachers' responses to the items that make up the "instructionally focused mentoring" scale presented in Chapter 2.

Exhibit B-2. Percent of first-year beginning teachers receiving support monthly or more often, by treatment status



Note. All control and Cohort 2 treatment responses only include teachers who reported that they were formally assigned a mentor or coach. Every Cohort 1 treatment teacher who responded to the survey is included.
*$p < .05$, **$p < .01$, ***$p < .001$.

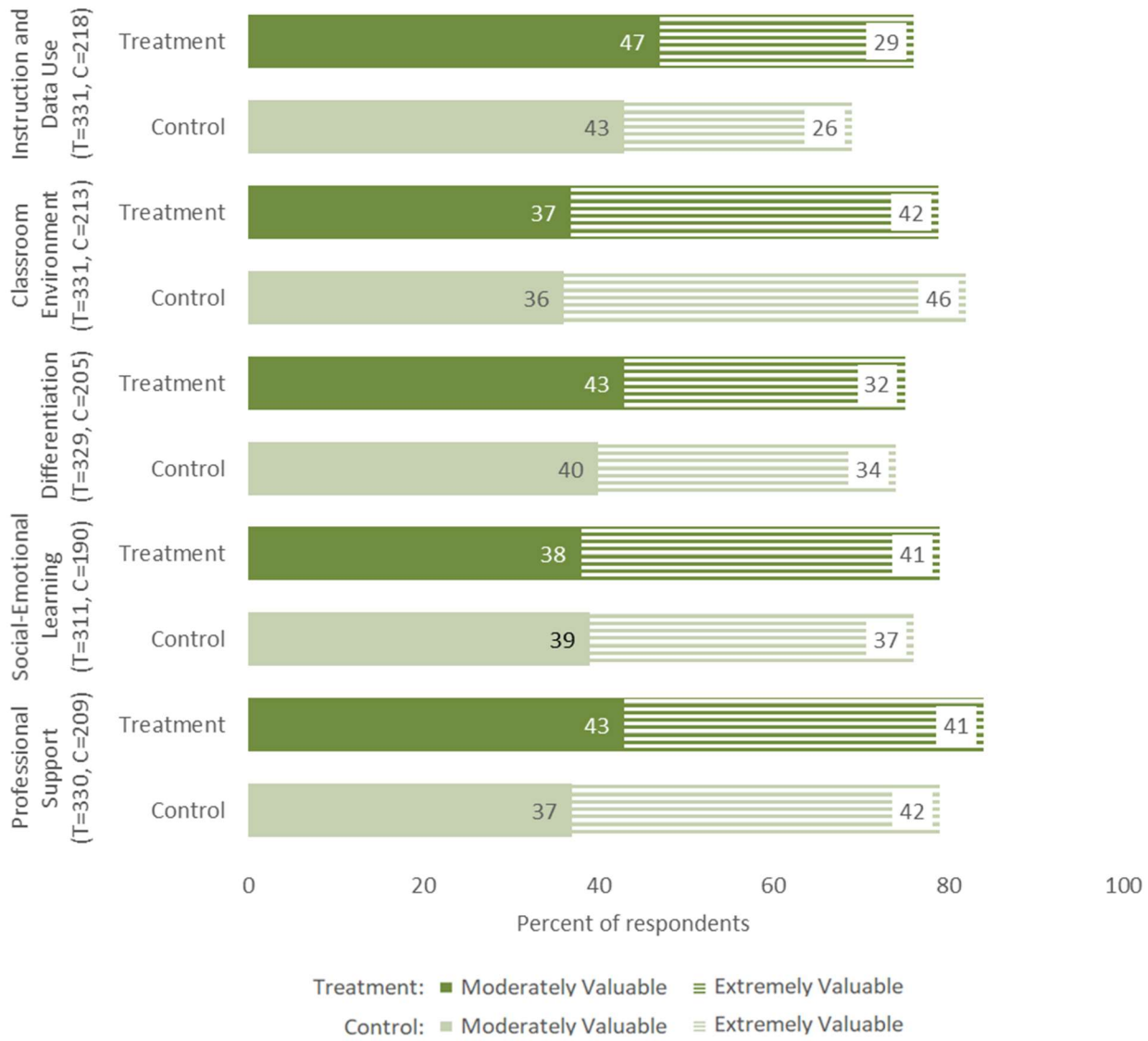
Exhibit B-3. Percent of second-year beginning teachers receiving support monthly or more often, by treatment status



Note. Responses only include teachers who reported that they were formally assigned a mentor or coach.
*$p < .05$, **$p < .01$, ***$p < .001$.

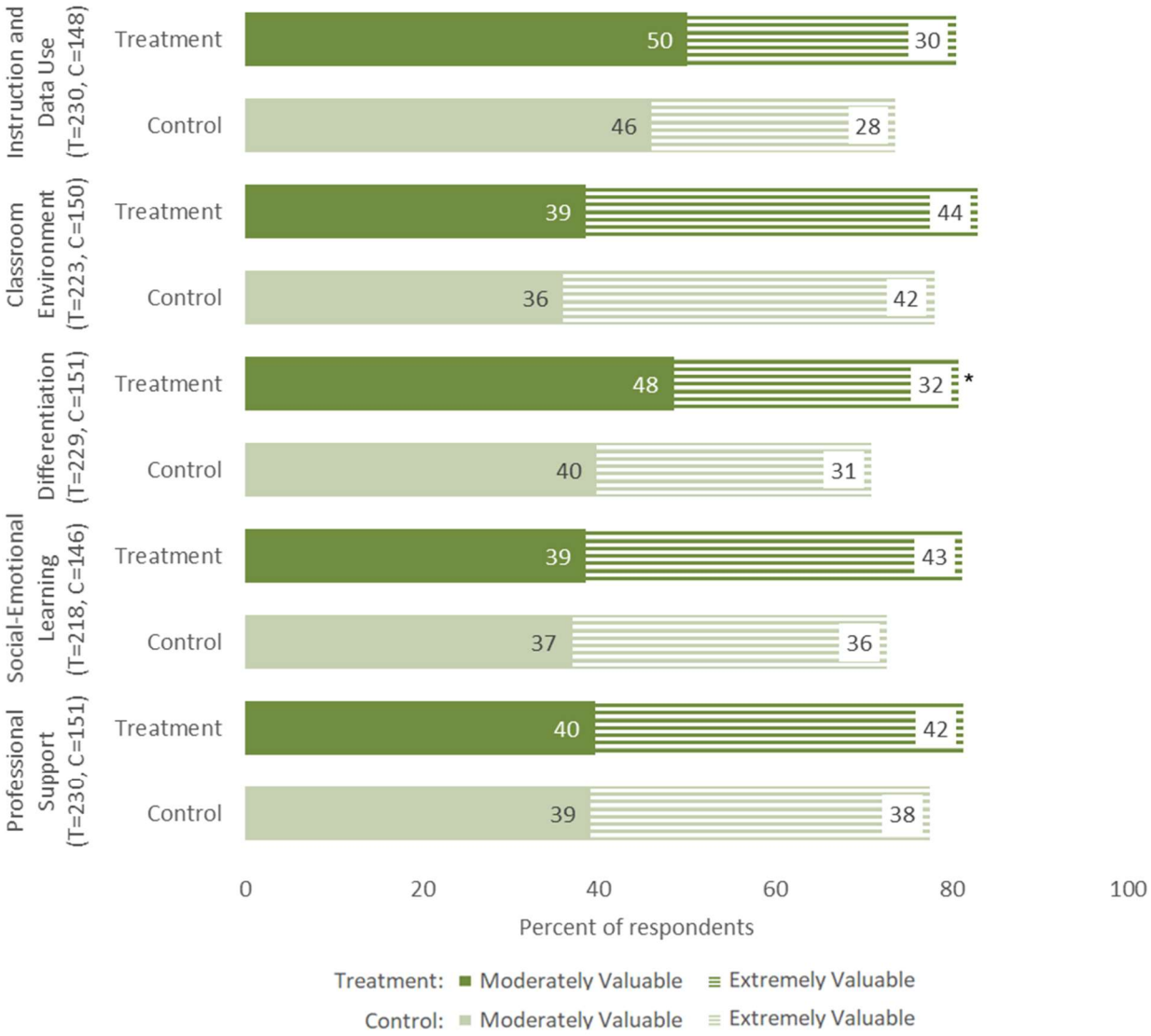
Exhibits B-4 and B-5 provide the item-level differences between treatment and control teachers' responses to the items that make up the "Perceived value of mentoring" scale presented in Chapter 2, as well as items on the value of mentoring to Classroom Environment and Professional Support. Although there was only one statistically significant difference between treatment and control in the proportion of first- and second-year teachers who found mentoring moderately or extremely valuable, there was a consistent pattern of small, positive differences that favored treatment.

Exhibit B-4. Percent of first-year beginning teachers reporting each support was moderately or more valuable, by treatment status



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

Exhibit B-5. Percent of second-year beginning teachers reporting each support was moderately or more valuable, by treatment status



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

Appendix C. Teacher Practice Supplementary Analysis Results

As a supplement to the overall analyses reported in Chapter 3, SRI also examined variation in the impact of NTC induction supports on teacher practice by mentor type and estimated the models excluding late joiner teachers.

Results by Mentor Type

A key scaling strategy NTC used under this grant was to allow sites to employ school-based mentors, rather than or in addition to centrally deployed mentors. School-based mentors were usually teachers or coaches already employed by the school, who had smaller caseloads of beginning teachers due to their other duties outside of NTC mentoring. Two sites (Garfield City and Fillmore County) used exclusively school-based mentors, two sites (Hayes Park and Taftville) used exclusively centrally deployed mentors, and one site (Chesterton) used a mix of both models.

To test whether the impact of NTC induction supports on teacher practice was significantly different for teachers with school-based mentors than for teachers with centrally deployed mentors, the SRI evaluation team ran models with the same specifications as those discussed in Chapter 3, but with an interaction between treatment status and mentor type.

As shown in Exhibit C-1, the main impact in Communicating with Students was evident among teachers with school-based mentors, but not among teachers with centrally deployed mentors. However, the difference in impacts between these two models was not statistically significant. Additionally, while there was no overall impact on teachers' practice in creating an Environment of Respect and Rapport, there was a significant negative impact for teachers with centrally deployed mentors, a significant positive impact for teachers with school-based mentors, and a significant difference between the two. Finally, while neither the overall impact on Managing Classroom Procedures nor the impact for centrally deployed or school-based mentors was statistically significant, the *difference* in impacts was statistically significant. This indicates that the impact of NTC induction supports on teacher practice in this component was also more positive in school-based than in centrally deployed sites.

Note that all observed teachers in Chesterton had school-based mentors, so there was no variation in mentor type within site, though there was variation in site within mentor type. Because of the strong overlap between mentor type and site, it is not possible to entirely separate the effect of mentor type from the effect of site. Additionally, because sites chose whether to implement school-based or centrally deployed mentors, this difference in conditions was not randomly assigned. Thus, the difference in impacts shown here cannot be causally attributed to the difference in mentor type. However, these results suggest that the mentor deployment model a site chooses may be associated with the outcomes its teachers experience, such that sites that chose to deploy school-based mentors saw a more positive impact on teacher practice.

Exhibit C-1. Impact of NTC induction supports on teacher practice, by mentor type

	Centrally deployed		School-based		Difference in impacts	
Measure	**Impact**	***p* value**	**Impact**	***p* value**	**Impact**	***p* value**
Environment of Respect and Rapport	-0.23*	0.01	0.17*	0.04	-0.40**	0.001
Establishing a Culture for Learning	-0.03	0.74	0.01	0.93	-0.03	0.75
Managing Classroom Procedures	-0.13	0.10	0.10	0.14	-0.23*	0.03
Managing Student Behavior	-0.12	0.26	0.08	0.41	-0.20	0.16
Communicating with Students	0.07	0.31	0.13*	0.02	-0.06	0.41
Questioning and Discussion Techniques	0.06	0.54	0.04	0.64	0.02	0.89
Engaging Students in Learning	0.04	0.58	0.09	0.24	-0.05	0.65
Using Assessment in Instruction	0.06	0.35	0.03	0.68	0.03	0.76
Sample	**Schools**	**Teachers**	**Schools**	**Teachers**	**Schools**	**Teachers**
n	83	157	118	199	201	356

*$p < 0.05$, **$p < 0.01$.

Results Excluding Late Joiner Teachers

The SRI evaluation team also examined models that excluded "late joiner" teachers. Late joiner teachers were Cohort 2 teachers who were hired into schools nearly one year after random assignment (hired in summer or fall 2017 into schools that were randomly assigned in fall 2016). The models estimating the impact of NTC on teacher practice included 41 such teachers in 29 schools. Due to the inclusion of these teachers in the analysis, the impacts reported in Chapter 3 may be due, in part, to a change in the composition of beginning teachers in the treatment and control schools during the course of the study, in addition to a change in teachers' classroom practice. For the inclusion of joiners to bias the results, the type of teacher who joined treatment schools must be systematically different from the type of teacher who joined control schools, usually because teachers chose to enter a school based on the presence or absence of the intervention. The What Works Clearinghouse (WWC) guidelines categorize interventions by the level of risk of bias due to joiners (WWC, 2022). Because this intervention was low profile and beginning teachers often have very little choice in where they are hired, the risk in this study is likely low. Nonetheless, to test the sensitivity of the conclusions to joiner teachers, the SRI evaluation team also ran the analyses excluding these teachers.

When late joiner teachers were removed from the models, none of the components of the FFT were significantly related to treatment status (Exhibit C-2). Thus, although the full sample shows a positive, statistically significant relationship between NTC induction supports and teachers' practice in the Communicating with Students component, when the sample is restricted to exclude joiners, this conclusion does not hold. This may be due to the reduction in sample size, or it may be that the impact in this component is due, in part, to a change in the composition of the teachers at the treatment or

control schools between the two years (i.e., Cohort 2 teachers who were stronger in Communicating with Students were more likely to get hired in treatment than in control schools).

Exhibit C-2. Impact of the NTC model on teacher practice outcomes using the non-joiner sample

Overall teacher practice		
	Impact	***p* value**
Overall teacher practice	0.02	0.73
Domain 2: The Classroom Environment		
Classroom Environment Domain	-0.03	0.59
Creating an Environment of Respect and Rapport	-0.05	0.43
Establishing a Culture for Learning	-0.02	0.76
Managing Classroom Procedures	0.01	0.85
Managing Student Behavior	-0.06	0.44
Domain 3: Instruction		
Instruction Domain	0.08	0.12
Communicating with Students	0.08	0.09
Using Questioning and Discussion Techniques	0.07	0.27
Engaging Students in Learning	0.05	0.44
Using Assessment in Instruction	0.07	0.24
Sample size		
n schools	172	
n teachers	315	

Appendix D. Student Achievement Supplementary Analysis Results

This appendix presents the results of three types of analyses as a supplement to the student achievement analyses reported in Chapters 3 and 4. The first section examines variation in the impact of NTC induction supports on student achievement by mentor type. The second section tests the sensitivity of the overall results to variations in the sample used and in the specifications of the model. Finally, the third section presents tests to further establish the relationship between the level and type of mentoring and student achievement.

Results by Mentor Type

To test whether the impact of NTC induction supports on student achievement was significantly different for teachers with school-based than for teachers with centrally deployed mentors, SRI ran models with the same specifications as those discussed in Chapter 3, but with an interaction between treatment status and mentor type. As shown in Exhibit D-1 there was no statistically significant impact of NTC induction supports on student achievement in mathematics among students of teachers with either centrally deployed or school-based mentors. Additionally, there was no statistically significant difference in the effect on student achievement based on mentor type.

Exhibit D-1. Impact of NTC induction supports on teacher practice, by mentor type

	Centrally deployed		School-based		Difference in impacts	
Mathematics	**Impact**	***p* value**	**Impact**	***p* value**	**Impact**	***p* value**
Impact	0.11	0.22	-0.02	0.72	-0.14	0.41
n students	2,671		4,341		7,012	
n teachers	39		103		142	
n schools	31		76		107	
ELA	**Impact**	***p* value**	**Impact**	***p* value**	**Impact**	***p* value**
Impact	-0.14	0.10	0.07	0.13	0.13	0.15
n students	2,789		5,471		8,260	
n teachers	43		122		165	
n schools	32		85		117	

Sensitivity Tests

To test whether the "no effect" findings for the main student achievement analysis were robust to variations in the specification of the model or sample, SRI also examined models that:

- Excluded late joiner teachers

- Used only students of teachers who were also observed for the teacher practice outcomes.[27]
- Excluded interactions between site-level fixed effects and other covariates
- Added a covariate for teacher certification (full versus partially certified)

The main student achievement analysis includes both student and teacher joiners—those who joined the study more than six weeks after randomization. This includes late student joiners—those who joined schools in the fall of 2017 that had been randomly assigned in fall 2016. Thus, any impacts may be due, in part, to a change in the composition of students in the treatment and control schools during the course of the study, in addition to a change in students' average achievement. As in the teacher practice analysis, the risk of bias due to joiners is likely low, particularly for student joiners, because students were unlikely to be aware of the intervention or to switch schools based on the presence or absence of an NTC mentor serving their teacher.

Note that the confirmatory models do not include teacher certification as a covariate because the data on this variable at baseline was unreliable. Teachers may earn their full certification during their first year of teaching (after the start of the intervention), and most sites did not store historical data on teachers' status as of their hire date. Thus, this variable may be endogenous with treatment (i.e., treatment may have affected teachers' likelihood of earning their full certification).

As shown in Exhibit D-2, varying the sample and model specifications did not substantially change the estimate of the impact of NTC induction supports on student achievement in either mathematics or ELA. Regardless of these changes, there was no overall impact detected.

[27] This analysis was included to confirm whether the impact on teacher practice, which was not present in student achievement, was due to a difference in the sample of teachers observed as compared to the sample of teachers included in the achievement analysis.

Exhibit D-2. Sensitivity of the main effects to variations in the specification of the sample and models

Mathematics					
	Impact	***p* value**	***n* schools**	***n* teachers**	***n* students**
Confirmatory outcome (reported in Chapter 3)	0.02	0.74	107	142	7,012
Exclude late joiner teachers	-0.01	0.87	96	131	6,632
Sample that was also observed	0.08	0.18	86	103	5,136
Exclude interaction terms	-0.004	0.93	107	142	7,012
Add teacher certification covariate	0.01	0.92	107	142	7,012
ELA					
	Impact	***p* value**	***n* schools**	***n* teachers**	***n* students**
Confirmatory outcome (reported in Chapter 3)	0.03	0.50	117	165	8,260
Exclude late joiner teachers	0.01	0.88	105	150	7,521
Sample that was also observed	0.05	0.31	95	123	6,155
Exclude interaction terms	0.02	0.62	117	165	8,260
Add teacher certification covariate	0.03	0.92	117	165	8,260

Level and Type of Mentoring and Student Achievement

As discussed in Chapter 4, the models testing the relationship between the level and type of mentoring and student achievement are correlational, not causal. One reason for the correlations may be if mentors chose to focus on instruction more with teachers whose students were already higher achieving. While this hypothesis cannot be definitively disproven, SRI chose to test its plausibility by predicting the achievement of teachers' students in Year 1 using the level and type of mentoring teachers received in Year 2. This analysis assumes that mentors would base their mentoring choices in Year 2 on the performance of a teacher's students in the prior year. As shown in Exhibit D-3, there was no association between measures of mentoring in Year 2 and teacher-level average achievement in Year 1, indicating that the mentoring teachers received in their second year was not significantly related to how their students performed the year before. This finding supports the conclusion that instructionally focused mentoring activities are associated with higher student achievement in mathematics.

Exhibit D-3. Sensitivity of the main effects to variations in the specification of the sample and models

	Measure	Effect size	*p* value
Math	Met minimum minutes and months	-0.096	0.23
	Used any tool in 85 percent of interactions	0.016	0.83
	Used the PCG tool at least three times	-0.026	0.69
	Used the ASL tool at least three times	0.059	0.36
	Completed an observation cycle at least three times	-0.051	0.42
	Met with a mentor weekly for an average of an hour or more	-0.006	0.90
	Frequency of instructionally focused mentoring activities–-monthly or more	-0.003	0.93
	Mentoring was moderately or more valuable	0.007	0.89
ELA	Met minimum minutes and months	-0.023	0.59
	Used any tool in 85 percent of interactions	0.063	0.09
	Used the PCG tool at least three times	-0.025	0.50
	Used the ASL tool at least three times	-0.018	0.64
	Completed an observation cycle at least three times	< 0.001	0.99
	Met with a mentor weekly for an average of an hour or more	< 0.001	0.99
	Frequency of instructionally focused mentoring activities–-monthly or more	0.051	0.08
	Mentoring was moderately or more valuable	0.016	0.65

Appendix E. Teacher Retention Supplementary Analysis Results

As a supplement to the overall analyses reported in Chapter 3, SRI also examined the impact of NTC induction supports on teacher retention separately for each teacher cohort, for teachers supported by school-based mentors, and for teachers supported by centrally deployed mentors.

As shown in Exhibit E-1, there was no significant impact of NTC induction supports on teacher retention in any of these subgroups.

Exhibit E-1. Variation in retention effects by cohort and mentor type

By cohort				
	Impact	***p* value**	***n* schools**	***n* teachers**
Cohort 1	0.25	0.43	184	357
Cohort 2	0.14	0.61	225	438
By mentor type				
	Impact	***p* value**	***n* schools**	***n* teachers**
Centrally deployed	-0.01	0.96	115	337
School-based	0.19	0.46	184	458

Appendix F. Supplemental Tables to Exploratory Findings

In addition to the findings reported in Chapter 4, SRI also explored eight student characteristics as moderators, which showed no differential effects in either mathematics or ELA (Exhibit F-1).

Exhibit F-1. Results of models examining the moderation effect of student-level demographic variables

		Mathematics		ELA	
	Sample	Impact	*p* value	Impact	*p* value
Main results	Full sample	0.03	0.55	0.05	0.25
Grade level	Elementary (grades 4 & 5)	-0.04	0.53	0.04	0.49
	Middle (grades 6–8)	0.08	0.18	0.05	0.33
	Difference	0.13	0.17	0.01	0.88
White	Non-White	0.03	0.51	0.05	0.21
	White	-0.01	0.94	0.03	0.56
	Difference	-0.04	0.55	-0.02	0.73
Female	Non-female	0.01	0.82	0.04	0.30
	Female	0.04	0.37	0.05	0.22
	Difference	0.03	0.37	0.01	0.80
FRPL	Non-FRPL	-0.03	0.66	0.07	0.20
	FRPL	0.04	0.44	0.04	0.29
	Difference	0.07	0.29	-0.03	0.58
SPED	Non-SPED	0.02	0.63	0.06	0.12
	SPED	0.04	0.51	0.00	0.94
	Difference	0.02	0.75	-0.07	0.14
EL	Non-EL	0.02	0.64	0.04	0.38
	EL	0.04	0.48	0.09	0.10
	Difference	0.02	0.68	0.06	0.28
Low prior achievement	At or above average prior achievement	0.00	0.93	0.03	0.57
	Below average prior achievement	0.06	0.23	0.07	0.13
	Difference	0.06	0.09	0.04	0.38

Exhibit F-1, concluded. Results of models examining the moderation effect of student-level demographic variables

		Mathematics		ELA	
	Sample	**Impact**	***p* value**	**Impact**	***p* value**
Very low prior achievement	At or above the 25th percentile in prior achievement	0.02	0.68	0.04	0.31
	Below the 25th percentile in prior achievement	0.04	0.47	0.06	0.21
	Difference	0.02	0.08	0.02	0.65
	n schools	107		117	
	n teachers	142		165	
	n students	7012		8,260	

Likewise, SRI also explored two additional school-level characteristics as moderators, which showed no differential effects in either mathematics or ELA: the percent of the students in the school whose race/ethnicity was Black and/or Hispanic and school-level baseline percent meeting or exceeding standards. The results of these models, as well as the other school-level characteristics discussed in Chapter 4, are shown in Exhibit F-2.

Exhibit F-2. Results of models examining the moderation effect of school-level demographic variables

		Mathematics		ELA	
	Sample	**Impact**	***p* value**	**Impact**	***p* value**
	Full sample	0.03	0.55	0.05	0.25
Poverty	Below average percent in poverty	-0.01	0.95	-0.08	0.24
	Above average percent in poverty	0.04	0.46	0.12*	0.02
	Difference	0.05	0.63	0.20*	0.03
English learners	Below average percent EL	0.02	0.77	-0.05	0.35
	Above average percent EL	0.02	0.78	0.14*	0.01
	Difference	0.00	0.99	0.19*	0.02
Race	Below average percent Black or Hispanic	0.04	0.64	0.04	0.57
	Above average percent Black or Hispanic	0.02	0.72	0.05	0.33
	Difference	-0.01	0.89	0.01	0.90
Prior proficiency	Below average percent proficient	-0.05	0.42	0.05	0.38
	Above average percent proficient	0.09	0.15	0.05	0.39
	Difference	0.14	0.12	0.01	0.93

Exhibit F-2, concluded. Results of models examining the moderation effect of school-level demographic variables

		Mathematics		ELA	
	Sample	**Impact**	***p* value**	**Impact**	***p* value**
Prior scale score	Below average prior achievement	0.02	0.73	0.06	0.23
	Above average prior achievement	0.03	0.69	0.03	0.70
	Difference	0.01	0.90	-0.03	0.71
Sample size	*n* schools		107		117
	n teachers		142		165
	n students		7,012		8,260

Appendix G. Additional Data for What Works Clearinghouse Review

The impact study discussed in this report used a randomized controlled trial design with schools randomly assigned within blocks. All schools had a 50 percent chance of assignment to treatment or control, within their block. Dummy variables for the blocks were included in all models to account for the design. The analyses use multilevel modeling to account for the nesting of students within teachers and teachers within schools. All analyses used multiple imputation to fill in missing values for the covariates used in the model. Outcomes were not imputed.

Impact on Teacher Practice

Outcome measure

SRI measured the impact of NTC induction supports on teacher practice using teacher scores on the Danielson Framework for Teaching. This measure is commonly used by schools and districts to assess teaching quality, indicating that it is not over-aligned with the intervention. Observers were blind to teachers' treatment status, and they scored teachers using the same procedures in both treatment and control classrooms. In analysis, SRI combined the element-level scores to create eight component-level scales, two domain-level scales, and an "overall teacher practice" scale using a simple average of the non-missing element-level scores.[28] Exhibit G-1 shows the alpha reliability coefficients of these scales.

Exhibit G-1. Alpha reliability coefficients for FFT component-level, domain-level, and overall scales

Measure	Alpha
Overall teacher practice	0.90
Classroom environment domain	0.85
Creating an Environment of Respect and Rapport	0.65
Establishing a Culture for Learning	0.52
Managing Classroom Procedures	0.77
Managing Student Behavior	0.84
Instruction domain	0.84
Communicating with Students	0.60
Using Questioning and Discussion Techniques	0.71
Engaging Students in Learning	0.80
Using Assessment in Instruction	0.70

[28] Element-level scores were only missing when they could not be scored because the practice was not observed in the classroom. For example, teachers' "response to student misbehavior" was not scored when no student misbehavior was observed.

A total of 356 teachers were observed at both baseline and follow-up, across two cohorts. Of these, 33 (9.3 percent) were observed and rated by more than one observer on the same lesson and examined for inter-rater reliability. The average reliability of ratings between observers within one point was 95 percent, while reliability at the element level ranged from 67 percent (performance of classroom routines) to 100 percent (multiple elements) (Exhibit G-2).

Exhibit G-2. Inter-rater reliability overall and for each element

Component	Element	Percent agreement within one point
Pooled across components	**Pooled across elements**	95
Creating an Environment of Respect and Rapport	Teacher interactions with students	100
	Student interactions with other students	97
Establishing a Culture for Learning	Importance of the content and of learning	100
	Expectations for learning and achievement	100
Managing Classroom Procedures	Management of instructional groups	94
	Management of transitions	97
	Management of materials and supplies	97
	Performance of classroom routines	67
Managing Student Behavior	Expectations	94
	Monitoring of student behavior	100
	Response to student misbehavior	94
Communicating with Students	Expectations for learning	88
	Directions for activities	94
	Explanations of content	100
	Use of oral and written language	100
Using Questioning and Discussion Techniques	Quality of questions/prompts	94
	Discussion techniques	88
	Student participation	100
Engaging Students in Learning	Activities and assignments	97
	Grouping of students	97
	Instructional materials and resources	97
	Structure and pacing	100
Using Assessment in Instruction	Assessment criteria	97
	Monitoring of student learning	100
	Feedback to students	88

Component	Element	Percent agreement within one point
Pooled across components	**Pooled across elements**	95
	Student self-assessment and monitoring of progress	91
n		33

Assignment to intervention and comparison conditions

As described in Chapter 1, schools were randomly assigned to treatment and control conditions within blocks. All schools had equal probabilities of selection within block, and block fixed effects were included in all impact models. Teachers in randomly assigned schools were eligible for the study if they were beginning teachers, and they were eligible to be observed if they taught a core subject (mathematics, reading/English language arts, social studies, or science), or if they taught in a multi-subject elementary classroom.

All eligible teachers were included in the teacher practice sample at baseline and included in the attrition calculations displayed in Chapter 3. However, not all randomly assigned schools had teachers eligible to be observed. As shown in Exhibit 15 in Chapter 3, of the 151 schools randomly assigned to control, 131 had teachers eligible to be observed, and of the 148 schools randomly assigned to treatment, 133 had teachers eligible to be observed. These schools with eligible teachers served as the reference sample for attrition calculations.

Teachers were analyzed in their original randomized condition, and the analysis included only randomly assigned units.

Compositional change

As shown in Exhibit 15 in Chapter 3, school-level overall attrition was 24 percent, with 4.9 percentage point differential attrition. Thus, attrition was below the cautious boundary at the cluster level.

The main impact analysis discussed in Chapter 3 includes late joiner teachers—those who were hired into study schools in the second year of the study. However, because the unit of assignment was schools and the unit of analysis was teachers, this study has a high risk of bias due to joiners. Therefore, SRI tested the sensitivity of the results to the inclusion of these joiner teachers, and results are shown in Appendix C. Overall, the results were similar, with the exception of the impact on the Communicating with Students component, which was not significant when late joiner teachers were excluded.

Representativeness

Because this analysis has low cluster-level attrition, but a high risk of bias due to joiners, SRI examined the representativeness of the clusters included in the analytic sample at follow up. As shown in Exhibit 15 in Chapter 3, within non-attrited schools, teacher-level overall attrition with respect to eligible teachers at follow-up was 28 percent, with 0.4 percentage point differential attrition. Thus, this analysis met the representativeness standard at the cautious boundary.

Baseline equivalence

Although this analysis stems from a low-attrition RCT and met WWC standards for representativeness, SRI also examined the equivalence of the sample on baseline measures of classroom practice. As shown in Exhibit G-3, the baseline differences between treatment and control were all less than 0.25 standard deviation. The models include these baseline measures as covariates in the multilevel regression models, thus meeting standards for baseline equivalence.

Exhibit G-3. Baseline equivalence of the analysis sample, impact on teacher practice

	Treatment mean	Control mean	Pooled *SD*	Standardized difference
Overall teacher practice	2.34	2.29	0.42	0.12
Classroom environment domain	2.44	2.41	0.48	0.06
Creating an Environment of Respect and Rapport	2.56	2.57	0.60	-0.01
Establishing a Culture for Learning	2.38	2.33	0.48	0.10
Managing Classroom Procedures	2.43	2.38	0.55	0.10
Managing Student Behavior	2.37	2.37	0.63	< 0.01
Instruction domain	2.24	2.17	0.43	0.18
Communicating with Students	2.45	2.53	0.43	0.05
Using Questioning and Discussion Techniques	2.12	2.02	0.59	0.18
Engaging Students in Learning	2.36	2.29	0.53	0.13
Using Assessment in Instruction	2.02	1.93	0.10	0.20

Additional information for calculating effect sizes

Exhibit G-4 shows the means of the outcome variables used in analysis. The first three columns show the unadjusted means of the treatment and control group and the pooled standard deviation of the analysis sample. The last two columns show the model-adjusted treatment and control means.

Exhibit G-4. Unadjusted and adjusted outcome means of FFT components and domains, analysis sample

	Unadjusted			Adjusted	
	Treatment mean	Control mean	Pooled *SD*	Treatment mean	Control mean
Overall teacher practice	2.50	2.43	0.41	2.45	2.47
Classroom environment domain	2.63	2.59	0.48	2.62	2.60
Creating an Environment of Respect and Rapport	2.74	2.72	0.62	2.75	2.72
Establishing a Culture for Learning	2.49	2.51	0.48	2.50	2.49
Managing Classroom Procedures	2.64	2.56	0.48	2.60	2.59
Managing Student Behavior	2.65	2.61	0.68	2.64	2.63
Instruction domain	2.41	2.31	0.41	2.33	2.39
Communicating with Students	2.66	2.53	0.40	2.55	2.64
Using Questioning and Discussion Techniques	2.27	2.18	0.61	2.20	2.25
Engaging Students in Learning	2.49	2.37	0.52	2.40	2.47
Using Assessment in Instruction	2.17	2.10	0.48	2.11	2.16

Impact on Student Achievement

Outcome measure

The impact of NTC induction supports on student achievement was measured using students' scaled scores on their state's annual standardized assessments in grades 4 through 8 in mathematics and reading/English language arts (ELA). SRI standardized the students' scores to their district mean and standard deviation (Hayes Park, Fillmore County, Chesterton, and Taftville) or to the sample mean and standard deviation, where district statistics were not available (Garfield City). State standardized assessments are assumed by WWC to be reliable and valid. These measures are also not over-aligned with the intervention, and consistently collected between treatment and control.

Assignment to intervention and comparison conditions

As described in Chapter 1, schools were randomly assigned to treatment and control conditions within blocks. All schools had equal probabilities of selection within block, and block fixed effects were included in all impact models. Students in randomly assigned schools were eligible for the analysis if they took mathematics or ELA classes with eligible teachers in grades 4 through 8 (the grades with state standardized testing in those subjects).

All eligible students were included in the student achievement sample at baseline and included in the attrition calculations displayed in Chapter 3. However, not all randomly assigned schools had eligible students. Of the 151 schools randomly assigned to control, 54 had eligible students in mathematics and 60 had eligible students in ELA. Likewise, of the 148 schools randomly assigned to treatment, 57 had eligible students in mathematics and 62 had eligible students in ELA. These schools with eligible students serve as the reference sample for attrition calculations.

Students were analyzed in their original randomized condition, and the analysis included only randomly assigned units.

Compositional change

SRI identified mathematics and ELA teachers in grades 4 through 8 using district extant data linking students to teachers in their individual classrooms at the time of testing. All eligible math and ELA teachers with students in grades 4 through 8 were included in the student achievement sample at baseline and included in the attrition calculations displayed in Chapter 3. As shown in Exhibit 21 in Chapter 3, school-level overall attrition for the mathematics sample was 4 percent, with 3.8 percentage point differential attrition. Overall attrition for the Ela sample was 4 percent, with 1.8 percentage point differential attrition.

The unit of assignment for this analysis was schools and the unit of analysis was students. Because students could not be identified at the time of random assignment, they were identified at the time of testing. Therefore, this analysis includes joiner students. SRI tested the sensitivity of the results to the inclusion of joiner teachers, and results are shown in Appendix D. Overall, the results were similar, with the exception of the impact on the Communicating with Students component, which was not significant when late joiner teachers were excluded. The districts were not able to provide adequate data to identify joiner students.

Representativeness

Because this analysis has low cluster-level attrition, but a high risk of bias due to joiners, SRI examined the representativeness of the clusters included in the analytic sample at follow up. As shown in Exhibit 22 in Chapter 3, within non-attrited schools, the overall student-level attrition was 36.6 percent, with differential attrition of 0.9 percentage points in mathematics. In ELA the overall student-level attrition within non-attrited schools was 34.4% with differential attrition of 2.7 percentage points. These are both within the WWC's "cautious boundary." Thus, the students with complete baseline and outcome data are representative of clusters at follow-up. (WWC, 2022).

Exhibit G-5 shows the same calculations, including students with imputed baseline data, thus reflecting the full analysis sample in the numerator of the calculations. Attrition by this measure is also within WWC's "cautious boundary."

Exhibit G-5. Representativeness of clusters at follow-up, including imputed data – student achievement

Mathematics				
	Treatment	**Control**	**Total**	**Differential**
Eligible students in non-attrited schools	4,321	3,448	7,769	
Students included in the final analysis	3,801	3,211	7,012	
Percent of students attrited from non-attrited schools	12.0%	6.9%	9.7%	5.2 points
		Met attrition standards		
ELA				
	Treatment	**Control**	**Total**	**Differential**
Eligible students in non-attrited schools	5,163	3,730	8,893	
Students included in the final analysis	4,754	3,506	8.260	
Percent of students attrited from non-attrited schools	7.9%	6.0%	7.1%	1.9 points
		Met attrition standards		

Baseline equivalence

Although this analysis stems from a low-attrition RCT that meets WWC standards for representativeness, SRI also examined the equivalence of the sample on baseline measures of student achievement, which are included in the models. Exhibit G-5 shows the baseline difference using unimputed data, and these differences were both less than 0.25 standard deviations. The models include these baseline measures as covariates in the multilevel regression models, thus meeting standards for baseline equivalence.

Exhibit G-5. Baseline equivalence of the sample, impact on student achievement, unimputed data

	Treatment mean	Control mean	Pooled *SD*	Standardized difference	N schools	N teachers	N students
Mathematics	-0.11	-0.05	0.96	-0.07	63	83	4,928
English language arts	-0.02	0.02	0.96	-0.04	71	100	5,831

Exhibit G-6 shows the baseline equivalence of the full analysis sample, using imputed baseline data. These differences were also less than 0.25 standard deviations.

Exhibit G-6. Baseline equivalence of the analysis sample, impact on student achievement, imputed data

	Treatment mean	Control mean	Pooled *SD*	Standardized difference	N schools	N teachers	N students
Mathematics	-0.11	-0.04	0.99	-0.08	107	142	7,012
English language arts	-0.03	0.01	1.00	-0.04	117	165	8,260

Additional information for calculating effect sizes

Exhibit G-7 shows the means of the outcome variables used in analysis. The first three columns show the unadjusted means of the treatment and control group and the pooled standard deviation of the analysis sample. The last two columns show the model-adjusted treatment and control means.

Exhibit G-7. Unadjusted and adjusted outcome means of achievement tests, analysis sample

	Unadjusted			Adjusted	
	Treatment	Control	Pooled *SD*	Treatment	Control
Mathematics	-0.05	-0.03	0.97	-0.06	-0.04
English language arts	-0.08	-0.003	0.98	-0.10	-0.06

Impact on Teacher Retention

Outcome measure

The impact of NTC induction supports on teacher retention in instructional positions was measured using district human resources records. SRI collected employment records from each site and identified which study teachers were still employed full-time in instructional positions as of October 1, 2018 (Cohort 1) or October 1, 2019 (Cohort 2).

Assignment to intervention and comparison conditions

As described in Chapter 1, schools were randomly assigned to treatment and control conditions within blocks. All schools had equal probabilities of selection within block, and block fixed effects were included in all impact models. Teachers in randomly assigned schools were eligible for the study if they were beginning teachers, as defined in Chapter 1.

Compositional change and representativeness

All study teachers were included in the teacher retention sample at baseline and included in the attrition calculations displayed in Chapter 3. There was no attrition from this analysis. Thus, attrition and representativeness were both within the WWC's cautious boundary (WWC, 2022). However, the analysis includes beginning teachers hired in fall of 2017 into schools that were randomized in fall 2016 (Cohort 2 teachers within Cohort 1 schools). Thus, this analysis does include joiner teachers.

Baseline equivalence

Although the impact of NTC's induction supports on student achievement stem from a low-attrition RCT with a representative sample, SRI also examined the equivalence of the sample on baseline measures recommended by the WWC Review Protocol for Teacher Excellence, version 4.0 (WWC, 2019): average attendance of the teacher's students at baseline, school-average teacher retention at baseline, student race/ethnicity, and economic disadvantage. As shown in Exhibit G-8, the baseline differences between

treatment and control were less than 0.05 standard deviation, except for the percent students of color, which was 0.1 standard deviation. All four baseline measures were included as covariates in the models estimating the impact of NTC induction supports on teacher retention.

Exhibit G-8. Baseline equivalence of the analysis sample, impact on teacher retention

	Treatment mean	Control mean	Pooled *SD*	Standardized difference
Classroom average student attendance	0.932	0.931	0.03	0.03
School average teacher retention	0.887	0.886	0.08	0.02
Percent students of color	90.7	91.8	10.3	-0.10
Percent of students eligible for free or reduced-price meals	75.9	76.7	19.4	-0.03

Additional information for calculating effect sizes

Exhibit G-9 shows the means of the outcome variables used in analysis. The first two columns show the unadjusted means of the treatment and control group. The last two columns show the model-adjusted treatment and control means.

Exhibit G-9. Unadjusted and adjusted outcome means of achievement tests, analysis sample

	Unadjusted		Adjusted	
	Treatment	Control	Treatment	Control
Teacher retention rate	0.760	0.757	0.807	0.791