



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 by:

SRI Education

Viki M. Young

Rebecca Schmidt

Haiwen Wang

Lauren Cassidy

Katrina Laguarda

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

December 2017

Submitted to:

Ali Picucci
Vice President of Impact and Improvement
New Teacher Center

Prepared by:

SRI Education
Viki M. Young
Rebecca Schmidt
Haiwen Wang
Lauren Cassidy
Katrina Laguarda

Acknowledgments: The findings reported here culminated from the efforts of a large team over 5 years. The authors thank Marjorie Wechsler, Paul Hu, Hannah Cheever, Hannah Kistler, Andrew Ezekoye, Chi Nguyen, Bonnee Groover, Francine Biscocho, Juliet Tiffany-Morales, Jennifer Bland, Matt McCracken, Tiffany Hsieh, and Janelle Sands. The authors also thank Ali Picucci and the district teams at the New Teacher Center for their strong engagement with and feedback on the evaluation. Not least, the authors are grateful to the three partner sites—Broward County Public Schools, Chicago Public Schools, and Grant Wood Area Education Agency—for their participation in the evaluation.

Suggested citation: Young, V. M., Schmidt, R., Wang, H., Cassidy, L., & Laguarda, K. (2017, December). *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. Menlo Park, CA: SRI International.

Copyright 2017 SRI International. All rights reserved.

CONTENTS

Executive Summary	iii
Introduction.....	1
Program Description	1
Study Design Summary	7
Implementation Findings.....	12
Implementation Fidelity	12
Contrast in Induction Supports Between Treatment and Control Schools	14
Impact on Teacher Outcomes	27
Impacts on Teacher Practice	27
Impact on Teacher Retention.....	32
Impact on Student Outcomes.....	33
Randomized Controlled Trials of the Impact on Student Achievement.....	33
Quasi-experimental Study of the Impact on Student Achievement	35
Conclusions and Implications	39
References.....	41
Appendix	
Appendix A. Implementation Fidelity Measures.....	A-1
Appendix B. Teacher Survey Methods and Measures.....	B-1
Appendix C. Randomized Controlled Trials Methods.....	C-1
Appendix D. Teacher Practice Impact Analysis and Model Results.....	D-1
Appendix E. Teacher Retention Impact Analysis and Model Results	E-1
Appendix F. Student Achievement Model Results for RCT Districts	F-1
Appendix G. Sensitivity Tests for RCT Results	G-1
Appendix H. QED Study Methods and Student Achievement Model Results	H-1

EXHIBITS

Exhibit 1. Logic Model for the New Teacher Center i3 Validation Grant.....	4
Exhibit 2. Average District and Study Sample Characteristics at the Time of Random Assignment, RCT Districts.....	6
Exhibit 3. Average Consortium and Study Sample Characteristics, Year Before Intervention, QED Site	6
Exhibit 4. Data Sources for Implementation Analysis of Key Components	7
Exhibit 5. School Year by Cohort and Years of Experience	8
Exhibit 6. Overall Survey Response Rates by Years of Teaching Experience, Combined Cohorts	8
Exhibit 7. Data Sources and Their Purposes	10
Exhibit 8. Data Collection Activities by Cohort and Year for BCPS and CPS.....	11
Exhibit 9. Data Collection by Cohort and Year for GWAEA.....	11
Exhibit 10. Implementation Fidelity by Key Component	13
Exhibit 11. New Teacher Interactions with Mentors, RCT Combined Sample, 2014–16.....	15
Exhibit 12. New Teacher Interactions with Mentors, QED Site, 2013–16	16
Exhibit 13. Frequency of Mentoring Activities, RCT Combined Sample, 2014–16	17
Exhibit 14. Frequency of Mentoring Activities, QED Site, 2013–16.....	18
Exhibit 15. Focus on Instruction in Mentoring and Other Induction Supports, RCT Combined Sample, 2014–16.....	19
Exhibit 16. Focus on Instruction in Mentoring and Other Induction Supports, QED Site, 2013–16.....	19
Exhibit 17. Focus of Mentoring and Other Induction Supports, RCT Combined Sample, 2014–16.....	20
Exhibit 18. Focus of Mentoring and Other Induction Supports, QED Site, 2013–16	22
Exhibit 19. Frequency of Other Induction Supports, RCT Combined Sample, 2014–16	24
Exhibit 20. Frequency of Other Induction Supports, QED Site, 2013–16	24
Exhibit 21. Value of Mentoring Activities and Other Induction Supports, RCT Combined Sample, 2014–16.....	25
Exhibit 22. Value of Mentoring Activities and Other Induction Supports, QED Site, 2013–16.....	25
Exhibit 23. New Teacher Ratings of Self-Efficacy and Impact of Supports, RCT Combined Sample, 2014–16.....	26
Exhibit 24. New Teacher Ratings of Self-Efficacy and Impact of Supports, QED Site, 2013–16	26
Exhibit 25. Overall School Observation Sample Selection and Attrition, Cohort 1, Cohort 2, and Combined, RCT Districts	28
Exhibit 26. Framework for Teaching Domains, Components, and Elements Observed	29
Exhibit 27. Impact of the NTC Model on Teacher Practice Outcomes, Combined RCT Sample	30
Exhibit 28. Model-Implied Means on Teacher Practice Outcomes for Treatment and Control Groups Overall	31
Exhibit 29. Second-Year Impact on Student Achievement, Combined RCT Sites.....	35
Exhibit 30. Timing of Prior Achievement and Outcome Scores, QED Site	36
Exhibit 31. Numbers of Schools, Teachers, and Students Included in Cohort 1 Year 2 Achievement Analyses, QED Site	37
Exhibit 32. Cohort 1 Year 2 Baseline Student Test Scores, by Groups of Teachers, QED Site	37
Exhibit 33. Estimated Impact on Student Achievement for Cohort 1 Year 2	38

EXECUTIVE SUMMARY

Teacher induction strategies aim to provide novice teachers with crucial supports as they first confront the realities of the classroom, shoring up essential management and instructional skills, improving retention in the profession, and ultimately bolstering student learning. The New Teacher Center (NTC) received an Investing in Innovation (i3) Validation grant in 2013 to implement its induction model in three sites: Broward County Public Schools (BCPS) in Florida, Chicago Public Schools (CPS) in Illinois, and the Grant Wood Area Education Agency (GWAEA), a consortium of rural districts in Iowa. Across the three sites, NTC trained full-time released mentors and served two cohorts of beginning teachers for 2 years each. Through the grant, NTC formalized four key components of its comprehensive induction model: (1) build the capacity of districts and school leaders to support the mentoring program, (2) select and assign full-time release mentors to caseloads of no more than 15 teachers each, (3) provide mentors more than 100 hours of intensive training through institutes and in-field support from lead coaches, and (4) provide regular, high-quality mentoring to first- and second-year teachers using a system of NTC-developed online formative assessment tools.

SRI Education conducted the evaluation of NTC's i3 Validation grant, examining the implementation and impact of NTC's induction model. The evaluation used a rigorous mixed-methods design to measure implementation fidelity and impact on teacher and student outcomes across the three participating sites. To account for different local contexts and needs, SRI used two methods to study impact: (1) randomized controlled trials (RCT) in BCPS and CPS with schools randomly assigned to NTC mentoring and control groups and (2) a quasi-experimental design (QED) in GWAEA. In each site, the evaluation team followed two cohorts of new teachers for 2 years each—Cohort 1 began teaching in 2013–14 and Cohort 2 in 2014–15—for a total implementation period of 3 years (2013–14 through 2015–16). The evaluation measured implementation across all 3 years and teacher and student impacts after teachers had participated in 2 years of induction.

Implementation Findings

Using teacher and mentor surveys, interviews, and NTC's online formative assessment system including a coaching log and tool data, the evaluation team annually examined each site's fidelity to the NTC key components. Additionally, SRI measured the extent to which the NTC model as implemented in treatment schools differed from the business-as-usual supports that new teachers received in control schools. The level of implementation fidelity and treatment-control differences helped indicate whether to expect an impact of the NTC induction model on teacher and student outcomes.

Implementation Fidelity

The fidelity of implementation analysis comprised four key program components: (1) NTC supports for the sites, (2) selection and assignment of high-quality mentors, (3) mentor development and accountability, and (4) provision of high-quality mentoring. Each component comprised three to eight indicators, each with defined thresholds for full (i.e., high), medium, and low implementation. Each site received a fidelity score for each indicator, and indicator-level scores were combined to create a site-level score for each key component. Each site's component scores were aggregated across all three sites for a program-level score.

The results across the 3 years showed high implementation fidelity for all sites. The sites improved their implementation of Component 3, mentor development and accountability, and Component 4, provision of high-quality mentoring, which had been scored as "medium" across the three sites in the first year (2013–14). This level of implementation fidelity in the first year was not

surprising, representing typical challenges of organizing the new induction strategy for newly selected and trained mentors and establishing relationships with schools and beginning teachers in that first year. In the second and third years (2014–15 and 2015–16), implementation fidelity was high on all key components, reflecting local focus and growth on the indicators central to NTC induction model.

Treatment–Control Contrast

On the annual surveys of teachers in treatment and control schools, treatment teachers consistently reported more robust induction supports than control teachers. Treatment teachers were more likely to report having a mentor than control teachers. Of those teachers who reported having mentors, treatment teachers met with mentors more frequently and for more time than control teachers and focused more on instruction during their meetings with mentors. Treatment teachers also rated the value of mentoring activities higher than control teachers and were more likely than control teachers to report that the induction supports they received helped them grow as teachers. These multiple measures of beginning teachers’ induction experiences indicate that the NTC induction model indeed provided substantially different supports and experiences to treatment teachers from those reported by control teachers.

Teacher Impact Findings

The evaluation examined the extent to which the NTC induction model had an impact on teacher instructional practices and teacher retention in the RCT districts.¹

Teacher Practice

To determine whether participating in the NTC induction model for 2 years resulted in better teaching practices, the evaluation team measured teacher practice outcomes through structured classroom observations using the Framework for Teaching (Danielson, 2013). Teachers of core subjects (mathematics, reading/English language arts, social studies, science, or self-contained elementary classrooms) in treatment and control schools were randomly selected and observed at two time points (baseline—at the start of their first year of teaching—and at the end of their second year of teaching).

The evaluation found no statistically significant differences between observed treatment and control teachers on the four measures of Domain 2: Classroom Management and the four measures of Domain 3: Instruction. However, because of attrition over time, the number of schools remaining in the analysis sample was relatively low, as was the number of teachers in each school, even when both cohorts and RCT districts were combined. The reduced sample size limited our ability to detect the effects of the NTC model on teacher practice using the Framework for Teaching, particularly if those effects were small or variability in practice among teachers was considerable.

Teacher Retention

Using district administrative data, SRI assessed the impacts of the NTC induction model on teachers’ retention into their third year of teaching. Across both cohorts, 79 percent of treatment teachers and 78 percent of control teachers in the RCT district were retained; the difference was not statistically significant. The retention rates across both treatment and control teachers were lower than those found in a national sample of teachers beginning teaching in 2007–08, among whom 85 percent remained in teaching 3 years later (Gray & Taie, 2015). This difference raises the

¹ Teacher instructional practice could not be measured at baseline in the QED study because the comparison cohort began teaching before the start of the NTC grant. The teacher retention analysis from the QED site is a purely descriptive off-year comparison; therefore, we conducted only descriptive, not causal, analysis to inform NTC.

possibility that local factors and/or more recent trends may be influencing retention patterns that induction might not address.

Student Impact Findings

We examined whether the student achievement of teachers participating in the full NTC induction model for 2 years improved, specifically in English language arts (ELA) and mathematics among students in grades 4 through 8.² We used the Florida State Assessment (FSA) for BCPS and the Measures of Academic Progress (MAP) for CPS; CPS administered the MAP to have a consistent assessment that bridged the years during which Illinois switched state tests. For GWAEA, we used the state test, the Iowa Assessment.

RCT Sites ³

The evaluation team found that NTC's induction program had overall positive effects on student achievement in ELA and mathematics in the two RCT districts (Exhibit ES-1).⁴ The students in NTC-supported teachers' classroom for 1 year during the teachers' second year of support demonstrated higher achievement than students of teachers in the control group. In ELA, the average student achievement of teachers in the second year who participated in NTC induction for 2 years was approximately 0.05, compared with -0.04 for students of control teachers. This difference equals an effect size of 0.09 standard deviation ($p < .05$)—equivalent to moving from the 48th to the 52nd percentile—and represents the equivalent of approximately 2 to 3.5 additional months of learning, depending on the student's grade level (Lipsey, Puzio, Yun, Hebert, Steinka-Fry, Cole, et al., 2012).

In mathematics, students in grades 4 through 8 of teachers in the second year who participated in NTC induction for 2 years scored 0.15 standard deviation ($p < .01$) higher on average than students of control teachers. These impacts are equivalent to moving from the 46th to the 52nd percentile and represent the equivalent of approximately 2.4 to 4.5 additional months of learning, depending on the student's grade level.

Exhibit ES-1. Second-Year Impact on Student Achievement, Combined RCT Sites

Subject	Adjusted Mean Test Scores		Difference (effect size)	Students	Sample Sizes	
	Treatment	Control			Teachers	Schools
ELA	0.05	-0.04	0.09*	6,147	149	99
Math	0.06	-0.09	0.15**	4,972	129	86

Note: The effect on student achievement is a 1-year effect as the districts provided current and prior achievement data annually but did not consistently provide identifiers to link students across the data sets given to researchers each year.

The 1-year impact after 2 years of mentoring includes achievement in 2014–15 for Cohort 1 teachers and 2015–16 for Cohort 2 teachers.

Adjusted mean test scores are in standard deviation units.

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

² Students in third grade take state assessments in Florida, Illinois, and Iowa. The third-grade scores serve as the measure of prior achievement for fourth-grade students. As the lowest tested grade, however, third-grade students do not have a measure of prior achievement and could not be included in the analysis. Fourth grade was the lowest grade that we could include in the sample.

³ SRI released a findings brief in June 2017 with the student achievement results from the RCTs. Schmidt, R., Young, Cassidy, L., Wang, H., & Laguarda, K. (2017, June). *Impact of the New Teacher Center's new teacher induction model on teachers and students*. Menlo Park, CA: SRI International. <https://www.sri.com/work/publications/impact-new-teacher-centers-new-teacher-induction-model-teachers-and-students>

⁴ District results varied; see Appendix F for methods and district results.

QED Site

In the quasi-experimental study, SRI used a difference-in-differences approach to estimate the impact of participating in the 2-year NTC induction program. The study compared the difference in student achievement between beginning teachers who started teaching in 2013–14 and received NTC induction support for 2 years and a cohort of comparison beginning teachers who started teaching in 2012–13 and did not receive NTC induction support with the difference in the student achievement of veteran teachers in the same years.⁵

The impact estimate for teachers beginning teaching in 2013–14 and in their second year of induction support was not statistically significant, suggesting no detected NTC impact on this cohort of teachers in the QED site. However, the sample size of beginning teachers teaching ELA or mathematics in grades 4 through 8 that resulted from the participating districts' hiring patterns and testing schedules was very small, with 8 comparison and 19 treatment teachers in the ELA analysis and 7 and 23, respectively, in the mathematics analysis. The QED was extremely constrained in being able to detect any effects. As a result, the QED results are inconclusive—we do not know whether the NTC induction model had an impact in the QED site and these results should be interpreted with caution.

Conclusions and Implications

The high implementation fidelity levels and contrasts in induction experiences between treatment and control teachers indicate that the NTC induction model can be implemented well in a range of district contexts and even during times of budget cutbacks, as was the case in CPS. NTC induction did not yield differences in teacher practice as measured through classroom observations, although the sample sizes were small, and in teacher retention rates between treatment and control groups.

The positive student impacts in the RCT sites suggest that the NTC induction model can improve the ELA and mathematics achievement of students in beginning teachers' classrooms. The QED using a differences-in-differences approach did not bear out positive impacts on student outcomes, but it was limited by the small sample size and we do not know statistically whether the NTC induction model had an impact in the QED site.

The mixed results of positive impact on student outcomes but not on teacher practices warrants further investigation. The lack of impact on teacher practices was most likely due to attrition and small sample size. In addition, it is possible that the measures of teacher practice were not fine-grained enough to capture the nature of NTC effects on instruction.

Building on these results under an i3 Scale Up grant, NTC is currently implementing its model in five urban districts across the country and SRI is conducting RCTs in each district. Although NTC successfully achieved high implementation fidelity under the i3 Validation grant, scaling up to more districts and more diverse contexts necessitated adaptations to enhance sustainability and applicability. The evaluation of the i3 Scale Up grant will examine further whether and to what extent the NTC induction model incorporating certain adaptations, such as school-based and part-time mentors and classroom video tools, can achieve high implementation fidelity in larger and more diverse district settings. It will also explore whether, across these varying contexts, the NTC induction model has positive effects on teacher practice, teacher retention, and student achievement.

⁵ When comparing different cohorts of new teachers/their students across years, the intervention and comparison conditions are completely aligned with different time periods, and the estimated impact is confounded with policy or environmental changes from one year to the next that might have affected achievement. By including the veteran teachers from each time period as extra comparison groups for the intervention and comparison new teachers, respectively, this difference-in-differences design attempts to address this issue by controlling for changes that may have occurred between time periods, therefore eliminating the confounding with time.

Few professions demand that from their first days on the job, novices perform the same duties at the same level as seasoned veterans. The teaching profession routinely does. In the first years of their careers, teachers experiment—and often struggle alone—with managing student behavior, mastering the curriculum, engaging students in their own learning, pacing activities, and attending to differences in how students learn and to their diverse academic and social needs. Reformers have argued that during their formative years, new teachers also set the foundation for habits and dispositions that persist through their careers (Snyder & Bristol, 2015). Concerns about instructional quality and students’ resulting academic performance (Hanushek, 1992; National Commission on Teaching and America’s Future, 1997, 2016; Sanders & Rivers, 1996), as well as turnover (Smith & Ingersoll, 2004), further underscore the need for robust induction supports for beginning teachers.

The New Teacher Center (NTC), headquartered in Santa Cruz, California, was at the forefront of developing comprehensive induction strategies as California put in place a statewide induction policy, Beginning Teacher Support and Assessment (BTSA), in the late 1990s. Over the years, NTC has refined a comprehensive mentor-based induction model and served beginning teachers in districts across the country. NTC received an Investing in Innovation (i3) Validation grant in 2013 to implement its induction model in three sites: Broward County Public Schools in Florida, Chicago Public Schools in Illinois, and the Grant Wood Area Education Agency, a consortium of rural districts in Iowa. Through the grant, NTC trained full-time released mentors and served two cohorts of beginning teachers for 2 years each across the three sites.

SRI Education conducted the evaluation of NTC’s induction model under the i3 Validation grant. The evaluation featured a rigorous mixed-methods design to measure implementation fidelity and impact on teacher and student outcomes in the three participating sites. To accommodate local program needs, SRI conducted randomized controlled trials in Broward County Public Schools and Chicago Public Schools and a quasi-experimental study in Grant Wood Area Education Agency. The evaluation team used multiple measures to capture implementation fidelity and provide timely feedback to NTC and to the sites and to inform the outcomes analyses.

This report begins with an overview of NTC’s induction program, the study design, and evaluation activities. It then presents results from the implementation study, gleaned from analyses of implementation fidelity and the degree of contrast in mentoring that teachers in the treatment group received compared with that of teachers in the control group. Finally, the report examines the effect of the NTC induction model on teacher outcomes, including retention and classroom practice, and on student achievement in English language arts and mathematics after 2 years of induction support for teachers.⁶ Comprehensive methods description with supporting tables are in the appendices.

Program Description

NTC has long worked with district partners to implement a high-quality mentoring and induction program. Under the i3 Validation grant, NTC formalized key components of its induction model. NTC provides professional development, research-based resources, and online formative

⁶ SRI released a findings brief in June 2017 with the student achievement results from the RCTs. Schmidt, R., Young, Cassidy, L., Wang, H., & Laguarda, K. (2017, June). *Impact of the New Teacher Center’s new teacher induction model on teachers and students*. Menlo Park, CA: SRI International. <https://www.sri.com/work/publications/impact-new-teacher-centers-new-teacher-induction-model-teachers-and-students>

assessment tools for beginning teachers, mentors, and school leaders, as well as technical assistance and capacity building for program leaders.

Logic Model

As depicted in the logic model (Exhibit 1), the NTC induction model featured carefully selected full-time mentors housed in district-level teacher development offices. These mentors received more than 100 hours of training annually from NTC program staff, during institutes, and through in-field support from local induction program leaders and lead coaches. The mentors, who were supervised centrally, supported first- and second-year teachers in multiple schools at a ratio of 15 beginning teachers to 1 mentor. New teachers received 2 years of coaching, meeting with their assigned mentors weekly for a minimum of 180 minutes per month. Mentors and teachers worked through a system of NTC-developed online formative assessments, including tools to guide observation cycles and to develop teachers' skills in planning lessons and analyzing student work. These features are consistent with Ingersoll and Strong's (2011) meta-analysis identifying characteristics of high-quality induction, including mentoring, common and regular planning time with the mentor, and mentor training, similarly cited by Hobson et al. (2009).⁷ The three components of NTC induction—mentor selection and assignment, mentor development and accountability, and high-quality mentoring—were intended to improve instructional effectiveness, increase teacher retention, and ultimately improve student achievement (Exhibit 1, far right boxes). Supporting these three components are NTC's efforts to build district capacity to sustain teacher induction over the long term and resources, tools, and convenings led by NTC's national office (Exhibit 1, far left box). Key district and school conditions provided the context for mentors' opportunities to work with beginning teachers and shaped teachers' evolving practice (Exhibit 1, top box).

Several elements distinguished the NTC induction model from traditional district mentoring programs. The teacher to mentor ratio was intentionally low to enable mentors to work with new teachers frequently, intensively during each meeting, and consistently during the school year. The induction model encompassed the first 2 years in the classroom, when novices need to rapidly master classroom management and pedagogical skills and build the foundation for a sustainable teaching career. It is at this time that they are also at high risk of leaving the profession, which sustained induction support is intended to mitigate (Smith & Ingersoll, 2004).

Comprehensive mentor training and a system of formative assessment tools shaped the content and quality of mentor-teacher interactions under the NTC induction model. Mentors received a series of 12 professional learning days over 2 years.⁸ These Mentor Academies introduced the mentors to key tenets of coaching (e.g., taking a collaborative stance with beginning teachers, focusing on equitable instruction), mentoring skills (e.g., observing and giving feedback), and formative assessment tools intended to aid mentors in enacting those tenets in their work with new teachers. Mentor Academies developed mentors' expertise in identifying effective teacher practice, using data to inform instruction, creating classroom conditions to foster equitable learning, supporting language development, and differentiating instruction for diverse learners.⁹ The training focused mentors' interactions with beginning teachers on instruction, in contrast to mentoring that often provides logistical and emotional support for teachers and lacks sufficient emphasis on the instruction demanded by high content standards (Wang & Odell, 2002). Monthly Mentor Forums provided additional opportunity for mentors to reflect on how effectively they used

⁷ NTC induction was one of the programs in the 15 studies Ingersoll and Strong (2011) reviewed.

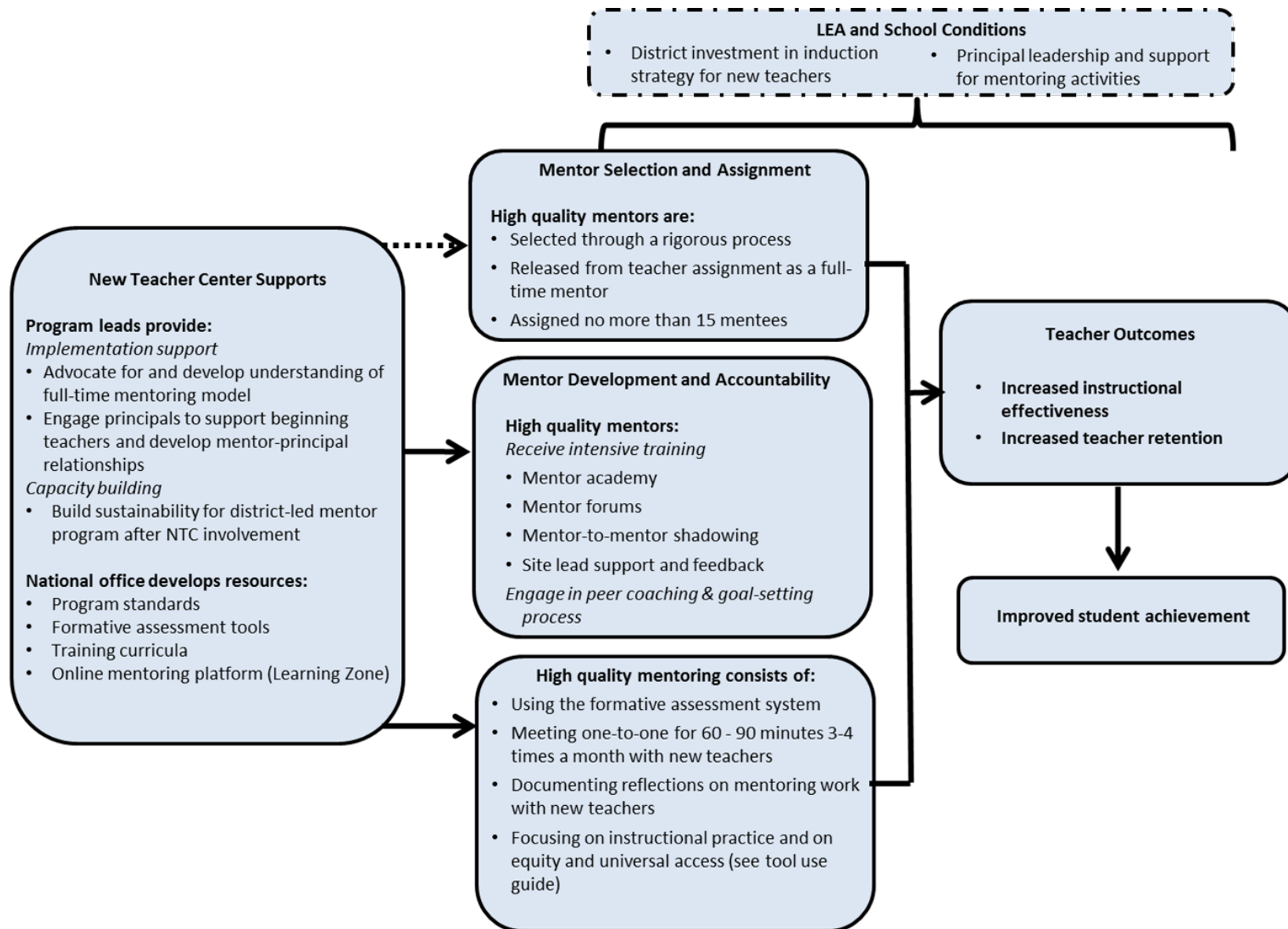
⁸ NTC has since revised its mentor training and now offers the 8-day Professional Learning Series (PLS) for Mentors, as well as Mentor Forums and in-field coaching.

⁹ See, for example, <https://newteachercenter.org/wp-content/uploads/Professional-Learning-Series-Product-Sheet.pdf>

tools aligned with specific instructional strategies, to raise questions about how to use particular tools, and to brainstorm and problem-solve with fellow mentors on how to meet specific teachers' needs or common needs across many teachers. In-field coaching, where a lead coach observed a mentor working with a beginning teacher and provided the mentor with feedback on that interaction, further supported mentors in refining their coaching practice and instructional support.

The formative assessment system was a comprehensive set of NTC-developed, instructionally focused tools and protocols. The tools support mentors in structuring their mentoring sessions with beginning teachers and ensure that the conversation and activities drove toward a specific instructionally focused objective for that coaching session. NTC highlighted lesson planning, guiding purposeful classroom observations and conferencing, and analyzing student work as high-leverage skills with aligned tools. All tools resided in NTC's online system, Learning Zone, in which mentors or new teachers entered the content as they worked through a tool. Learning Zone provided summary data on mentoring frequency and duration, numbers of tools mentors used, and which tools they used during each mentor-new teacher meeting. Additionally, mentors and teachers together could reference their prior work with the tools to understand their progress. NTC provided Learning Zone data monthly to the sites as ongoing feedback on the extent to which mentors were meeting expectations in mentoring frequency, duration, and tool use.

Exhibit 1. Logic Model for the New Teacher Center i3 Validation Grant



Implementing Sites

Three sites participated in the study—Broward County Public Schools (BCPS) in Florida, Chicago Public Schools (CPS) in Illinois, and Grant Wood Area Education Agency (GWAEA), a consortium of mainly rural districts in Iowa. Through the grant, NTC provided funds to hire full-time released mentors in each site and site-based lead coaches responsible for mentor training and support at their respective sites. The grant also covered district administrators' time in overseeing program implementation and data collection at the sites. In each site, the evaluation followed two cohorts of new teachers for 2 years each—Cohort 1 began teaching in 2013–14 and Cohort 2 in 2014–15—for a total implementation period of 3 years (2013–14 through 2015–16). Designated NTC staff members served as site-specific client liaisons to communicate with site executives and administrators, integrate induction into the sites' overall teacher development strategy, and build local capacity to sustain induction over the long term.

In the urban sites (BCPS and CPS), the grant funds provided the capacity to serve new teachers in a subset of schools, capped at 15 new teachers for 2 years each per funded mentor. Fifteen mentors were funded throughout the implementation period in CPS; BCPS had 9 mentors in the first year of implementation, and the number increased to 15 in the second year. In GWAEA, the participating districts were relatively small and had fewer schools, so they pooled resources to support mentors to work with beginning teachers across the districts. Any given district in the consortium had few new teachers each year, so the funded mentors in GWAEA served all new teachers during the implementation period.

These differences in program implementation necessitated two approaches to evaluating effectiveness. Because the number of new teachers in BCPS and CPS exceeded the capacity for induction support under the grant, we were able to use randomized controlled trials (RCTs), with random assignment as the mechanism for allocating NTC induction support to schools. In GWAEA, random assignment was not practical, so we adopted a quasi-experimental design (QED) to determine the effectiveness of NTC mentoring on teacher and student outcomes.

Within each site, the characteristics of the schools in the study generally reflected those of the district or the consortium, in the case of GWAEA, overall, with some slight differences. On average, the CPS study schools had higher proportions of English learners than the district overall and had higher school ratings. The BCPS study schools had slightly higher average percentages of students eligible for free or reduced-price lunch and racial/ethnic minority students than the district overall (Exhibit 2). In GWAEA, study schools tended to have slightly higher proportions of students eligible for free or reduced-priced lunch, and treatment schools in the student outcomes analysis were lower performing on average in English language arts (ELA) and mathematics compared with the state (Exhibit 3).

Exhibit 2. Average District and Study Sample Characteristics at the Time of Random Assignment, RCT Districts

		No. of Schools ^a	No. of Teachers in Study	School Rating	% English Learners	% Special Education	% Free or Reduced- Price Lunch Eligible	% Minority
BCPS	Treatment	43	193	2.2	10	13	71	81
	Control	44	148	2.6	11	16	68	79
	District	213	--	2.7	12	13	66	77
CPS	Treatment	65	149	2.4	22	13	84	92
	Control	75	139	2.5	24	12	84	88
	District	536	--	2.1	14	14	84	91

Note: "School Rating" refers to the state report card of quality ratings applied to each school. At the time of random assignment, BCPS assigned all schools a letter grade (A to F), and CPS used whole numbers between 1 and 3. These ratings were put on a common scale where 0 = F in BCPS and 3 in CPS, 1 = D in BCPS, 2 = C in BCPS and 2 in CPS, 3 = B in BCPS, and 4 = A in BCPS and 1 in CPS.

^a School count does not include high schools.

Source: <http://cps.edu/SchoolData/Pages/SchoolData.aspx> (CPS); <http://www.broward.k12.fl.us/dsa/counts/1213/20DayCount1213.shtml>; and schoolgrades.fldoe.org/xls/1213/SGbasic_2013.xls (BCPS)

Exhibit 3. Average Consortium and Study Sample Characteristics, Year Before Intervention, QED Site

	No. of Schools	No. of Teachers in Study ^a	No. of New Teachers in Study	% Passing ELA	% Passing Math	% English Learners	% Special Education	% Free or Reduced- Price Lunch Eligible	% Minority
Treatment sample in student outcomes analysis	24	159	34	67	69	2	16	39	13
Control sample in student outcomes analysis	14	95	16	73	77	1	14	35	10
Treatment group with new teachers	57	112	112	77	72	2	14	37	12
Control group with new teachers	51	113	113	78	73	1	12	36	10
Consortium ^b	88	NA	NA	NA	NA	1	--	33	11

Note: Data for Cohort 1 teachers and their schools only.

State-level passing rate was 70.6% for reading and 76.8% for math across grades 3 through 8.

^a Number of teachers in study includes veteran comparison teachers.

^b Consortium data do not include high schools. Special education data were not available at the consortium level.

Source: <https://portal.ed.iowa.gov/iowalandingpage/Landing.aspx>, and additional individual school-level data from Iowa Department of Education.

Study Design Summary

The i3 Validation grant requires that evaluations include an implementation study to examine implementation fidelity in the participating sites and an impact study to determine whether the program as implemented had an impact on student outcomes and relevant teacher outcomes.

Implementation Study

In the implementation study, we examined the extent to which each site implemented the full induction program as described in the NTC logic model (Exhibit 1) and aggregated implementation fidelity scores across the three sites, addressing the research question: *What is the level of implementation fidelity to the NTC model in the three participating sites?*

Implementation Fidelity

We measured fidelity of the implementation of the four key program components depicted in the logic model: (1) NTC supports for the sites, (2) selection and assignment of high-quality mentors, (3) mentor development and accountability, and (4) provision of high-quality mentoring. Each component comprised three to eight indicators, which we developed in collaboration with NTC staff. Some indicators were measured at the site level, others at the individual level (e.g., by principal, mentor, or new teacher). For each indicator, SRI worked with NTC staff to set thresholds for full (i.e., high), medium, and low implementation. Each site received a fidelity score for each indicator, and the scores were combined to create a site-level score for each key component using the following rules:

- High implementation fidelity—60 percent or more of the indicators were scored as high, and no more than 20 percent of the indicators were scored as low.
- Medium implementation fidelity—Individual indicator scores did not reach the threshold for high fidelity, and less than 50 percent of indicators were scored as low.
- Low implementation fidelity—50 percent or more of the indicators were scored as low.

At the program level (i.e., across all sites in the study), NTC achieved implementation with fidelity under each key component if at least two sites achieved high implementation and no site achieved low implementation.

Multiple data sources were required to measure the constituent indicators for each component (Exhibit 4); we collected implementation data and calculated fidelity measures annually.

Exhibit 4. Data Sources for Implementation Analysis of Key Components

Component	No. of Indicators	Data Sources
NTC supports	Year 1: 7 Years 2 & 3: 8	Attendance log at half-day principal training; logs of one-on-one meetings between site leads and principals; copies of program standards, formative assessment tools and mentor training materials; Learning Zone data
Mentor selection and assignment	3	Mentor application materials; mentor survey; rosters of teacher assignments to mentors
Mentor development and accountability	7	Attendance log at mentor academies and mentor forums; logs of mentor-to-mentor shadowing, meetings between site leads and mentors, peer coaching, and goal setting
Provision of high-quality mentoring	5	Learning Zone data; teacher survey

Appendix A contains a full matrix that defines the specific implementation measures, data sources, and fidelity scores across the three sites for the 3 implementation years.

Teacher Survey

Annual spring surveys of the new teachers being served by NTC-trained mentors and their counterparts in control schools provided measures of the extent to which induction activities fundamental to the NTC induction model differed between treatment and control schools (treatment-control contrast). The survey contained items about the mentoring new teachers received (frequency and duration of meetings, focus of mentors' work with new teachers), other kinds of induction supports, school environment, and beginning teachers' self-evaluation. To understand treatment-control contrast across the 3 years of the program, we analyzed the surveys administered between spring 2013 and spring 2016. The analysis combined teachers across cohorts in their first and second years of teaching as shown in Exhibit 5, which matched the impact analyses combining both teacher cohorts in the RCT sites.

Exhibit 5. School Year by Cohort and Years of Experience

New Teacher Cohort	Years of Teaching Experience	
	1	2
1	2013–14	2014–15
2	2014–15	2015–16

A total of 860 first-year teachers across both cohorts and 660 second-year teachers across both cohorts in all three sites responded to the survey. Overall, 90 percent of treatment teachers and 64 percent of control teachers responded to the survey (Exhibit 6). The survey sample sizes and response rates by site are presented in Appendix B.

Exhibit 6. Overall Survey Response Rates by Years of Teaching Experience, Combined Cohorts

Site	Years of Experience—Both Cohorts						Overall		
	Treatment	Year 1 ^a Control ^b	Subtotal	Treatment	Year 2 Control	Subtotal			
Surveyed	608	417	1025	518	373	891	1126	790	1916
Responded	563	297	860	450	210	660	1013	507	1520
Response rate	93%	71%	84%	87%	56%	74%	90%	64%	79%

Source: NTC New Teacher Survey, spring 2013–2016.

^a Attrition data were not available for the Year 1 calculations.

^b The control group was administered the teacher survey 1 year before the treatment group in GWAEA.

Impact Studies

As noted, because of differences in local contexts we used two approaches to examine the impact of NTC's induction model, RCTs in BCPS and CPS and a QED in GWAEA.

Randomized Controlled Trials in BCPS and CPS

The RCTs in CPS and BCPS featured school-level random assignment to estimate the impact of the NTC model on teacher and student outcomes. In both districts, we randomly assigned a sample of participating schools employing beginning teachers in summer 2013, before NTC began serving the new teachers. The schools in each district were blocked on grades served and the most relevant

local factor—geographic area in CPS and Teacher Incentive Fund status in BCPS. All schools were assigned and all teachers identified by October 1. Within each block, schools were assigned to the NTC program or to the usual district supports for new teachers until the target number of beginning teachers was reached (before October 1). In the second year (2014–15), all incoming first-year teachers in treatment schools were added to the treatment group and all incoming first-year teachers in control schools were added to the control group. To reach the target number of teachers served, previously unassigned schools in each district were assigned in summer 2014 following the established blocks if they employed beginning teachers before October 1. See Appendix C for assignment details.

In both sites, NTC served all new teachers in treatment schools unless they were served by other programs with induction support, such as Teach For America. Teachers covered under other induction programs were excluded from both the treatment and control conditions.

The RCTs in BCPS and CPS address the following research questions about NTC impacts:

1. Does participating in the full NTC induction model for 2 years result in better practices on eight components of teaching? (*Confirmatory*)
2. Does participating in the full NTC induction model for 2 years result in improved student achievement in reading and math among students in grades 4–8? (*Confirmatory*)
3. Does participating in the full NTC induction model result in improved teacher retention after 2 years? (*Exploratory*)¹⁰

Quasi-experimental Design in GWAEA

Sixteen districts in the GWAEA consortium agreed to participate in the study in 2013–14. Because all beginning teachers in the GWAEA participating districts began receiving NTC mentoring in 2013–14, we could not use random assignment to study impact. Instead, we used a quasi-experimental difference-in-differences approach to estimate impact on student achievement. That is, we compared beginning teachers in GWAEA participating districts who began teaching in 2012–13 and did not receive NTC mentoring with beginning teachers in the same districts who began teaching in 2013–14, when NTC mentoring was offered to all new teachers, adjusting for differences between veteran teachers in those years. We included veteran teachers in the analysis because when comparing different cohorts of new teachers/their students across years, the intervention and comparison conditions are completely aligned with different time periods, and the estimated impact is confounded with policy or environmental changes from one year to the next that might have affected achievement. By including the veteran teachers from each time period as extra comparison groups for the intervention and comparison new teachers, respectively, this difference-in-differences design attempted to address this issue by controlling for changes that may have occurred between time periods, therefore eliminating the confounding with time.¹¹ Therefore, to meet this standard, we could compare the differences only between the first cohort of NTC-

¹⁰ Confirmatory questions are those related to impact that the evaluation team defines a priori, before examining outcome data. Exploratory questions are not specified in advance, may be more responsive to program developers' information needs, and may change or be developed to investigate questions that arise after seeing findings.

¹¹ From communication between the National Evaluation of i3 staff and the SRI research team, this difference-in-differences approach has better validity and is more likely to meet the What Works Clearinghouse (WWC) standards when treatment and comparison groups are no more than 1 year apart, although WWC changed the standards for difference-in-differences approaches in 2014, after we had designed the impact study for GWAEA and the site had begun serving new teachers. SRI conducted student outcomes analysis for GWAEA teachers in the second cohort to inform NTC about its program. The results are in Appendix H and are not intended for WWC review.

served teachers in the 16 districts that volunteered to participate in the study in 2013–14 and the beginning teachers in 2012–13 in these districts who were not served by NTC.

The impact study in GWAEA addresses the research question: *Does participating in the full NTC induction model for 2 years result in improved student achievement in reading and math among students in grades 4–8?*

Data Collection Activities

Over the course of the study, we collected and analyzed data from multiple and varied sources to gain a comprehensive view of implementation and impact. Exhibit 7 indicates the data sources by purpose. Exhibits 8 and 9 detail the data collection activities by cohort, implementation year, and type of school for the two impact studies. For the implementation study, we analyzed Learning Zone data on mentoring activities for each year of implementation (2013–14 to 2015–16), NTC-administered teacher and mentor survey data, school- and district-level interviews, and extant data. For the impact studies, we conducted and analyzed classroom observations and teacher retention data from the RCT sites and collected and analyzed student-level achievement and demographic data from all participating sites.

Exhibit 7. Data Sources and Their Purposes

Data Source	Purpose					
	Randomly Assign Schools	Teacher Eligibility	Teacher Outcomes Analysis	Student Outcomes Analysis	Sensitivity and Follow-up Analyses	Implementation and Treatment-Control Contrast
School demographic and achievement data	✓		✓	✓	✓	✓
Human resources data		✓	✓	✓	✓	
Teacher observations			✓		✓	
Teacher, mentor, school leader, and district leader interviews						✓
Teacher and mentor surveys					✓	✓
Student demographics and achievement data				✓	✓	

Exhibit 8. Data Collection Activities by Cohort and Year for BCPS and CPS

Random Assignment Group	2013-14		2014-15				2015-16				2016-17	
	Cohort 1		Cohort 1		Cohort 2		Cohort 1		Cohort 2		Cohort 2	
	T	C	T	C	T	C	T	C	T	C	T	C
Learning Zone	✓		✓		✓				✓			
Teacher survey	✓	✓	✓	✓	✓	✓			✓	✓		
Mentor survey	✓		✓		✓				✓			
Interviews			✓		✓				✓	✓		
Classroom observations	✓	✓	✓	✓	✓	✓			✓	✓		
Student achievement & demographic data	✓	✓	✓	✓	✓	✓			✓	✓		
Human resources data	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Cohort 1 = Teachers beginning teaching in 2013–14.

Cohort 2 = Teachers beginning teaching in 2014–15.

T = treatment group.

C = control group.

Exhibit 9. Data Collection by Cohort and Year for GWAEA

	2012–13			2013–14			2014–15			2015–16		
	Com	C1	C2	Com	C1	C2	Com	C1	C2	Com	C1	C2
Implementation Data												
Learning Zone		--	--		✓	--		✓	✓	--	--	✓
Teacher survey	✓	--	--	✓	✓	--		✓	✓	--	--	✓
Mentor survey		--	--		✓	--		✓	✓	--	--	✓
Interviews		--	--			--		✓		--	--	✓
Impact Data												
Classroom observations	✓	--	--	✓		--		✓		--	--	
Student achievement & demographic data ^a	✓	--	--	✓		--	✓	✓		--	--	✓
Human resources data	✓	--	--	✓	✓	--	✓	✓	✓		✓	✓

Comparison (Com) = Cohort of teachers beginning teaching in 2012–13.

C1 = Cohort 1 treatment teachers beginning teaching in 2013–14.

C2 = Cohort 2 treatment teachers beginning teaching in 2014–15.

^a We also collected student achievement and demographic data for veteran teachers in the same years as the comparison cohort and Cohorts 1 and 2 for the difference-in-differences analysis.

IMPLEMENTATION FINDINGS

Broward County Public Schools in Florida, Chicago Public Schools, and the Grant Wood Area Education Agency in Iowa—the participating sites—had differing contexts that influenced how the NTC induction model was implemented locally. Therefore, understanding the extent to which each site was able to implement the key components of the NTC model given varying local conditions was the first step in establishing whether the overall model was implemented to a level of fidelity that would predict an impact on the target teacher and student outcomes.

Throughout the study, we provided NTC with information on implementation as we completed data analysis to help program leaders and sites identify specific areas for improvement, such as barriers to mentors being able to meet with their assigned beginning teachers regularly or supports that mentors and teachers might need to use the formative assessment tools well. In addition to measuring implementation fidelity and supporting program improvement, we collected and analyzed data from beginning teachers in treatment and control schools to understand any differences in their induction experiences. These treatment-control differences signaled whether to expect an impact of the NTC induction model on teacher and student outcomes compared with the status quo supports that control teachers received in their first and second years of teaching. This chapter presents the implementation fidelity data for all 3 years of implementation and discusses the evidence on the extent to which treatment and control teachers' induction experiences differed.

Implementation Fidelity

The NTC induction logic model identified four key components for which we measured implementation fidelity:

1. NTC supports—Eight indicators of the supports NTC provided in launching the program in each site. One indicator (capacity-building by site leads) was measured only in years 2 and 3.
2. Mentor selection and assignment—Three indicators addressing mentor hiring and allocation to new teachers.
3. Mentor development and accountability—Seven indicators of the site-level supports and training for mentors.
4. Provision of high-quality mentoring—Five indicators reflecting the joint activities mentors and beginning teachers engaged in and teachers' perceptions of the quality of their mentoring experience.

Exhibit 10 provides the scores for each component, aggregated across the three sites for a program-level score. The results pertain to the overall program serving both first- and second-year teachers in 2013–14, 2014–15, and 2015–16. Appendix A presents the full matrix that defines the specific measures, data sources, and fidelity scores across the sample for the 3 study years.

Exhibit 10. Implementation Fidelity by Key Component

Key Component	Number of Indicators	Year 2 (2014–15)					
		Year 1 (2013–14)		Cohort 1, second year teaching		Year 3 (2015–16)	
		Cohort 1, first year teaching		Cohort 2, first year teaching		Cohort 2, second year teaching	
		Number of Sites Meeting Fidelity Threshold	Fidelity, Program Level (High/ Medium/ Low)	Number of Sites Meeting Fidelity Threshold	Fidelity, Program Level (High/ Medium/ Low)	Number of Sites Meeting Fidelity Threshold	Fidelity, Program Level (High/ Medium/ Low)
1. New Teacher Center supports	8	High: 3 Medium: 0 Low: 0	High	High: 3 Medium: 0 Low: 0	High	High: 2 Medium: 1 Low: 0	High
2. Mentor selection and assignment	3	High: 3 Medium: 0 Low: 0	High	High: 3 Medium: 0 Low: 0	High	High: 3 Medium: 0 Low: 0	High
3. Mentor development and accountability	7	High: 0 Medium: 3 Low: 0	Medium	High: 2 Medium: 1 Low: 0	High	High: 2 Medium: 1 Low: 0	High
4. Provision of high-quality mentoring	5	High: 1 Medium: 2 Low: 0	Medium	High: 3 Medium: 0 Low: 0	High	High: 3 Medium: 0 Low: 0	High

Note: “High” implementation at program level (three sites) for each key component = at least two sites scored high and none scored low for that component.

The results across 3 years show that implementation fidelity was generally high. However, the implementation fidelity of Component 3, mentor development and accountability, and Component 4, provision of high-quality mentoring, was lower in the first year (2013–14), scoring “medium” across the three sites. This level of implementation fidelity was not surprising for the first year of program launch, when all sites were starting a brand-new program.

In the second and third years (2014–15 and 2015–16), implementation fidelity was high on all key components, showing local focus and growth on the indicators that define the NTC induction model. For Component 3 in particular, mentors in all three sites were more consistent in attending Mentor Academies and Mentor Forums. All sites met the definition for high implementation fidelity for that indicator (80 percent of mentors attended forums and academies offered). All three sites also improved in having mentors meet regularly with beginning teachers and using the NTC formative assessment tools during their meetings with mentees, driving improvement under Component 4.

Despite generally high implementation, some inconsistencies in the first three components in the third year (2015–16) reflected local challenges that were outside NTC’s control. For example, intensifying local budget constraints affected mentor workloads and retention and created uncertainty about continued funding for the induction program beyond the grant period. In the last year of implementation, the sites were less consistent in the following indicators under each component:

- Component 1, New Teacher Center supports—Engaging principals was less consistent, with two districts not meeting the high definition for the indicator relating to annual one-on-one meetings between site leads and principals.
- Component 2, mentor selection and assignment—Maintaining a caseload of 15 teachers to each coach was less consistent. Two districts met this caseload ratio for more than 90 percent of mentors in the second year, but in the third year two districts were able to meet this caseload ratio for only over 80 percent of mentors.
- Component 3, mentor development and accountability—Mentor-to-mentor shadowing and mentors’ receiving feedback from site leads was less consistent, reflecting the general pressure on mentors’ and program staff’s time.

Despite these challenges, teachers’ mentoring experiences remained of high quality. On Component 4, provision of high-quality mentoring, all sites met the standards for frequency and intensity of mentoring and for focus on instruction during mentoring. Program staff members and mentors placed a primacy on serving beginning teachers and preserving the level of attention they needed, choosing to scale back on some of the mentor supports and principal engagement instead.

Contrast in Induction Supports Between Treatment and Control Schools

Beyond establishing the level of implementation fidelity in treatment schools, putting the impact of the NTC model in perspective requires an understanding of any differences in the comprehensive induction that NTC provides in treatment schools with business-as-usual supports that districts and schools normally provide beginning teachers in control schools. Over the course of a multiyear study and particularly for a 2-year intervention such as the NTC induction model, the scope and the quality of status quo induction supports may change as implementing sites seek to improve their own programs, so the differences in induction experiences between teachers in treatment and control schools may fluctuate over time.

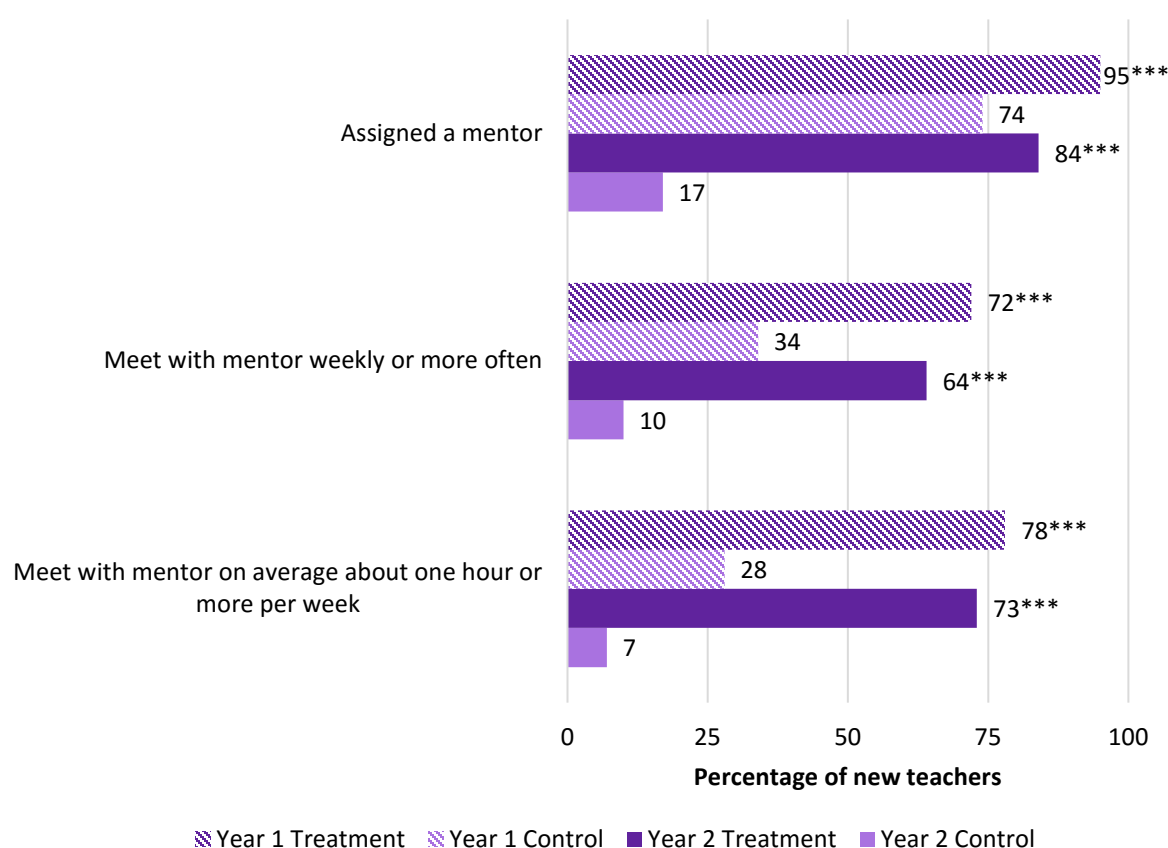
Characterizing this treatment-control contrast provides important context for understanding any impact of NTC induction on teachers’ practice and students’ achievement. Based on NTC’s logic

model, impacts would be expected to be greater in sites where NTC services are very different in scope and quality from the districts' status quo approach to induction. Similarly, where we find variations in the treatment-control contrast, we might also expect to see variations in impact.

To parallel and more accurately inform the impact findings, we conducted the analyses of sites in the RCT and QED separately. All results discussed here were statistically significant at the $p < .05$ level. Survey scale items are listed in Appendix B.

The survey results showed consistent differences between treatment and control teachers. Overall, treatment teachers were more likely to have a mentor than control teachers. In the RCT sites, treatment teachers were more likely than control teachers to have a formally assigned mentor in both years of teaching (Exhibit 11).¹² Treatment teachers across all three sites met with their mentors more frequently and for more time than control teachers (Exhibits 11 and 12).

Exhibit 11. New Teacher Interactions with Mentors, RCT Combined Sample, 2014–16



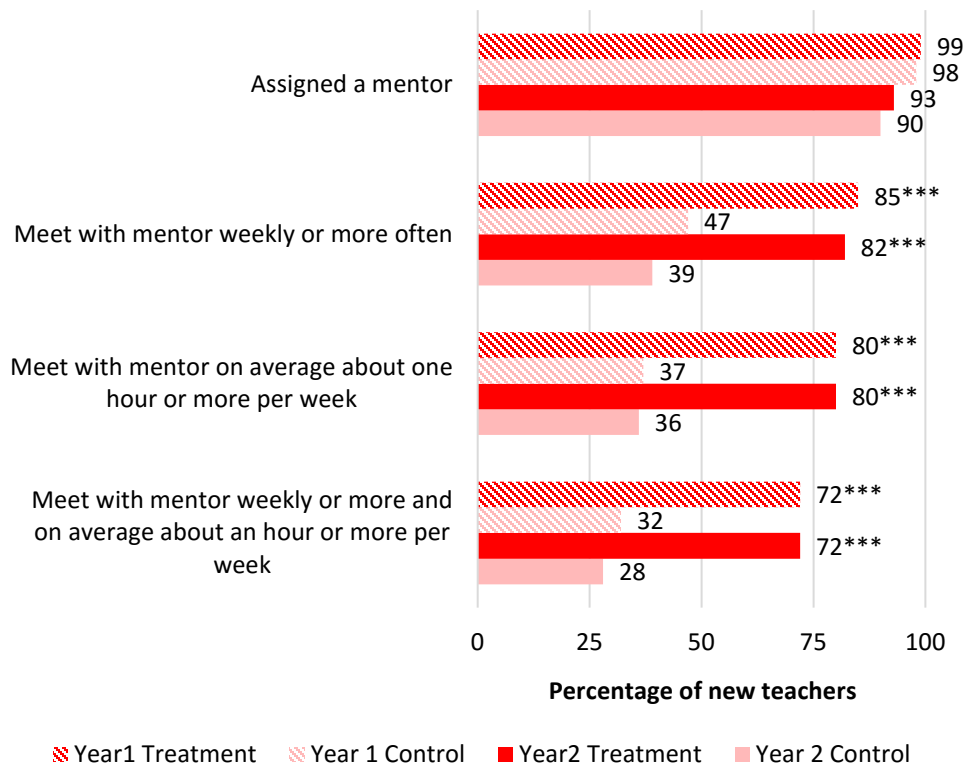
Source: NTC New Teacher Survey, spring 2014–2016.

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

Value labels are rounded to the nearest percentage point.

¹² Although all teachers in the treatment group were assigned mentors, a couple reasons might account for why only 95% of treatment teachers reported having a mentor. Self-report error is a potential risk in any survey; here, we assume that self-report error affects both treatment and control groups equally. Also, in a few rare cases reported through interviews, beginning teachers refused to meet regularly with their mentors and might have thus reported on the survey that they did not have a mentor.

Exhibit 12. New Teacher Interactions with Mentors, QED Site, 2013–16



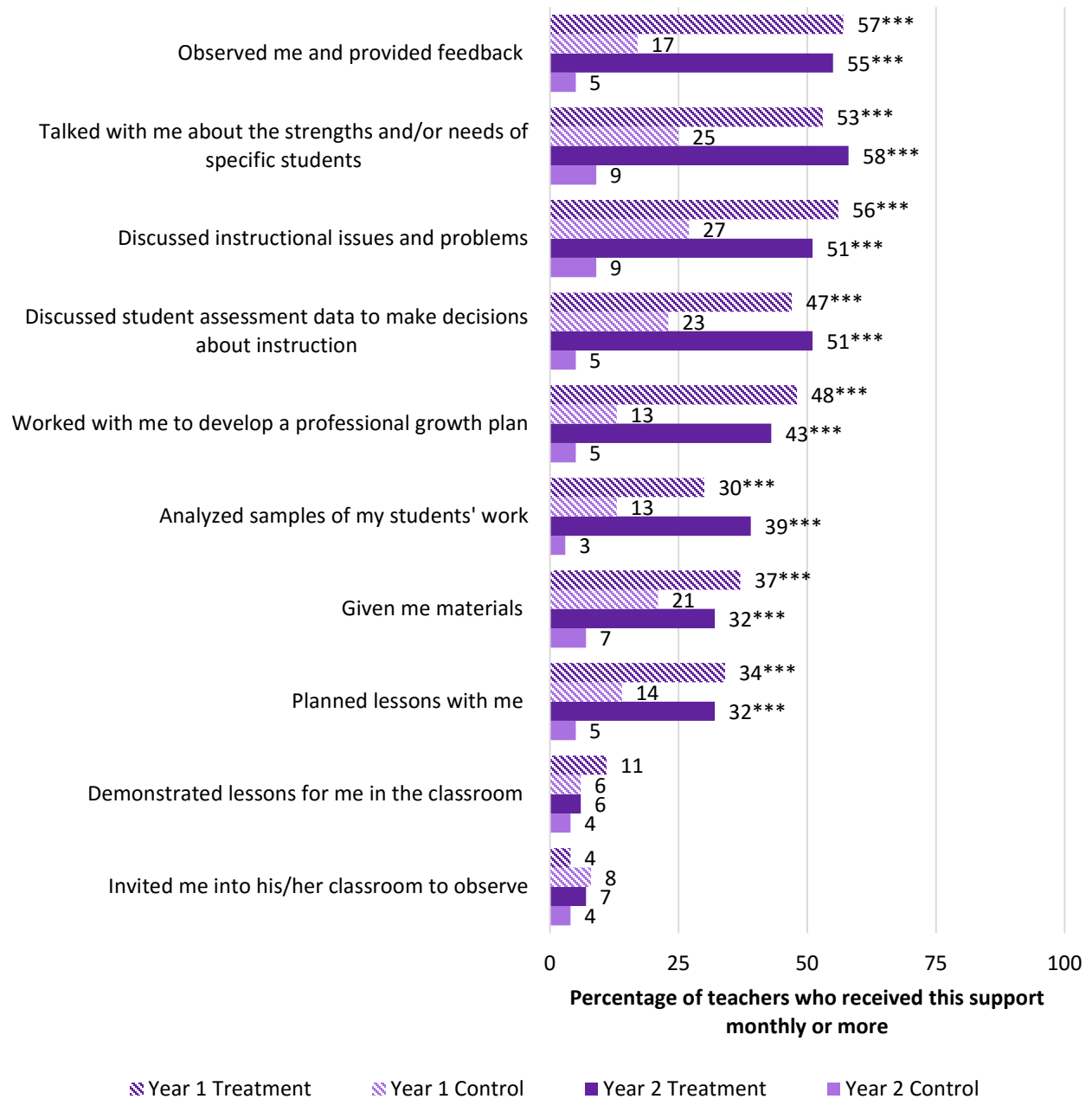
Source: NTC New Teacher Survey, spring 2013–2016.

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

Value labels are rounded to the nearest percentage point.

NTC mentors in all three sites and in both years worked with new teachers more consistently than non-NTC mentors in observing instruction and providing feedback, talking with teachers about the strengths or needs of specific students, discussing student assessment data to make decisions about instruction, and working with teachers to develop a professional growth plan (Exhibits 13 and 14). Interviews with treatment teachers corroborated that frequent and consistent mentoring mattered—teachers knew they had a knowledgeable and supportive colleague to rely on, who could observe and provide feedback regularly and get a feel for the classroom and respond to the teacher’s needs.

Exhibit 13. Frequency of Mentoring Activities, RCT Combined Sample, 2014–16

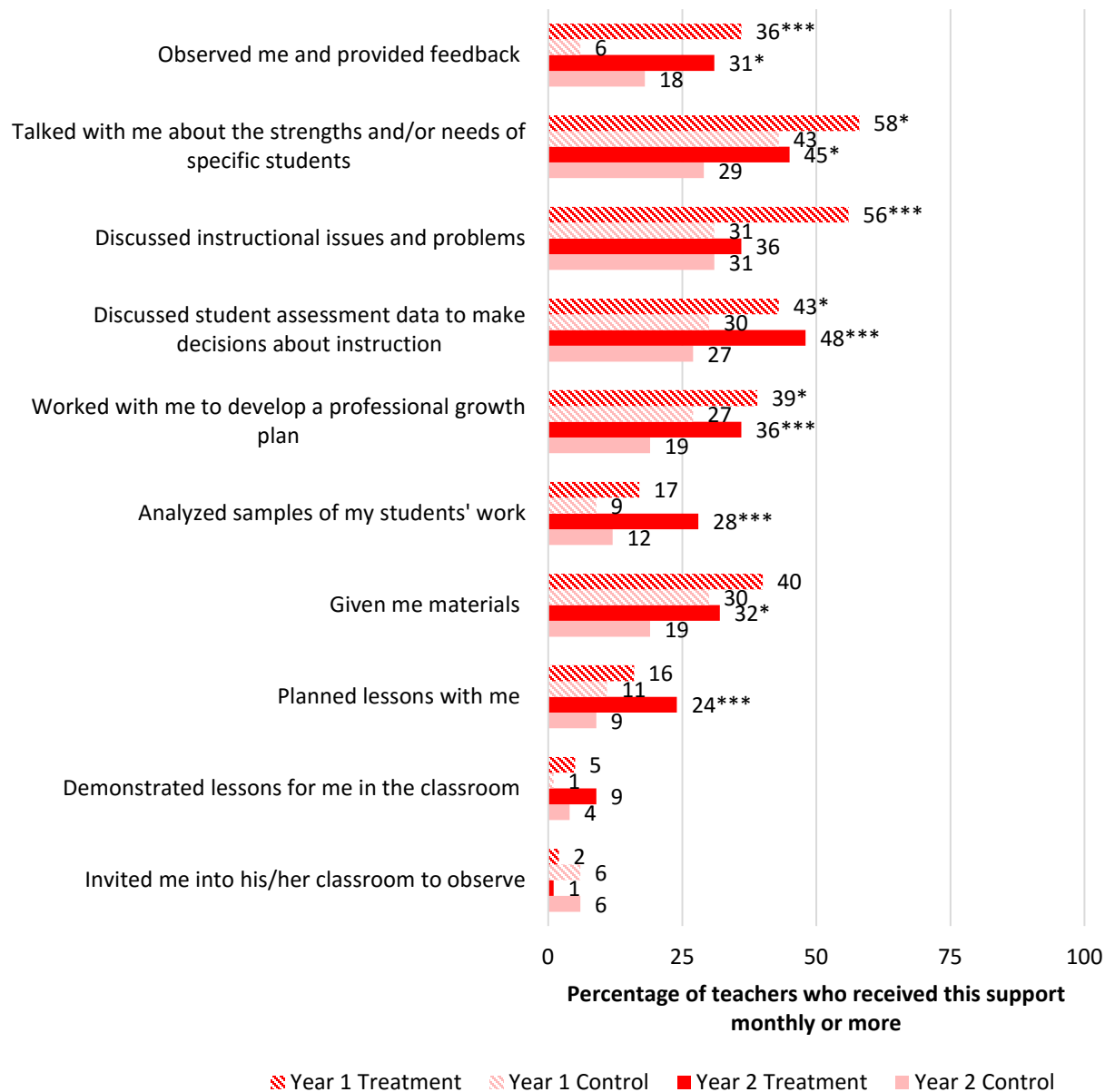


Source: NTC New Teacher Survey, spring 2014–2016.

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

Value labels are rounded to the nearest percentage point.

Exhibit 14. Frequency of Mentoring Activities, QED Site, 2013–16



Source: NTC New Teacher Survey, spring 2013–2016.

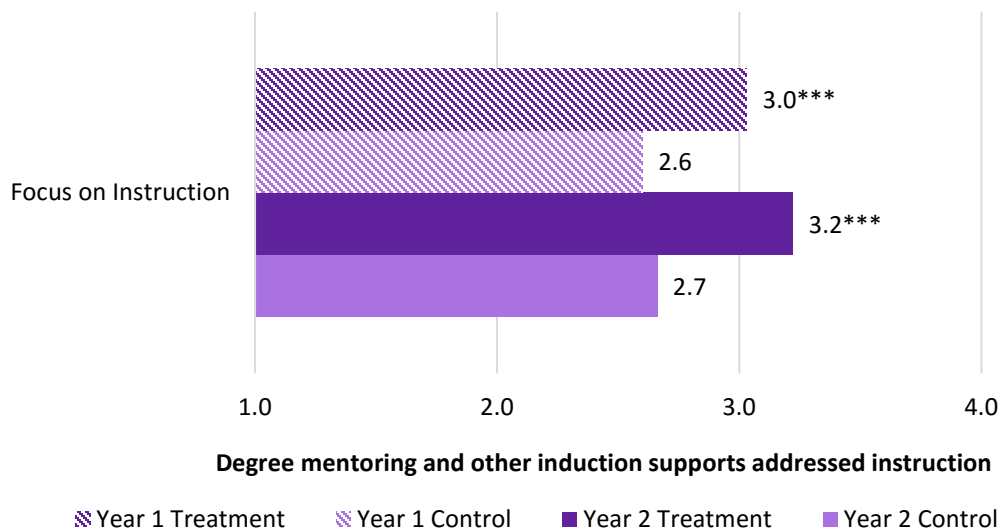
* $p < .05$, ** $p < .01$, *** $p < .001$.

Value labels are rounded to the nearest percentage point.

The mentoring treatment teachers received focused more on instruction than that control teachers received. Treatment teachers in all three sites and in both years were more likely than control teachers to report that induction supports addressed instruction (Exhibits 15 and 16). In the RCT sites in particular, NTC teachers were more likely to report that mentoring and other induction supports focused on evaluating and reflecting on their teaching practice, adapting

instruction to meet the needs of students at varying academic levels, instructional techniques appropriate to the grade level and subject matter they taught, and the use of assessment strategies in instruction (Exhibits 17 and 18).

Exhibit 15. Focus on Instruction in Mentoring and Other Induction Supports, RCT Combined Sample, 2014–16

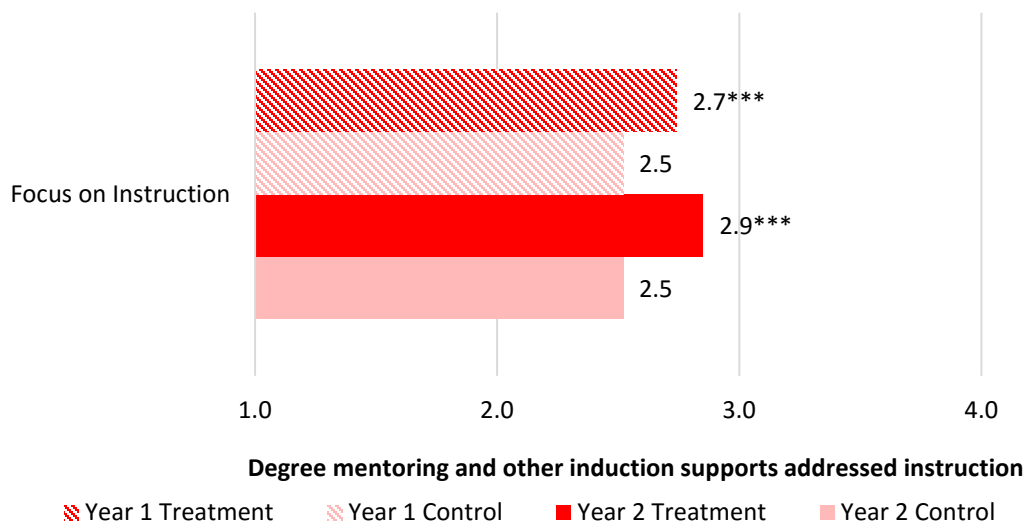


Source: NTC New Teacher Survey, spring 2014–2016.

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

Value labels are rounded to the nearest tenth.

Exhibit 16. Focus on Instruction in Mentoring and Other Induction Supports, QED Site, 2013–16

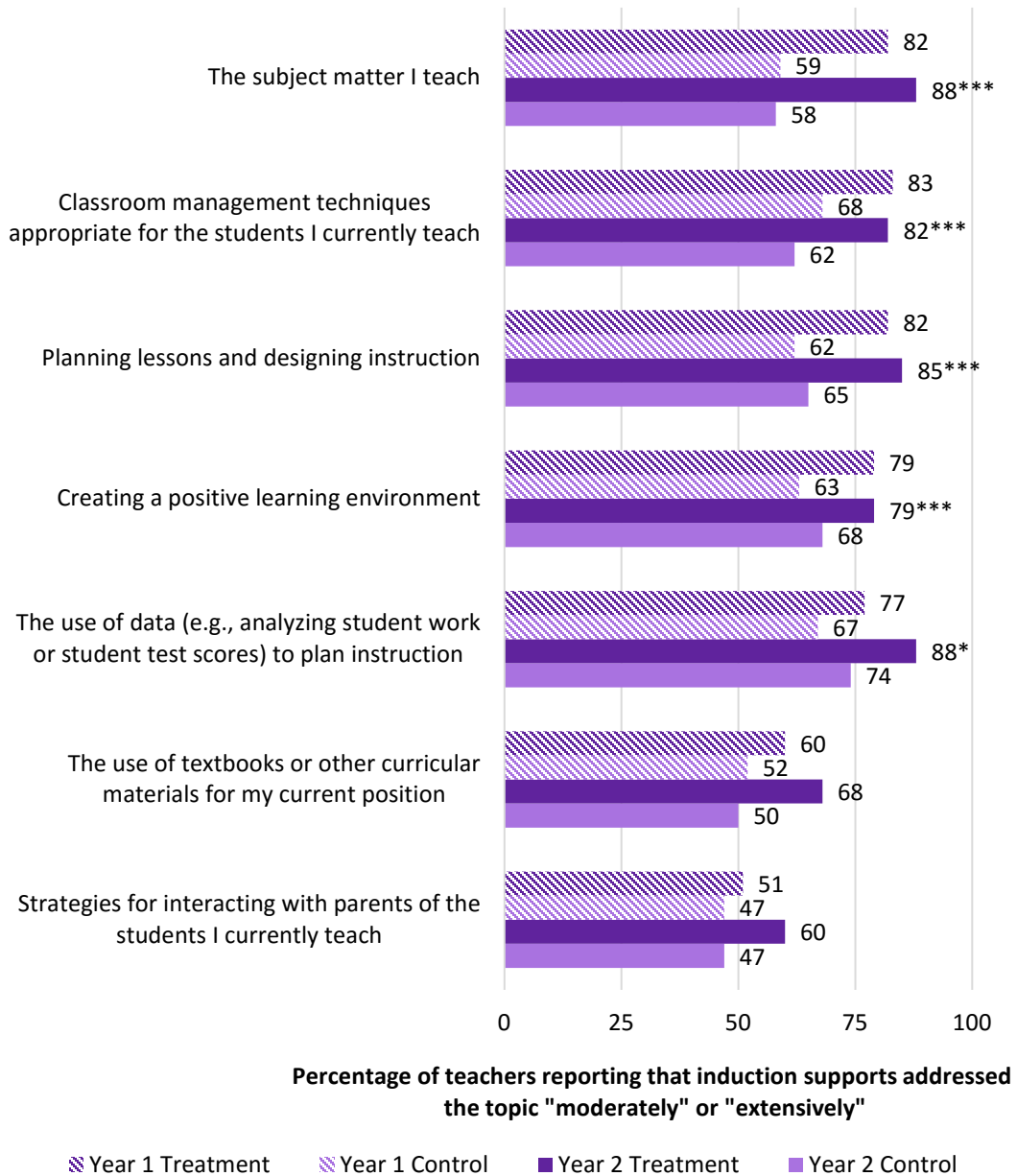


Source: NTC New Teacher Survey, spring 2013–2016.

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

Value labels are rounded to the nearest tenth.

**Exhibit 17. Focus of Mentoring and Other Induction Supports,
RCT Combined Sample, 2014–16**

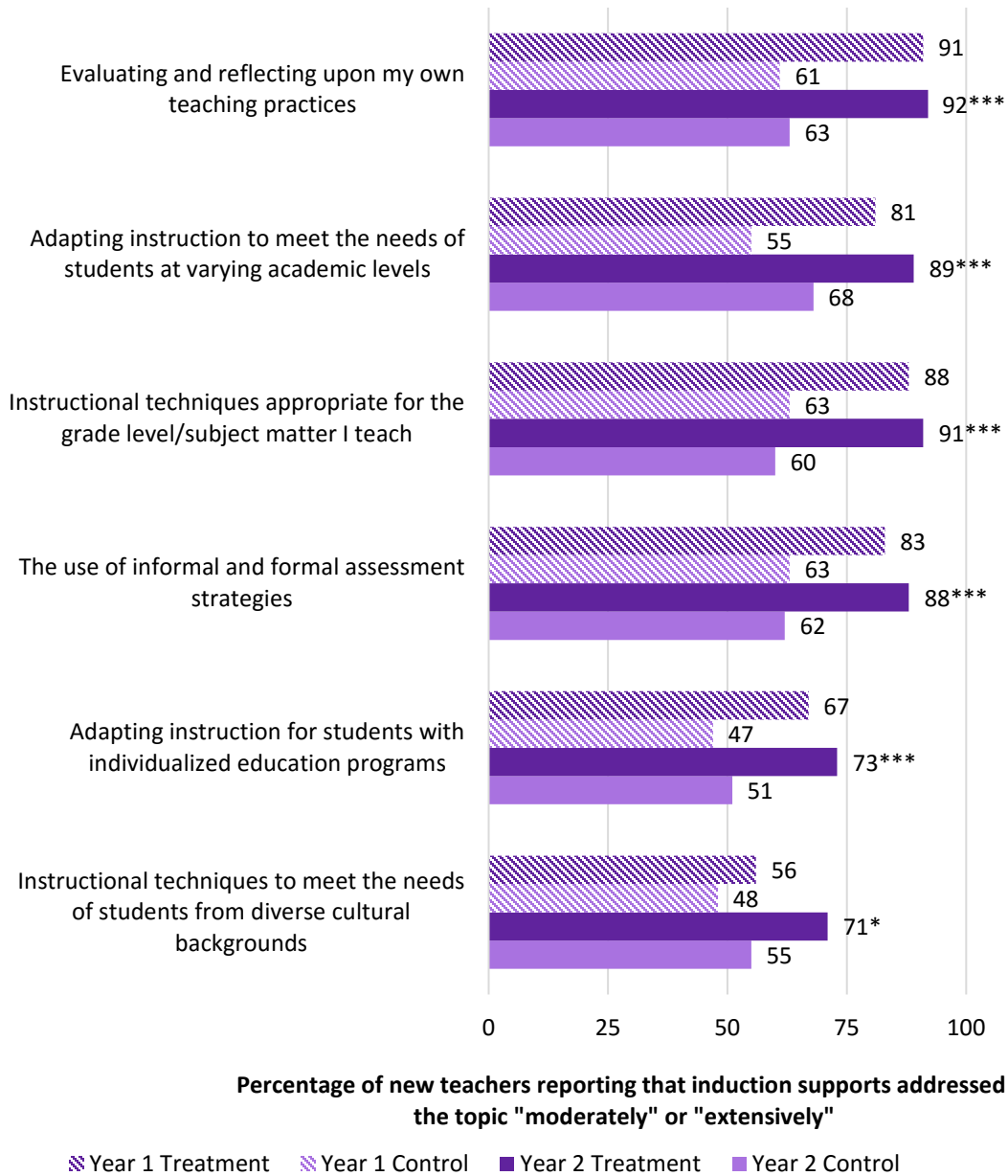


Source: NTC New Teacher Survey, spring 2014–2016.

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

Values labels are rounded to the nearest percentage point

**Exhibit 17. Focus of Mentoring and Other Induction Supports,
RCT Combined Sample, 2014–16 (concluded)**

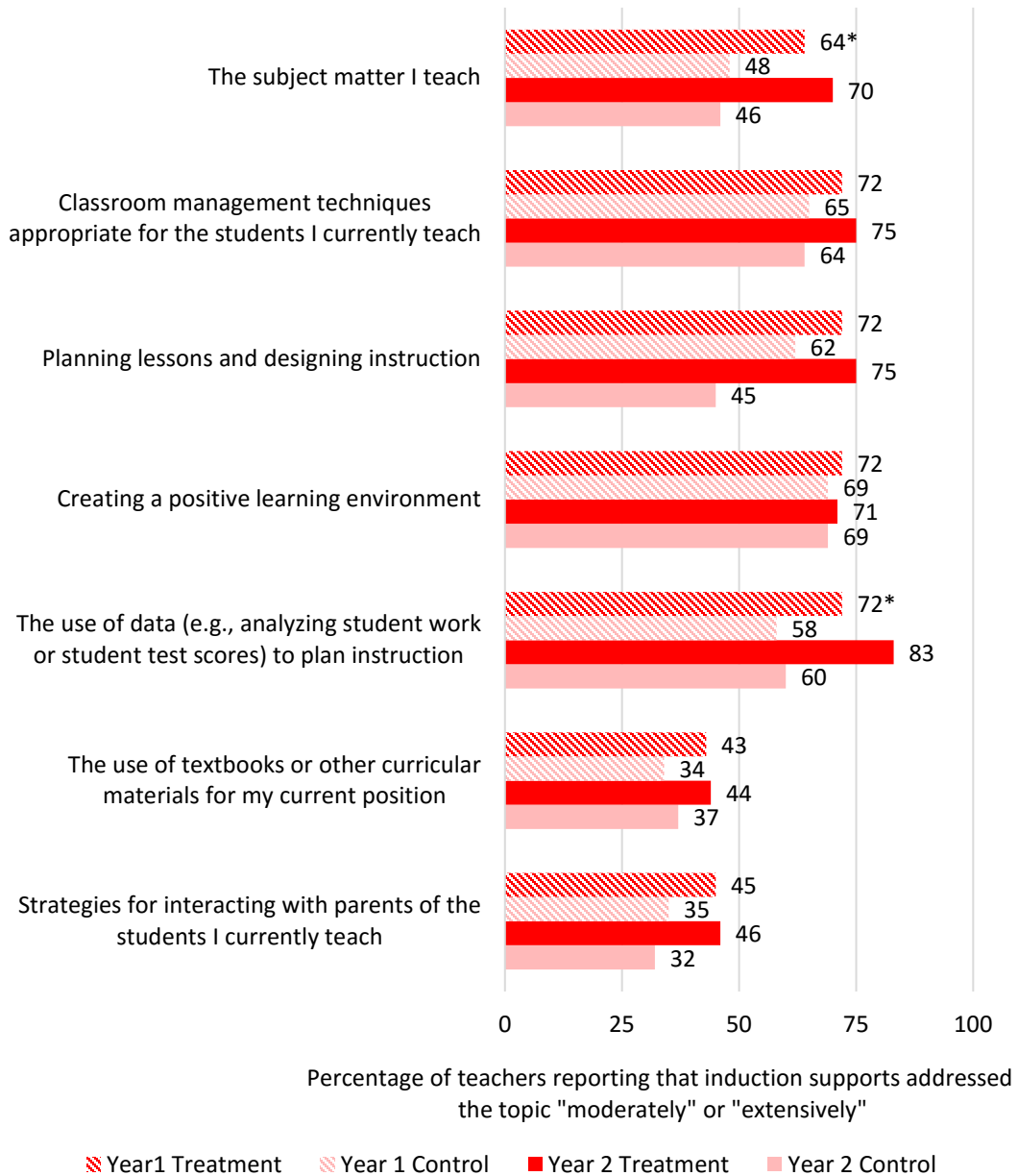


Source: NTC New Teacher Survey, spring 2014–2016.

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

Value labels are rounded to the nearest percentage point.

**Exhibit 18. Focus of Mentoring and Other Induction Supports,
QED Site, 2013–16**

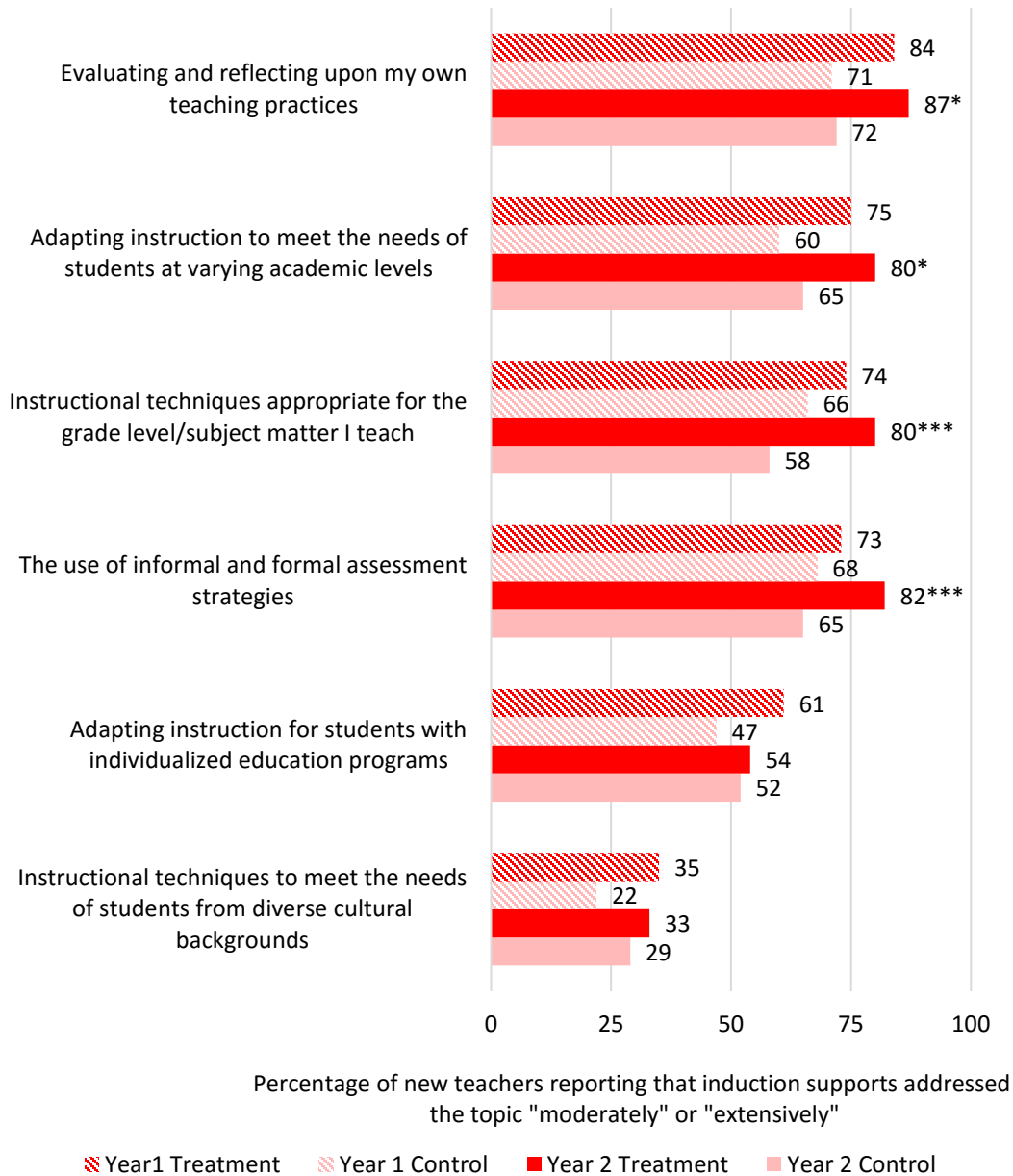


Source: NTC New Teacher Survey, spring 2013–2016.

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

Value labels are rounded to the nearest percentage point.

**Exhibit 18. Focus of Mentoring and Other Induction Supports,
QED Site, 2013–16 (concluded)**



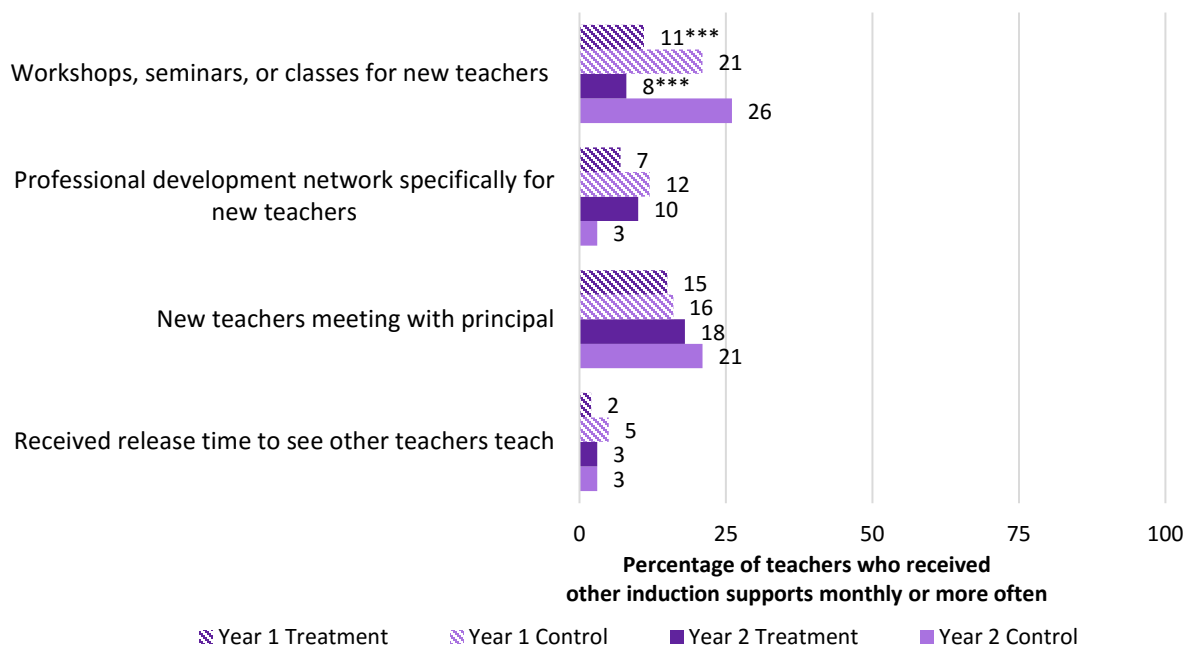
Source: NTC New Teacher Survey, spring 2013–2016.

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

Value labels are rounded to the nearest percentage point.

Access to induction supports other than mentors was limited for both treatment and control teachers. Overall, the percentages of teachers who received at least monthly induction supports other than mentoring were small (Exhibits 19 and 20), although in both years control teachers in the RCT districts were more likely than treatment teachers to attend workshops, seminars, or classes for new teachers (Exhibit 19).

Exhibit 19. Frequency of Other Induction Supports, RCT Combined Sample, 2014–16

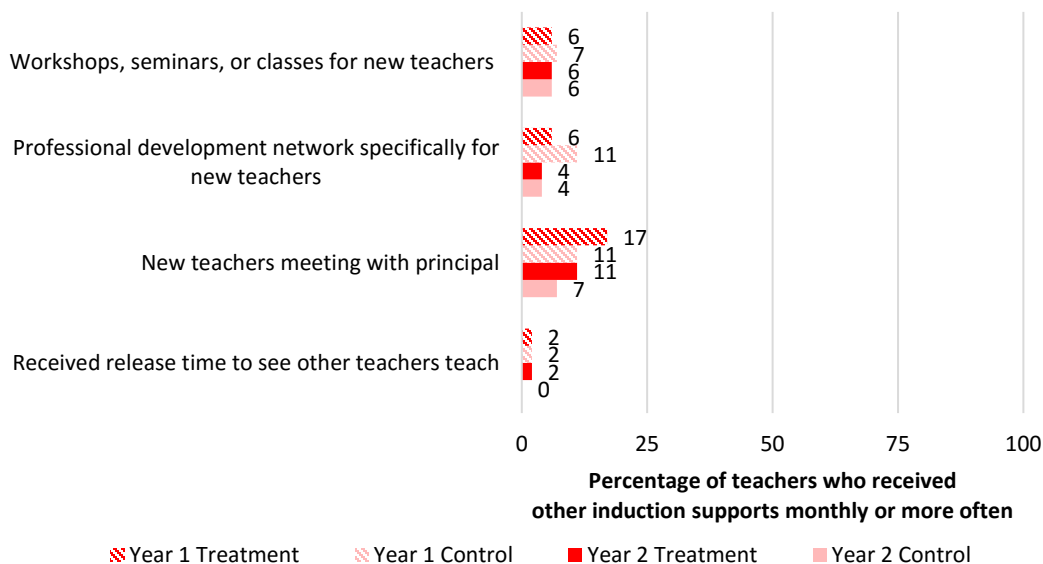


Source: NTC New Teacher Survey, spring 2014–2016.

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

Value labels are rounded to the nearest percentage point.

Exhibit 20. Frequency of Other Induction Supports, QED Site, 2013–16



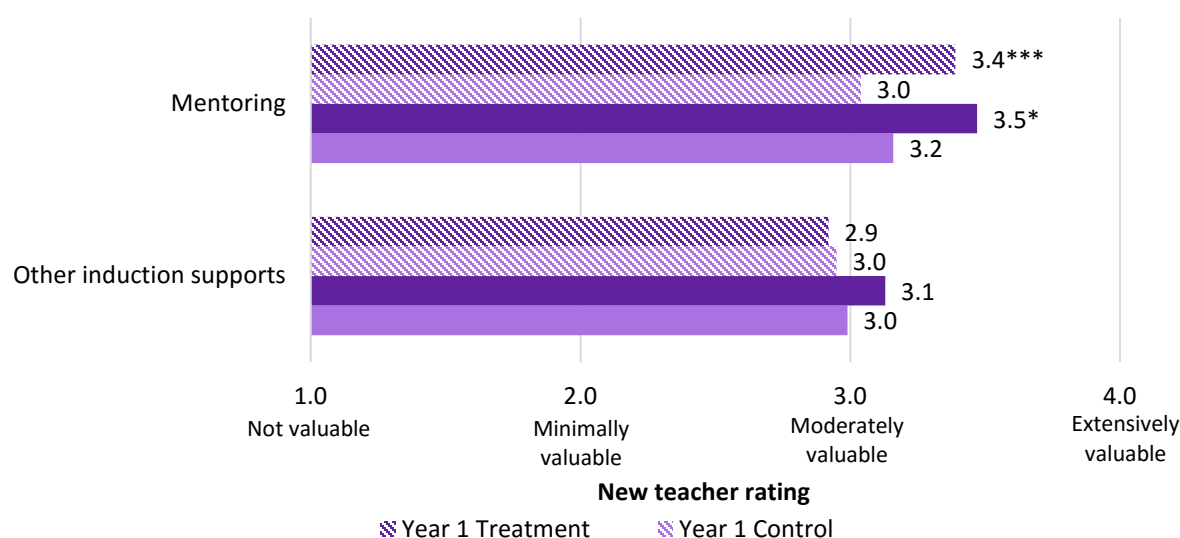
Source: NTC New Teacher Survey, spring 2013–2016.

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

Value labels are rounded to the nearest percentage point.

Treatment teachers, compared with control teachers, rated the value of the mentoring activities they engaged in higher. Treatment teachers in the RCT sites in both years perceived more value in their mentoring activities than control teachers, as did QED site teachers in their second year of teaching (Exhibits 21 and 22).

**Exhibit 21. Value of Mentoring Activities and Other Induction Supports,
RCT Combined Sample, 2014–16**

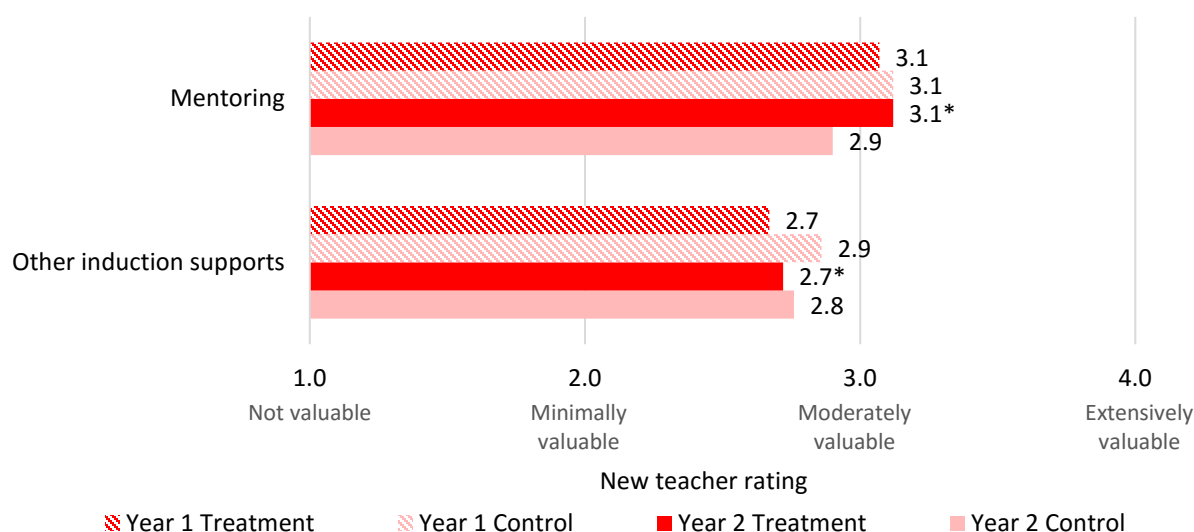


Source: NTC New Teacher Survey, spring 2014–2016.

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

Value labels are rounded to the nearest tenth.

**Exhibit 22. Value of Mentoring Activities and Other Induction Supports,
QED Site, 2013–16**



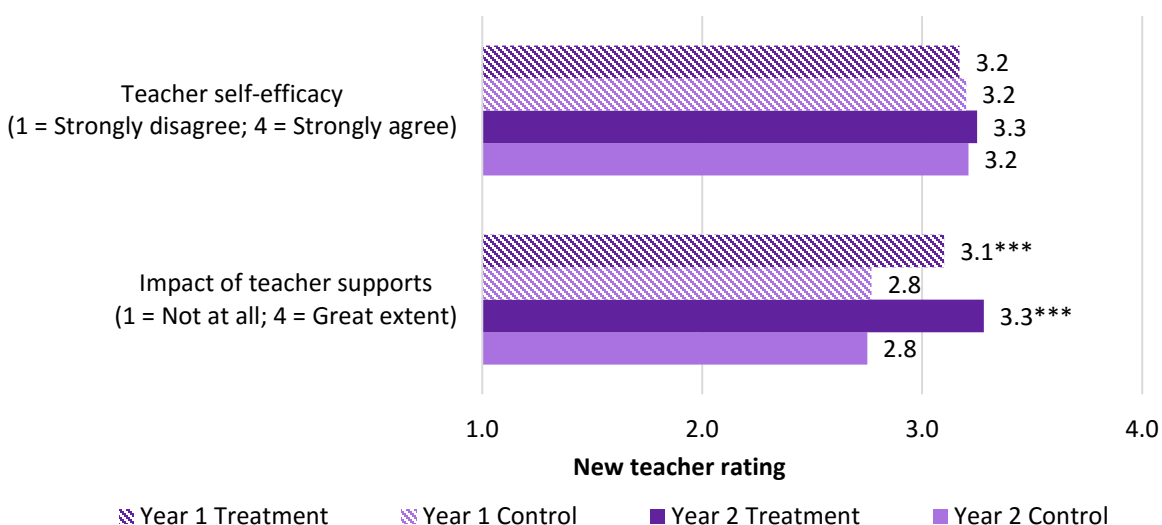
Source: NTC New Teacher Survey, spring 2013–2016.

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

Value labels are rounded to the nearest tenth.

Further reflecting teachers' perception of the value of mentoring, treatment teachers in all sites were more likely than control teachers to report that induction supports helped them grow as teachers, even though treatment and control teachers had similarly positive assessments of their own efficacy in the classroom (Exhibits 23 and 24). Interviewed mentors and teachers elaborated that the 2-year model allowed mentors to use a gradual release approach in having teachers, over time, more independently use high-leverage protocols and tools such as analyzing student work and data to drive their own reflection and growth. Teachers built confidence and independence while continuing to receive support for 2 years.

Exhibit 23. New Teacher Ratings of Self-Efficacy and Impact of Supports, RCT Combined Sample, 2014–16

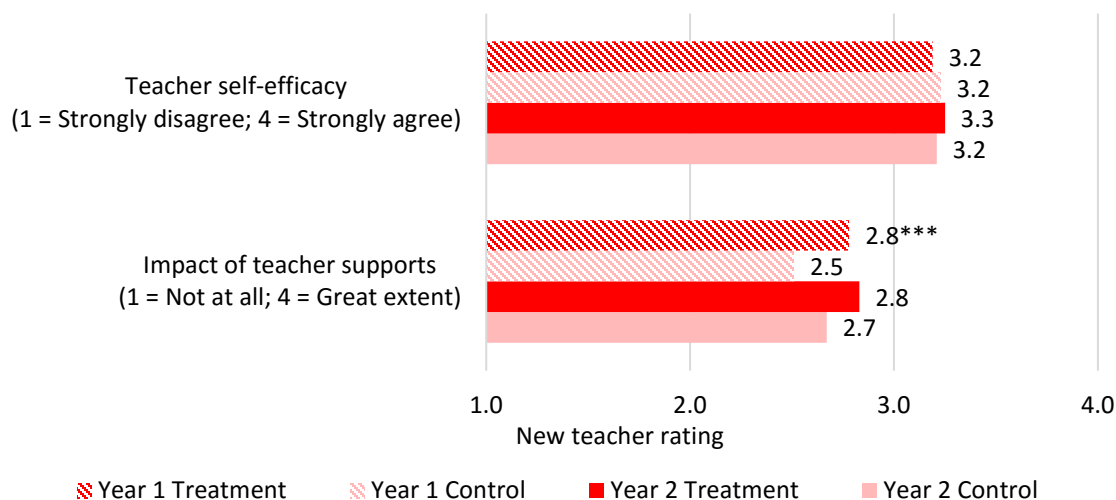


Source: NTC New Teacher Survey, spring 2014–2016.

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

Value labels are rounded to the nearest tenth.

Exhibit 24. New Teacher Ratings of Self-Efficacy and Impact of Supports, QED Site, 2013–16



Source: NTC New Teacher Survey, spring 2013–2016.

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

Value labels are rounded to the nearest tenth.

These results suggest that the significantly different levels of mentoring that treatment teachers and control teachers received on instructional priorities was consistent and persistent. Moreover, control teachers did not participate more than treatment teachers in other induction supports that might have compensated for the differences in their mentoring experiences. Thus, the overall high levels of implementation fidelity and treatment-control contrast together suggest that the participating sites provided a good testing ground for the impact of the NTC induction model on teacher and student outcomes.

IMPACT ON TEACHER OUTCOMES

NTC induction, as depicted in the logic model, is intended to improve beginning teachers' practice through its comprehensive support system as the path toward improved student learning. More robust support during those critical beginning years in the teaching profession is also intended to retain teachers at higher rates than districts would achieve if beginning teachers were left to struggle on their own. We present results for impacts on teacher practice and teacher retention for the RCT sites after 2 years of induction supports.¹³

Impacts on Teacher Practice

We measured teacher practice outcomes through structured classroom observations using the Framework for Teaching (Danielson, 2013), capturing dimensions of classroom management and culture and instructional quality on the same sample at baseline (at the beginning of the teachers' first year of teaching) and at the end of the teachers' second year of teaching.

Final Observation Analysis Sample

The analysis of classroom observations of teacher practice included treatment and control teachers who were randomly selected and observed in fall 2013 (baseline) and spring 2015 (after 2 years of teaching) for Cohort 1 and in fall 2014 (baseline) and spring 2016 (after 2 years of teaching) for Cohort 2. Teachers were eligible for the sample if they taught core subjects (mathematics, reading/English language arts, social studies, science, or self-contained elementary classrooms). We conducted all observations during instruction in the core subjects.¹⁴

Attrition

We measured attrition from the observation sample at the school level. Schools attrited from the sample when all teachers who were selected for observation in the school attrited, i.e., no teachers selected for observation were observed at both time periods. Exhibit 25 displays the number of treatment and control schools with teachers selected for observation in each cohort,¹⁵ the number of schools with teachers observed at each time period, and the school-level attrition by condition in each district and overall. WWC standards for attrition take into account both overall attrition and the difference in attrition between treatment and control groups. In Cohort 1, overall school-level attrition was 36 percent, with differential attrition of 3 percentage points. In Cohort 2, overall attrition was 21 percent, with differential attrition of 6 percentage points. When combined,

¹³ We were not able to include the QED site in the teacher outcomes. For teacher practice, we could not observe the comparison cohort of beginning teachers when they first started teaching because that period preceded the start of the grant. Given that veteran teachers (needed in the difference-in-differences design) do not serve as a sound comparison group for new teacher retention because they inherently have different retention patterns from new teachers, the teacher retention analysis from the QED site is a purely descriptive off-year comparison; therefore, we conducted only descriptive, not causal, analysis to inform NTC.

¹⁴ Observers were trained and calibrated on the Framework for Teaching before each round of observations. After calibrations, observers achieved interrater reliability where over 90% of scores were within one point across all elements scored across multiple test videos.

¹⁵ This number includes all schools with teachers selected, including those who declined to participate and the teachers who replaced them. In some cases, the teachers selected as replacements also declined to participate. Therefore, the total number of schools selected may have been larger in one district or condition than in the other, with the aim of obtaining a final sample that was balanced across treatment and control in each district.

the attrition for both cohorts was 28 percent, with differential attrition of 2 percentage points. These rates are all within the range of acceptable attrition.

After attrition, the number of schools remaining in the analysis sample was relatively low, as was the number of teachers in each of these schools, even when both cohorts and districts were combined. Lower sample size limited our ability to detect the effects of the NTC model on teacher practice, particularly if those effects were small or variability in practice among teachers was great. A second consequence of attrition, particularly when differential attrition occurs, is that the schools and teachers remaining in the sample may differ in both measurable and unmeasurable ways from those who attrited. Because the differential attrition was relatively low, this implication was not particularly problematic in the observation sample.

Exhibit 25. Overall School Observation Sample Selection and Attrition, Cohort 1, Cohort 2, and Combined, RCT Districts

	Cohort 1			Cohort 2			Combined		
	Treat- ment	Control	Over- all	Treat- ment	Control	Over- all	Treat- ment	Control	Over- all
Schools selected for observation (with replacement)	46	41	87	45	42	87	91	83	174
Observed at time 1	44	37	81	45	42	87	89	79	168
Stayed and were observed at time 2	29	27	56	37	32	69	66	59	125
Percentage attrited	37%	34%	36%	18%	24%	21%	27%	29%	28%
	<i>Met standard</i>			<i>Met standard</i>			<i>Met standard</i>		

Baseline Equivalence

To address the concern that teachers who remained in the analysis may be different from those who attrited, both on observable and unobservable factors, WWC requires that we show the baseline equivalence on the outcome measures of treatment and control teachers if attrition exceeds acceptable levels. This step was not necessary because the attrition levels met WWC standards. However, a full discussion of baseline equivalence is presented in Appendix D.

Measures of Teacher Practice

The teacher practice outcomes were eight measures on the Framework for Teaching (Danielson, 2013), four components each under Domain 2: Classroom Environment and Domain 3: Instruction. Trained observers scored each observed teacher on the 12 elements representing the four components under Domain 2 and the 15 elements representing four components under Domain 3, as shown in Exhibit 26.

The scores used in the analysis were factor variables combining the element-level scores into one variable representing each component; each component was based on two to four elements (Exhibit 26).¹⁶ Each factor variable was continuous, had a mean of 0 and a standard deviation of 1, and the majority of teachers scored in the range from -2 to 2. A score of zero on each component therefore is equivalent to being at the average score for teachers in this sample at baseline. A change in these variables of 1.0 is a change of 1 standard deviation, which is roughly equivalent to 0.5 or 0.6 point on the original 1 to 4 scale on the Framework for Teaching.

¹⁶ The factor variable reflects the structure of the correlations between the elements. It is similar to a weighted average of the elements, where the weights include the strength of the relationship between the elements as well as teachers' scores on the elements.

Exhibit 26. Framework for Teaching Domains, Components, and Elements Observed

Domain 2. The Classroom Environment	
Component (Factor Variable)	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
Managing Student Behavior	Expectations Monitoring of student behavior Response to student misbehavior
Domain 3: Instruction	
Component (Factor Variable)	Elements
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

Source: Excerpted from the Framework for Teaching (Danielson, 2013).

Controls

The teacher practice impact estimates were derived from a two-level hierarchical model with teacher and school levels that controlled for school- and teacher-level variables, the baseline observation measure for each teacher observed, and the blocking variables used in the random selection and assignment of schools. Using these blocking variables ensured representation of different types of schools by geography (in CPS), Teacher Incentive Fund status (in BCPS), and grade levels served (in both districts).

Results

Examining the teacher practice outcomes for the combined RCT districts, we found no statistically significant differences between observed treatment and control teachers on any of the measured components in either cohort or when the two cohorts were combined. Exhibit 27 displays the impact estimates for each component for the combined RCT districts.¹⁷ The small sample size in

¹⁷ Because the relationship between the control variables and the outcomes differed by cohort, as did the structure of the variation at the school and teacher levels, the estimates found in the models by cohort do not average to the estimates found in the analysis including both cohorts.

the end impeded our ability to find impacts. Post hoc power tests indicated that the final sample size had minimum detectable effect sizes of 0.40 to 0.46 across the eight teacher practice outcomes. Full model tables are in Appendix D.

Exhibit 27. Impact of the NTC Model on Teacher Practice Outcomes, Combined RCT Sample

		Cohort 1	Cohort 2	Both Cohorts
Creating an Environment of Respect and Rapport	Estimate	0.16	-0.10	0.04
	SE	(0.21)	(0.25)	(0.16)
	N teachers	71	88	159
	N schools	56	69	108
Establishing a Culture for Learning	Estimate	-0.10	-0.14	-0.24
	SE	(0.23)	(0.34)	(0.20)
	N teachers	71	88	159
	N schools	56	69	108
Managing Classroom Procedures ^a	Estimate	0.23	0.07	0.13
	SE	(0.25)	(0.23)	(0.16)
	N teachers	59	80	139
	N schools	51	68	102
Managing Student Behavior ^b	Estimate	0.34	0.18	0.28
	SE	(0.24)	(0.24)	(0.18)
	N teachers	71	88	159
	N schools	56	69	108
Communicating with Students	Estimate	-0.06	0.08	0.01
	SE	(0.24)	(0.27)	(0.19)
	N teachers	70	87	159
	N schools	56	69	108
Using Questioning and Discussion Techniques	Estimate	-0.04	0.28	0.21
	SE	(0.23)	(0.28)	(0.18)
	N teachers	69	88	157
	N schools	55	69	107
Engaging Students in Learning	Estimate	-0.11	0.43	0.15
	SE	(0.24)	(0.28)	(0.17)
	N teachers	70	87	157
	N schools	56	68	108
Using Assessment in Instruction	Estimate	-0.21	0.18	0.06
	SE	(0.22)	(0.30)	(0.18)
	N teachers	70	88	158
	N schools	55	69	107

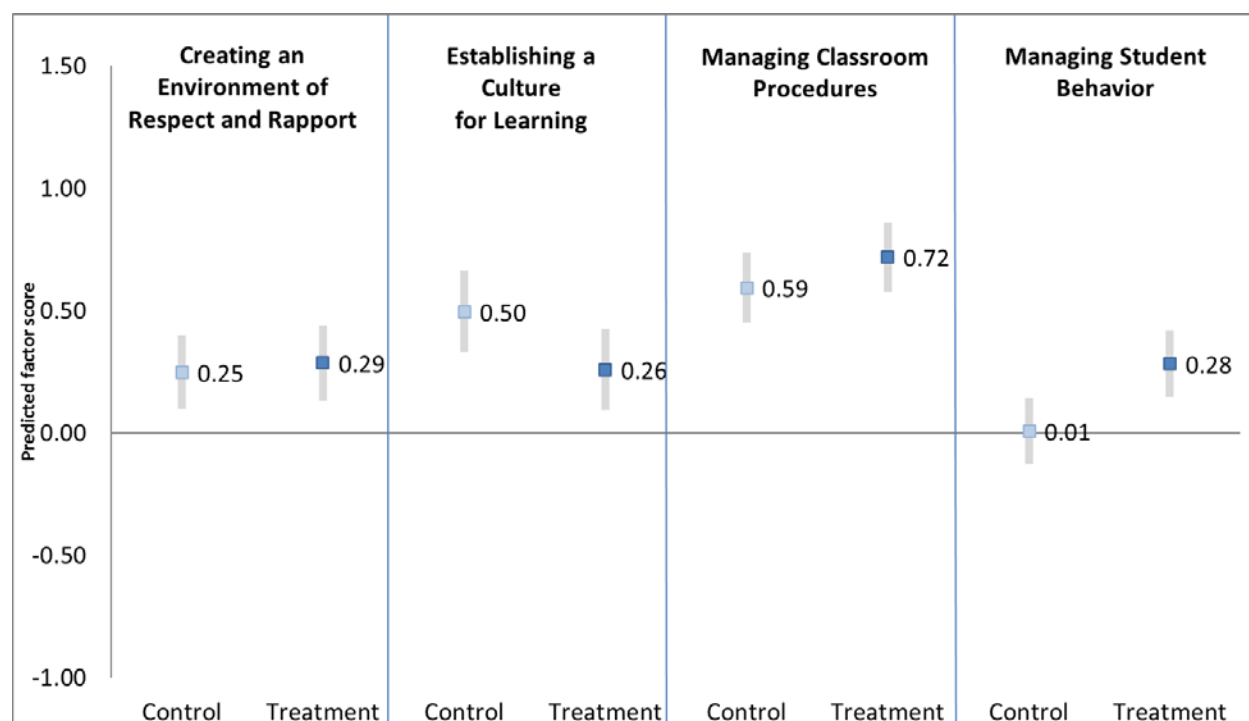
* $p < .05$, ** $p < .01$, *** $p < .001$.

^a This variable excludes the elements that had many blanks because *Management of Instructional Groups* and *Performance of Classroom Routines* were not observed during the observation period for a number of teachers.

^b This variable excludes the elements that had many blanks because *Response to Student Misbehavior* was not observed during the observation period for a number of teachers.

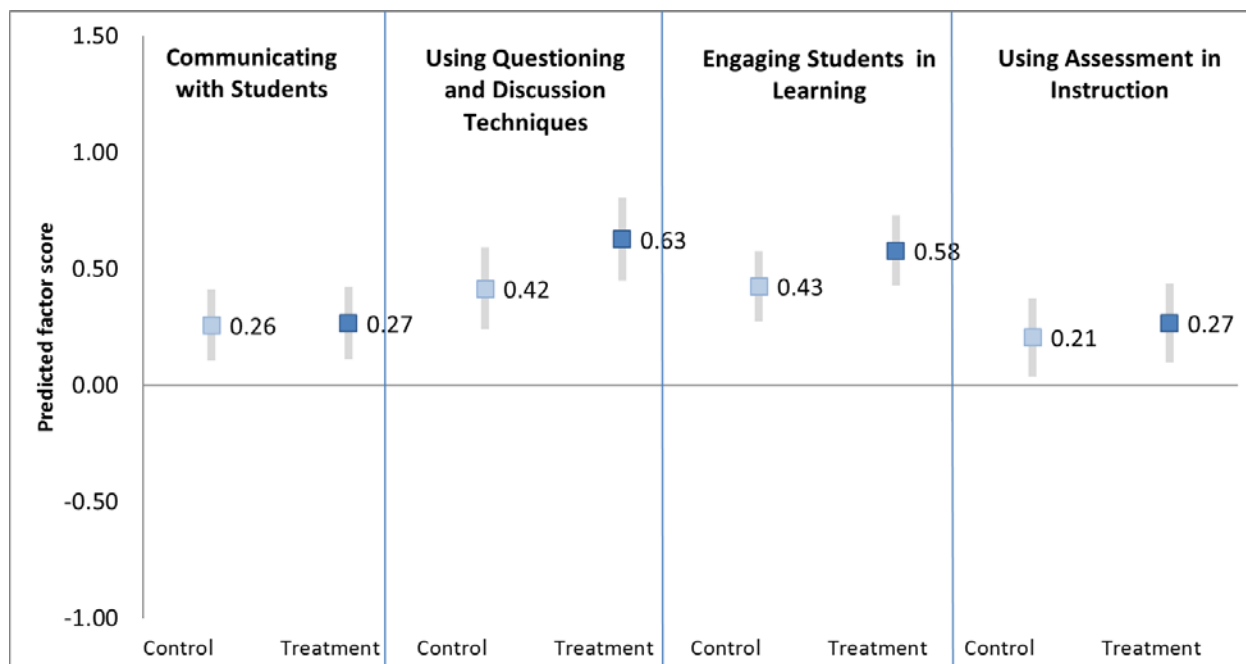
Exhibit 28 displays the results of the impact analysis combining both districts and cohorts. The estimates for treatment and control account for the baseline observation score for teachers, as well as teacher and school demographics. As discussed, factor scores were created at baseline to have a mean of 0 and a standard deviation of 1. Therefore, the positive estimates for both treatment and control indicate that on average the sample of teachers at the end of 2 years scored higher on the Framework for Teaching measures than the sample at baseline. However, the amount of growth on these measures was not significantly different between teachers in treatment and control groups. Additionally, these estimates confound the impact of teacher growth and the impact of attrition; the average score may be higher at the end of 2 years because of attrition of lower performing teachers from the sample. On three factors, *Creating an Environment of Respect and Rapport*, *Communicating with Students*, and *Using Questioning and Discussion Techniques*, the baseline scores of teachers who attrited were significantly lower than the baseline scores of teachers who remained in the sample. Such a pattern suggests that weaker teachers may have attrited from the sample overall, consistent with the pattern of higher mean scores compared with baseline across all teacher practice measures and for both treatment and control groups.

Exhibit 28. Model-Implied Means on Teacher Practice Outcomes for Treatment and Control Groups Overall



Note: The blue boxes depict the estimated treatment and control means, and the grey lines depict the standard error around those means. The standard error is a measure of confidence in the estimate of the mean, and when the grey lines of treatment and control overlap, we cannot say with confidence that the means are significantly different between the two groups.

Exhibit 28. Model-Implied Means on Teacher Practice Outcomes for Treatment and Control Groups Overall (concluded)



Note: The blue boxes depict the estimated treatment and control means, and the grey lines depict the standard error around those means. The standard error is a measure of confidence in the estimate of the mean, and when the grey lines of treatment and control overlap, we cannot say with confidence that the means are significantly different between the two groups.

Impact on Teacher Retention

Using district administrative data, SRI assessed the impacts of the NTC induction model on teachers' retention into their third year of teaching (a 2-year retention rate) in the RCT districts.

Analysis Sample

Human resources data were complete for all study teachers in Cohort 1 and Cohort 2 in treatment and control schools in the RCT sites. Therefore, no attrition occurred in the sample for the teacher retention analysis.

Outcome Measure

Because summers are a natural time for teachers to change jobs, we counted teachers as "retained" if they were still employed by the district at the beginning of their third year, i.e., fall 2015 for Cohort 1 and fall 2016 for Cohort 2.

Results

We estimated the impact of the NTC induction model on teacher retention using a two-level hierarchical model with the same controls as those in the teacher practice models. Across both cohorts and districts, 79 percent of treatment teachers and 78 percent of control teachers were retained; the difference was not statistically significant. The retention rates for both treatment and control teachers are lower than the rate found for a national sample of teachers that began teaching in 2007–08, 85 percent of whom remained in the profession 3 years later (Gray & Taie, 2015). Local factors, such as one participating district being commonly thought of as a teacher training ground for suburban districts, might be stronger than the any mitigation offered by stronger induction supports. See Appendix E for full teacher retention models.

The NTC induction model aims not only to provide novice teachers with instructional skills, but also to cultivate in them the reflective skills to critically analyze their own practice and use data as evidence of instruction that leads to deep student learning. This chapter reports on the studies of impacts on student achievement, the RCTs in two districts and the QED in the third site.

Randomized Controlled Trials of the Impact on Student Achievement

Through the RCTs, SRI examined the impact of the NTC induction model on English language arts and mathematics achievement in grades 4 through 8 in teachers' second year of participating in NTC's 2-year induction support. Special education teachers who taught reading and/or mathematics and who could be linked to students in district data sets were included in the analysis, along with regular education teachers.

Attrition

Overall attrition at the school level for the combined sample across the RCT districts was 1 percent for mathematics and 10 percent for ELA. Both the overall and differential attrition levels were within acceptable thresholds under WWC (see Appendix F).

Baseline Equivalence in RCT Districts

We examined baseline equivalence in student achievement scores between treatment and control schools. Baseline equivalence ensures that any differences in the outcomes between treatment and control groups are due to the treatment and not to systematic differences between the groups that were present before the intervention. A baseline difference of less than 0.05 standard deviation is considered equivalent. A baseline difference between 0.05 and 0.25 can be considered equivalent if prior achievement is included in the model. We included prior achievement in the models for all student achievement analyses to obtain better precision for the estimated impact.

For the second-year impact, with the two RCT sites and both cohorts combined, the difference in baseline achievement scores between students with treatment teachers and students with control teachers was 0.01 standard deviation in mathematics and 0.10 standard deviation in ELA. As the difference in mathematics was below 0.05 standard deviation, this analysis achieved baseline equivalence. The difference in ELA was between 0.05 standard deviation and 0.25 standard deviation; therefore, it achieved baseline equivalence with prior achievement included in the model.

Student Achievement Measures

We used scale scores from state assessments of ELA and mathematics for grades 4 through 8 as measures of student achievement.¹⁸ In BCPS, we used scale scores from the 2014–15 and 2015–16 Florida Standards Assessment (FSA) (for Cohort 1 and Cohort 2 teachers in their second year of teaching, respectively), and similarly in CPS we used scale scores from the 2014–15 and 2015–16 Measures of Academic Progress (MAP).¹⁹ To combine test results across grade levels and also

¹⁸ Students in third grade take state assessments in Florida, Illinois, and Iowa. The third-grade scores serve as the measure of prior achievement for fourth-grade students. As the lowest tested grade, however, third-grade students do not have a measure of prior achievement and could not be included in the analysis. Fourth grade was the lowest grade that we could include in the sample.

¹⁹ Because the state of Illinois changed assessments, CPS administered the MAP in 2014–15 and 2015–16 to bridge the two time periods under different assessments. CPS needed continuity in student achievement measurements for the district's teacher evaluation system.

across districts, we standardized each scale score at each grade level using a common metric, the z score, which has a mean of 0 and a standard deviation of 1.²⁰

We compared the test scores of students of treatment teachers with those of control teachers, controlling for the prior achievement of each student (as measured by their test scores in 2013–14 or 2014–15), student background, teacher background, school characteristics, and the blocking variables used in the random assignment of schools into treatment and control groups. These blocking variables ensured that the sample represented different types of schools by geography (in CPS), Teacher Incentive Fund status (in BCPS), and grade levels served (in both sites). The hierarchical models that we used to estimate the impact of NTC induction accounted for the nesting of students within classrooms and of teachers within schools. The blocking variables accounted for both sampling design and district-level nesting because schools in each site were blocked on a different set of variables. Because this analysis combined data collected in two different years (2014–15 for Cohort 1 and 2015–16 for Cohort 2), a centered year variable was also included to account for any historical changes in test scores between the two years. Finally, we included interactions between the district and cohort indicators and all background characteristics to account for the different relationships between background characteristics and student outcomes by district and cohort.

Second-Year Impact, Combined RCT Sites

Exhibit 29 shows the difference between the adjusted mean test scores among treatment teachers' students and control teachers' students at the end of the second year of teaching, controlling for prior achievement, student characteristics, teacher characteristics, and school characteristics. This difference, measured in standard deviations of the underlying distribution of student scale scores across both sites, represents the 1-year impact of NTC induction on the student achievement of teachers in their second year of induction support.

In ELA, the average student achievement of teachers in the second year who participated in NTC induction for 2 years was approximately 0.05, compared with -0.04 for students of control teachers. This difference equals an effect size of 0.09 standard deviation ($p < .05$)—equivalent to moving from the 48th to the 52nd percentile. On broad-scope standardized tests of reading like the FSA and the MAP, an effect size of 0.09 is equivalent to an approximately 23 to 39 percent greater annual gain than otherwise expected for students in grades 4 through 8 and represents the equivalent of approximately 2 to 3.5 additional months of learning, depending on the student's grade level (Lipsey et al., 2012).

The NTC induction model also showed significant and positive impacts on mathematics achievement in grades 4 through 8. Students in grades 4 through 8 of teachers in their second year who participated in NTC induction for 2 years scored 0.15 standard deviation ($p < .01$) higher on average than students of control teachers. These impacts are equivalent to moving from the 46th to the 52nd percentile. On broad-scope standardized tests like the FSA and the MAP, an effect size of 0.15 is equivalent to an approximately 27–50 percent greater annual gain than otherwise expected for students in grades 4 through 8 and represents the equivalent of approximately 2.4 to 4.5 additional months of learning, depending on the student's grade level.

²⁰ To calculate z scores, we first computed the mean and standard deviation of scale scores separately for ELA and for mathematics in each site and at each grade level, based on the full set of scores that we received for students of treatment and control teachers. We converted scale scores to z scores by taking the scale score in ELA or mathematics, subtracting the overall sample mean for that subject, and dividing by the pooled standard deviation for students of treatment and control teachers. A z score of 0 means that the student scored at the mean for his or her grade level in the study schools in his or her district in ELA or in mathematics. A z score of 1 means that the student scored 1 full standard deviation above the mean, and a z score of -1 means that the student scored 1 full standard deviation below the mean.

Exhibit 29. Second-Year Impact on Student Achievement, Combined RCT Sites

Subject	Adjusted Mean Test Scores Treatment	Control	Difference (effect size)	Students	Sample Sizes Teachers	Schools
ELA	0.05	-0.04	0.09*	6,147	149	99
Mathematics	0.06	-0.09	0.15**	4,972	129	86

Note: The effect on student achievement is a 1-year effect as the districts provided current and prior achievement data annually but did not consistently provide identifiers to link students across the data sets given to researchers each year.

The 1-year impact after 2 years of mentoring includes achievement in 2014–15 for Cohort 1 teachers and 2015–16 for Cohort 2 teachers.

Adjusted mean test scores are in standard deviation units.

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

We further tested the robustness of these findings by running several sensitivity analyses, including removing late joiners from the analytic sample (Appendix F), which did not change the student achievement results substantively. Appendix G provides other exploratory analyses on the RCT districts, including examining whether the results differ for elementary versus middle school students, for students taking ELA or mathematics with more than one teacher, or by school characteristics.

Quasi-experimental Study of the Impact on Student Achievement

In the quasi-experimental study, SRI used a differences-in-differences approach to estimate the impact of participating in the 2-year NTC induction program in one site. The study compared the difference in the 2014–15 achievement of students of Cohort 1 beginning teachers receiving NTC induction support from 2013–14 to 2014–15 and the 2013–14 student achievement of a prior cohort of comparison beginning teachers who started teaching in 2012–13 and did not receive NTC induction support with the difference in the student achievement of veteran teachers in the same years.²¹

Outcome Measures and Timing

The outcomes used in the QED were Iowa Assessment scores in ELA and mathematics in grades 4 through 8.²² To combine test results across grade levels, we standardized each scale score at each grade level using the z score.²³ The distribution of standardized outcome scores for each of the four groups of teachers included in each of the four impact analyses are presented in Appendix H.

The state testing schedule posed an unexpected constraint on the analysis. GWAEA districts varied in when they chose to administer the state test in ELA and mathematics (fall, winter, or spring). The time period of the achievement data therefore did not align perfectly with the treatment period. For example, for districts testing in the fall, the baseline measure was taken at the

²¹ From communication between NEi3 and SRI research team, this difference-in-differences approach has better validity and is more likely to meet the WWC standards when treatment and comparison groups are no more than 1 year apart, although WWC changed the standards for difference-in-differences approaches in 2014, after we had designed the impact study for GWAEA and the site had begun serving new teachers. SRI conducted student outcomes analysis for GWAEA teachers in the second cohort to inform NTC about its program. The results are in Appendix H and are not intended for WWC review.

²² Students in third grade take state assessments in Iowa. The third-grade scores serve as the measure of prior achievement for fourth-grade students. As the lowest tested grade, however, third-grade students do not have a measure of prior achievement and could not be included in the analysis. Fourth grade was the lowest grade that we could include in the sample.

²³ To calculate z scores, we computed the mean and standard deviation of scale scores separately for reading and for mathematics at each grade level, based on the whole sample of students in GWAEA for a given year. We converted scale scores to z scores by taking the scale score in reading or mathematics, subtracting the overall sample mean for that subject, and dividing by the overall sample standard deviation. A z score of 0 means that the student scored at the mean for his or her grade level among the sample of students included in the study. A z score of 1 means that the student scored one full standard deviation above the mean.

beginning or slightly after treatment teachers started receiving supports, and the first-year outcome measure was taken at the beginning of the following school year, with whatever effect summer loss or transition to a new teacher or a new school level might have had on the students' test scores. If the testing window did not vary between the comparison cohort and the treatment cohort, we could assume that any effects of the testing window would equally affect the comparison and treatment teachers. We included indicators of the testing schedule in the analytic model to adjust for its effect on the outcome analysis.

Students in most participating districts were tested in the spring of each year, while a few districts tested in the fall or winter. Some districts switched testing from fall in one year to spring in the next during the evaluation period.²⁴ As an example, Exhibit 30 details the timing of prior achievement and second-year outcomes for Cohort 1 and comparison teachers in their second year of teaching. The fall to spring testing scenario was included in the Cohort 1 Year 2 impact analysis because the comparison cohort had a similar testing window. In all analyses, there was fall to fall, spring to spring, and winter to winter testing in both groups; we therefore included these testing patterns in all analyses.

Exhibit 30. Timing of Prior Achievement and Outcome Scores, QED Site

	District Testing Window	Baseline Achievement	Year 2 Outcome
Comparison teachers (Teachers beginning teaching in 2012–13)	Winter	Winter 2013	Winter 2014
	Spring	Spring 2013	Spring 2014
	Fall	Fall 2013	Fall 2014
	Fall to spring	Fall 2012	Spring 2014
Cohort 1 teachers (Teachers beginning teaching in 2013–14)	Winter	Winter 2014	Winter 2015
	Spring	Spring 2014	Spring 2015
	Fall	Fall 2014	Fall 2015
	Fall to spring	Fall 2013	Spring 2015

Analysis Sample

The student achievement analysis sample for Cohort 1 included all NTC and comparison new teachers who taught reading and/or mathematics in grades 4 through 8, together with their corresponding comparison veteran teachers (with 3 or more years of experience) who taught the same grade levels as the NTC/comparison new teachers in the same school. As described, all included students must have had aligned pre- and post-test scores. Special education teachers who taught reading and/or mathematics and who could be linked to students in district data sets were included in the analysis, along with regular education teachers.

The small number of new teachers in tested grades and subjects who had students with aligned pre- and post-test scores limited the student achievement analysis. There were more than 100 new teachers in the Cohort 1 and comparison groups. Of those, only a fraction—much lower than we expected—taught reading or mathematics in grades 4 through 8 in each year. The number further dropped and was somewhat uneven across years after we excluded teachers who had no students with aligned pre- and post-test scores.²⁵ Exhibit 31 displays the numbers of teachers and students included in reading and mathematics analyses for Cohort 1.

²⁴ A few schools that did not test at all during a given year of study were excluded from the analysis.

²⁵ Cohort 1 Year 2 impact analysis had fall to spring testing in both treatment and comparison groups, while Cohort 2 Year 2 analysis did not. Therefore, although Cohort 1 and Cohort 2 analyses were supposed to use the same new teacher comparison group, there was a larger comparison group sample for Cohort 1 than for Cohort 2 in the Year 2 impact analysis because for comparability, comparison teachers with fall to spring testing were dropped in the Cohort 2 Year 2 analysis

**Exhibit 31. Numbers of Schools, Teachers, and Students Included in Cohort 1 Year 2
Achievement Analyses, QED Site**

	Comparison					Treatment				
	New Teachers			Veteran Teachers		New Teachers			Veteran Teachers	
	No. of Schools	No. of Teachers	No. of Students	No. of Teachers	No. of Students	No. of Schools	No. of Teachers	No. of Students	No. of Teachers	No. of Students
ELA	8	8	194	35	1305	15	19	340	59	1189
Math	6	7	299	26	1140	17	23	533	68	1822

Baseline Equivalence

We examined baseline equivalence among students of the four groups of teachers for each of the difference-in-differences analyses. The top two panels of Exhibit 32 provide summary information on baseline student reading and mathematics standardized scores for the Cohort 1 Year 2 impact analysis. When the differences in baseline student achievement scores among these four groups of teachers were all under 0.25 standard deviation, the subsequent analysis achieved baseline equivalence once prior achievement was included in the models. However, for the Cohort 1 Year 2 mathematics analysis, there were differences in baseline scores larger than the 0.25 standard deviation threshold among students of the four groups of teachers. We took an additional precaution in accounting for the differences between these groups of teachers. For each analysis where there was a greater than 0.25 standard deviation difference in prior achievement, we created a propensity score weight by predicting the likelihood of being a student of a Cohort 1 teacher based on the prior achievement variable. Then we used this propensity score as a weight on students in the other three groups. This weighting approach added an extra correction for imbalances in baseline student achievement among the four groups of teachers. The bottom panels of Exhibit 32 present the descriptive information of the post-weighting baseline scores of the four groups of students. After weighting there was no differential baseline score larger than 0.25 between any groups for any analysis. This weight was therefore included in the subsequent impact analysis of the outcome score.

Exhibit 32. Cohort 1 Year 2 Baseline Student Test Scores, by Groups of Teachers, QED Site

	Cohort 1 Teachers (Year 2 treated, 2014–15)	Comparison Teachers (Year 2, 2013–14)	Comparison Veteran Teachers for Cohort 1 Teachers (2014–15)	Comparison Veteran Teachers for Comparison Teachers (2013–14)
ELA				
Mean	-0.18	0.04	-0.16	0.01
SD	1.02	0.92	1.04	1.03
N students	340	194	1189	1305
Mathematics				
Mean	-0.26	0.09	-0.01	-0.09
SD	0.98	0.99	1.01	0.95
N students	533	299	1822	1140
Mathematics (weighted)				
Mean	-0.26	-0.11	-0.22	-0.27
SD	0.98	1.02	1.00	0.93
N students	533	299	1822	1140

Second-Year Impact, QED Site

For the Cohort 1 Year 2 impact analysis in the QED site, we compared the test scores of students of Cohort 1 teachers in Year 2 with those of comparison teachers in Year 2, adjusting for the difference with their corresponding comparison veteran teachers, and controlling for the prior achievement of each student, student background, teacher background, and school characteristics. The hierarchical models that we used to estimate the impact of NTC induction accounted for the nesting of students within teachers, and teachers within schools. (Because most districts included in the analysis had only one school in the study, a district level in the hierarchical model was not possible.)

Exhibit 33 shows the difference-in-differences estimate for Cohort 1 teachers in their second year of NTC induction, together with sample sizes at the student, teacher, and school levels. This difference-in-differences estimate, measured in standard deviations of the underlying distribution of student scale scores, represents the impact of NTC induction on student achievement. The impact estimate for Cohort 1 teachers in their second year of induction support is not statistically significant, suggesting no detected NTC impact on Cohort 1 teachers in the QED site.

Given the very small number of new teachers in Cohort 1 with tested students and the varying testing schedules, only a small sample of NTC and comparison new teachers could be included in each of the analyses. Such a small number of NTC and comparison new teachers may not be representative of all new teachers in the QED site over the years, and the small sample itself prohibits a reliable estimation of the NTC impact. In view of these concerns with the small sample and its representativeness, this analysis should be interpreted with caution, and the results are not conclusive—we do not know whether the NTC induction model had an impact on student achievement in the QED site.

Exhibit 33. Estimated Impact on Student Achievement for Cohort 1 Year 2, QED Site

	Estimate	SE	p-value	No. of Schools	No. of Teachers	No. of Students
ELA	-0.078	0.065	0.258	23	121	3028
Mathematics	0.011	0.042	0.752	22	125	3794

CONCLUSIONS AND IMPLICATIONS

NTC implemented its induction model in three sites with high fidelity, and the model had a strong positive influence on teachers' induction experiences. Differences in the induction received by teachers in the treatment schools and control schools were notable. Relative to control teachers, NTC teachers met with their mentors more often, worked on tasks more directly related to instruction, placed a higher value on mentoring activities, and were more likely to credit their induction experience with contributing to the development of their knowledge and skills as teachers. Interviewed teachers elaborated on the value of the NTC mentors in helping them learn to manage their classrooms, plan thoughtfully, differentiate instruction, reflect on their practice, and gradually become more confident in their teaching.

The impacts on teacher and student outcomes were more mixed. From the RCTs, we detected no differences between treatment and control teachers in measures of classroom environment and instruction domains using the Framework for Teaching (Danielson, 2013). The lack of impact was likely due to attrition and small sample size. In addition, it is possible that the measures of teacher practice were not fine-grained enough to capture the nature of NTC effects on instruction. Retention rates into the third year of teaching were also similar between treatment and control teachers in the RCT sites, a result that differs from other research identifying participation in induction and having a mentor as related to higher retention (Borman & Dowling, 2008). On the whole, the retention rates for teachers in the RCT study (79 percent for treatment and 78 percent for control) were lower than those found among a national sample of teachers beginning teaching in 2007–08, among whom 85 percent remained in teaching 3 years later (Gray & Taie, 2015). This difference raises the possibility that local factors and/or more recent trends may be influencing retention patterns that induction might not address.

We did find significant and positive impacts on student achievement in ELA and mathematics after 2 years of NTC induction support for the combined RCT sites. These results suggest that the NTC induction model can improve the ELA and mathematics achievement of students in beginning teachers' classrooms. The QED using a differences-in-differences approach did not bear out positive impacts on student outcomes. The sample size of beginning teachers teaching ELA or mathematics in grades 4 through 8 that resulted from the participating districts' hiring patterns and testing schedules was very small, and the QED was extremely constrained in being able to detect any effects. The positive impacts found in the RCTs presented here contrast with a 2010 study of comprehensive teacher induction that reported no effects on student achievement after 2 years of induction support (Glazerman et al., 2010). In that study, effects in the third year of teaching, lagging the induction period, were positive and statistically significant but inconsistent under different approaches to estimating impact. Although we tested different scenarios than in the 2010 study, in all cases the student achievement impacts remained similar across the sensitivity analyses we conducted, indicating these findings are robust and stable (see Appendix G).

The results from this evaluation of the i3 Validation grant point to several implications for NTC's i3 Scale Up grant and for the field more broadly.

- High risk of attrition among the observation sample (due to both burden and turnover) means that the initial sample needs to account for an attrition rate higher than average turnover in the profession.
- While classroom observation tools such as the Framework for Teaching and the Classroom Assessment Scoring System (CLASS) have achieved reliability as measures for use in research studies, they may still apply to broad aspects of classroom practice that might not reflect instructional dimensions impacted by the NTC model on beginning teachers. For

instance, measures of how beginning teachers interrogate their own practice, how they make instructional decisions using data, how they think through integrating strategies for diverse learners, or how they exemplify mindsets of a reflective practitioner might round out the aspects of teacher practice NTC aims to improve.

- Developing the logic model and defining the implementation fidelity indicators provided an opportunity for NTC to clarify the aspects of the model program developers believed would be most meaningful in achieving results, and the indicators provided NTC a tool with which to establish common expectations and monitor progress with local program leaders.
- The NTC model by design is a comprehensive one supporting multiple levels of the system—teacher, mentor, school leader, and district. As NTC continues engaging districts across the country in its induction strategy, local contexts that vary at each of those levels will pose new opportunities to refine and adapt the model and raise new questions about which components are nonnegotiable and which can tolerate more flexibility and to what effect.

Under an i3 Scale Up grant that began in 2016, NTC is currently implementing its model in five urban districts across the country. SRI is conducting RCTs in each district. While NTC successfully achieved high implementation fidelity under the i3 Validation grant, scaling up to more districts and more diverse contexts necessitated adaptations to enhance sustainability and applicability. For example, full-time release mentors can be cost-prohibitive for some districts; the Scale Up sample includes several sites using school-based mentors who are retaining their own classroom full or part time as a more cost-efficient staffing model. Additionally, classroom observations as one of the primary sources of data by which mentors support their beginning teachers in examining instructional practice has practical limits if the observations need to be conducted in person. NTC integrated a classroom video tool into Learning Zone to promote more efficient and more frequent observation and feedback by mentors. The expectations for beginning teachers' induction experiences—frequency, intensity, content—have not changed, however. The i3 Scale Up study will build on the results reported here to determine whether and to what extent the NTC induction model with some specific adaptations can achieve high implementation fidelity in larger and more diverse district settings and whether, across these varying contexts, it has positive effects on teacher practice, teacher retention, and student achievement.

REFERENCES

- Borman, G., & Dowling, N. M. (2008, September). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research, 78*(3), 367–409.
- Danielson, C. (2013). *The framework for teaching evaluation instrument: 2013 edition*. Princeton, NJ: The Danielson Group.
- Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M., & Jacobus, M. (2010). *Impacts of comprehensive teacher induction: Final results from a randomized controlled study* (NCEE 2010-4028). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Gray, L., & Taie, S. (2015). *Public school teacher attrition and mobility in the first five years: Results from the first through fifth waves of the 2007–08 Beginning Teacher Longitudinal Study* (NCES 2015-337). U.S. Department of Education. Washington, DC: National Center for Education Statistics. <http://nces.ed.gov/pubsearch>
- Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*, 84–117.
- Hobson, A. J., Ashby, P., Malderez, A., & Tomlinson, P. D. (2009). Mentoring beginning teachers: What we know and what we don't. *Teaching and Teacher Education, 25*(1), 207–216.
- Ingersoll, R., & Strong, M. (2011, June). The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of Education Research, 81*(2), 201–233.
- 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*. (NCSE 2013–3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences (IES), U.S. Department of Education. This report is available on the IES website, <http://ies.ed.gov/ncser/>
- National Commission on Teaching and America's Future. (1997). *Doing what matters most: Investing in quality teaching*. New York, NY: Author.
- National Commission on Teaching and America's Future. (2016). *What matters now: A new compact for teaching and learning*. New York, NY: Author.
- Sanders, W., & Rivers, J. (1996, November). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville: University of Tennessee, Value- Added Research and Assessment Center.
- 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). Retrieved from <http://files.eric.ed.gov/fulltext/EJ1070474.pdf>
- Wang, J., & Odell, S. (2002, Autumn). Mentored learning to teach according to standards-based reform: A critical review. *Review of Educational Research, 72*(3), 481–546.

A Comprehensive Model of Teacher Induction: Implementation and Impact on Teachers and Students

Evaluation of New Teacher Center's i3 Validation Grant Final Comprehensive Report Appendix

December 2017

CONTENTS

Appendix A. Implementation Fidelity Measures	A-1
Calculating Fidelity Scores Across the Full Sample	A-8
Decision rules for rolling up to the component level for each study site:	A-8
Decision rules for rolling up to the component level across study sites:	A-8
Appendix B. Teacher Survey Methods and Measures	B-1
Sample	B-1
Data and Measures	B-2
Appendix C. Randomized Controlled Trials Methods	C-1
Selection and Random Assignment of Schools and Teacher Eligibility	C-1
Appendix D. Teacher Practice Impact Analysis and Model Results	D-1
Analysis	D-1
Sample	D-1
Data and Measures	D-3
Attrition	D-5
Baseline equivalence	D-8
Statistical Analysis	D-13
Results	D-14
Full Model Tables	D-17
Both Cohorts and RCT Districts Combined	D-17
Both Cohorts Combined, District A	D-33
Both Cohorts Combined, District B	D-41
Cohort 1, Both RCT Districts Combined	D-49
Cohort 1, District A	D-57
Cohort 1, District B	D-65
Cohort 2, Both RCT Districts Combined	D-73
Cohort 2, District A	D-81
Cohort 2, District B	D-89
Appendix E. Teacher Retention Impact Analysis and Model Results	E-1
Analysis	E-1
Full Model Tables	E-5
Appendix F. Student Achievement Model Results for RCT Districts	F-1
Sample	F-1
Data and Measures	F-3
Attrition	F-5
Baseline Equivalence	F-8
Statistical Analysis	F-9
Results	F-10
Sensitivity to Teacher Joiners	F-13

Full Model Tables	F-14
Both Cohorts and RCT Districts Combined	F-14
Both Cohorts Combined, District A	F-20
Both Cohorts Combined, District B	F-24
Cohort 1, Both RCT Districts Combined	F-28
Cohort 1, District A	F-32
Cohort 1, District B	F-34
Cohort 2, Both Districts Combined	F-36
Cohort 2, District A	F-40
Cohort 2, District B	F-42
Appendix G. Sensitivity Tests for RCT Results	G-1
Is the estimated impact of the NTC model on student achievement different for elementary school students than for middle school students?	G-1
Is the relationship between the NTC induction model and student achievement affected by students taking mathematics or ELA classes with more than one study teacher?	G-2
Does the relationship between the NTC induction model and student achievement vary by school characteristics?	G-3
Are measures of the frequency, duration, or quality of the mentoring experience associated with higher student achievement?	G-4
Appendix H. QED Study Methods and Student Achievement Model Results.....	H-1
Difference-in-Differences Design	H-1
Baseline Equivalence	H-2
Distribution of Outcomes.....	H-3
Statistical Analysis.....	H-5
Full Model Results	H-6

APPENDIX A. IMPLEMENTATION FIDELITY MEASURES

Exhibit A-1 provides the definition and metric for each element under each key component included for implementation fidelity.

Exhibit A-1: Implementation Fidelity Indicators by Component—Definition, Data Sources, and Thresholds

Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Implementation Score Thresholds
Component 1. New Teacher Center Supports Decision rules for rolling up to the component level for each study site: • High fidelity if greater than 60% of individual indicators are scored as high (3) <u>and</u> less than or equal to 20% of indicators are scored as low (1). • Medium fidelity if individual indicator scores do not reach the threshold for high fidelity (3) <u>and</u> less than 50% of indicators are scored as low (1). • Low fidelity if greater than or equal to 50% of indicators are scored as low (1).			
1.1a	Implementation support by site leads—developing and advocating for full-time mentoring model.	Whether site lead provides support for implementation at the study site by advocating for and developing understanding of a full-time mentoring model.	Interviews with district staff
			3 = present (based on interview data identifying at least one activity with this indicator as the goal) 1 = absent
1.1b	Implementation support by NTC leads—engaging principals.	Whether NTC lead provides support for implementation at the study site by conducting a half-day “role of the principal” training session for principals in schools receiving NTC services.	Attendance records from NTC
			3 = 80%+ of principals in their first year at an NTC- served school participated in a half- day training session; 2 = 60–79% of principals in their first year at an NTC- served school participated in a half- day training session; 1 = Less than 60% of principals in their first year at an NTC- served school participated in a half- day training session
1c	Implementation support by site leads—engaging principals.	Whether site lead provides support for implementation at the study site by engaging principals through annual 1:1 meetings.	Attendance records from NTC
			3 = 80%+ of principals met 1:1 with site lead at least once during the academic year; 2 = 60–79% of principals met 1:1 with site lead at least once during the academic year; 1 = Less than 60% of principals met 1:1 with site lead at least once during the academic year.

Exhibit A-1: Implementation Fidelity Indicators by Component—Definition, Data Sources, and Thresholds (continued)

Key Elements of Component		Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Implementation Score
1.2	Capacity building by site leads.	Whether site lead builds sustainability for district-led mentor program after NTC involvement by attending NTIN to conduct goal-setting along with 2 additional LEA staff in year 2; attending presenters' academy along with those same LEA staff in year 2; leading all mentor forums in year 2; and co-presenting all mentor academies in year 3.	Attendance records from NTC	3 = All requirements for that year completed; 1 = Not all requirements for that year completed
1.3a	Development of program standards by NTC national office.	Whether the national NTC office creates and distributes clearly defined program standards to support high-quality mentoring and mentor development.	Copies of the standards	3 = present; 1 = absent
1.3b	Development of formative assessment tools by NTC national office.	Whether the national NTC office creates and distributes formative assessment tools to support high-quality mentoring and mentor development.	Copies of the formative assessment tools	3 = present; 1 = absent
1.3c	Development of mentor training materials by NTC national office.	Whether the national NTC office creates and distributes training materials to support high-quality mentoring and mentor development.	Copies of the mentor training materials	3 = present; 1 = absent
1.3d	Development of online mentoring platform by NTC national office.	Whether the national NTC office creates and grants mentors access to an online platform to support high-quality mentoring and mentor development.	Access to data from the online mentoring platform	3 = present; 1 = absent

Exhibit A-1: Implementation Fidelity Indicators by Component—Definition, Data Sources, and Thresholds (continued)

Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Implementation Score
Component 2. Mentor Selection and Assignment			
Decision rules for rolling up to the component level for each study site:			
• High fidelity if greater than 60% of individual indicators are scored as high (3) <u>and</u> less than or equal to 20% of indicators are scored as low (1). • Medium fidelity if individual indicator scores do not reach the threshold for high fidelity (3) <u>and</u> less than 50% of indicators are scored as low (1). • Low fidelity if greater than or equal to 50% of indicators are scored as low (1).			
2.1	Rigorous process for selecting high- quality mentors.	Whether study site follows a rigorous mentor selection process that includes explicit selection criteria, formal job posting, and multiple rounds of screening.	Interviews with NTC leads and site leads/ Staff
			3 = High (all steps completed); 2 = Medium (2 steps completed); 1 = Low (0 or 1 step(s) completed)
2.2	Mentors released full time from teaching assignments.	Whether the study site releases all of its mentors from their classroom teaching responsibilities to be full-time mentors.	Mentor survey
			3 = 80%+ of mentors released full time; 2 = 60–79% of mentors released full time; 1 = Less than 60% of mentors are released full time.
2.3	Mentors assigned no more than 15 mentees.	Whether each mentor is assigned less than or equal to 15 mentees.	Rosters from each study site and mentor survey
			3 = 80%+ of mentors assigned 15 or fewer mentees; 2 = 60–79% of mentors assigned 15 or fewer mentees; 1 = Less than 60% of mentors assigned 15 or fewer mentees.

Exhibit A-1: Implementation Fidelity Indicators by Component—Definition, Data Sources, and Thresholds (continued)

Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Implementation Score
Component 3. Mentor Development and Accountability Decision rules for rolling up to the component level for each study site: • High fidelity if greater than 60% of individual indicators are scored as high (3) <u>and</u> less than or equal to 20% of indicators are scored as low (1). • Medium fidelity if individual indicator scores do not reach the threshold for high fidelity (3) <u>and</u> less than 50% of indicators are scored as low (1). • Low fidelity if greater than or equal to 50% of indicators are scored as low (1).			
3.1a	Mentors participate in mentor academy.	Whether the mentor participates in mentor academies (4 per year).	Attendance records from NTC 3 = 80%+ of mentors attended 80%+ of offered mentor academy days; 2 = 60–79% of mentors attended 80%+ of offered mentor academy days; 1 = Less than 60% of mentors attended 80%+ of offered mentor academy days.
3.1b	Site holds the number of mentor academy days specified by NTC.	Whether the site held all 12 of the mentor academy days specified by NTC.	Attendance records from NTC 3 = Site held at least 90% of mentor academy days; 2 = Site held 76–89% of mentor academy days; 1 = Site held less than or equal to 75% of mentor academy days.
3.2a	Mentors participate in mentor forums.	Whether the mentor participates in mentor forums (12 per year for each year serving as an NTC mentor).	Attendance records from NTC 3 = 80%+ of mentors attended 80%+ of offered mentor forum days; 2 = 60–79% of mentors attended 80%+ of offered mentor forum days; 1 = Less than 60% of mentors attended 80%+ of offered mentor forum days.
3.2b	Site holds the number of mentor forums specified by NTC.	Whether the site held all 12 of the mentor forum meetings specified by NTC.	Attendance records from NTC 3 = Site held at least 90% of mentor forum days; 2 = Site held 76–89% of mentor forum days; 1 = Site held less than or equal to 75% of mentor forum days

Exhibit A-1: Implementation Fidelity Indicators by Component—Definition, Data Sources, and Thresholds (continued)

Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Implementation Score
3.3 Mentors participate in mentor-to-mentor shadowing,	Whether the mentor participates in mentor- to-mentor shadowing at least once per year in their first year as an NTC mentor and at least twice per year in subsequent years.	Learning Zone (MAGA Tool—Mentor Field Observation Tool)	3 = 80%+ of mentors conducted required shadowing; 2 = 60–79% of mentors conducted required shadowing; 1 = Less than 60% of mentors conducted required shadowing.
3.4 Mentors receive support and feedback from site leads.	Whether the mentor meets with the site lead one on one for at least 3 hours per quarter. Operationalized Y2: 2 full observation cycles and 6 1-1 meetings with program lead.	Learning Zone (MAGA Interaction Log—Time field/ date field—Ad hoc reports)	3 = 80%+ of mentors met with site lead at least 3 hours/quarter for each quarter in year; 2 = 60–79% of mentors met with site lead at least 3 hours/quarter for each quarter in year; 1 = Less than 60% of mentors met with site lead at least 3 hours/quarter for each quarter in year.
3.5 Mentors engage in peer coaching and goal-setting process.	Mentors complete mentor collaborative assessment log; set mentor professional goals using self-assessment summary and ILP; complete mid-year review; and complete mentor professional growth reflection. Y2 operationalized: Peer Coaching: Complete Mentor CAL with Peer Mentor (LM) Goal Setting: Complete three Mentor Assessment: Individual Learning Plans (one each for initial, mid year and end of year)	Learning Zone or hard copy (MAGA Tool Count Report for specific tools of interest— hard copies will be noted in Learning Zone retroactively)	3 = 80%+ of mentors completed peer coaching and goal- setting requirements; 2 = 60–79% of mentors completed peer coaching and goal-setting requirements; 1 = Less than 60% of mentors completed peer coaching and goal-setting requirements.

Exhibit A-1: Implementation Fidelity Indicators by Component—Definition, Data Sources, and Thresholds (continued)

Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Implementation Score
Component 4. Provision of High-Quality Mentoring Decision rules for rolling up to the component level for each study site: • High fidelity if greater than 60% of individual indicators are scored as high (3) <u>and</u> less than or equal to 20% of indicators are scored as low (1). • Medium fidelity if individual indicator scores do not reach the threshold for high fidelity (3) <u>and</u> less than 50% of indicators are scored as low (1). • Low fidelity if greater than or equal to 50% of indicators are scored as low (1).			
4.1	Mentors meet regularly with new teachers.	Whether mentors meet with each of their new teachers at least 180 minutes per month over at least 7 months.	Learning Zone (Ad Hoc Report – Admin Section – Date and Time Fields)
			3 = 80%+ of mentees met with mentor the required amount of time in 7+ months; 2 = 60–79% of mentees met with mentor the required amount of time in 7+ months; 1 = Less than 60% of mentees met with mentor the required amount of time in 7+ months.
4.2	Mentors use NTC’s formative assessment system.	Whether mentors use a Formative Assessment System tool for at least 85% of their interactions with beginning teachers.	Learning Zone (Interaction Log – Tool Use Field & Tool Report Count)
			3 = 80%+ of mentees had a formative assessment tool completed during at least 85% of meetings with mentor; 2 = 60–79% of mentees had a formative assessment tool completed during at least 85% of meetings with mentor; 1 = Less than 60% of mentees had a formative assessment tool completed during at least 85% of meetings with mentor.
4.3	Mentors document their reflections on mentoring work.	Whether mentors fill out the “goal” field in the Learning Zone interaction log after interacting with beginning teachers.	Learning Zone (Interaction Log— Reflection Section, Item 1— Goal. This is a text field for which we will be looking for non-missing data.)
			3 = 80%+ of mentees had “goal” field filled out for 80%+ of interactions with their mentor; 2 = 60–79% of mentees had “goal” field filled out for 80%+ of interactions with their mentor; 1 = Less than 60% of mentees had “goal” field filled out for 80%+ of interactions with their mentor.

Exhibit A-1: Implementation Fidelity Indicators by Component—Definition, Data Sources, and Thresholds (concluded)

	Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Implementation Score
4.4	Mentors focus on instructional practice and on equity and universal access in their work with mentees.	Whether mentors focus on equity and universal access in their instructional mentoring.	Teacher survey	3 = 80%+ of mentees agree or strongly agree that mentoring focused on equity and universal access; 2 = 60–79% of mentees agree or strongly agree that mentoring focused on equity and universal access; 1 = Less than 60% of mentees agree or strongly agree that mentoring focused on equity and universal access.
4.5	New teachers' perceptions of value of mentoring	Extent to which new teachers find the support that they receive from their NTC mentors to be valuable.	Teacher survey	3 = 80%+ of mentees' scores are over 3; 2 = 60–79% of mentees' scores are over 3; 1 = Less than 60% of mentees' scores are over 3.

Calculating Fidelity Scores Across the Full Sample

Fidelity scores were calculated annually for each key component. For specific indicators that were not measured every year (1.1a, 1.1b, 1.2, 1.3a–d, 2.1, as specified above), those individual indicators were excluded from the component-level analysis in the years when they were not measured.

Decision rules for rolling up to the component level for each study site:

- High fidelity if > 60% of individual indicators are scored as high and ≤20 percent of indicators are scored as low.
- Medium fidelity if individual indicator scores do not reach the threshold for high fidelity and < 50 percent of indicators are scored as low.
- Low fidelity if ≥ 50 percent of indicators are scored as low.

Decision rules for rolling up to the component level across study sites:

- High fidelity if at least two sites are scored as high and no sites are scored as low (this is the threshold for implementation “with fidelity” at the program level).
- Medium fidelity if the combination of scores across the three sites does not reach the threshold for high fidelity and no more than one site is scored as low.
- Low fidelity if at least two sites are scored as low.

APPENDIX B. TEACHER SURVEY METHODS AND MEASURES

To measure the contrast between the induction experiences of treatment and control teachers, NTC administered surveys to all teachers and mentors served by their program as well as control teachers in the spring of each year of implementation.

Sample

The survey was administered to all eligible study teachers with no restrictions. However, response rates were substantially better among treatment than control teachers. Exhibit B-1 shows the site sample sizes by years of teaching and Exhibit B-2 provides response rates by site and overall, by years of teaching experience. .

Exhibit B-1. Survey Sample by Years of Teaching Experience, Combined Cohorts

Site	Year 1			Year 2		
	Treatment	Control	Subtotal	Treatment	Control	Subtotal
BCPS	187	103	290	148	43	191
CPS	146	95	241	109	83	192
GWAEA	230	99	329	193	84	277
<i>Totals</i>	<i>563</i>	<i>297</i>	<i>860</i>	<i>450</i>	<i>210</i>	<i>660</i>

Exhibit B-2. Survey Response Rates

Site	Years of Experience—Both Cohorts					
	Year 1 ^a		Year 2		Overall	
	Treatment	Control	Treatment	Control	Treatment	Control
BCPS						
Surveyed	213	161	161	126	374	287
Responded	187	103	148	43	335	146
<i>Response Rate (%)</i>	<i>88%</i>	<i>64%</i>	<i>92%</i>	<i>34%</i>	<i>90%</i>	<i>51%</i>
CPS						
Surveyed	147	139	129	130	276	269
Responded	146	95	109	83	255	178
<i>Response Rate (%)</i>	<i>99%</i>	<i>68%</i>	<i>84%</i>	<i>64%</i>	<i>92%</i>	<i>66%</i>
GWAEA^b						
Surveyed	248	117	228	117	476	234
Responded	230	99	193	84	423	183
<i>Response Rate (%)</i>	<i>93%</i>	<i>85%</i>	<i>85%</i>	<i>72%</i>	<i>89%</i>	<i>78%</i>
Totals						
Surveyed	608	417	518	373	1126	790
Responded	563	297	450	210	1013	507
<i>Response Rate (%)</i>	<i>93%</i>	<i>71%</i>	<i>87%</i>	<i>56%</i>	<i>90%</i>	<i>64%</i>

Source: NTC New Teacher Survey, spring 2013–2016

^a Attrition data was not available for the Year 1 calculations.

^b The control group was administered the teacher survey one year prior to the treatment group in GWAEA.

Data and Measures

Exhibit B-3 displays the items comprising each survey scale.

Exhibit B-3. Items Comprising Each Survey Scale

Items	Scale
Availability of Materials	
To what extent do you agree or disagree with the following statements about your school?	Strongly
Please choose the response that best reflects your experience at your current school.	disagree
I have the necessary textbooks and print resources to teach.	Disagree
I can get instructional materials (e.g., lab supplies, math manipulatives, classroom library books) without buying them myself.	Agree
I can get the classroom supplies (e.g., paper, pencils, staples, tape) I need without buying them myself.	Strongly agree
Supportive Teaching Environment	
To what extent do you agree or disagree with the following statements about your school?	Strongly
Please choose the response that best reflects your experience at your current school.	disagree
Teachers in this school trust each other.	Disagree
I feel supported by colleagues to try out new ideas.	Agree
Teachers in this school feel responsible to help each other do their best.	Strongly agree
A conscious effort is made by faculty to make new teachers feel welcome here.	
Teacher Collaboration Around Instruction	
How often this year do you do each of the following activities with teachers in your school other than with a mentor teacher or consultant?	Never or once
Analyze samples of work done by your students	A few times
Work together to develop teaching materials or activities for particular classes	At least monthly
Seek each other's advice about instructional issues and problems	At least weekly
Observe each other's classrooms to offer feedback and/or learn strategies (excluding observation for the purpose of formal evaluation)	
Discuss student assessment data to make decisions about instruction	
Instructional Leadership	
Please rate how strongly you agree or disagree with the following statements about school leadership in your school? (Mark one answer per question.)	
The faculty and leadership have a shared vision.	Strongly
There is an atmosphere of trust and mutual respect in this school.	disagree
The majority of teachers in your school feel comfortable raising issues and concerns that are important to them.	Disagree
The school leadership consistently supports teachers.	Agree
Teacher performance is assessed objectively.	Strongly agree
Teachers receive feedback that can help them improve teaching.	
The procedures for teacher evaluation are consistent.	
The school improvement team provides effective leadership at the school.	
The faculty are recognized for accomplishments.	

Exhibit B-3. Items Comprising Each Survey Scale (continued)

Items	Scale
Student Conduct	
Please rate how strongly you agree or disagree with the following statements about managing student conduct in your school? (Mark one answer per question.)	
Students at this school understand expectations for their conduct.	Strongly disagree
Students at this school follow rules of conduct.	Disagree
Policies and procedures about student conduct are clearly understood by the faculty.	Agree
School administrators consistently enforce rules for student conduct.	Strongly agree
School administrators support teachers' efforts to maintain discipline in the classroom.	
Teachers consistently enforce rules for student conduct.	
The faculty work in a school environment that is safe.	
Frequency of High-Leverage Mentoring Activities	
For the following set of questions, think about the mentoring you have received this school year (including the summer of 2013) through your new teacher support program.	
(a) How often have you received the following supports from your mentor(s) or consultant(s)?	
My mentor/consultant has observed me teaching and provided feedback	Never or once
My mentor/consultant has worked with me to develop a professional growth plan	A few times
My mentor/consultant has demonstrated lessons for me in my classroom	At least monthly
My mentor/consultant has given me materials	At least weekly
My mentor/consultant has planned lessons with me	
My mentor/consultant has analyzed samples of my students' work	
My mentor/consultant has talked with me about the strengths and/or needs of specific students	
My mentor/consultant has discussed instructional issues and problems	
My mentor/consultant has discussed student assessment data to make decisions about instruction	
Value of Mentoring Activities	
For the following set of questions, think about the mentoring you have received this school year (including the summer of 2013) through your new teacher support program.	
(b) If you received the support, how valuable has it been to your development as a teacher?	
My mentor/consultant has observed me teaching and provided feedback	Not valuable
My mentor/consultant has worked with me to develop a professional growth plan	Minimally valuable
My mentor/consultant has demonstrated lessons for me in my classroom	Moderately valuable
My mentor/consultant has given me materials	Extremely valuable
My mentor/consultant has planned lessons with me	
My mentor/consultant has analyzed samples of my students' work	
My mentor/consultant has talked with me about the strengths and/or needs of specific students	
My mentor/consultant has discussed instructional issues and problems	
My mentor/consultant has discussed student assessment data to make decisions about instruction	

Exhibit B-3. Items Comprising Each Survey Scale (continued)

Items	Scale
Focus on Instruction	
Thinking about all the new teacher supports you have received during the 2013–14 school year (including summer 2013), to what extent have they addressed the following topics?	
The subject matter I teach	
Instructional techniques appropriate for the grade level/subject matter I teach	Not at all
Classroom management techniques appropriate for the students I currently teach	addressed
The use of textbooks or other curricular materials for my current position	Minimally
Strategies for interacting with parents of the students I currently teach	addressed
The use of data (e.g., analyzing student work or student test scores) to plan instruction	Moderately
Adapting instruction to meet the needs of students at varying academic levels	addressed
Adapting instruction for students with individualized education programs	Extensively
Instructional techniques to meet the needs of students from diverse cultural backgrounds	addressed
Planning lessons and designing instruction	
Creating a positive learning environment	
The use of informal and formal assessment strategies	
Evaluating and reflecting upon my own teaching practices	
Using culturally responsive pedagogical practices	
Frequency of Other Induction Supports	
Think about the services and support you have received this school year (including the summer of 2013) through your new teacher support program.	
How often have you participated in new teacher meetings with the principal at your school?	Never or once
	A few times
How often have you participated in workshops, seminars, or classes for new teachers (excluding an initial orientation)?	At least monthly
	At least weekly
How often have you received release time to see other teachers teach?	
How often have you participated in a professional development network specifically for new teachers?	
Value of Other Induction Supports	
Think about the services and support you have received this school year (including the summer of 2013) through your new teacher support program.	
How valuable was this to your development as a teacher... in new teacher meetings with the principal at your school?	Not valuable
	Minimally valuable
How valuable was this to your development as a teacher... in workshops, seminars, or classes for new teachers (excluding an initial orientation)?	Moderately
	valuable
How often have you received release time to see other teachers teach?	Extremely valuable
How valuable was this to your development as a teacher... in a professional development network specifically for new teachers?	

Exhibit B-3. Items Comprising Each Survey Scale (continued)

Items	Scale
Need for Instructional Support	
Thinking about the 2013–14 school year, indicate the level of support you have needed in the following areas. (Mark one answer per question.)	
The subject matter I teach	No support needed
Instructional techniques appropriate for the grade level/subject matter I teach	Minimal support
The use of data (e.g., analyzing student work or student test scores) to plan instruction	needed
Adapting instruction to meet the needs of students at varying academic levels	Moderate support
Adapting instruction for students with individualized education programs	needed
Instructional techniques to meet the needs of students from diverse cultural backgrounds	Extensive support
Planning lessons and designing instruction	needed
Creating a positive learning environment	
The use of informal and formal assessment strategies	
Teacher-Reported Growth	
Thinking about all the new teacher supports you have received during the 2013-14 school year (including summer 2013), to what extent have they improved your knowledge and skills in the following areas?	
Deepened my grasp of the subject matter I teach.	
Increased my knowledge of instructional techniques appropriate for the grade level/subject matter I teach.	
Improved my classroom management.	
Increased my effectiveness in using textbooks or other curricular materials.	
Improved my interactions with parents.	
Improved my ability to use data (e.g., analyzing student work or student test scores) to plan instruction.	Not at all
Improved my ability to adapt instruction to meet the needs of students at varying academic levels.	Minimal extent
Increased my ability to adapt instruction for students with individualized education programs.	Moderate extent
Improved my ability to meet the instructional needs of students from diverse cultural backgrounds.	Great extent
Improved my ability to meet instructional needs of English language learners.	
Improved my ability to plan lessons and design instruction.	
Increased my ability to create a positive learning environment.	
Increased my effectiveness in using informal and formal assessment strategies.	
Improved my ability to evaluate and reflect upon my own teaching practices.	
Influenced my decision to stay in the profession.	

Exhibit B-3. Items Comprising Each Survey Scale (concluded)

Items	Scale
Teacher Self-Efficacy	
To what extent do you agree or disagree with each of the following statements? (Mark one answer per question.)	
I am confident in my ability to teach effectively.	
I can handle a range of challenging classroom management and discipline situations.	
If a student in my class becomes disruptive and noisy, I know techniques to redirect him/her quickly.	Strongly disagree
I am equally successful in helping students from all racial/ethnic backgrounds to learn.	Disagree
I have the knowledge and skills I need to address the needs of special education students.	Agree
If I really try hard, I can get through to even the most difficult or unmotivated students.	Strongly agree
If a student did not remember information I gave in a previous lesson, I would know how to increase his/her retention in the next lesson.	
If one of my students couldn't do a class assignment, I would be able to accurately assess whether the assignment was at the correct level of difficulty.	
I am able to adapt instruction so that I meet the needs of students at varying academic levels equally well.	
Mentor Support for New Teacher Participation in Other Professional Development	
To what extent has your mentor teacher or consultant supported your participation in the following activities?	
Common planning time with colleagues	Not at all
Small learning communities focused on instruction	Minimal extent
Other professional development that the district offers	Moderate extent
Work with an instructional coach, in addition to the work I do with my mentor/consultant	Great extent

We created 14 survey scales from items in the teacher survey, capturing school environment, mentoring and other induction supports, and teacher self-evaluation. Using 2013–14 survey data from all respondents, we carried out a factor analysis to examine the properties of the 14 scale variables. This analysis was used to ensure that combining survey items results in conceptually relevant scales that are also reliable and useful for analysis. All the scales are highly reliable, with most achieving alphas of over 0.8 (Exhibit B-3). After the factor analysis, we created scale variables using a weighted average approach. This approach keeps the composite variables in the same scale as the original items, which makes them easier to interpret.

Exhibit B-4. Properties of Teacher Survey Scale Variables

	Eigenvalue	Alpha	Number of items	<i>n</i>
School Environment				
Student Conduct	3.87	0.88	7	454
Instructional Leadership	5.37	0.93	9	454
Teacher Collaboration around Instruction	2.74	0.81	5	454
Supportive Teaching Environment	2.50	0.82	4	454
Availability of Materials	2.07	0.77	3	454
Mentoring and Induction				
Frequency of Mentoring Activities	4.39	0.89	9	374
Frequency of Other Induction Supports	1.93	0.58	4	455
Value of Other Induction Supports	2.23	0.77	4	180
Focus on Instruction	7.53	0.95	14	358
Value of Mentoring Activities	5.80	0.90	9	125
Mentor Support for New Teacher Participation in Other Professional Development ^a	2.75	0.85	4	189
Self-Evaluation				
Teacher Self-Efficacy	4.27	0.88	9	455
Teacher-Reported Growth	8.76	0.96	15	167
Need for Instructional Support	4.35	0.88	9	455

Source: NTC New Teacher Survey, spring 2014 and spring 2016.

^aFactor based on spring 2016 survey. All other factors are based on spring 2014 survey.

APPENDIX C. RANDOMIZED CONTROLLED TRIALS METHODS

The evaluation used a randomized controlled trial (RCT) in two of the three participating districts—Broward County Public Schools (BCPS) and Chicago Public Schools (CPS)—to study the impact of NTC’s new teacher induction model on teacher practice and student achievement.¹

Selection and Random Assignment of Schools and Teacher Eligibility

This study included two cohorts of schools, the first randomly assigned in 2013 and the second randomly assigned in 2014. All schools serving grades K–8 in CPS and all schools serving grades K–12 in BCPS who hired beginning teachers before October 1, 2013 (Cohort 1) or October 1, 2014 (Cohort 2) were eligible to participate in the study. Starting in the late summer, we collected data from the school districts on hiring in all district schools and identified those schools with eligible first-year teachers.

Teachers were eligible for this study if they were in their first year of teaching and were hired into an instructional position within a study school by October 1, 2013, (Cohort 1) or October 1, 2014, (Cohort 2). Teachers were defined as “first-year” teachers if they had less than 2 continuous months of prior teaching experience, excluding experience as a substitute teacher and experience in foreign countries. While each district stored data on teachers’ prior experience, this data varied in its type, breadth, quality, and consistency. On the teacher survey in each year, we asked teachers to provide their self-reported years of experience, but this data also proved to be both incomplete (response rates were low) and inconsistent (teachers provided different answers on the survey than they did to their mentors). Therefore, we relied on NTC staff to verify eligibility with both the treatment and control groups.

In the first year (2013–14) in CPS, we randomly assigned participating schools employing beginning teachers to treatment and control conditions, blocking on geographic area (18 categories) and grades served (K–5 or K–8 school). These blocking variables were selected because school context and student achievement vary substantially by both geography and school level in CPS. Because layoffs and school closings delayed hiring in summer 2013, we randomly assigned schools in waves. As each new group of schools with beginning teachers was identified, we randomly assigned them to the treatment or control condition until we reached the target number of beginning teachers to be served (135 treatment and control teachers). NTC hired nine mentors in CPS in the first year, each with the capacity to serve 15 teachers. However, mentors served both first- and second-year teachers. Therefore, the total number of study treatment teachers (first-year teachers) they were able to serve in CPS in Cohort 1 was 68. The study then included 68 control teachers for a balanced sample. See Exhibit C-1 for a graphic depiction of this process.

In the second year of the study (2014–15), all Cohort 1 teachers who remained in CPS were included in the study, and treatment teachers continued to be served by the NTC mentors. However, in addition to all second-year (non-study) teachers who were served by NTC mentors during the prior year exiting the induction program as planned, 14 treatment teachers left the district and/or attrited from the study. Therefore, NTC mentors had additional capacity in their rosters, and were able to add more than 68 additional first-year teachers in Cohort 2. All first-year teachers newly hired into the existing treatment and control schools in 2014–15 were added to the study first. Then, to reach the target number of beginning teachers, we randomly assigned additional schools to the treatment and control conditions. The first and second cohorts thus

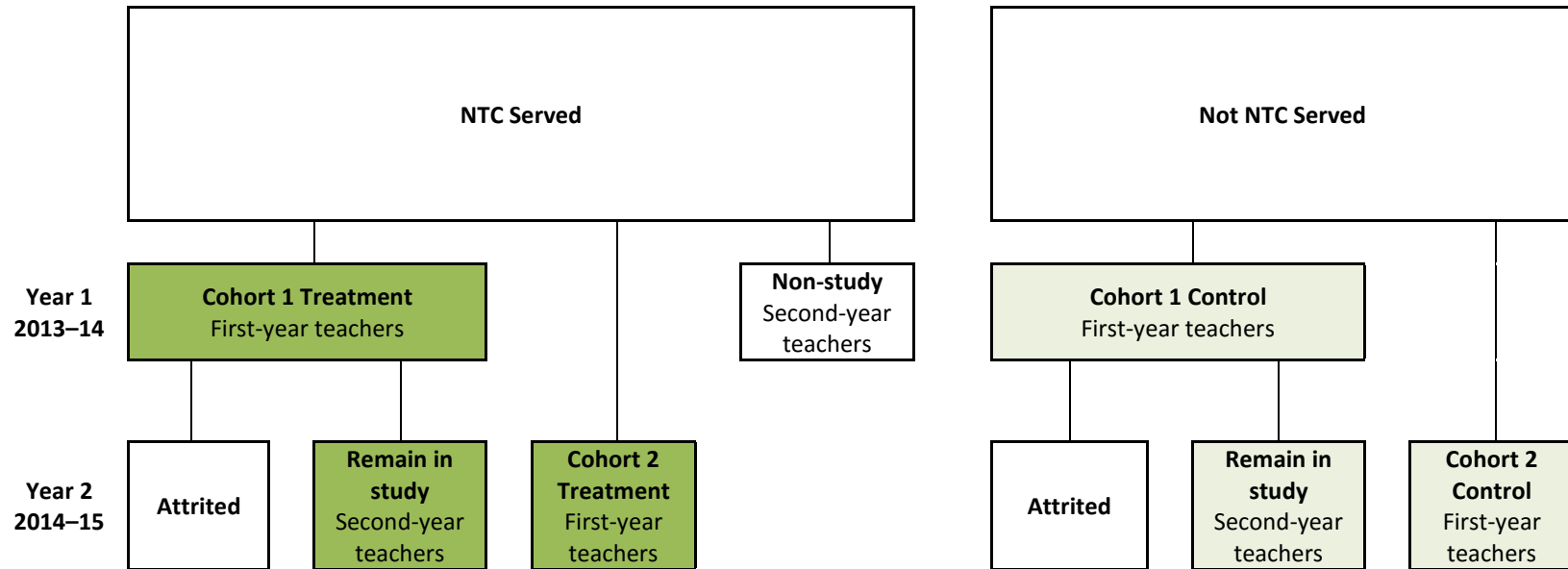
1 Grant Wood Area Education Agency (GWAEA), a consortium of rural districts in Iowa, also implemented the NTC model. Because the model was implemented with all new teachers, SRI used a difference-in-differences approach to study impact in GWAEA. Methods for the difference-in-differences approach are in Appendix H.

totaled 149 treatment and 139 control teachers before attrition. The approach of assigning schools as they hired new teachers and setting a cut-off date of October 1 meant that the study sample included only schools that completed at least a portion of their hiring early in the school year. However, as discussed below, the study schools generally reflected the demographics of the district, so selecting schools with hiring completed early likely did not impact generalizability.

In BCPS, we randomly assigned a sample of participating schools employing new teachers to treatment and control conditions, blocking on Teacher Incentive Fund (TIF) status and grades served (elementary, middle, and high schools). These blocking variables were selected because school context and student achievement vary substantially by both TIF status and school level in BCPS. Within each block, we assigned schools to the NTC program or to the usual district supports for new teachers, until the target number of beginning teachers was reached. NTC hired seven mentors in BCPS in the first year, each with the capacity to serve 15 teachers. Unlike CPS, BCPS mentors served only first-year teachers in that first year of implementation. Therefore, the total number of study treatment teachers (first-year teachers) they were able to serve in BCPS in Cohort 1 was 105. The study was only able to identify 88 teachers in control schools, so the sample in the first cohort was unbalanced at the teacher level, though balanced at the school level.

In the second year (2014–15), all incoming first-year teachers in BCPS treatment schools were added to the treatment group and all incoming first-year teachers in control schools were added to the control group. To reach the target number of beginning teachers for Cohort 2, we also randomly assigned new schools to the treatment and control conditions. The first and second cohorts thus totaled 193 treatment and 148 control teachers. The number of control teachers was lower than the number of treatment teachers because the schools were uneven in the number of first-year teachers they hired before October 1. (Exhibit C-1 portrays this same process for BCPS as for CPS.)

Exhibit C-1. Schematic of Teachers in Schools Randomly Assigned to Treatment and Control, Years 1 and 2



In both districts, NTC served all new teachers in treatment schools unless they were served by other comprehensive induction programs with an intensive mentoring component such as Teach For America. To the extent possible, teachers covered under other induction programs with intensive mentoring components were excluded from both the treatment and control conditions.

At the time of random assignment, the schools included in this study had high average proportions of non-white students (86 percent) and students receiving free or reduced-price lunch (78 percent). They were distributed across the categories of school quality assigned by their districts, with slightly more schools in the highest categories (Exhibit C-2).²

Exhibit C-2. Descriptive Statistics for All Study Schools³

	Overall		Treatment		Control	
	Mean	SD	Mean	SD	Mean	SD
Site						
District A	38	49	40	49	37	48
District B	62	49	60	49	63	48
Cohort						
Cohort 1 Only	26	44	20	40	31	46
Cohort 2 Only	39	49	36	48	42	50
Cohort 1 and 2	35	48	44	50	27	45
School Demographics						
Percent free or reduced-price lunch	78	22	79	21	78	23
Percent non-white	86	18	88	17	85	18
Report Card Rating of F or 3	18	38	19	40	16	37
Report Card Rating of D	5	22	5	21	6	24
Report Card Rating of C or 2	30	46	31	46	30	46
Report Card Rating of B	9	28	11	32	7	25
Report Card Rating of A or 1	38	49	34	48	41	49
Percent English Language Learners	18	18	17	18	19	17
Percent Special Education	13	11	13	7	13	14
Blocking Variables						
District A block 1	4	21	5	21	4	20
District A block 2	0	7	0	0	1	9
District A block 3	21	41	18	38	24	43
District A block 4	5	22	6	25	3	18
District A block 5	12	33	14	35	11	31
District A block 6	2	13	2	14	2	13
District B block 1	15	36	16	37	15	36
District B block 2	0	7	0	0	1	9
District B block 3	1	9	0	0	2	13
District B block 4	15	35	14	35	15	36
District B block 5	8	28	8	28	8	28
District B block 6	8	27	9	29	7	25
District B block 7	4	18	5	21	3	16
District B block 8	2	15	2	14	3	16
District B block 9	2	13	2	14	2	13
<i>n</i> Schools	227		108		119	

² At the time of random assignment, BCPS assigned all schools a letter grade from A to F, while CPS used whole numbers between 1 and 3. This baseline “Report Card Rating” was used as a control for school quality in the models.

³ All district-level outcomes analyses are labeled District A and District B to preserve confidentiality.

In BCPS, the study sample included high school teachers. In CPS, NTC served only teachers in K-8 so the study sample was limited to those grades. The study sample was primarily white (56 percent) and female (78 percent) and most study teachers held a Bachelor's degree as their highest degree (72 percent; Exhibit C-3).

Exhibit C-3. Descriptive Statistics for All Study Teachers

	Overall		Treatment		Control	
	Mean	SD	Mean	SD	Mean	SD
Site						
District A	54%	50	56%	50	52%	50
District B	46%	50	44%	50	48%	50
Cohort						
Cohort 1	50%	50	46%	50	54%	50
Cohort 2	50%	50	54%	50	46%	50
Teacher race/ethnicity						
White	56%	50	53%	50	59%	49
Black/African American	25%	44	27%	44	24%	43
Asian/Pacific Islander	3%	18	3%	18	3%	17
Hispanic	13%	34	14%	35	11%	32
Native American	1%	10	2%	13	0%	6
Other ⁴	2%	13	1%	12	2%	13
Female	78%	42	77%	42	78%	41
Teacher highest degree						
Associate's	2%	15	3%	17	2%	13
Bachelor's	72%	45	73%	45	71%	45
Master's	24%	42	22%	41	26%	44
Only partial certification	27%	44	28%	45	25%	43
<i>n</i> Teachers	629		342		287	

⁴ The "other" race category was only available in one of the districts.

APPENDIX D. TEACHER PRACTICE IMPACT ANALYSIS AND MODEL RESULTS

Appendix D provides details on the analysis of the impact of NTC induction on teacher practice.

Analysis

The analysis of the impact of the NTC induction support on teacher practice combined teachers across cohorts in their first and second years of teaching as detailed in Exhibit D-1. The purpose of this approach was to maximize the size of the sample.

Exhibit D-1. Cohort by School Year and Years of Experience

New Teacher Cohort	Years of Teaching Experience	
	1	2
Cohort 1	2013–14	2014–15
Cohort 2	2014–15	2015–16

Sample

The analysis of teacher practice included treatment and control teachers who were randomly selected and were observed in fall 2013 (time 1, baseline) and spring 2015 (time 2) for Cohort 1 or fall 2014 (time 1, baseline) and spring 2016 (time 2) for Cohort 2. Teachers were eligible for the sample if they taught core subjects (mathematics, reading/English language arts, social studies, science, or self-contained elementary classrooms), and all observations were conducted during instruction in the core subjects. In BCPS, the observation sample included high school teachers. In CPS, NTC served only teachers in K–8 so the observation sample was limited to those grades.

An approximately equal number of teachers was observed at baseline across districts and cohorts (Exhibit D-2). As in the larger sample, observed teachers were predominantly white (50 percent) and female (81 percent) and held a Bachelor's (69 percent) as their highest degree. On average, these teachers taught in schools with high proportions of non-white students (85 percent) and students receiving FRPL (77 percent). Their schools were distributed across the categories of school quality assigned by their districts, with slightly more schools in the highest categories. Unlike the larger study sample, substantially more observed teachers in treatment were only partially certified (35 percent) compared to control (20 percent; Exhibit D-2).

Exhibit D-2. Descriptive Statistics for Sample of Teachers Observed at Baseline

	Overall			Treatment			Control		
	Mean	SD	n obs	Mean	SD	n obs	Mean	SD	n obs
Site									
District A	50%	50	233	50%	0.50	121	49%	50	112
District B	50%	50	233	50%	0.50	121	51%	50	112
Cohort									
Cohort 1	50%	50	233	48%	0.50	121	52%	50	112
Cohort 2	50%	50	233	52%	0.50	121	48%	50	112

Exhibit D-2. Descriptive Statistics for Sample of Teachers Observed at Baseline (concluded)

	Overall			Treatment			Control		
	Mean	SD	n obs	Mean	SD	n obs	Mean	SD	n obs
Teacher race/ethnicity									
White	49%	50	233	44%	50	121	55%	50	112
Black/African American	26%	44	233	26%	44	121	26%	44	112
Asian/Pacific Islander	3%	17	233	4%	20	121	2%	13	112
Hispanic	16%	37	233	19%	39	121	13%	34	112
Other	5%	22	233	7%	25	121	4%	19	112
Female	49%	50	232	80%	40	121	83%	38	111
Teacher highest degree									
Associate's	2%	14	211	2%	13	109	2%	14	102
Bachelor's	69%	46	213	69%	46	111	70%	46	102
Masters	27%	45	217	27%	44	112	28%	45	105
Only partial certification	28%	45	228	35%	48	119	20%	40	109
School demographics									
Percent free or reduced-price lunch	77%	21	233	77%	19	121	78%	23	112
Percent non-white	85%	16	233	86%	17	121	85%	16	112
Report Card Rating of F or 3	13%	34	233	12%	33	121	13%	34	112
Report Card Rating of D	5%	22	233	2%	16	121	8%	27	112
Report Card Rating of C or 2	34%	48	233	37%	49	121	31%	47	112
Report Card Rating of B	14%	34	233	17%	37	121	11%	31	112
Report Card Rating of A or 1	34%	47	233	31%	47	121	37%	48	112
Percent English language learners	17%	16	233	17%	17	121	18%	16	112
Percent special education	12%	4	233	13%	4	121	12%	5	112
Blocking Variables									
District A block 1	15%	36	233	12%	33	121	19%	39	112
District A block 2	12%	32	233	12%	32	121	12%	32	112
District A block 3	14%	35	233	16%	37	121	13%	33	112
District A block 4	5%	21	233	6%	23	121	4%	19	112
District A block 5	3%	18	233	4%	20	121	3%	16	112
District B block 6	12%	32	233	12%	32	121	12%	32	112
District B block 7	6%	23	233	7%	25	121	4%	21	112
District B block 8	7%	26	233	8%	28	121	6%	24	112
District B block 9	3%	17	233	3%	18	121	3%	16	112

Data and Measures

Teacher practice was measured through structured classroom observations using the Framework for Teaching (Danielson Group, 2013).⁵ Trained observers scored the observed teachers on each of the 12 elements under four components of Classroom Environment and the 15 elements under four components of Instruction on the Framework for Teaching. Each element has a scale from 1 to 4 where 1 is “unsatisfactory,” 2 is “basic,” 3 is “proficient,” and 4 is “distinguished.” Observers were blind to teachers’ treatment or control condition when they conducted the observations.

For descriptive purposes, we combined the element-level scores into component-level scores using a simple average approach. The average scores at baseline on each component of the Framework for Teaching for this group of beginning teachers ranged from 1.9 (just below “basic”) to 2.6 (halfway between “basic” and “proficient”) (Exhibit D-3).

Exhibit D-3. Average Observation Scores at Baseline

	Mean	SD	<i>n obs</i>
Environment of Respect and Rapport	2.6	0.62	233
Establishing a Culture for Learning	2.4	0.54	233
Managing Classroom Procedures	2.3	0.59	232
Managing Student Behavior	2.4	0.59	233
Communicating with Students	2.5	0.42	233
Questioning and Discussion Techniques	1.9	0.59	233
Engaging Students in Learning	2.1	0.53	233
Using Assessment in Instruction	2.0	0.48	233

Instead of using these simple averages in analysis, we created component-level variables by combining elements via factor analysis. Factors combined element-level scores on each component of the Framework for Teaching into one component-level variable. The factor variable reflects the structure of the correlations between the elements. It is similar to a weighted average of the elements, where the weights include the strength of the relationship between the elements as well as teachers’ scores on those elements.

Each factor variable is continuous, has a mean of 0 and a standard deviation of 1, and the majority of teachers score in the range from -2 to 2. A score of zero on each component therefore is equivalent to being at the average score for all teachers observed in fall 2013 (for Cohort 1) or fall 2014 (for Cohort 2). A change of 1.0 in these variables is a change of 1 standard deviation, which is roughly equivalent to 0.5 or 0.6 point on the original 1 to 4 scale of the elements, where 1 is “unsatisfactory,” 2 is “basic,” 3 is “proficient,” and 4 is “distinguished.” A difference of 1.0 on the factor scores, therefore, is equivalent to about half the distance between “basic” and “proficient” on the original scale. Exhibit D-4 shows the results of the factor analysis and the variables that were created. The eigenvalues for all factors were over 1.5 and the reliabilities (Cronbach’s alpha) for each of these scales were acceptable, though not high.

⁵ Danielson, C. (2013). The framework for teaching evaluation instrument: 2013 edition. Princeton, NJ: The Danielson Group.

Exhibit D-4. Properties of Factor Observation Variables

	Elements	Eigenvalue	Alpha
Creating an environment of respect and rapport	1. Teacher interactions with students 2. Student interactions with one another	1.5	0.64
Establishing a culture for learning	1. Importance of content 2. Expectations for learning and achievement	1.5	0.66
Managing classroom procedures	1. Management of instructional groups 2. Management of transitions 3. Management of materials and supplies 4. Performance of classroom routines	2.5	0.81
Managing student behavior	1. Expectations 2. Monitoring of student behavior 3. Response to student misbehavior	2.3	0.84
Communicating with students	1. Expectations for learning 2. Directions for activities 3. Explanations of content 4. Use of oral and written language	1.9	0.60
Questioning and discussion techniques	1. Quality of questions/prompts 2. Discussion techniques 3. Student participation	2	0.73
Engaging students in learning	1. Activities and assignments 2. Grouping of students 3. Instructional materials and resources 4. Structure and pacing	2.6	0.82
Using assessment in instruction	1. Assessment criteria 2. Monitoring of student learning 3. Feedback to students 4. Student self-assessment and monitoring of progress	2.1	0.68

Attrition

Under the National Evaluation of i3 (NEi3), an RCT is considered eligible for the Meets i3 Criteria without Reservations designation if each analysis meets the WWC “liberal attrition standards.” These standards track attrition first at the cluster (school) level, followed by the individual (teacher) level.

Schools attrited from the sample when all teachers who were selected for observation within the school attrited, i.e., were not observed at both time periods. Exhibit D-5 displays the number of treatment and control schools with teachers selected for observation in each cohort,⁶ the number of schools with teachers observed at both time periods, and the school-level attrition by condition in each district and overall. WWC standards for attrition take into account both overall attrition and the difference in attrition between treatment and control groups. In Cohort 1, overall school-level attrition was 36 percent, with differential attrition of 0 percentage points. In Cohort 2, overall attrition was 21 percent, with differential attrition of 6 percentage points. When combined, the attrition for both cohorts was 23 percent, with differential attrition of 2 percentage points. These were all within the range of acceptable attrition. For district-specific analyses, the school-level attrition and differential attrition met WWC attrition standards in Cohort 1 and with the two cohorts combined, but exceeded the acceptable thresholds in Cohort 2 (Exhibit D-5).

Individual (teacher) level attrition was calculated only on teachers that remained in non-attrited schools. Teachers attrited from this analysis if they left the profession or the district, or refused to be observed, or we were unable to schedule them for an observation either at baseline or at follow-up shows the teacher-level attrition, which was within acceptable levels for all analyses (Exhibit D-6).

⁶ This number includes all schools with teachers selected, including those who declined to participate, and the teachers who replaced them. In some cases, the teachers selected as replacements also declined to participate. Therefore, the total number of schools selected may have been larger in one district or condition than in the others, with the aim of obtaining a final sample that was balanced across treatment and control in each district.

Exhibit D-5. Cluster-Level Attrition for Observation Analysis

		Cohort 1				Cohort 2				Combined			
		Treat- ment	Control	Differential	Overall	Treat- ment	Control	Differential	Overall	Treat- ment	Control	Differential	Overall
District A	Selected for observation	21	20			24	17			32	30		
	Stayed and were observed at Time 2	12	12			18	15			24	23		
	Percent attrited	43%	40%	3%	41%	25%	12%	13%	20%	25%	23%	2%	24%
	<i>Attrition standard</i>	<i>Met standard</i>				<i>Did not meet standard</i>				<i>Met standard</i>			
District B	Selected for observation ⁷	24	22			21	25			41	38		
	Stayed and were observed at Time 2	17	15			19	17			31	30		
	Percent attrited	29%	32%	3%	30%	10%	32%	22%	22%	24%	21%	3%	23%
	<i>Attrition standard</i>	<i>Met standard</i>				<i>Did not meet standard</i>				<i>Met standard</i>			
Overall	Selected for observation	45	42			45	42			71	70		
	Stayed and were observed at Time 2	29	27			37	32			55	53		
	Percent attrited	36%	36%	0%	36%	18%	24%	6%	21%	23%	24%	2%	23%
	<i>Attrition standard</i>	<i>Met Standard</i>				<i>Met Standard</i>				<i>Met Standard</i>			

Note: Cohorts do not sum to the overall total because some schools include both Cohort 1 and Cohort 2 teachers.

⁷ This number includes schools that attrited both before the baseline observation and between observations.

Exhibit D-6. Teacher-Level Attrition for Observation Analysis

		Cohort 1				Cohort 2				Combined			
		Treat- ment	Control	Differential	Overall	Treat- ment	Control	Differential	Overall	Treat- ment	Control	Differential	Overall
District A	Selected for observation	21	30			28	25			49	55		
	Stayed and were observed at Time 2	14	19			23	21			37	40		
	Percent attrited	33%	37%	3%	35%	18%	16%	2%	17%	24%	27%	3%	26%
	<i>Attrition standard</i>	<i>Met standard</i>				<i>Met standard</i>				<i>Met Standard</i>			
District B	Selected for observation	21	26			29	22			50	48		
	Stayed and were observed at Time 2	18	20			25	19			43	39		
	Percent attrited	14%	23%	9%	19%	14%	14%	0%	14%	14%	19%	5%	16%
	<i>Attrition standard</i>	<i>Met standard</i>				<i>Met standard</i>				<i>Met Standard</i>			
Overall	Selected for observation	42	56			57	47			99	103		
	Stayed and were observed at Time 2	32	39			48	40			80	79		
	Percent attrited	24%	30%	7%	28%	16%	15%	1%	15%	19%	23%	4%	21%
	<i>Attrition standard</i>	<i>Met standard</i>				<i>Met standard</i>				<i>Met Standard</i>			

Attrition had two main consequences. First, the number of schools remaining in the analysis sample was low, as was the number of teachers in each of these schools, even when both cohorts and districts were combined. Lower sample size limited our ability to detect the effects of the NTC model on teacher practice, particularly if those effects were small or there was great variability in teacher practice among teachers.

Second, the schools and teachers who remained in the sample may have differed in both measurable and unmeasurable ways from those who attrited. This implication was particularly problematic with differential attrition, as treatment teachers and schools who left may have been substantially different from control teachers and schools who left. The second issue was partly addressed through measuring the baseline equivalence of the teachers who remained in the sample, as discussed below.

Baseline equivalence

Because differential attrition between the treatment and control groups was evident at the district-within-cohort level, we examined baseline equivalence in the observation scores of the teachers included in the analysis. Baseline equivalence was measured using the same model as was used to measure outcomes (discussed below), applied only to the baseline measures. Where attrition exceeds acceptable standards, an outcome analysis can still Meet Criteria with Reservations if the baseline difference between treatment and control is less than 0.05 standard deviation, or if the difference is between 0.05 and 0.25 standard deviation and a baseline measure is included in the model.

The samples that combined the two RCT sites met acceptable thresholds for attrition; nonetheless, we tested baseline equivalence within these samples. With the two RCT sites and both cohorts combined, the difference in baseline observation scores ranged from 0.07 standard deviation on creating an environment of respect and rapport to 0.26 standard deviation on communicating with students and using assessment in instruction (Exhibit D-7). Where the absolute value of each of these differences was less than 0.25 standard deviation, baseline observation scores were included in the model, and thus the analysis achieved baseline equivalence. For the two components that exceeded acceptable standards for baseline difference (managing student behavior and using assessment in instruction), we applied a propensity score weight to the models, predicting treatment from each of these components at baseline, as a correction for this lack of equivalence. These components are marked with a cross (†).

In the cohort-specific analyses, the baseline difference between treatment and control in Cohort 1 was greater than 0.25 standard deviation in creating an environment of respect and rapport. Therefore, we applied the propensity score weight approach to this model. In Cohort 2 the baseline difference was greater than 0.25 standard deviation in six of the eight components. We also applied propensity weights to these models.

**Exhibit D-7. Baseline Observation Difference Between Treatment and Control,
by Cohort, RCT Sites Combined**

		Treatment Mean (SD)	Comparison Mean (SD)	Unstandardized Difference (Pooled SD)	Standardized difference
Cohorts combined	Creating an environment of respect and rapport	0.123 (0.913)	0.059 (0.977)	0.063 (0.944)	0.067
	Establishing a culture for learning	0.128 (1.013)	0.043 (1.050)	0.085 (1.029)	0.083
	Managing classroom procedures	0.173 (0.952)	-0.060 (0.934)	0.234 (0.946)	0.247
	Managing student behavior	0.182 (1.020)	-0.077 (0.964)	0.259 (0.997)	0.260 ⁺
	Communicating with students	0.042 (1.039)	0.131 (0.916)	-0.089 (0.976)	-0.091
	Using questioning and discussion techniques	0.220 (1.037)	0.050 (0.989)	0.170 (1.014)	0.168
	Engaging students in learning	0.192 (0.986)	-0.013 (1.002)	0.205 (0.996)	0.206
	Using assessment in instruction	0.226 (1.001)	-0.041 (1.010)	0.267 (1.011)	0.264 ⁺
	Creating an environment of respect and rapport	-0.069 (0.794)	0.209 (0.964)	-0.279 (0.879)	-0.317 ⁺
Cohort 1	Establishing a culture for learning	-0.056 (1.182)	0.189 (1.073)	-0.245 (1.133)	-0.216
	Managing classroom procedures	0.141 (0.878)	0.112 (0.977)	0.029 (0.920)	0.032
	Managing student behavior	0.036 (0.955)	0.074 (0.951)	-0.038 (0.947)	-0.040
	Communicating with students	0.083 (1.053)	0.119 (0.985)	-0.036 (1.016)	-0.035
	Using questioning and discussion techniques	0.273 (1.019)	0.164 (0.996)	0.109 (1.003)	0.108
	Engaging students in learning	0.045 (0.984)	0.089 (1.072)	-0.044 (1.018)	-0.043
	Using assessment in instruction	0.240 (0.982)	0.024 (0.946)	0.216 (0.965)	0.224

**Exhibit D-7. Baseline Observation Difference Between Treatment and Control,
by Cohort, RCT Sites Combined (concluded)**

		Treatment Mean (SD)	Comparison Mean (SD)	Unstandardized Difference (Pooled SD)	Standardized difference
Cohort 2	Creating an environment of respect and rapport	0.309 (0.991)	-0.041 (0.984)	0.350 (0.997)	0.351 [†]
	Establishing a culture for learning	0.307 (0.790)	-0.055 (1.034)	0.362 (0.944)	0.383 [†]
	Managing classroom procedures	0.200 (1.021)	-0.174 (0.898)	0.374 (0.968)	0.387 [†]
	Managing student behavior	0.324 (1.072)	-0.178 (0.969)	0.502 (1.042)	0.482 [†]
	Communicating with students	0.002 (1.036)	0.139 (0.877)	-0.137 (0.948)	-0.144
	Using questioning and discussion techniques	0.170 (1.065)	-0.024 (0.987)	0.194 (1.022)	0.189
	Engaging students in learning	0.332 (0.979)	-0.082 (0.957)	0.415 (0.984)	0.422 [†]
	Using assessment in instruction	0.212 (1.031)	-0.083 (1.057)	0.296 (1.050)	0.282 [†]

[†] Difference between treatment and control was greater than 0.25 standard deviation.

Cells where differential attrition exceeded WWC standards and baseline difference was greater than 0.25 standard deviation are shaded grey.

In the district-specific analyses, attrition exceeded acceptable thresholds for each district in Cohort 2 (Exhibit D-8 & D-9). Therefore, the analyses must meet baseline equivalence standards to meet WWC standards. The baseline difference between treatment and control was greater than 0.25 standard deviation in a large number of the components at the district level (marked with a [†]), particularly when further broken down by cohort. This, combined with the attrition in these samples, resulted in a number of estimates that did not meet WWC standards. The impact estimates in these cells are not reported (Exhibit D-11).

Exhibit D-8. Baseline Difference Between Treatment and Control Analysis Samples in Teacher Practice Scores for District A

		Treatment Mean (SD)	Comparison Mean (SD)	Unstandardized Difference (Pooled SD)	Standardized difference
Cohorts combined	Creating an environment of respect and rapport	0.061 (0.829)	-0.124 (0.977)	0.185 (0.902)	0.205
	Establishing a culture for learning	-0.095 (1.061)	-0.171 (0.945)	0.077 (1.001)	0.077
	Managing classroom procedures	-0.046 (1.002)	-0.226 (0.824)	0.180 (0.909)	0.198
	Managing student behavior	-0.101 (1.021)	-0.411 (0.877)	0.310 (0.961)	0.323 [†]
	Communicating with students	-0.078 (1.093)	-0.034 (1.000)	-0.044 (1.041)	-0.042
	Using questioning and discussion techniques	-0.189 (0.881)	-0.203 (0.758)	0.014 (0.818)	0.017
	Engaging students in learning	-0.146 (0.951)	-0.236 (0.949)	0.090 (0.945)	0.095
	Using assessment in instruction	-0.035 (1.003)	-0.261 (0.862)	0.226 (0.939)	0.241
	Creating an environment of respect and rapport	0.013 (0.714)	0.062 (0.977)	-0.049 (0.822)	-0.059
	Establishing a culture for learning	-0.360 (1.242)	-0.012 (0.844)	-0.348 (1.090)	-0.320 [†]
Cohort 1	Managing classroom procedures	-0.097 (0.909)	-0.165 (0.662)	0.068 (0.778)	0.087
	Managing student behavior	-0.234 (0.930)	-0.370 (0.922)	0.136 (0.915)	0.149
	Communicating with students	0.026 (1.106)	0.014 (1.241)	0.012 (1.147)	0.010
	Using questioning and discussion techniques	-0.025 (0.996)	-0.088 (0.693)	0.063 (0.864)	0.073
	Engaging students in learning	-0.221 (0.931)	-0.159 (0.935)	-0.061 (0.918)	-0.067
	Using assessment in instruction	0.072 (0.894)	0.182 (0.909)	-0.109 (0.888)	-0.123
	Creating an environment of respect and rapport	0.104 (0.937)	-0.237 (0.980)	0.342 (0.964)	0.354 [†]
Cohort 2	Establishing a culture for learning	0.146 (0.825)	-0.268 (1.007)	0.414 (0.938)	0.442 [†]
	Managing classroom procedures	-0.010 (1.088)	-0.263 (0.922)	0.253 (0.996)	0.254 [†]
	Managing student behavior	0.020 (1.105)	-0.436 (0.869)	0.456 (1.003)	0.454 [†]
	Communicating with students	-0.172 (1.101)	-0.064 (0.852)	-0.108 (0.965)	-0.112
	Using questioning and discussion techniques	-0.329 (0.765)	-0.273 (0.801)	-0.056 (0.775)	-0.073
Cohort 2	Engaging students in learning	-0.079 (0.986)	-0.285 (0.976)	0.206 (0.975)	0.211
	Using assessment in instruction	-0.131 (1.105)	-0.531 (0.726)	0.399 (0.937)	0.426 [†]

[†] Standardized difference is greater than 0.25 standard deviations. Propensity score weight will be applied.

Exhibit D-9. Baseline Difference Between Treatment and Control Analysis Samples in Teacher Practice Scores for District B

		Treatment Mean (SD)	Comparison Mean (SD)	Unstandardized Difference (Pooled SD)	Standardized difference
Cohorts combined	Creating an environment of respect and rapport	0.186 (0.999)	0.217 (0.961)	-0.031 (0.974)	-0.032
	Establishing a culture for learning	0.356 (0.919)	0.227 (1.111)	0.129 (1.020)	0.126
	Managing classroom procedures	0.345 (0.887)	0.069 (1.003)	0.276 (0.954)	0.290 [†]
	Managing student behavior	0.472 (0.947)	0.209 (0.952)	0.262 (0.953)	0.275 [†]
	Communicating with students	0.159 (0.983)	0.273 (0.821)	-0.114 (0.898)	-0.127
	Using questioning and discussion techniques	0.629 (1.030)	0.273 (1.117)	0.356 (1.084)	0.329 [†]
	Engaging students in learning	0.549 (0.903)	0.174 (1.018)	0.375 (0.978)	0.383 [†]
	Using assessment in instruction	0.493 (0.938)	0.153 (1.098)	0.341 (1.032)	0.330 [†]
Cohort 1	Creating an environment of respect and rapport	-0.147 (0.874)	0.324 (0.965)	-0.472 (0.936)	-0.504 [†]
	Establishing a culture for learning	0.233 (1.074)	0.345 (1.223)	-0.112 (1.133)	-0.099
	Managing classroom procedures	0.299 (0.844)	0.307 (1.127)	-0.007 (0.977)	-0.008
	Managing student behavior	0.292 (0.929)	0.419 (0.844)	-0.127 (0.880)	-0.144
	Communicating with students	0.133 (1.030)	0.200 (0.760)	-0.067 (0.901)	-0.074
	Using questioning and discussion techniques	0.542 (0.987)	0.373 (1.170)	0.169 (1.063)	0.159
	Engaging students in learning	0.311 (0.988)	0.282 (1.157)	0.029 (1.058)	0.027
	Using assessment in instruction	0.399 (1.057)	-0.106 (0.983)	0.505 (1.041)	0.485 [†]
Cohort 2	Creating an environment of respect and rapport	0.536 (1.025)	0.140 (0.971)	0.396 (1.002)	0.395 [†]
	Establishing a culture for learning	0.485 (0.730)	0.141 (1.040)	0.343 (0.925)	0.371 [†]
	Managing classroom procedures	0.388 (0.947)	-0.100 (0.890)	0.488 (0.938)	0.520 [†]
	Managing student behavior	0.660 (0.953)	0.059 (1.012)	0.601 (1.021)	0.589 [†]
	Communicating with students	0.185 (0.958)	0.325 (0.874)	-0.140 (0.903)	-0.155
	Using questioning and discussion techniques	0.721 (1.093)	0.205 (1.098)	0.516 (1.114)	0.463 [†]
	Engaging students in learning	0.787 (0.763)	0.096 (0.922)	0.691 (0.915)	0.755 [†]
	Using assessment in instruction	0.592 (0.811)	0.329 (1.155)	0.264 (1.019)	0.259 [†]

† Standardized difference is greater than 0.25 standard deviations. Propensity score weight will be applied.

Exhibit D-10. Treatment and Control Analysis Samples in Teacher Practice for RCT Sites Combined

		Schools			Teachers		
		Treatment	Control	Total	Treatment	Control	Total
Cohorts Combined	Creating an environment of respect and rapport	53	55	108	79	80	159
	Establishing a culture for learning	53	55	108	79	80	159
	Managing classroom procedures	50	52	102	66	73	139
	Managing student behavior	53	55	108	79	80	159
	Communicating with students	53	55	108	77	80	157
	Using questioning and discussion techniques	53	54	107	78	79	157
	Engaging students in learning	53	55	108	78	79	157
	Using assessment in instruction	53	54	107	79	79	158
Cohort 1	Creating an environment of respect and rapport	27	29	56	39	32	71
	Establishing a culture for learning	27	29	56	39	32	71
	Managing classroom procedures	24	27	51	30	29	59
	Managing student behavior	27	29	56	39	32	71
	Communicating with students	27	29	56	38	32	70
	Using questioning and discussion techniques	27	28	55	38	31	69
	Engaging students in learning	27	29	56	38	32	70
	Using assessment in instruction	27	28	55	39	31	70
Cohort 2	Creating an environment of respect and rapport	32	37	69	40	48	88
	Establishing a culture for learning	32	37	69	40	48	88
	Managing classroom procedures	32	36	68	36	44	80
	Managing student behavior	32	37	69	40	48	88
	Communicating with students	32	37	69	39	48	87
	Using questioning and discussion techniques	32	37	69	40	48	88
	Engaging students in learning	32	36	68	40	47	87
	Using assessment in instruction	32	37	69	40	48	88

Statistical Analysis

To analyze the impact of the NTC model on teacher practice, we estimated the following two-level model, with teachers nested within schools:

Level 1 (Teachers): $y_{jk} = \beta_{0k} + \beta_1 Treatment_{jk} + \beta_{2k} Z_{jk} + \beta_{3k} Coh_{jk} + \beta_{4k} Z_{jk} Coh_{jk} + r_{jk}$

Level 2 (Schools): $\beta_{0k} = \gamma_{00} + \gamma_{01} W_k + \gamma_{02} Dist_k + \gamma_{03} W_k Dist_k + e_k$

$$\beta_{1k} = \gamma_{10}$$

$$\beta_{2k} = \gamma_{20} + \gamma_{21} Dist_k$$

$$\beta_{3k} = \gamma_{30} + \gamma_{31} W_k + \gamma_{32} Dist_k + \gamma_{33} W_k Dist_k$$

$$\beta_{4k} = \gamma_{40} + \gamma_{41} Dist_k$$

Mixed:

$$y_{jk} = \gamma_{00} + \gamma_{10} Treatment_{jk} + \gamma_{01} W_k + \gamma_{20} Z_{jk} + \gamma_{02} Dist_k + \gamma_{30} Coh_{jk} + \gamma_{03} W_k Dist_k + \gamma_{21} Z_{jk} Dist_k + \gamma_{31} W_k Coh_{jk} + \gamma_{40} Z_{jk} Coh_{jk} + \gamma_{32} Dist_k Coh_{jk} + \gamma_{33} W_k Coh_{jk} Dist_k + \gamma_{41} Z_{jk} Coh_{jk} Dist_k + r_{jk} + e_k$$

In this model, y_{jk} represents the teacher's score on one of eight components of the Framework for Teaching, discussed above, standardized using the full baseline sample mean and standard deviation. Scores were standardized to create a truly linear scale and to estimate an effect size in standard deviation units. The coefficient γ_{10} estimates the impact of the NTC model on teacher practice for each component. Z_{jk} represents a vector of teacher-level controls (including baseline observation score on the same component), W_k represents a vector of school-level controls (including the blocking variables used in randomization), Coh_{jk} is a cohort indicator, and $Dist_k$ is a district indicator. All variables were centered to the analysis sample. The model also included interactions with district, cohort, and a district-by-cohort three-way interaction for each teacher and school control variable. We included school- and teacher-level covariates, cohort and district effects, and interactions between the covariates and these effects for the same reasons we included them in the retention analysis.

This analysis was an intent-to-treat analysis, in which all teachers who were randomly selected to be observed were included in their original assigned condition, regardless of cross-over, noncompliance, or level of treatment received. Teachers who changed schools between the baseline and follow-up observations were followed into their new schools and observed where possible. We used their original school location for the school control variables in the models.

Results

There were no statistically significant differences in teacher practice scores between treatment and control teachers overall. However, there were a few differences by district. In District A, treatment teachers had significantly higher scores in establishing a culture for learning and managing student behavior in the first cohort, but in the second cohort the analysis did not meet WWC standards for attrition and baseline equivalence, and when the cohorts were combined, these differences were not apparent. In District B, treatment teachers had significantly *lower* scores in establishing a culture for learning in Cohort 1, but the Cohort 2 analysis did not meet standards for attrition and baseline equivalence, and the combined analysis did not show a statistically significant difference. Treatment teachers in District B also scored significantly higher on managing classroom procedures when the cohorts were combined. These inconsistent and largely non-significant results can be attributed in part to the small sample size, particularly when looking by cohort and district. Full model tables are included below.

Exhibit D-11. Impact of the NTC Model on Teacher Practice by District and Cohort

	District A			District B			Districts Combined		
	Cohort 1	Cohort 2	Combined	Cohort 1	Cohort 2	Combined	Cohort 1	Cohort 2	Combined
Creating an environment of respect and rapport	0.29	♦	-0.05	0.1†	♦	0.15	-0.05†	-0.1†	0.04
Establishing a culture for learning	0.73**†	♦	-0.11	-0.68*	♦	-0.34	-0.11	-0.14†	-0.24
Managing classroom procedures	-0.04	♦	-0.36	0.43	♦	0.56*†	-0.36	0.07†	0.13
Managing student behavior	1.08***	♦	0.26†	-0.32	♦	0.28†	0.26	0.18†	0.28†
Communicating with students	0.40	0.11	0.19	-0.48	-0.10	-0.29	0.19	0.08	0.01
Using questioning and discussion techniques	0.13	0.01	0.12	-0.19	♦	0.14†	0.12	0.28	0.21
Engaging students in learning	0.15	0.34	0.32	-0.35	♦	-0.04†	0.32	0.43†	0.15
Using assessment in instruction	0.01	♦	0.21	-0.3†	♦	-0.21†	0.21	0.18†	0.06†

*** $p < 0.001$

** $p < 0.01$

* $p < 0.05$

†Propensity score weight applied to account for inequivalent baseline

♦ Outcome is not reported because the measure did not meet WWC standards for attrition and baseline significance

Exhibit D-12. Unadjusted Treatment and Comparison Outcome Means in Analysis Sample for Teacher Practice by District and Cohort

	Overall		District A		District B	
	Comparison Mean (SD)	Treatment Mean (SD)	Comparison Mean (SD)	Treatment Mean (SD)	Comparison Mean (SD)	Treatment Mean (SD)
Cohorts combined	0.31 (0.86)	0.15 (1.13)	0.11 (0.85)	-0.09 (1.07)	0.52 (0.82)	0.35 (1.14)
	0.27 (0.99)	0.29 (1.20)	-0.08 (1.09)	0.24 (1.29)	0.63 (0.74)	0.34 (1.12)
	0.70 (0.78)	0.62 (0.98)	0.73 (0.93)	0.52 (0.99)	0.69 (0.66)	0.69 (0.98)
	0.15 (0.94)	0.19 (1.02)	-0.07 (0.98)	0.10 (0.96)	0.37 (0.86)	0.27 (1.07)
	0.19 (1.10)	0.23 (1.13)	-0.31 (1.01)	0.09 (1.09)	0.68 (0.96)	0.35 (1.17)
	0.37 (1.17)	0.56 (1.07)	-0.18 (1.03)	0.29 (1.12)	0.91 (1.04)	0.80 (0.97)
	0.43 (1.11)	0.53 (0.98)	0.02 (1.18)	0.38 (1.03)	0.86 (0.86)	0.66 (0.92)
	0.34 (1.07)	0.17 (1.17)	-0.17 (0.88)	-0.06 (1.11)	0.86 (0.99)	0.38 (1.19)

Exhibit D-12. Unadjusted Treatment and Comparison Outcome Means in Analysis Sample for Teacher Practice by District and Cohort (concluded)

		Overall		District A		District B	
		Comparison	Treatment	Comparison	Treatment	Comparison	Treatment
		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Cohort 1	Creating an environment of respect and rapport	0.39 (0.92)	0.37 (1.22)	0.12 (0.83)	-0.07 (1.18)	0.65 (0.94)	0.70 (1.17)
	Establishing a culture for learning	-0.08 (0.94)	-0.02 (1.09)	-0.61 (0.85)	0.12 (1.07)	0.42 (0.74)	-0.13 (1.13)
	Managing classroom procedures	0.56 (0.69)	0.57 (1.11)	0.39 (0.71)	0.26 (1.19)	0.67 (0.66)	0.80 (1.03)
	Managing student behavior	0.06 (0.98)	0.15 (1.00)	-0.40 (0.81)	0.09 (0.85)	0.49 (0.94)	0.20 (1.12)
	Communicating with students	-0.10 (1.05)	-0.05 (0.89)	-0.63 (0.95)	-0.09 (0.93)	0.39 (0.92)	-0.01 (0.89)
	Using questioning and discussion techniques	0.13 (1.13)	0.31 (0.97)	-0.42 (0.81)	0.10 (1.07)	0.64 (1.15)	0.49 (0.86)
	Engaging students in learning	0.16 (1.10)	0.21 (0.96)	-0.25 (1.06)	0.08 (1.06)	0.57 (0.99)	0.32 (0.90)
	Using assessment in instruction	0.18 (0.97)	0.05 (1.04)	-0.28 (0.75)	-0.01 (0.96)	0.61 (0.97)	0.10 (1.12)
Cohort 2	Creating an environment of respect and rapport	0.24 (0.80)	0.00 (1.05)	♦	♦	♦	♦
	Establishing a culture for learning	0.61 (0.94)	0.50 (1.23)	♦	♦	♦	♦
	Managing classroom procedures	0.82 (0.85)	0.65 (0.90)	♦	♦	♦	♦
	Managing student behavior	0.24 (0.90)	0.21 (1.05)	♦	♦	♦	♦
	Communicating with students	0.46 (1.08)	0.41 (1.24)	-0.03 (1.00)	0.20 (1.18)	0.98 (0.91)	0.61 (1.29)
	Using questioning and discussion techniques	0.59 (1.17)	0.72 (1.11)	0.03 (1.16)	0.41 (1.16)	♦	♦
	Engaging students in learning	0.68 (1.08)	0.75 (0.93)	0.26 (1.26)	0.57 (0.98)	♦	♦
	Using assessment in instruction	0.51 (1.14)	0.25 (1.25)	♦	♦	♦	♦

♦ Outcome is not reported because the measure did not meet WWC standards for attrition and baseline significance

Full Model Tables⁸

Both Cohorts and RCT Districts Combined

Exhibit D-13. Impact of the NTC Model on Creating an Environment of Respect and Rapport – Both Cohorts and RCT Districts Combined

	Estimate	Standard Error	p-value
Treatment status	0.04	0.16	0.81
Baseline score	0.10	0.09	0.26
Teacher controls			
Teacher is female CENTERED	<0.001	0.22	0.98
Teacher race is black CENTERED	-0.02	0.20	0.90
Teacher has only partial certification CENTERED	<0.001	0.28	0.99
School controls			
School percent FRPL CENTERED	<0.001	0.01	0.70
School report card rating - on one scale CENTERED	<0.001	0.10	0.98
School percent non-white CENTERED	<0.001	0.02	0.90
School percent ELL CENTERED	0.01	0.01	0.26
School percent Special Education CENTERED	-0.01	0.02	0.55
Blocking variables			
District A block 3	0.52	0.35	0.14
District A block 2	-0.11	0.35	0.75
District B block 5	-0.60	0.34	0.08
District B block 7	-1.17	0.47	0.01
District B block 8	-0.29	0.33	0.39
District B block 6	-1.21	0.36	<0.001
Interactions with district			
Centered district indicator	-1.05	0.34	<0.001
District interaction: Baseline score	0.38	0.17	0.03
District interaction: Teacher is female	0.13	0.43	0.76
District interaction: Teacher race is black	0.05	0.41	0.91
District interaction: Teacher has only partial certification	-0.58	0.54	0.29
District interaction: School percent FRPL	-0.03	0.02	0.15
District interaction: School report card rating - on one scale	-0.29	0.21	0.16
District interaction: School percent non-white	0.04	0.04	0.30
District interaction: School percent ELL	0.01	0.02	0.50
District interaction: School percent Special Education	0.05	0.05	0.33
Interactions with cohort			
Centered cohort indicator	-0.69	0.33	0.04
Cohort interaction: Baseline score	-0.24	0.17	0.17
Cohort interaction: Teacher is female	-0.47	0.43	0.28
Cohort interaction: Teacher race is black	-0.44	0.41	0.28
Cohort interaction: Teacher has only partial certification	0.12	0.54	0.83
Cohort interaction: School percent FRPL	-0.03	0.02	0.24
Cohort interaction: School report card rating - on one scale	-0.19	0.21	0.36

⁸ Blocking variable names are being kept consistent across model results, which resulted in some of the blocking variables appearing out of order in Appendix D exhibits.

**Exhibit D-13. Impact of the NTC Model on Creating an Environment of Respect and Rapport –
Both Cohorts and RCT Districts Combined (concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: School percent non-white	0.02	0.04	0.54
Cohort interaction: School percent ELL	<0.001	0.02	0.83
Cohort interaction: School percent Special Education	-0.12	0.05	0.02
Cohort interaction: District A block 3	-1.19	0.70	0.09
Cohort interaction: District A block 2	0.12	0.70	0.87
Cohort interaction: District B block 5	1.20	0.68	0.08
Cohort interaction: District B block 7	1.29	0.96	0.18
Cohort interaction: District B block 8	0.87	0.65	0.18
Cohort interaction: District B block 6	0.45	0.70	0.52
District-by cohort interactions			
District by cohort interaction	1.60	0.68	0.02
District by cohort interaction: Baseline score	-0.37	0.35	0.29
District by cohort interaction: Teacher is female	-0.61	0.86	0.48
District by cohort interaction: Teacher race is black	0.40	0.83	0.63
District by cohort interaction: Teacher has only partial certification	-0.45	1.09	0.68
District by cohort interaction: School percent FRPL	0.01	0.05	0.86
District by cohort interaction: School report card rating – on one scale	-0.29	0.41	0.49
District by cohort interaction: School percent non-white	-0.07	0.08	0.40
District by cohort interaction: School percent ELL	0.03	0.03	0.43
District by cohort interaction: School percent Special Education	-0.08	0.10	0.39
Constant	0.61	0.17	<0.001
Random effects			
School	<0.001		
Teacher	0.57		
n			
Schools	108		
Teachers	159		

**Exhibit D-14. Impact of the NTC Model on Establishing a Culture for Learning –
Both Cohorts and RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	-0.24	0.20	0.22
Baseline score	0.06	0.07	0.44
Teacher controls			
Teacher is female CENTERED	-0.19	0.22	0.40
Teacher race is black CENTERED	0.02	0.20	0.91
Teacher has only partial certification CENTERED	-0.39	0.26	0.14
School controls			
School percent FRPL CENTERED	<0.001	0.01	0.81
School report card rating - on one scale CENTERED	0.07	0.12	0.52
School percent non-white CENTERED	-0.01	0.02	0.72
School percent ELL CENTERED	-0.01	0.01	0.27
School percent Special Education CENTERED	-0.03	0.03	0.26
Blocking variables			
District A block 3	-0.83	0.39	0.03
District A block 2	-0.58	0.39	0.13
District B block 5	-0.85	0.42	0.04
District B block 7	-0.55	0.59	0.36
District B block 8	-0.16	0.43	0.71
District B block 6	-0.99	0.44	0.02
Interactions with district			
Centered district indicator	-0.70	0.40	0.08
District interaction: Baseline score	0.36	0.15	0.01
District interaction: Teacher is female	0.19	0.44	0.67
District interaction: Teacher race is black	0.51	0.41	0.21
District interaction: Teacher has only partial certification	0.50	0.52	0.34
District interaction: School percent FRPL	-0.03	0.03	0.22
District interaction: School report card rating - on one scale	<0.001	0.23	0.99
District interaction: School percent non-white	0.06	0.04	0.14
District interaction: School percent ELL	-0.01	0.02	0.63
District interaction: School percent Special Education	-0.05	0.06	0.41
Interactions with cohort			
Centered cohort indicator	-0.12	0.36	0.74
Cohort interaction: Baseline score	-0.09	0.15	0.57
Cohort interaction: Teacher is female	-0.07	0.44	0.87
Cohort interaction: Teacher race is black	-0.30	0.41	0.46
Cohort interaction: Teacher has only partial certification	0.90	0.55	0.10
Cohort interaction: School percent FRPL	-0.03	0.03	0.30
Cohort interaction: School report card rating – on one scale	-0.44	0.22	0.05
Cohort interaction: School percent non-white	0.01	0.04	0.88
Cohort interaction: School percent ELL	0.02	0.02	0.33
Cohort interaction: School percent Special Education	-0.03	0.05	0.58
Cohort interaction: District A block 3	1.87	0.71	0.01

**Exhibit D-14. Impact of the NTC Model on Establishing a Culture for Learning –
Both Cohorts and RCT Districts Combined (concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: District A block 2	1.74	0.63	0.01
Cohort interaction: District B block 5	0.14	0.82	0.86
Cohort interaction: District B block 7	-0.17	1.17	0.89
Cohort interaction: District B block 8	-1.44	0.78	0.07
Cohort interaction: District B block 6	0.12	0.81	0.88
District-by-cohort interactions			
District by cohort interaction	-0.97	0.73	0.18
District by cohort interaction: Baseline score	-0.57	0.30	0.06
District by cohort interaction: Teacher is female	2.25	0.89	0.01
District by cohort interaction: Teacher race is black	1.46	0.82	0.07
District by cohort interaction: Teacher has only partial certification	-1.30	1.10	0.24
District by cohort interaction: School percent FRPL	0.01	0.05	0.91
District by cohort interaction: School report card rating - on one scale	-0.67	0.44	0.12
District by cohort interaction: School percent non-white	-0.11	0.09	0.20
District by cohort interaction: School percent ELL	0.01	0.03	0.77
District by cohort interaction: School percent Special Education	-0.13	0.10	0.21
Constant	0.92	0.21	<0.001
Random effects			
School	0.40		
Teacher	0.33		
n			
School	108		
Teacher	159		

**Exhibit D-15. Impact of the NTC Model on Managing Classroom Procedures –
Both Cohorts and RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.13	0.16	0.45
Baseline score	0.08	0.09	0.43
Teacher controls			
Teacher is female CENTERED	0.01	0.23	0.95
Teacher race is black CENTERED	0.07	0.20	0.71
Teacher has only partial certification CENTERED	-0.07	0.26	0.78
School controls			
School percent FRPL CENTERED	-0.01	0.01	0.27
School report card rating - on one scale CENTERED	0.04	0.10	0.66
School percent non-white CENTERED	0.02	0.02	0.34
School percent ELL CENTERED	0.01	0.01	0.44
School percent Special Education CENTERED	-0.04	0.03	0.10
Blocking variables			
District A block 3	-0.40	0.36	0.27
District A block 2	-0.43	0.36	0.23
District B block 5	-0.32	0.31	0.30
District B block 7	-0.09	0.56	0.87
District B block 8	0.36	0.31	0.23
District B block 6	<0.001	0.33	1.00
Interactions with district			
Centered district indicator	0.14	0.33	0.68
District interaction: Baseline score	<0.001	0.19	0.98
District interaction: Teacher is female	-0.53	0.46	0.25
District interaction: Teacher race is black	-0.02	0.41	0.97
District interaction: Teacher has only partial certification	-0.85	0.51	0.10
District interaction: School percent FRPL	-0.04	0.02	0.07
District interaction: School report card rating – on one scale	<0.001	0.20	0.99
District interaction: School percent non-white	0.07	0.04	0.05
District interaction: School percent ELL	0.02	0.02	0.33
District interaction: School percent Special Education	-0.04	0.05	0.40
Interactions with cohort			
Centered cohort indicator	0.36	0.32	0.26
Cohort interaction: Baseline score	0.23	0.19	0.23
Cohort interaction: Teacher is female	1.02	0.47	0.03
Cohort interaction: Teacher race is black	-0.17	0.39	0.67
Cohort interaction: Teacher has only partial certification	-0.59	0.51	0.25
Cohort interaction: School percent FRPL	-0.03	0.02	0.26
Cohort interaction: School report card rating – on one scale	-0.22	0.20	0.28
Cohort interaction: School percent non-white	0.01	0.04	0.72
Cohort interaction: School percent ELL	0.03	0.02	0.05
Cohort interaction: School percent Special Education	<0.001	0.05	0.94
Cohort interaction: District A block 3	0.18	0.71	0.80

**Exhibit D-15. Impact of the NTC Model on Managing Classroom Procedures –
Both Cohorts and RCT Districts Combined (concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: District A block 2	0.22	0.70	0.76
Cohort interaction: District B block 5	-0.11	0.61	0.85
Cohort interaction: District B block 7	-1.11	1.17	0.34
Cohort interaction: District B block 8	-0.95	0.59	0.11
Cohort interaction: District B block 6	-1.40	0.64	0.03
District-by-cohort interactions			
District by cohort interaction	0.77	0.65	0.24
District by cohort interaction: Baseline score	0.37	0.38	0.34
District by cohort interaction: Teacher is female	0.09	0.92	0.92
District by cohort interaction: Teacher race is black	-0.88	0.81	0.27
District by cohort interaction: Teacher has only partial certification	0.66	1.02	0.51
District by cohort interaction: School percent FRPL	0.03	0.05	0.49
District by cohort interaction: School report card rating – on one scale	-0.79	0.41	0.06
District by cohort interaction: School percent non-white	-0.10	0.08	0.21
District by cohort interaction: School percent ELL	0.03	0.03	0.33
District by cohort interaction: School percent Special Education	0.05	0.11	0.67
Constant	0.86	0.17	<0.001
Random effects			
School	<0.001		
Teacher	0.45		
n			
School	102		
Teacher	139		

**Exhibit D-16. Impact of the NTC Model on Managing Student Behavior –
Both Cohorts and RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.28	0.18	0.12
Baseline score	0.11	0.09	0.21
Teacher controls			
Teacher is female CENTERED	0.28	0.20	0.16
Teacher race is black CENTERED	-0.03	0.23	0.90
Teacher has only partial certification CENTERED	0.04	0.20	0.85
School controls			
School percent FRPL CENTERED	-0.01	0.01	0.71
School report card rating - on one scale CENTERED	0.18	0.10	0.07
School percent non-white CENTERED	0.01	0.02	0.47
School percent ELL CENTERED	<0.001	0.01	0.92
School percent Special Education CENTERED	-0.01	0.02	0.77
Blocking variables			
District A block 3	-0.14	0.31	0.65
District A block 2	0.31	0.31	0.30
District B block 5	-0.92	0.34	0.01
District B block 7	-1.29	0.41	<0.001
District B block 8	-0.10	0.23	0.66
District B block 6	-1.31	0.32	<0.001
Interactions with district			
Centered district indicator	-0.96	0.27	<0.001
District interaction: Baseline score	0.03	0.18	0.87
District interaction: Teacher is female	-0.15	0.39	0.70
District interaction: Teacher race is black	0.79	0.45	0.08
District interaction: Teacher has only partial certification	-0.80	0.39	0.04
District interaction: School percent FRPL	<0.001	0.03	0.99
District interaction: School report card rating – on one scale	0.16	0.20	0.41
District interaction: School percent non-white	0.03	0.04	0.43
District interaction: School percent ELL	0.01	0.02	0.44
District interaction: School percent Special Education	0.02	0.04	0.70
Interactions with cohort			
Centered cohort indicator	0.25	0.26	0.33
Cohort interaction: Baseline score	-0.09	0.18	0.62
Cohort interaction: Teacher is female	-0.17	0.41	0.67
Cohort interaction: Teacher race is black	-0.26	0.43	0.54
Cohort interaction: Teacher has only partial certification	-0.09	0.37	0.81
Cohort interaction: School percent FRPL	-0.03	0.02	0.19
Cohort interaction: School report card rating – on one scale	-0.26	0.19	0.17
Cohort interaction: School percent non-white	0.03	0.04	0.50
Cohort interaction: School percent ELL	-0.02	0.01	0.24
Cohort interaction: School percent Special Education	-0.08	0.04	0.05
Cohort interaction: District A block 3	-1.67	0.61	0.01

**Exhibit D-16. Impact of the NTC Model on Managing Student Behavior –
Both Cohorts and RCT Districts Combined (concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: District A block 2	-1.66	0.57	<0.001
Cohort interaction: District B block 5	1.11	0.68	0.10
Cohort interaction: District B block 7	0.38	0.85	0.65
Cohort interaction: District B block 8	-1.24	0.48	0.01
Cohort interaction: District B block 6	-1.66	0.63	0.01
District-by-cohort interactions			
District by cohort interaction	1.11	0.58	0.06
District by cohort interaction: Baseline score	0.33	0.35	0.35
District by cohort interaction: Teacher is female	0.12	0.81	0.88
District by cohort interaction: Teacher race is black	-0.78	0.92	0.39
District by cohort interaction: Teacher has only partial certification	1.11	0.74	0.13
District by cohort interaction: School percent FRPL	0.06	0.05	0.23
District by cohort interaction: School report card rating – on one scale	-0.55	0.38	0.15
District by cohort interaction: School percent non-white	-0.12	0.08	0.11
District by cohort interaction: School percent ELL	-0.03	0.03	0.33
District by cohort interaction: School percent Special Education	-0.12	0.08	0.13
Constant	0.45	0.15	<0.001
Random effects			
School	<0.001		
Teacher	0.61		
n			
School	108		
Teacher	159		

**Exhibit D-17. Impact of the NTC Model on Communicating with Students –
Both Cohorts and RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.01	0.19	0.96
Baseline score	0.09	0.09	0.29
School controls			
Teacher is female CENTERED	-0.12	0.23	0.62
Teacher race is black CENTERED	-0.41	0.23	0.08
Teacher has only partial certification CENTERED	0.04	0.32	0.91
School controls			
School percent FRPL CENTERED	<0.001	0.01	0.76
School report card rating - on one scale CENTERED	0.08	0.12	0.51
School percent non-white CENTERED	<0.001	0.02	0.95
School percent ELL CENTERED	<0.001	0.01	0.71
School percent Special Education CENTERED	0.01	0.03	0.68
Blocking variables			
District A block 3	-0.16	0.41	0.70
District A block 2	-0.39	0.39	0.32
District B block 5	-0.96	0.40	0.02
District B block 7	-0.55	0.54	0.31
District B block 8	-0.41	0.41	0.32
District B block 6	-1.14	0.42	0.01
Interactions with district			
Centered district indicator	-1.08	0.40	0.01
District interaction: Baseline score	0.26	0.18	0.14
District interaction: Teacher is female	0.04	0.47	0.94
District interaction: Teacher race is black	0.35	0.48	0.46
District interaction: Teacher has only partial certification	-0.02	0.62	0.97
District interaction: School percent FRPL	-0.01	0.03	0.75
District interaction: School report card rating – on one scale	<0.001	0.24	0.99
District interaction: School percent non-white	0.05	0.05	0.29
District interaction: School percent ELL	-0.02	0.02	0.40
District interaction: School percent Special Education	0.06	0.06	0.27
Interactions with cohort			
Centered cohort indicator	0.42	0.39	0.28
Cohort interaction: Baseline score	0.18	0.18	0.31
Cohort interaction: Teacher is female	-0.44	0.47	0.35
Cohort interaction: Teacher race is black	0.04	0.47	0.93
Cohort interaction: Teacher has only partial certification	0.54	0.62	0.39
Cohort interaction: School percent FRPL	0.03	0.03	0.31
Cohort interaction: School report card rating – on one scale	-0.02	0.24	0.94
Cohort interaction: School percent non-white	-0.03	0.05	0.46
Cohort interaction: School percent ELL	0.01	0.02	0.50
Cohort interaction: School percent Special Education	-0.02	0.06	0.77
Cohort interaction: District A block 3	0.90	0.80	0.26

**Exhibit D-17. Impact of the NTC Model on Communicating with Students –
Both Cohorts and RCT Districts Combined (concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: District A block 2	1.10	0.75	0.14
Cohort interaction: District B block 5	-0.68	0.81	0.40
Cohort interaction: District B block 7	-0.08	1.09	0.94
Cohort interaction: District B block 8	-1.84	0.79	0.02
Cohort interaction: District B block 6	-0.83	0.81	0.31
District-by-cohort interactions			
District by cohort interaction	-1.23	0.79	0.12
District by cohort interaction: Baseline score	0.14	0.35	0.70
District by cohort interaction: Teacher is female	2.05	0.94	0.03
District by cohort interaction: Teacher race is black	0.28	0.96	0.77
District by cohort interaction: Teacher has only partial certification	-0.60	1.25	0.63
District by cohort interaction: School percent FRPL	0.02	0.05	0.75
District by cohort interaction: School report card rating – on one scale	-0.45	0.48	0.34
District by cohort interaction: School percent non-white	-0.06	0.09	0.50
District by cohort interaction: School percent ELL	0.02	0.04	0.64
District by cohort interaction: School percent Special Education	-0.02	0.11	0.84
Constant	0.64	0.20	<0.001
Random effects			
School	0.07		
Teacher	0.66		
n			
School	107		
Teacher	157		

**Exhibit D-18. Impact of the NTC Model on Using Questioning and Discussion Techniques –
Both Cohorts and RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.21	0.18	0.25
Baseline score	0.05	0.10	0.62
Teacher variables			
Teacher is female CENTERED	-0.03	0.23	0.91
Teacher race is black CENTERED	0.13	0.22	0.57
Teacher has only partial certification CENTERED	-0.38	0.30	0.20
School variables			
School percent FRPL CENTERED	-0.01	0.01	0.47
School report card rating - on one scale CENTERED	0.13	0.12	0.26
School percent non-white CENTERED	0.01	0.02	0.58
School percent ELL CENTERED	<0.001	0.01	0.72
School percent Special Education CENTERED	0.01	0.03	0.76
Blocking variables			
District A block 3	-0.78	0.45	0.08
District A block 2	0.18	0.39	0.65
District B block 5	-0.77	0.39	0.05
District B block 7	-0.99	0.56	0.08
District B block 8	-0.50	0.39	0.20
District B block 6	-1.34	0.40	<0.001
Interactions by district			
Centered district indicator	-0.94	0.39	0.02
District interaction: Baseline score	-0.18	0.19	0.35
District interaction: Teacher is female	0.19	0.45	0.67
District interaction: Teacher race is black	0.04	0.45	0.93
District interaction: Teacher has only partial certification	-0.13	0.59	0.83
District interaction: School percent FRPL	-0.03	0.03	0.31
District interaction: School report card rating – on one scale	0.16	0.23	0.50
District interaction: School percent non-white	0.06	0.04	0.16
District interaction: School percent ELL	<0.001	0.02	0.84
District interaction: School percent Special Education	0.09	0.05	0.11
Interactions by cohort			
Centered cohort indicator	0.17	0.37	0.65
Cohort interaction: Baseline score	-0.06	0.19	0.76
Cohort interaction: Teacher is female	0.38	0.45	0.41
Cohort interaction: Teacher race is black	-0.23	0.44	0.61
Cohort interaction: Teacher has only partial certification	0.82	0.60	0.17
Cohort interaction: School percent FRPL	0.02	0.03	0.34
Cohort interaction: School report card rating – on one scale	-0.10	0.23	0.68
Cohort interaction: School percent non-white	-0.04	0.04	0.40
Cohort interaction: School percent ELL	-0.01	0.02	0.74
Cohort interaction: School percent Special Education	-0.12	0.05	0.02
Cohort interaction: District A block 3	0.96	0.88	0.27

**Exhibit D-18. Impact of the NTC Model on Using Questioning and Discussion Techniques –
Both Cohorts and RCT Districts Combined (concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: District A block 2	0.28	0.74	0.70
Cohort interaction: District B block 5	0.50	0.77	0.51
Cohort interaction: District B block 7	0.11	1.12	0.92
Cohort interaction: District B block 8	-0.39	0.74	0.60
Cohort interaction: District B block 6	-0.64	0.77	0.41
District-by-cohort interactions			
District by cohort interaction	-0.83	0.77	0.28
District by cohort interaction: Baseline score	0.69	0.39	0.07
District by cohort interaction: Teacher is female	-0.07	0.91	0.93
District by cohort interaction: Teacher race is black	0.71	0.90	0.43
District by cohort interaction: Teacher has only partial certification	-0.74	1.20	0.54
District by cohort interaction: School percent FRPL	0.03	0.05	0.59
District by cohort interaction: School report card rating – on one scale	0.03	0.46	0.95
District by cohort interaction: School percent non-white	-0.08	0.09	0.37
District by cohort interaction: School percent ELL	0.01	0.04	0.69
District by cohort interaction: School percent Special Education	-0.12	0.11	0.27
Constant	0.89	0.20	<0.001
Random effects			
School	0.10		
Teacher	0.59		
n			
School	107		
Teacher	157		

**Exhibit D-19. Impact of the NTC Model on Engaging Students in Learning –
Both Cohorts and RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.15	0.17	0.37
Baseline score	0.03	0.09	0.70
Teacher variables			
Teacher is female CENTERED	0.03	0.22	0.90
Teacher race is black CENTERED	0.01	0.21	0.97
Teacher has only partial certification CENTERED	-0.54	0.29	0.07
School variables			
School percent FRPL CENTERED	-0.02	0.01	0.16
School report card rating - on one scale CENTERED	0.04	0.11	0.71
School percent non-white CENTERED	0.02	0.02	0.34
School percent ELL CENTERED	0.01	0.01	0.32
School percent Special Education CENTERED	<0.001	0.03	0.86
Blocking variables			
District A block 3	-0.53	0.39	0.17
District A block 2	-0.16	0.36	0.66
District B block 5	-0.19	0.39	0.62
District B block 7	-0.97	0.50	0.05
District B block 8	-0.14	0.40	0.72
District B block 6	-0.72	0.38	0.06
Interactions by district			
Centered district indicator	-0.32	0.39	0.41
District interaction: Baseline score	0.03	0.18	0.87
District interaction: Teacher is female	-0.04	0.45	0.92
District interaction: Teacher race is black	0.51	0.43	0.24
District interaction: Teacher has only partial certification	-0.04	0.58	0.95
District interaction: School percent FRPL	-0.02	0.02	0.47
District interaction: School report card rating – on one scale	0.06	0.22	0.80
District interaction: School percent non-white	0.02	0.04	0.68
District interaction: School percent ELL	0.03	0.02	0.09
District interaction: School percent Special Education	0.10	0.05	0.07
Interactions by cohort			
Centered cohort indicator	0.40	0.38	0.28
Cohort interaction: Baseline score	-0.25	0.18	0.17
Cohort interaction: Teacher is female	0.69	0.45	0.12
Cohort interaction: Teacher race is black	-0.44	0.42	0.30
Cohort interaction: Teacher has only partial certification	0.80	0.58	0.17
Cohort interaction: School percent FRPL	<0.001	0.02	0.97
Cohort interaction: School report card rating – on one scale	-0.34	0.22	0.12
Cohort interaction: School percent non-white	-0.02	0.04	0.56
Cohort interaction: School percent ELL	0.01	0.02	0.57
Cohort interaction: School percent Special Education	-0.06	0.05	0.23
Cohort interaction: District A block 3	0.66	0.77	0.39

**Exhibit D-19. Impact of the NTC Model on Engaging Students in Learning –
Both Cohorts and RCT Districts Combined (concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: District A block 2	0.09	0.69	0.89
Cohort interaction: District B block 5	0.46	0.78	0.56
Cohort interaction: District B block 7	0.37	1.01	0.72
Cohort interaction: District B block 8	-1.57	0.78	0.04
Cohort interaction: District B block 6	-0.81	0.73	0.27
District-by-cohort interactions			
District by cohort interaction	-0.69	0.77	0.37
District by cohort interaction: Baseline score	0.40	0.36	0.26
District by cohort interaction: Teacher is female	0.83	0.89	0.35
District by cohort interaction: Teacher race is black	0.37	0.87	0.67
District by cohort interaction: Teacher has only partial certification	-0.23	1.17	0.84
District by cohort interaction: School percent FRPL	0.02	0.05	0.73
District by cohort interaction: School report card rating – on one scale	-0.47	0.44	0.28
District by cohort interaction: School percent non-white	-0.07	0.08	0.38
District by cohort interaction: School percent ELL	0.01	0.03	0.88
District by cohort interaction: School percent Special Education	-0.08	0.10	0.43
Constant	0.70	0.20	<0.001
Random effects			
School	0.04		
Teacher	0.59		
n			
School	108		
Teacher	157		

**Exhibit D-20. Impact of the NTC Model on Using Assessment in Instruction –
Both Cohorts and RCT Districts Combined (continued)**

	Estimate	Standard Error	p-value
Treatment status	0.06	0.18	0.74
Baseline score	-0.01	0.07	0.92
Teacher controls			
Teacher is female CENTERED	0.15	0.19	0.41
Teacher race is black CENTERED	0.02	0.25	0.93
Teacher has only partial certification CENTERED	-0.11	0.20	0.57
School controls			
School percent FRPL CENTERED	0.01	0.01	0.72
School report card rating - on one scale CENTERED	-0.02	0.09	0.82
School percent non-white CENTERED	-0.02	0.02	0.41
School percent ELL CENTERED	<0.001	0.01	0.99
School percent Special Education CENTERED	-0.03	0.02	0.20
Blocking variables			
District A block 3	-0.90	0.30	<0.001
District A block 2	-0.27	0.39	0.49
District B block 5	-0.99	0.33	<0.001
District B block 7	-1.10	0.56	0.05
District B block 8	-0.01	0.23	0.98
District B block 6	-1.78	0.38	<0.001
Interactions with district			
Centered district indicator	-1.22	0.34	<0.001
District interaction: Baseline score	0.11	0.14	0.45
District interaction: Teacher is female	-0.35	0.37	0.35
District interaction: Teacher race is black	-0.02	0.51	0.97
District interaction: Teacher has only partial certification	-0.33	0.37	0.38
District interaction: School percent FRPL	-0.01	0.03	0.75
District interaction: School report card rating – on one scale	0.14	0.19	0.46
District interaction: School percent non-white	0.05	0.05	0.29
District interaction: School percent ELL	0.02	0.02	0.32
District interaction: School percent Special Education	0.04	0.05	0.40
Interactions with cohort			
Centered cohort indicator	-0.27	0.28	0.32
Cohort interaction: Baseline score	0.09	0.14	0.53
Cohort interaction: Teacher is female	0.26	0.37	0.48
Cohort interaction: Teacher race is black	-0.95	0.47	0.04
Cohort interaction: Teacher has only partial certification	-0.46	0.38	0.22
Cohort interaction: School percent FRPL	-0.01	0.03	0.72
Cohort interaction: School report card rating – on one scale	0.15	0.19	0.45
Cohort interaction: School percent non-white	0.06	0.05	0.22
Cohort interaction: School percent ELL	0.01	0.02	0.48
Cohort interaction: School percent Special Education	0.01	0.04	0.89
Cohort interaction: District A block 3	0.39	0.60	0.52

**Exhibit D-20. Impact of the NTC Model on Using Assessment in Instruction –
Both Cohorts and RCT Districts Combined (concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: District A block 2	1.46	0.95	0.12
Cohort interaction: District B block 5	1.06	0.63	0.09
Cohort interaction: District B block 7	-1.62	1.21	0.18
Cohort interaction: District B block 8	-1.55	0.52	<0.001
Cohort interaction: District B block 6	-2.55	0.64	<0.001
District-by-cohort interactions			
District by cohort interaction	-0.01	0.62	0.99
District by cohort interaction: Baseline score	0.40	0.28	0.15
District by cohort interaction: Teacher is female	0.15	0.74	0.84
District by cohort interaction: Teacher race is black	1.98	1.00	0.05
District by cohort interaction: Teacher has only partial certification	0.96	0.76	0.21
District by cohort interaction: School percent FRPL	0.08	0.06	0.20
District by cohort interaction: School report card rating – on one scale	0.04	0.35	0.90
District by cohort interaction: School percent non-white	-0.19	0.10	0.06
District by cohort interaction: School percent ELL	0.02	0.03	0.61
District by cohort interaction: School percent Special Education	-0.05	0.09	0.53
Constant	1.05	0.16	<0.001
Random effects			
School	<0.001		
Teacher	0.66		
n			
School	107		
Teacher	158		

Both Cohorts Combined, District A

Exhibit D-21. Impact of the NTC Model on Creating an Environment of Respect and Rapport – Both Cohorts Combined, District A

	Estimate	Standard Error	p-value
Treatment status	-0.05	0.22	0.83
Baseline score	0.31	0.13	0.01
Teacher variables			
Teacher is female CENTERED	0.05	0.26	0.84
Teacher race is black CENTERED	-0.03	0.23	0.91
Teacher has only partial certification CENTERED	-0.28	0.23	0.22
School variables			
School percent FRPL CENTERED	-0.02	0.01	0.11
School report card rating - on one scale CENTERED	-0.13	0.19	0.48
School percent non-white CENTERED	0.02	0.02	0.24
School percent ELL CENTERED	0.01	0.01	0.33
School percent Special Education CENTERED	0.01	0.05	0.77
Blocking variables			
District A block 3	0.52	0.37	0.16
District A block 2	-0.08	0.37	0.84
Interactions with cohort			
Centered cohort indicator	0.10	0.33	0.76
Cohort interaction: Baseline score	-0.46	0.25	0.07
Cohort interaction: Teacher is female	-0.81	0.52	0.12
Cohort interaction: Teacher race is black	-0.25	0.45	0.58
Cohort interaction: Teacher has only partial certification	-0.10	0.44	0.82
Cohort interaction: School percent FRPL	-0.02	0.03	0.33
Cohort interaction: School report card rating – on one scale	-0.36	0.38	0.34
Cohort interaction: School percent non-white	-0.01	0.03	0.76
Cohort interaction: School percent ELL	0.01	0.03	0.78
Cohort interaction: School percent Special Education	-0.16	0.09	0.08
Cohort interaction: District A block 3	-1.21	0.73	0.10
Cohort interaction: District A block 2	0.04	0.71	0.96
Constant	0.13	0.20	0.53
Random effects			
School	0.04	0.14	0.75
Teacher	0.57	0.16	<0.001
n			
School	36		
Teacher	44		

**Exhibit D-22. Impact of the NTC Model on Establishing a Culture for Learning –
Both Cohorts Combined, District A**

	Estimate	Standard Error	p-value
Treatment status	-0.11	0.31	0.72
Baseline score	0.23	0.11	0.03
Teacher variables			
Teacher is female CENTERED	-0.08	0.25	0.76
Teacher race is black CENTERED	0.28	0.22	0.20
Teacher has only partial certification CENTERED	-0.18	0.24	0.47
School variables			
School percent FRPL CENTERED	-0.01	0.02	0.40
School report card rating - on one scale CENTERED	0.06	0.22	0.80
School percent non-white CENTERED	0.02	0.02	0.27
School percent ELL CENTERED	-0.01	0.02	0.45
School percent Special Education CENTERED	-0.06	0.06	0.30
Blocking variables			
District A block 3	-0.81	0.44	0.06
District A block 2	-0.58	0.43	0.18
Interactions by cohort			
Centered cohort indicator	-0.57	0.34	0.10
Cohort interaction: Baseline score	-0.33	0.23	0.17
Cohort interaction: Teacher is female	1.05	0.50	0.03
Cohort interaction: Teacher race is black	0.41	0.44	0.35
Cohort interaction: Teacher has only partial certification	0.22	0.45	0.63
Cohort interaction: School percent FRPL	-0.02	0.03	0.44
Cohort interaction: School report card rating – on one scale	-0.78	0.42	0.07
Cohort interaction: School percent non-white	-0.05	0.03	0.11
Cohort interaction: School percent ELL	0.02	0.03	0.48
Cohort interaction: School percent Special Education	-0.09	0.10	0.41
Cohort interaction: District A block 3	1.84	0.79	0.02
Cohort interaction: District A block 2	1.74	0.72	0.02
Constant	0.50	0.27	0.06
Random effects			
School	0.46	0.19	0.02
Teacher	0.43	0.12	<0.001
n			
School	47		
Teacher	77		

**Exhibit D-23. Impact of the NTC Model on Managing Classroom Procedures –
Both Cohorts Combined, District A**

	Estimate	Standard Error	p-value
Treatment status	-0.36	0.28	0.20
Baseline score	-0.19	0.08	0.01
Teacher variables			
Teacher is female CENTERED	-0.54	0.17	<0.001
Teacher race is black CENTERED	0.06	0.16	0.71
Teacher has only partial certification CENTERED	-0.47	0.17	0.01
School variables			
School percent FRPL CENTERED	-0.04	0.01	0.01
School report card rating - on one scale CENTERED	0.38	0.19	0.05
School percent non-white CENTERED	0.08	0.02	<0.001
School percent ELL CENTERED	0.01	0.02	0.68
School percent Special Education CENTERED	-0.21	0.05	<0.001
Blocking variables			
District A block 3	-1.07	0.41	0.01
District A block 2	-0.08	0.37	0.83
Interactions with cohort			
Centered cohort indicator	0.87	0.22	<0.001
Cohort interaction: Baseline score	0.35	0.20	0.07
Cohort interaction: Teacher is female	0.74	0.38	0.05
Cohort interaction: Teacher race is black	-1.14	0.32	<0.001
Cohort interaction: Teacher has only partial certification	-0.22	0.30	0.46
Cohort interaction: School percent FRPL	-0.04	0.01	<0.001
Cohort interaction: School report card rating – on one scale	-1.54	0.29	<0.001
Cohort interaction: School percent non-white	-0.03	0.01	0.02
Cohort interaction: School percent ELL	0.01	0.02	0.61
Cohort interaction: School percent Special Education	0.20	0.08	0.01
Cohort interaction: District A block 3	1.24	0.63	0.05
Cohort interaction: District A block 2	-1.04	0.42	0.01
Constant	1.09	0.25	<0.001
Random effects			
School	0.67		
Teacher	0.05		
n			
School	43		
Teacher	61		

**Exhibit D-24. Impact of the NTC Model on Managing Student Behavior –
Both Cohorts Combined, District A**

	Estimate	Standard Error	p-value
Treatment status	0.26	0.24	0.27
Baseline score	0.12	0.12	0.32
Teacher variables			
Teacher is female CENTERED	0.20	0.24	0.42
Teacher race is black CENTERED	0.36	0.20	0.07
Teacher has only partial certification CENTERED	-0.35	0.20	0.09
School variables			
School percent FRPL CENTERED	<0.001	0.02	0.88
School report card rating - on one scale CENTERED	0.27	0.17	0.12
School percent non-white CENTERED	0.03	0.02	0.06
School percent ELL CENTERED	<0.001	0.01	0.76
School percent Special Education CENTERED	<0.001	0.03	1.00
Blocking variables			
District A block 3	-0.13	0.32	0.70
District A block 2	0.33	0.31	0.29
Interacts with cohorts			
Centered cohort indicator	0.77	0.31	0.01
Cohort interaction: Baseline score	0.07	0.23	0.77
Cohort interaction: Teacher is female	-0.10	0.51	0.84
Cohort interaction: Teacher race is black	-0.64	0.43	0.14
Cohort interaction: Teacher has only partial certification	0.49	0.40	0.21
Cohort interaction: School percent FRPL	-0.01	0.02	0.77
Cohort interaction: School report card rating – on one scale	-0.54	0.33	0.11
Cohort interaction: School percent non-white	-0.03	0.02	0.10
Cohort interaction: School percent ELL	-0.03	0.02	0.19
Cohort interaction: School percent Special Education	-0.14	0.07	0.05
Cohort interaction: District A block 3	-1.69	0.63	0.01
Cohort interaction: District A block 2	-1.63	0.58	0.01
Constant	-0.05	0.21	0.83
Random effects			
School	<0.001		
Teacher	0.59		
n			
School	47		
Teacher	77		

**Exhibit D-25. Impact of the NTC Model on Communicating with Students –
Both Cohorts Combined, District A**

	Estimate	Standard Error	p-value
Treatment status	0.19	0.26	0.48
Baseline score	0.21	0.11	0.07
Teacher variables			
Teacher is female CENTERED	-0.09	0.26	0.74
Teacher race is black CENTERED	-0.23	0.25	0.36
Teacher has only partial certification CENTERED	-0.08	0.25	0.76
School variables			
School percent FRPL CENTERED	-0.01	0.01	0.60
School report card rating - on one scale CENTERED	0.04	0.22	0.84
School percent non-white CENTERED	0.02	0.02	0.20
School percent ELL CENTERED	-0.01	0.02	0.53
School percent Special Education CENTERED	0.05	0.05	0.38
Blocking variables			
District A block 3	-0.04	0.42	0.93
District A block 2	-0.38	0.41	0.35
Interactions with cohort			
Centered cohort indicator	-0.21	0.35	0.55
Cohort interaction: Baseline score	0.19	0.23	0.40
Cohort interaction: Teacher is female	0.57	0.53	0.28
Cohort interaction: Teacher race is black	0.21	0.49	0.67
Cohort interaction: Teacher has only partial certification	0.35	0.47	0.46
Cohort interaction: School percent FRPL	0.04	0.03	0.16
Cohort interaction: School report card rating – on one scale	-0.17	0.42	0.68
Cohort interaction: School percent non-white	-0.07	0.03	0.03
Cohort interaction: School percent ELL	0.03	0.03	0.44
Cohort interaction: School percent Special Education	-0.01	0.10	0.93
Cohort interaction: District A block 3	0.81	0.82	0.32
Cohort interaction: District A block 2	1.09	0.73	0.14
Constant	0.01	0.24	0.96
Random effects			
School	0.21		
Teacher	0.56		
n			
School	47		
Teacher	75		

**Exhibit D-26. Impact of the NTC Model on Using Questioning and Discussion Techniques –
Both Cohorts Combined, District A**

	Estimate	Standard Error	p-value
Treatment status	0.12	0.31	0.68
Baseline score	-0.10	0.14	0.48
Teacher variables			
Teacher is female CENTERED	-0.17	0.22	0.44
Teacher race is black CENTERED	0.01	0.20	0.96
Teacher has only partial certification CENTERED	-0.35	0.23	0.12
School variables			
School percent FRPL CENTERED	-0.03	0.02	0.09
School report card rating - on one scale CENTERED	0.20	0.22	0.36
School percent non-white CENTERED	0.05	0.02	0.01
School percent ELL CENTERED	<0.001	0.02	0.97
School percent Special Education CENTERED	0.07	0.05	0.20
Blocking variables			
District A block 3	-0.82	0.47	0.08
District A block 2	0.02	0.44	0.96
Interactions with cohort			
Centered cohort indicator	-0.61	0.32	0.05
Cohort interaction: Baseline score	0.22	0.27	0.41
Cohort interaction: Teacher is female	0.12	0.44	0.78
Cohort interaction: Teacher race is black	0.29	0.39	0.46
Cohort interaction: Teacher has only partial certification	0.77	0.42	0.07
Cohort interaction: School percent FRPL	0.06	0.02	0.01
Cohort interaction: School report card rating – on one scale	0.17	0.39	0.66
Cohort interaction: School percent non-white	-0.11	0.03	<0.001
Cohort interaction: School percent ELL	0.01	0.03	0.86
Cohort interaction: School percent Special Education	-0.16	0.10	0.09
Cohort interaction: District A block 3	1.16	0.86	0.18
Cohort interaction: District A block 2	-0.08	0.66	0.91
Constant	0.62	0.27	0.02
Random effects			
School	0.66		
Teacher	0.26		
n			
School	47		
Teacher	76		

**Exhibit D-27. Impact of the NTC Model on Engaging Students in Learning –
Both Cohorts Combined, District A**

	Estimate	Standard Error	p-value
Treatment status	0.32	0.26	0.21
Baseline score	0.05	0.14	0.71
Teacher variables			
Teacher is female CENTERED	<0.001	0.27	0.99
Teacher race is black CENTERED	0.27	0.26	0.30
Teacher has only partial certification CENTERED	-0.63	0.26	0.02
School variables			
School percent FRPL CENTERED	-0.03	0.01	0.07
School report card rating - on one scale CENTERED	0.03	0.21	0.87
School percent non-white CENTERED	0.03	0.02	0.14
School percent ELL CENTERED	0.02	0.02	0.14
School percent Special Education CENTERED	0.06	0.05	0.29
Blocking variables			
District A block 3	-0.48	0.44	0.27
District A block 2	-0.12	0.40	0.76
Interactions with cohort			
Centered cohort indicator	0.08	0.37	0.82
Cohort interaction: Baseline score	-0.05	0.28	0.86
Cohort interaction: Teacher is female	1.03	0.55	0.06
Cohort interaction: Teacher race is black	-0.29	0.51	0.58
Cohort interaction: Teacher has only partial certification	0.70	0.50	0.16
Cohort interaction: School percent FRPL	0.01	0.03	0.68
Cohort interaction: School report card rating – on one scale	-0.57	0.42	0.18
Cohort interaction: School percent non-white	-0.06	0.03	0.05
Cohort interaction: School percent ELL	0.02	0.03	0.61
Cohort interaction: School percent Special Education	-0.10	0.11	0.32
Cohort interaction: District A block 3	0.57	0.86	0.51
Cohort interaction: District A block 2	0.06	0.76	0.94
Constant	0.47	0.23	0.04
Random effects			
School	0.10		
Teacher	0.68		
n			
School	47		
Teacher	76		

**Exhibit D-28. Impact of the NTC Model on Using Assessment in Instruction –
Both Cohorts Combined, District A**

	Estimate	Standard Error	p-value
Treatment status	0.21	0.26	0.42
Baseline score	0.05	0.13	0.69
Teacher variables			
Teacher is female CENTERED	-0.12	0.25	0.63
Teacher race is black CENTERED	<0.001	0.23	1.00
Teacher has only partial certification CENTERED	-0.48	0.24	0.05
School variables			
School percent FRPL CENTERED	<0.001	0.01	0.77
School report card rating - on one scale CENTERED	0.01	0.20	0.95
School percent non-white CENTERED	0.01	0.02	0.69
School percent ELL CENTERED	0.01	0.02	0.63
School percent Special Education CENTERED	<0.001	0.05	0.94
Blocking variables			
District A block 3	-0.90	0.40	0.02
District A block 2	-0.30	0.39	0.45
Interactions with cohort			
Centered cohort indicator	-0.31	0.37	0.39
Cohort interaction: Baseline score	0.29	0.24	0.22
Cohort interaction: Teacher is female	0.38	0.50	0.44
Cohort interaction: Teacher race is black	0.08	0.45	0.85
Cohort interaction: Teacher has only partial certification	<0.001	0.44	0.99
Cohort interaction: School percent FRPL	0.03	0.03	0.17
Cohort interaction: School report card rating – on one scale	0.04	0.39	0.92
Cohort interaction: School percent non-white	-0.05	0.03	0.10
Cohort interaction: School percent ELL	0.03	0.03	0.38
Cohort interaction: School percent Special Education	<0.001	0.10	0.97
Cohort interaction: District A block 3	0.58	0.76	0.45
Cohort interaction: District A block 2	1.09	0.71	0.13
Constant	0.41	0.24	0.09
Random effects			
School	0.18	0.24	0.47
Teacher	0.52	0.20	0.01
n			
School	47		
Teacher	77		

Both Cohorts Combined, District B

Exhibit D-29. Impact of the NTC Model on Creating an Environment of Respect and Rapport – Both Cohorts Combined, District B

	Estimate	Standard Error	p-value
Treatment status	0.15	0.24	0.55
Baseline score	-0.10	0.12	0.40
Teacher variables			
Teacher is female CENTERED	-0.05	0.34	0.89
Teacher race is black CENTERED	-0.10	0.34	0.78
Teacher has only partial certification CENTERED	0.24	0.49	0.62
School variables			
School percent FRPL CENTERED	0.01	0.02	0.44
School report card rating - on one scale CENTERED	0.14	0.10	0.15
School percent non-white CENTERED	-0.03	0.04	0.43
School percent ELL CENTERED	<0.001	0.01	0.70
School percent Special Education CENTERED	-0.04	0.02	0.03
Blocking variables			
District B block 5	-0.65	0.34	0.05
District B block 7	-1.21	0.46	0.01
District B block 8	-0.34	0.33	0.31
District B block 6	-1.28	0.37	<0.001
Interactions with cohort			
Centered cohort indicator	-1.57	0.58	0.01
Cohort interaction: Baseline score	-0.02	0.24	0.92
Cohort interaction: Teacher is female	-0.20	0.68	0.76
Cohort interaction: Teacher race is black	-0.53	0.69	0.44
Cohort interaction: Teacher has only partial certification	0.36	0.96	0.71
Cohort interaction: School percent FRPL	-0.04	0.04	0.35
Cohort interaction: School report card rating – on one scale	-0.02	0.21	0.93
Cohort interaction: School percent non-white	0.07	0.07	0.33
Cohort interaction: School percent ELL	-0.02	0.02	0.29
Cohort interaction: School percent Special Education	-0.08	0.04	0.03
Cohort interaction: District B block 5	1.30	0.68	0.05
Cohort interaction: District B block 7	1.18	0.94	0.21
Cohort interaction: District B block 8	0.84	0.63	0.18
Cohort interaction: District B block 6	0.38	0.69	0.58
Constant	1.12	0.28	<0.001
Random effects			
School	<0.001		
Teacher	0.53		
n			
School	61		
Teacher	82		

**Exhibit D-30. Impact of the NTC Model on Establishing a Culture for Learning –
Both Cohorts Combined, District B**

	Estimate	Standard Error	p-value
Treatment status	-0.34	0.24	0.16
Baseline score	-0.16	0.09	0.08
Teacher variables			
Teacher is female CENTERED	-0.24	0.32	0.45
Teacher race is black CENTERED	-0.16	0.31	0.60
Teacher has only partial certification CENTERED	-0.69	0.39	0.08
School variables			
School percent FRPL CENTERED	0.02	0.02	0.33
School report card rating - on one scale CENTERED	0.08	0.10	0.46
School percent non-white CENTERED	-0.04	0.03	0.25
School percent ELL CENTERED	<0.001	0.01	0.59
School percent Special Education CENTERED	-0.01	0.02	0.72
Blocking variables			
District B block 5	-0.82	0.38	0.03
District B block 7	-0.50	0.53	0.35
District B block 8	-0.11	0.40	0.78
District B block 6	-0.94	0.40	0.02
Interactions with cohort			
Centered cohort indicator	0.41	0.58	0.48
Cohort interaction: Baseline score	0.20	0.19	0.28
Cohort interaction: Teacher is female	-1.32	0.65	0.04
Cohort interaction: Teacher race is black	-1.12	0.62	0.07
Cohort interaction: Teacher has only partial certification	1.75	0.86	0.04
Cohort interaction: School percent FRPL	-0.02	0.04	0.53
Cohort interaction: School report card rating – on one scale	-0.11	0.21	0.59
Cohort interaction: School percent non-white	0.06	0.07	0.42
Cohort interaction: School percent ELL	0.01	0.02	0.51
Cohort interaction: School percent Special Education	0.04	0.03	0.31
Cohort interaction: District B block 5	0.08	0.74	0.91
Cohort interaction: District B block 7	-0.19	1.04	0.85
Cohort interaction: District B block 8	-1.46	0.70	0.04
Cohort interaction: District B block 6	0.17	0.72	0.81
Constant	1.31	0.31	<0.001
Random effects			
School	0.41		
Teacher	0.18		
n			
School	61		
Teacher	82		

**Exhibit D-31. Impact of the NTC Model on Managing Classroom Procedures –
Both Cohorts Combined, District B**

	Estimate	Standard Error	p-value
Treatment status	0.56	0.25	0.03
Baseline score	0.05	0.12	0.69
Teacher variables			
Teacher is female CENTERED	0.56	0.42	0.18
Teacher race is black CENTERED	-0.28	0.40	0.48
Teacher has only partial certification CENTERED	0.18	0.34	0.59
School variables			
School percent FRPL CENTERED	0.03	0.03	0.26
School report card rating - on one scale CENTERED	0.04	0.08	0.60
School percent non-white CENTERED	-0.06	0.05	0.20
School percent ELL CENTERED	<0.001	0.01	0.55
School percent Special Education CENTERED	-0.03	0.02	0.06
Blocking variables			
District B block 5	-0.65	0.31	0.04
District B block 7	0.17	0.73	0.82
District B block 8	0.17	0.24	0.48
District B block 6	-0.28	0.35	0.43
Interactions with cohort			
Centered cohort indicator	-0.26	0.45	0.57
Cohort interaction: Baseline score	0.05	0.29	0.87
Cohort interaction: Teacher is female	0.50	0.92	0.58
Cohort interaction: Teacher race is black	0.67	0.83	0.42
Cohort interaction: Teacher has only partial certification	-1.05	0.61	0.09
Cohort interaction: School percent FRPL	-0.10	0.06	0.12
Cohort interaction: School report card rating – on one scale	0.28	0.19	0.13
Cohort interaction: School percent non-white	0.17	0.10	0.11
Cohort interaction: School percent ELL	0.01	0.01	0.31
Cohort interaction: School percent Special Education	-0.03	0.03	0.28
Cohort interaction: District B block 5	0.32	0.58	0.58
Cohort interaction: District B block 7	-2.49	1.52	0.10
Cohort interaction: District B block 8	-1.25	0.51	0.01
Cohort interaction: District B block 6	-1.91	0.74	0.01
Constant	0.77	0.23	<0.001
Random effects			
School	<0.001		
Teacher	0.51		
n			
School	59		
Teacher	78		

**Exhibit D-32. Impact of the NTC Model on Managing Student Behavior –
Both Cohorts Combined, District B**

	Estimate	Standard Error	p-value
Treatment status	0.28	0.28	0.31
Baseline score	0.10	0.13	0.44
Teacher variables			
Teacher is female CENTERED	0.36	0.32	0.26
Teacher race is black CENTERED	-0.42	0.43	0.33
Teacher has only partial certification CENTERED	0.43	0.35	0.22
School variables			
School percent FRPL CENTERED	-0.01	0.02	0.81
School report card rating - on one scale CENTERED	0.10	0.09	0.30
School percent non-white CENTERED	<0.001	0.04	0.98
School percent ELL CENTERED	-0.01	0.01	0.37
School percent Special Education CENTERED	-0.01	0.02	0.49
Blocking variables			
District B block 5	-0.90	0.35	0.01
District B block 7	-1.30	0.42	<0.001
District B block 8	-0.09	0.24	0.70
District B block 6	-1.28	0.34	<0.001
Interactions with cohort			
Centered cohort indicator	-0.33	0.48	0.50
Cohort interaction: Baseline score	-0.26	0.26	0.32
Cohort interaction: Teacher is female	-0.23	0.65	0.72
Cohort interaction: Teacher race is black	0.16	0.84	0.85
Cohort interaction: Teacher has only partial certification	-0.59	0.63	0.35
Cohort interaction: School percent FRPL	-0.06	0.04	0.19
Cohort interaction: School report card rating – on one scale	0.01	0.20	0.95
Cohort interaction: School percent non-white	0.08	0.08	0.26
Cohort interaction: School percent ELL	<0.001	0.01	0.90
Cohort interaction: School percent Special Education	-0.02	0.03	0.52
Cohort interaction: District B block 5	1.12	0.71	0.11
Cohort interaction: District B block 7	0.42	0.87	0.63
Cohort interaction: District B block 8	-1.22	0.48	0.01
Cohort interaction: District B block 6	-1.60	0.63	0.01
Constant	0.91	0.22	<0.001
Random effects			
School	<0.001		
Teacher	0.63		
n			
School	61		
Teacher	82		

**Exhibit D-33. Impact of the NTC Model on Communicating with Students –
Both Cohorts Combined, District B**

	Estimate	Standard Error	p-value
Treatment status	-0.29	0.27	0.29
Baseline score	<0.001	0.13	0.98
Teacher variables			
Teacher is female CENTERED	-0.17	0.38	0.65
Teacher race is black CENTERED	-0.44	0.40	0.27
Teacher has only partial certification CENTERED	0.25	0.57	0.66
School variables			
School percent FRPL CENTERED	-0.01	0.02	0.74
School report card rating - on one scale CENTERED	0.09	0.11	0.43
School percent non-white CENTERED	<0.001	0.04	0.96
School percent ELL CENTERED	<0.001	0.01	0.58
School percent Special Education CENTERED	-0.01	0.02	0.61
Blocking variables			
District B block 5	-0.74	0.40	0.06
District B block 7	-0.43	0.52	0.41
District B block 8	-0.20	0.41	0.62
District B block 6	-0.92	0.42	0.03
Interactions with cohort			
Centered cohort indicator	1.39	0.69	0.04
Cohort interaction: Baseline score	0.06	0.27	0.83
Cohort interaction: Teacher is female	-1.34	0.75	0.08
Cohort interaction: Teacher race is black	-0.49	0.82	0.55
Cohort interaction: Teacher has only partial certification	0.57	1.11	0.61
Cohort interaction: School percent FRPL	0.03	0.05	0.44
Cohort interaction: School report card rating – on one scale	0.10	0.24	0.67
Cohort interaction: School percent non-white	-0.05	0.09	0.57
Cohort interaction: School percent ELL	0.01	0.02	0.76
Cohort interaction: School percent Special Education	0.01	0.04	0.78
Cohort interaction: District B block 5	-1.08	0.79	0.17
Cohort interaction: District B block 7	0.07	1.05	0.95
Cohort interaction: District B block 8	-1.93	0.75	0.01
Cohort interaction: District B block 6	-0.84	0.77	0.28
Constant	1.14	0.32	<0.001
Random effects			
School	<0.001		
Teacher	0.68		
n			
School	61		
Teacher	82		

**Exhibit D-34. Impact of the NTC Model on Using Questioning and Discussion Techniques –
Both Cohorts Combined, District B**

	Estimate	Standard Error	p-value
Treatment status	0.14	0.24	0.57
Baseline score	0.15	0.10	0.15
Teacher variables			
Teacher is female CENTERED	-0.08	0.34	0.81
Teacher race is black CENTERED	-0.05	0.47	0.92
Teacher has only partial certification CENTERED	-0.25	0.39	0.51
School variables			
School percent FRPL CENTERED	<0.001	0.02	0.99
School report card rating - on one scale CENTERED	0.07	0.10	0.46
School percent non-white CENTERED	-0.01	0.05	0.90
School percent ELL CENTERED	<0.001	0.01	0.95
School percent Special Education CENTERED	-0.03	0.02	0.11
Blocking variables			
District B block 5	-0.87	0.36	0.02
District B block 7	-1.11	0.42	0.01
District B block 8	-0.55	0.26	0.04
District B block 6	-1.44	0.32	<0.001
Interactions with cohort			
Centered cohort indicator	0.73	0.60	0.22
Cohort interaction: Baseline score	-0.42	0.21	0.04
Cohort interaction: Teacher is female	0.49	0.67	0.47
Cohort interaction: Teacher race is black	-0.58	0.91	0.52
Cohort interaction: Teacher has only partial certification	0.92	0.65	0.16
Cohort interaction: School percent FRPL	0.01	0.05	0.91
Cohort interaction: School report card rating – on one scale	-0.26	0.20	0.19
Cohort interaction: School percent non-white	-0.01	0.09	0.91
Cohort interaction: School percent ELL	-0.01	0.01	0.42
Cohort interaction: School percent Special Education	-0.07	0.03	0.03
Cohort interaction: District B block 5	0.56	0.71	0.43
Cohort interaction: District B block 7	0.36	1.02	0.72
Cohort interaction: District B block 8	-0.29	0.67	0.67
Cohort interaction: District B block 6	-0.85	0.61	0.16
Constant	1.42	0.25	<0.001
Random effects			
School	<0.001		
Teacher	0.55		
n			
School	60		
Teacher	81		

**Exhibit D-35. Impact of the NTC Model on Engaging Students in Learning –
Both Cohorts Combined, District B**

	Estimate	Standard Error	p-value
Treatment status	-0.04	0.23	0.87
Baseline score	0.01	0.13	0.94
Teacher variables			
Teacher is female CENTERED	0.03	0.30	0.91
Teacher race is black CENTERED	-0.26	0.29	0.37
Teacher has only partial certification CENTERED	-0.38	0.35	0.28
School variables			
School percent FRPL CENTERED	-0.02	0.02	0.28
School report card rating - on one scale CENTERED	0.09	0.08	0.28
School percent non-white CENTERED	0.04	0.03	0.27
School percent ELL CENTERED	-0.01	0.01	0.28
School percent Special Education CENTERED	-0.03	0.02	0.05
Blocking variables			
District B block 5	-0.37	0.41	0.37
District B block 7	-1.23	0.41	<0.001
District B block 8	-0.19	0.28	0.51
District B block 6	-0.99	0.34	<0.001
Interactions with cohort			
Centered cohort indicator	0.93	0.60	0.12
Cohort interaction: Baseline score	-0.39	0.25	0.12
Cohort interaction: Teacher is female	0.37	0.61	0.54
Cohort interaction: Teacher race is black	-1.15	0.56	0.04
Cohort interaction: Teacher has only partial certification	0.43	0.61	0.48
Cohort interaction: School percent FRPL	0.01	0.03	0.78
Cohort interaction: School report card rating – on one scale	-0.12	0.16	0.47
Cohort interaction: School percent non-white	-0.02	0.07	0.75
Cohort interaction: School percent ELL	<0.001	0.01	0.88
Cohort interaction: School percent Special Education	<0.001	0.03	0.92
Cohort interaction: District B block 5	0.30	0.83	0.72
Cohort interaction: District B block 7	0.58	0.81	0.47
Cohort interaction: District B block 8	-1.50	0.68	0.03
Cohort interaction: District B block 6	-1.14	0.61	0.06
Constant	1.05	0.29	<0.001
Random effects			
School	<0.001		
Teacher	0.45		
n			
School	61		
Teacher	81		

**Exhibit D-36. Impact of the NTC Model on Using Assessment in Instruction –
Both Cohorts Combined, District B**

	Estimate	Standard Error	p-value
Treatment status	-0.21	0.26	0.42
Baseline score	-0.04	0.11	0.69
Teacher variables			
Teacher is female CENTERED	0.25	0.28	0.37
Teacher race is black CENTERED	0.17	0.46	0.71
Teacher has only partial certification CENTERED	0.12	0.35	0.72
School variables			
School percent FRPL CENTERED	<0.001	0.03	0.95
School report card rating - on one scale CENTERED	-0.07	0.09	0.43
School percent non-white CENTERED	-0.02	0.05	0.64
School percent ELL CENTERED	-0.01	0.01	0.35
School percent Special Education CENTERED	-0.04	0.02	0.07
Blocking variables			
District B block 5	-0.82	0.35	0.02
District B block 7	-1.14	0.56	0.04
District B block 8	0.14	0.24	0.57
District B block 6	-1.60	0.42	<0.001
Interactions with cohort			
Centered cohort indicator	-0.02	0.53	0.97
Cohort interaction: Baseline score	-0.13	0.21	0.55
Cohort interaction: Teacher is female	0.34	0.62	0.58
Cohort interaction: Teacher race is black	-2.24	0.87	0.01
Cohort interaction: Teacher has only partial certification	-0.92	0.56	0.10
Cohort interaction: School percent FRPL	-0.02	0.05	0.65
Cohort interaction: School report card rating – on one scale	0.02	0.21	0.91
Cohort interaction: School percent non-white	0.10	0.11	0.34
Cohort interaction: School percent ELL	<0.001	0.01	0.85
Cohort interaction: School percent Special Education	0.04	0.04	0.30
Cohort interaction: District B block 5	0.78	0.66	0.24
Cohort interaction: District B block 7	-1.15	1.32	0.38
Cohort interaction: District B block 8	-1.51	0.56	0.01
Cohort interaction: District B block 6	-2.40	0.67	<0.001
Constant	1.65	0.24	<0.001
Random effects			
School	<0.001		
Teacher	0.60		
n			
School	60		
Teacher	81		

Cohort 1, Both RCT Districts Combined

Exhibit D-37. Impact of the NTC Model on Creating an Environment of Respect and Rapport – Cohort 1, Both RCT Districts Combined

	Estimate	Standard Error	p-value
Treatment status	0.16	0.21	0.44
Baseline score	0.05	0.14	0.73
Teacher variables			
Teacher is female CENTERED	0.43	0.27	0.10
Teacher race is black CENTERED	0.16	0.27	0.54
Teacher has only partial certification CENTERED	-0.12	0.19	0.53
School variables			
School percent FRPL CENTERED	0.02	0.02	0.29
School report card rating - on one scale CENTERED	0.02	0.19	0.92
School percent non-white CENTERED	-0.04	0.04	0.27
School percent ELL CENTERED	0.01	0.01	0.50
School percent Special Education CENTERED	0.02	0.04	0.53
Blocking Variables			
District A block 3	1.10	0.61	0.07
District A block 2	-0.41	0.55	0.46
District B block 5	-1.23	0.45	0.01
District B block 7	-1.74	0.58	<0.001
District B block 8	-0.87	0.33	0.01
District B block 6	-1.26	0.30	<0.001
Interactions with district			
Centered district indicator	-1.92	0.46	<0.001
District interaction: Baseline score	0.59	0.30	0.05
District interaction: Teacher is female	-0.13	0.55	0.82
District interaction: Teacher race is black	-0.04	0.53	0.94
District interaction: Teacher has only partial certification	-0.40	0.39	0.30
District interaction: School percent FRPL	-0.07	0.04	0.08
District interaction: School report card rating – on one scale	-0.28	0.35	0.43
District interaction: School percent non-white	0.12	0.07	0.07
District interaction: School percent ELL	-0.01	0.02	0.61
District interaction: School percent Special Education	0.08	0.07	0.29
Constant	0.97	0.20	<0.001
Random effects			
School	<0.001		
Teacher	0.43		
n			
School	56		
Teacher	71		

**Exhibit D-38. Impact of the NTC Model on Establishing a Culture for Learning –
Cohort 1, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	-0.10	0.23	0.67
Baseline score	0.07	0.10	0.53
Teacher variables			
Teacher is female CENTERED	-0.17	0.34	0.61
Teacher race is black CENTERED	0.08	0.30	0.79
Teacher has only partial certification CENTERED	-0.31	0.42	0.46
School variables			
School percent FRPL CENTERED	0.02	0.02	0.45
School report card rating - on one scale CENTERED	0.31	0.15	0.04
School percent non-white CENTERED	-0.01	0.04	0.79
School percent ELL CENTERED	-0.02	0.01	0.08
School percent Special Education CENTERED	-0.03	0.04	0.41
Blocking variables			
District A block 3	-1.87	0.53	<0.001
District A block 2	-1.44	0.51	0.01
District B block 5	-0.98	0.51	0.05
District B block 7	-0.59	0.78	0.45
District B block 8	0.52	0.44	0.23
District B block 6	-1.03	0.37	0.01
Centered district indicator	-0.42	0.55	0.44
Interactions with district			
District interaction: Baseline score	0.41	0.21	0.05
District interaction: Teacher is female	-0.79	0.68	0.25
District interaction: Teacher race is black	-0.02	0.65	0.97
District interaction: Teacher has only partial certification	0.65	0.81	0.42
District interaction: School percent FRPL	-0.04	0.04	0.30
District interaction: School report card rating – on one scale	0.34	0.29	0.24
District interaction: School percent non-white	0.10	0.08	0.19
District interaction: School percent ELL	<0.001	0.02	0.87
District interaction: School percent Special Education	-0.01	0.07	0.89
Constant	0.83	0.25	<0.001
Random effects			
School	<0.001		
Teacher	0.47		
n			
School	56		
Teacher	71		

**Exhibit D-39. Impact of the NTC Model on Managing Classroom Procedures –
Cohort 1, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.23	0.25	0.37
Baseline score	-0.05	0.16	0.76
Teacher controls			
Teacher is female CENTERED	-0.44	0.42	0.29
Teacher race is black CENTERED	0.14	0.32	0.66
Teacher has only partial certification CENTERED	0.18	0.43	0.68
School controls			
School percent FRPL CENTERED	<0.001	0.02	0.86
School report card rating - on one scale CENTERED	0.14	0.16	0.39
School percent non-white CENTERED	<0.001	0.04	0.93
School percent ELL CENTERED	-0.01	0.01	0.40
School percent Special Education CENTERED	-0.04	0.05	0.34
Blocking Variables			
District A block 3	-0.47	0.62	0.45
District A block 2	-0.51	0.62	0.41
District B block 5	-0.37	0.51	0.47
District B block 7	0.58	1.04	0.58
District B block 8	0.80	0.43	0.06
District B block 6	0.67	0.37	0.07
Interactions with district			
Centered district indicator	-0.36	0.60	0.55
District interaction: Baseline score	-0.19	0.32	0.55
District interaction: Teacher is female	-0.62	0.80	0.44
District interaction: Teacher race is black	0.56	0.72	0.43
District interaction: Teacher has only partial certification	-1.19	0.83	0.15
District interaction: School percent FRPL	-0.07	0.05	0.14
District interaction: School report card rating – on one scale	0.41	0.32	0.21
District interaction: School percent non-white	0.14	0.08	0.07
District interaction: School percent ELL	<0.001	0.02	0.99
District interaction: School percent Special Education	-0.07	0.09	0.44
Constant	0.66	0.27	0.02
Random effects			
School	<0.001		
Teacher	0.44		
n			
School	51		
Teacher	59		

**Exhibit D-40. Impact of the NTC Model on Managing Student Behavior –
Cohort 1, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.34	0.24	0.15
Baseline score	0.23	0.13	0.08
Teacher controls			
Teacher is female CENTERED	0.45	0.34	0.18
Teacher race is black CENTERED	0.10	0.31	0.75
Teacher has only partial certification CENTERED	-0.07	0.44	0.87
School controls			
School percent FRPL CENTERED	<0.001	0.02	0.93
School report card rating - on one scale CENTERED	0.28	0.16	0.08
School percent non-white CENTERED	<0.001	0.04	0.92
School percent ELL CENTERED	<0.001	0.01	0.76
School percent Special Education CENTERED	0.05	0.04	0.25
Blocking variables			
District A block 3	0.60	0.59	0.30
District A block 2	1.21	0.54	0.03
District B block 5	-1.31	0.54	0.01
District B block 7	-1.63	0.76	0.03
District B block 8	0.54	0.49	0.27
District B block 6	-0.49	0.40	0.22
Interactions with district			
Centered district indicator	-1.55	0.59	0.01
District interaction: Baseline score	-0.11	0.27	0.69
District interaction: Teacher is female	-0.03	0.67	0.96
District interaction: Teacher race is black	1.19	0.68	0.08
District interaction: Teacher has only partial certification	-0.97	0.86	0.26
District interaction: School percent FRPL	-0.04	0.04	0.38
District interaction: School report card rating – on one scale	0.41	0.30	0.18
District interaction: School percent non-white	0.10	0.07	0.16
District interaction: School percent ELL	0.02	0.03	0.35
District interaction: School percent Special Education	0.08	0.08	0.31
Constant	0.27	0.27	0.32
Random effects			
School	0.03		
Teacher	0.50		
n			
School	56		
Teachers	71		

**Exhibit D-41. Impact of the NTC Model on Communicating with Students –
Cohort 1, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	-0.06	0.24	0.80
Baseline score	<0.001	0.12	0.97
Teacher controls			
Teacher is female CENTERED	0.13	0.33	0.69
Teacher race is black CENTERED	-0.41	0.32	0.20
Teacher has only partial certification CENTERED	-0.13	0.44	0.76
School controls			
School percent FRPL CENTERED	-0.02	0.02	0.32
School report card rating - on one scale CENTERED	0.10	0.16	0.54
School percent non-white CENTERED	0.03	0.04	0.51
School percent ELL CENTERED	-0.01	0.01	0.47
School percent Special Education CENTERED	0.02	0.04	0.57
Blocking variables			
District A block 3	-0.65	0.61	0.28
District A block 2	-1.00	0.55	0.07
District B block 5	-0.55	0.56	0.33
District B block 7	-0.57	0.73	0.44
District B block 8	0.48	0.53	0.36
District B block 6	-0.72	0.42	0.09
Interactions with district			
Centered district indicator	-0.36	0.61	0.55
District interaction: Baseline score	0.20	0.23	0.39
District interaction: Teacher is female	-0.94	0.65	0.15
District interaction: Teacher race is black	0.03	0.70	0.96
District interaction: Teacher has only partial certification	0.22	0.86	0.80
District interaction: School percent FRPL	-0.01	0.04	0.84
District interaction: School report card rating – on one scale	0.23	0.31	0.45
District interaction: School percent non-white	0.06	0.08	0.40
District interaction: School percent ELL	-0.03	0.02	0.29
District interaction: School percent Special Education	0.08	0.08	0.35
Constant	0.46	0.28	0.10
Random effects			
School	0.09		
Teacher	0.43		
n			
School	56		
Teacher	70		

**Exhibit D-42. Impact of the NTC Model on Using Questioning and Discussion Techniques –
Cohort 1, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	-0.04	0.23	0.86
Baseline score	0.10	0.13	0.42
Teacher controls			
Teacher is female CENTERED	-0.27	0.32	0.40
Teacher race is black CENTERED	0.30	0.31	0.32
Teacher has only partial certification CENTERED	-0.59	0.42	0.16
School controls			
School percent FRPL CENTERED	-0.02	0.02	0.21
School report card rating - on one scale CENTERED	0.24	0.16	0.13
School percent non-white CENTERED	0.04	0.04	0.22
School percent ELL CENTERED	0.01	0.01	0.54
School percent Special Education CENTERED	0.07	0.04	0.06
Blocking variables			
District A block 3	-1.33	0.69	0.05
District A block 2	<0.001	0.55	0.99
District B block 5	-0.74	0.52	0.15
District B block 7	-1.05	0.77	0.17
District B block 8	-0.16	0.45	0.72
District B block 6	-0.87	0.38	0.02
Interactions with district			
Centered district indicator	-0.29	0.58	0.62
District interaction: Baseline score	-0.52	0.26	0.04
District interaction: Teacher is female	0.34	0.64	0.59
District interaction: Teacher race is black	-0.56	0.67	0.40
District interaction: Teacher has only partial certification	0.16	0.82	0.85
District interaction: School percent FRPL	-0.02	0.04	0.55
District interaction: School report card rating – on one scale	0.12	0.30	0.69
District interaction: School percent non-white	0.06	0.07	0.39
District interaction: School percent ELL	<0.001	0.02	0.93
District interaction: School percent Special Education	0.15	0.08	0.06
Constant	0.78	0.27	<0.001
Random effects			
School	<0.001		
Teacher	0.49		
n			
School	55		
Teacher	69		

**Exhibit D-43. Impact of the NTC Model on Engaging Students in Learning –
Cohort 1, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	-0.11	0.24	0.65
Baseline score	0.16	0.13	0.22
Teacher controls			
Teacher is female CENTERED	-0.37	0.35	0.29
Teacher race is black CENTERED	0.31	0.33	0.34
Teacher has only partial certification CENTERED	-0.74	0.46	0.11
School controls			
School percent FRPL CENTERED	-0.02	0.02	0.29
School report card rating - on one scale CENTERED	0.26	0.16	0.11
School percent non-white CENTERED	0.05	0.04	0.24
School percent ELL CENTERED	<0.001	0.01	0.74
School percent Special Education CENTERED	0.03	0.04	0.44
Blocking variables			
District A block 3	-1.02	0.65	0.12
District A block 2	-0.31	0.56	0.58
District B block 5	-0.16	0.62	0.80
District B block 7	-1.15	0.76	0.13
District B block 8	0.75	0.55	0.17
District B block 6	-0.19	0.43	0.66
Interactions with district			
Centered district indicator	0.24	0.66	0.71
District interaction: Baseline score	-0.19	0.26	0.48
District interaction: Teacher is female	-0.44	0.70	0.53
District interaction: Teacher race is black	0.05	0.70	0.95
District interaction: Teacher has only partial certification	-0.03	0.90	0.98
District interaction: School percent FRPL	-0.01	0.04	0.80
District interaction: School report card rating – on one scale	0.30	0.31	0.34
District interaction: School percent non-white	0.02	0.08	0.76
District interaction: School percent ELL	0.03	0.03	0.29
District interaction: School percent Special Education	0.13	0.08	0.11
Constant	0.53	0.31	0.08
Random effects			
School	<0.001		
Teacher	0.55		
n			
School	56		
Teacher	70		

**Exhibit D-44. Impact of the NTC Model on Using Assessment in Instruction –
Cohort 1, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	-0.21	0.22	0.34
Baseline score	-0.03	0.11	0.82
Teacher controls			
Teacher is female CENTERED	-0.09	0.35	0.79
Teacher race is black CENTERED	0.72	0.31	0.02
Teacher has only partial certification CENTERED	0.13	0.39	0.73
School controls			
School percent FRPL CENTERED	<0.001	0.02	0.84
School report card rating - on one scale CENTERED	-0.01	0.15	0.96
School percent non-white CENTERED	-0.01	0.04	0.79
School percent ELL CENTERED	<0.001	0.01	0.64
School percent Special Education CENTERED	-0.03	0.04	0.45
Blocking variables			
District A block 3	-1.36	0.53	0.01
District A block 2	-1.01	0.50	0.04
District B block 5	-0.90	0.54	0.10
District B block 7	-0.24	0.99	0.81
District B block 8	0.76	0.49	0.12
District B block 6	-0.27	0.37	0.46
Interactions with district			
Centered district indicator	-0.62	0.60	0.30
District interaction: Baseline score	-0.20	0.23	0.38
District interaction: Teacher is female	-0.38	0.68	0.58
District interaction: Teacher race is black	-1.76	0.67	0.01
District interaction: Teacher has only partial certification	-0.74	0.77	0.33
District interaction: School percent FRPL	-0.02	0.05	0.66
District interaction: School report card rating – on one scale	0.13	0.28	0.64
District interaction: School percent non-white	0.08	0.08	0.33
District interaction: School percent ELL	<0.001	0.02	0.95
District interaction: School percent Special Education	0.06	0.07	0.36
Constant	1.06	0.27	<0.001
Random effects			
School	<0.001		
Teacher	0.42		
n			
School	55		
Teacher	70		

Cohort 1, District A

Exhibit D-45. Impact of the NTC Model on Creating an Environment of Respect and Rapport – Cohort 1, District A

	Estimate	Standard Error	p-value
Treatment status	0.29	0.37	0.44
Baseline score	1.01	0.13	<0.001
Teacher variables			
Teacher is female CENTERED	0.79	0.24	<0.001
Teacher race is black CENTERED	0.08	0.17	0.64
Teacher has only partial certification CENTERED	-0.20	0.20	0.31
School variables			
School percent FRPL CENTERED	-0.01	0.02	0.43
School report card rating – on one scale CENTERED	-0.35	0.27	0.19
School percent non-white CENTERED	0.01	0.02	0.78
School percent ELL CENTERED	0.02	0.02	0.48
School percent Special Education CENTERED	0.19	0.08	0.02
Blocking variables			
District A block 3	2.04	0.59	<0.001
District A block 2	0.02	0.60	0.97
Constant	-0.30	0.35	0.39
Random effects			
School	0.58		
Teacher	0.08		
n			
School	24		
Teacher	33		

**Exhibit D-46. Impact of the NTC Model on Establishing a Culture for Learning –
Cohort 1, District A**

	Estimate	Standard Error	p-value
Treatment status	0.73	0.28	0.01
Baseline score	0.06	0.13	0.68
Teacher variables			
Teacher is female CENTERED	-0.60	0.24	0.01
Teacher race is black CENTERED	0.28	0.28	0.33
Teacher has only partial certification CENTERED	-0.56	0.28	0.05
School variables			
School percent FRPL CENTERED	-0.03	0.01	0.05
School report card rating – on one scale CENTERED	0.25	0.31	0.42
School percent non-white CENTERED	0.04	0.02	0.05
School percent ELL CENTERED	-0.01	0.02	0.73
School percent Special Education CENTERED	-0.05	0.06	0.38
Blocking variables			
District A block 3	-1.24	0.55	0.02
District A block 2	-1.23	0.38	<0.001
Constant	0.05	0.26	0.84
Random effects			
School	<0.001		
Teacher	0.36		
n			
School	24		
Teacher	33		

Exhibit D-47. Impact of the NTC Model on Managing Classroom Procedures – Cohort 1, District A

		Standard	
	Estimate	Error	p-value
Treatment status	-0.04	0.33	0.89
Baseline score	-0.12	0.24	0.62
Teacher variables			
Teacher is female CENTERED	-0.83	0.35	0.02
Teacher race is black CENTERED	0.28	0.36	0.44
Teacher has only partial certification CENTERED	-0.30	0.34	0.38
School variables			
School percent FRPL CENTERED	-0.03	0.01	0.06
School report card rating – on one scale CENTERED	0.36	0.25	0.15
School percent non-white CENTERED	0.07	0.02	<0.001
School percent ELL CENTERED	-0.01	0.02	0.63
School percent Special Education CENTERED	-0.07	0.07	0.37
Blocking variables			
District A block 3	-0.51	0.53	0.33
District A block 2	-0.59	0.53	0.26
Constant	0.69	0.34	0.04
Random effects			
School	<0.001		
Teacher	0.31		
n			
School	21		
Teacher	24		

Exhibit D-48. Impact of the NTC Model on Managing Student Behavior – Cohort 1, District A

	Estimate	Standard Error	p-value
Treatment status	1.08	0.25	<0.001
Baseline score	0.32	0.15	0.04
Teacher variables			
Teacher is female CENTERED	0.64	0.26	0.01
Teacher race is black CENTERED	0.89	0.23	<0.001
Teacher has only partial certification CENTERED	-0.68	0.24	<0.001
School variables			
School percent FRPL CENTERED	-0.02	0.01	0.15
School report card rating – on one scale CENTERED	0.34	0.21	0.10
School percent non-white CENTERED	0.04	0.02	0.01
School percent ELL CENTERED	0.01	0.02	0.57
School percent Special Education CENTERED	0.07	0.06	0.20
Blocking variables			
District A block 3	1.07	0.45	0.02
District A block 2	1.28	0.40	<0.001
Constant	-1.02	0.25	<0.001
Random effects			
School	0.06		
Teacher	0.22		
n			
School	24		
Teacher	33		

Exhibit D-49. Impact of the NTC Model on Communicating with Students – Cohort 1, District A

	Estimate	Standard Error	p-value
Treatment status	0.40	0.36	0.26
Baseline score	0.28	0.05	<0.001
Teacher variables			
Teacher is female CENTERED	0.38	0.11	<0.001
Teacher race is black CENTERED	-0.83	0.09	<0.001
Teacher has only partial certification CENTERED	-0.62	0.11	<0.001
School variables			
School percent FRPL CENTERED	-0.02	0.01	0.16
School report card rating – on one scale CENTERED	<0.001	0.19	0.99
School percent non-white CENTERED	0.05	0.02	0.01
School percent ELL CENTERED	-0.03	0.02	0.19
School percent Special Education CENTERED	0.12	0.07	0.09
Blocking variables			
District A block 3	0.20	0.51	0.69
District A block 2	-0.57	0.54	0.29
Constant	0.16	0.32	0.62
Random effects			
School	0.60		
Teacher	0.02		
n			
School	24		
Teacher	32		

**Exhibit D-50. Impact of the NTC Model on Using Questioning and Discussion Techniques –
Cohort 1, District A**

	Estimate	Standard Error	p-value
Treatment status	0.13	0.25	0.60
Baseline score	-0.14	0.16	0.37
Teacher variables			
Teacher is female CENTERED	-0.07	0.25	0.78
Teacher race is black CENTERED	0.08	0.25	0.77
Teacher has only partial certification CENTERED	-0.56	0.23	0.02
School variables			
School percent FRPL CENTERED	-0.04	0.01	<0.001
School report card rating – on one scale CENTERED	0.27	0.21	0.21
School percent non-white CENTERED	0.07	0.02	<0.001
School percent ELL CENTERED	0.01	0.02	0.68
School percent Special Education CENTERED	0.14	0.06	0.01
Blocking variables			
District A block 3	-1.22	0.53	0.02
District A block 2	0.03	0.42	0.94
Constant	0.51	0.26	0.05
Random effects			
School	<0.001		
Teacher	0.28		
n			
School	24		
Teacher	32		

Exhibit D-51. Impact of the NTC Model on Engaging Students in Learning – Cohort 1, District A

	Estimate	Standard Error	p-value
Treatment status	0.15	0.33	0.64
Baseline score	0.08	0.19	0.69
Teacher variables			
Teacher is female CENTERED	-0.54	0.35	0.12
Teacher race is black CENTERED	0.42	0.33	0.21
Teacher has only partial certification CENTERED	-0.83	0.31	0.01
School variables			
School percent FRPL CENTERED	-0.03	0.02	0.10
School report card rating – on one scale CENTERED	0.36	0.28	0.20
School percent non-white CENTERED	0.05	0.02	0.01
School percent ELL CENTERED	0.02	0.02	0.40
School percent Special Education CENTERED	0.09	0.07	0.22
Blocking variables			
District A block 3	-0.89	0.63	0.16
District A block 2	-0.27	0.54	0.61
Constant	0.46	0.32	0.15
Random effects			
School	<0.001		
Teacher	0.50		
n			
School	24		
Teacher	33		

Exhibit D-52. Impact of the NTC Model on Using Assessment in Instruction – Cohort 1, District A

	Estimate	Standard Error	p-value
Treatment status	0.01	0.30	0.97
Baseline score	-0.02	0.14	0.89
Teacher variables			
Teacher is female CENTERED	-0.51	0.23	0.03
Teacher race is black CENTERED	-0.36	0.19	0.06
Teacher has only partial certification CENTERED	-0.12	0.21	0.56
School variables			
School percent FRPL CENTERED	-0.02	0.02	0.30
School report card rating – on one scale CENTERED	-0.14	0.23	0.54
School percent non-white CENTERED	0.02	0.02	0.24
School percent ELL CENTERED	-0.01	0.02	0.54
School percent Special Education CENTERED	-0.03	0.07	0.67
Blocking variables			
District A block 3	-1.34	0.50	0.01
District A block 2	-1.49	0.48	<0.001
Constant	0.75	0.30	0.01
Random effects			
School	0.30		
Teacher	0.12		
n			
School	24		
Teacher	33		

Cohort 1, District B

Exhibit D-53. Impact of the NTC Model on Creating an Environment of Respect and Rapport – Cohort 1, District B

	Estimate	Standard Error	p-value
Treatment status	0.10	0.30	0.75
Baseline score	-0.22	0.21	0.30
Teacher variables			
Teacher is female CENTERED	0.52	0.48	0.28
Teacher race is black CENTERED	0.26	0.52	0.62
Teacher has only partial certification CENTERED	0.11	0.25	0.65
School variables			
School percent FRPL CENTERED	0.05	0.03	0.13
School report card rating – on one scale CENTERED	0.16	0.13	0.20
School percent non-white CENTERED	-0.10	0.07	0.17
School percent ELL CENTERED	0.01	0.01	0.21
School percent Special Education CENTERED	-0.02	0.03	0.49
Blocking variables			
District B block 5	-1.14	0.51	0.03
District B block 7	-1.75	0.58	<0.001
District B block 8	-0.80	0.35	0.02
District B block 6	-1.21	0.33	<0.001
Constant	1.91	0.36	<0.001
Random effects			
School	<0.001		
Teacher	0.38		
n			
School	33		
Teacher	38		

Exhibit D-54. Impact of the NTC Model on Establishing a Culture for Learning – Cohort 1, District B

	Estimate	Standard Error	p-value
Treatment status	-0.68	0.31	0.03
Baseline score	-0.33	0.09	<0.001
Teacher variables			
Teacher is female CENTERED	0.68	0.54	0.21
Teacher race is black CENTERED	0.62	0.33	0.06
Teacher has only partial certification CENTERED	-0.59	0.82	0.47
School variables			
School percent FRPL CENTERED	0.03	0.03	0.35
School report card rating – on one scale CENTERED	0.16	0.13	0.23
School percent non-white CENTERED	-0.04	0.06	0.54
School percent ELL CENTERED	-0.01	0.01	0.35
School percent Special Education CENTERED	-0.03	0.03	0.24
Blocking variables			
District B block 5	-0.49	0.62	0.43
District B block 7	-0.26	0.82	0.75
District B block 8	0.59	0.61	0.33
District B block 6	-0.91	0.51	0.08
Constant	1.07	0.55	0.05
Random effects			
School	0.50		
Teacher	0.05		
n			
School	32		
Teacher	38		

Exhibit D-55. Impact of the NTC Model on Managing Classroom Procedures – Cohort 1, District B

	Estimate	Standard Error	p-value
Treatment status	0.43	0.36	0.24
Baseline score	0.03	0.18	0.86
Teacher variables			
Teacher is female CENTERED	0.02	0.79	0.98
Teacher race is black CENTERED	-0.30	0.62	0.63
Teacher has only partial certification CENTERED	0.70	0.81	0.39
School variables			
School percent FRPL CENTERED	0.05	0.05	0.27
School report card rating – on one scale CENTERED	-0.11	0.16	0.50
School percent non-white CENTERED	-0.10	0.09	0.27
School percent ELL CENTERED	-0.01	0.01	0.36
School percent Special Education CENTERED	-0.01	0.03	0.71
Blocking variables			
District B block 5	-0.57	0.60	0.34
District B block 7	0.80	1.16	0.49
District B block 8	0.72	0.47	0.13
District B block 6	0.61	0.41	0.14
Constant	0.89	0.51	0.08
Random effects			
School	<0.001		
Teacher	0.51		
n			
School	30		
Teacher	35		

Exhibit D-56. Impact of the NTC Model on Managing Student Behavior – Cohort 1, District B

	Estimate	Standard Error	p-value
Treatment status	-0.32	0.35	0.37
Baseline score	0.32	0.18	0.08
Teacher variables			
Teacher is female CENTERED	0.31	0.62	0.62
Teacher race is black CENTERED	0.13	0.66	0.84
Teacher has only partial certification CENTERED	0.75	0.88	0.40
School variables			
School percent FRPL CENTERED	-0.01	0.04	0.77
School report card rating – on one scale CENTERED	0.18	0.15	0.22
School percent non-white CENTERED	0.03	0.08	0.70
School percent ELL CENTERED	-0.01	0.01	0.65
School percent Special Education CENTERED	<0.001	0.03	0.88
Blocking variables			
District B block 5	-0.68	0.62	0.27
District B block 7	-1.73	0.82	0.04
District B block 8	0.78	0.54	0.15
District B block 6	-0.25	0.45	0.58
Constant	0.86	0.53	0.11
Random effects			
School	0.03		
Teacher	0.58		
n			
School	32		
Teacher	38		

Exhibit D-57. Impact of the NTC Model on Communicating with Students – Cohort 1, District B

	Estimate	Standard Error	p-value
Treatment status	-0.48	0.34	0.16
Baseline score	<0.001	0.19	0.99
Teacher variables			
Teacher is female CENTERED	0.43	0.57	0.46
Teacher race is black CENTERED	0.02	0.66	0.98
Teacher has only partial certification CENTERED	0.10	0.85	0.91
School variables			
School percent FRPL CENTERED	-0.04	0.04	0.38
School report card rating – on one scale CENTERED	0.07	0.15	0.61
School percent non-white CENTERED	0.05	0.08	0.55
School percent ELL CENTERED	<0.001	0.01	0.80
School percent Special Education CENTERED	-0.02	0.02	0.48
Blocking variables			
District B block 5	<0.001	0.61	0.99
District B block 7	-0.49	0.73	0.50
District B block 8	0.86	0.53	0.10
District B block 6	-0.43	0.41	0.29
Constant	0.37	0.52	0.48
Random effects			
School	<0.001		
Teacher	0.54		
n			
School	32		
Teacher	38		

**Exhibit D-58. Impact of the NTC Model on Using Questioning and Discussion Techniques –
Cohort 1, District B**

	Estimate	Standard Error	p-value
Treatment status	-0.19	0.36	0.60
Baseline score	0.37	0.18	0.04
Teacher variables			
Teacher is female CENTERED	-0.48	0.64	0.45
Teacher race is black CENTERED	0.74	0.69	0.29
Teacher has only partial certification CENTERED	-0.58	0.91	0.53
School variables			
School percent FRPL CENTERED	-0.02	0.04	0.65
School report card rating – on one scale CENTERED	0.20	0.17	0.23
School percent non-white CENTERED	0.03	0.09	0.71
School percent ELL CENTERED	0.01	0.01	0.45
School percent Special Education CENTERED	<0.001	0.03	0.92
Blocking variables			
District B block 5	-0.60	0.65	0.36
District B block 7	-1.08	0.90	0.23
District B block 8	-0.11	0.53	0.84
District B block 6	-0.81	0.45	0.07
Constant	0.88	0.56	0.12
Random effects			
School	<0.001		
Teacher	0.66		
n			
School	31		
Teacher	37		

Exhibit D-59. Impact of the NTC Model on Engaging Students in Learning – Cohort 1, District B

	Estimate	Standard Error	p-value
Treatment status	-0.35	0.34	0.30
Baseline score	0.27	0.18	0.13
Teacher variables			
Teacher is female CENTERED	-0.22	0.62	0.72
Teacher race is black CENTERED	0.53	0.64	0.41
Teacher has only partial certification CENTERED	-0.61	0.88	0.49
School variables			
School percent FRPL CENTERED	-0.03	0.04	0.48
School report card rating – on one scale CENTERED	0.15	0.15	0.31
School percent non-white CENTERED	0.07	0.08	0.42
School percent ELL CENTERED	-0.01	0.01	0.50
School percent Special Education CENTERED	-0.03	0.03	0.18
Blocking variables			
District B block 5	0.10	0.68	0.88
District B block 7	-1.19	0.78	0.13
District B block 8	0.84	0.57	0.14
District B block 6	-0.09	0.45	0.83
Constant	0.32	0.58	0.58
Random effects			
School	<0.001		
Teacher	0.58		
n			
School	32		
Teacher	37		

Exhibit D-60. Impact of the NTC Model on Using Assessment in Instruction – Cohort 1, District B

	Estimate	Standard Error	p-value
Treatment status	-0.30	0.30	0.31
Baseline score	0.02	0.12	0.87
Teacher variables			
Teacher is female CENTERED	0.03	0.50	0.96
Teacher race is black CENTERED	1.31	0.78	0.09
Teacher has only partial certification CENTERED	0.61	0.32	0.05
School variables			
School percent FRPL CENTERED	0.01	0.06	0.93
School report card rating – on one scale CENTERED	-0.06	0.17	0.73
School percent non-white CENTERED	-0.07	0.12	0.59
School percent ELL CENTERED	-0.01	0.01	0.34
School percent Special Education CENTERED	-0.06	0.03	0.03
Blocking variables			
District B block 5	-1.19	0.64	0.06
District B block 7	-0.70	1.11	0.53
District B block 8	0.98	0.26	<0.001
District B block 6	-0.40	0.23	0.08
Constant	1.71	0.49	<0.001
Random effects			
School	<0.001		
Teacher	0.43		
n			
School	31		
Teacher	37		

Cohort 2, Both RCT Districts Combined

Exhibit D-61. Impact of the NTC Model on Creating an Environment of Respect and Rapport – Cohort 2, Both RCT Districts Combined

	Estimate	Standard Error	p-value
Treatment status	-0.10	0.25	0.70
Baseline score	-0.05	0.12	0.68
Teacher variables			
Teacher is female CENTERED	-0.27	0.32	0.40
Teacher race is black CENTERED	-0.37	0.29	0.20
Teacher has only partial certification CENTERED	0.14	0.32	0.65
School variables			
School percent FRPL CENTERED	-0.02	0.01	0.05
School report card rating – on one scale CENTERED	-0.08	0.14	0.57
School percent non-white CENTERED	0.02	0.01	0.19
School percent ELL CENTERED	0.01	0.01	0.41
School percent Special Education CENTERED	-0.08	0.03	0.01
Blocking variables			
District A block 3	-0.05	0.34	0.89
District A block 2	0.03	0.49	0.95
District B block 5	-0.01	0.31	0.97
District B block 7	-0.26	0.59	0.65
District B block 8	0.24	0.40	0.54
District B block 6	-0.96	0.79	0.22
Interactions with district			
Centered district indicator	-0.28	0.33	0.40
District interaction: Baseline score	0.24	0.22	0.27
District interaction: Teacher is female	-0.37	0.65	0.56
District interaction: Teacher race is black	0.34	0.59	0.57
District interaction: Teacher has only partial certification	-0.69	0.62	0.27
District interaction: School percent FRPL	-0.03	0.02	0.18
District interaction: School report card rating – on one scale	-0.35	0.29	0.22
District interaction: School percent non-white	0.02	0.03	0.50
District interaction: School percent ELL	0.02	0.02	0.27
District interaction: School percent Special Education	-0.02	0.06	0.79
Constant	0.29	0.18	0.12
Random effects			
School	<0.001		
Teacher	0.73		
n			
School	61		
Teacher	88		

**Exhibit D-62. Impact of the NTC Model on Establishing a Culture for Learning –
Cohort 2, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	-0.14	0.34	0.68
Baseline score	0.07	0.09	0.44
Teacher variables			
Teacher is female CENTERED	-0.28	0.32	0.39
Teacher race is black CENTERED	-0.15	0.25	0.54
Teacher has only partial certification CENTERED	-0.10	0.38	0.79
School variables			
School percent FRPL CENTERED	-0.01	0.01	0.40
School report card rating – on one scale CENTERED	-0.12	0.17	0.48
School percent non-white CENTERED	-0.01	0.02	0.69
School percent ELL CENTERED	-0.01	0.01	0.57
School percent Special Education CENTERED	-0.05	0.03	0.10
Blocking variables			
District A block 3	-0.10	0.56	0.85
District A block 2	0.33	0.60	0.58
District B block 5	-1.16	0.62	0.06
District B block 7	-0.71	0.61	0.24
District B block 8	-1.16	0.54	0.03
District B block 6	-1.36	0.66	0.04
Interactions with district			
Centered district indicator	-1.44	0.47	<0.001
District interaction: Baseline score	0.01	0.18	0.96
District interaction: Teacher is female	1.24	0.64	0.05
District interaction: Teacher race is black	1.65	0.50	<0.001
District interaction: Teacher has only partial certification	0.19	0.74	0.80
District interaction: School percent FRPL	-0.02	0.02	0.28
District interaction: School report card rating – on one scale	-0.39	0.34	0.26
District interaction: School percent non-white	-0.01	0.03	0.87
District interaction: School percent ELL	-0.01	0.02	0.61
District interaction: School percent Special Education	-0.15	0.06	0.01
Constant	0.83	0.26	<0.001
Random effects			
School	0.16		
Teacher	0.74		
n			
School	69		
Teacher	88		

**Exhibit D-63. Impact of the NTC Model on Managing Classroom Procedures –
Cohort 2, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.07	0.23	0.76
Baseline score	0.13	0.12	0.28
Teacher variables			
Teacher is female CENTERED	0.58	0.30	0.05
Teacher race is black CENTERED	-0.16	0.23	0.50
Teacher has only partial certification CENTERED	-0.40	0.35	0.25
School variables			
School percent FRPL CENTERED	-0.02	0.01	0.01
School report card rating – on one scale CENTERED	-0.02	0.11	0.88
School percent non-white CENTERED	0.03	0.01	0.01
School percent ELL CENTERED	0.02	0.01	0.04
School percent Special Education CENTERED	-0.04	0.02	0.07
Blocking variables			
District A block 3	-0.05	0.35	0.89
District A block 2	-0.13	0.46	0.78
District B block 5	-0.45	0.32	0.17
District B block 7	-0.69	0.56	0.22
District B block 8	-0.18	0.45	0.69
District B block 6	-0.78	0.56	0.17
Interactions with district			
Centered district indicator	0.39	0.33	0.24
District interaction: Baseline score	0.19	0.22	0.39
District interaction: Teacher is female	-0.52	0.61	0.39
District interaction: Teacher race is black	-0.28	0.47	0.54
District interaction: Teacher has only partial certification	-0.39	0.67	0.57
District interaction: School percent FRPL	-0.03	0.02	0.15
District interaction: School report card rating – on one scale	-0.29	0.23	0.20
District interaction: School percent non-white	0.04	0.02	0.10
District interaction: School percent ELL	0.03	0.02	0.15
District interaction: School percent Special Education	-0.03	0.05	0.54
Constant	1.05	0.19	<0.001
Random effects			
School	<0.001		
Teacher	0.48		
n			
School	68		
Teacher	80		

**Exhibit D-64. Impact of the NTC Model on Managing Student Behavior –
Cohort 2, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.18	0.24	0.45
Baseline score	0.06	0.09	0.50
Teacher variables			
Teacher is female CENTERED	0.19	0.25	0.45
Teacher race is black CENTERED	-0.16	0.25	0.53
Teacher has only partial certification CENTERED	0.03	0.29	0.92
School variables			
School percent FRPL CENTERED	-0.02	0.01	0.09
School report card rating – on one scale CENTERED	0.03	0.15	0.86
School percent non-white CENTERED	0.03	0.01	0.04
School percent ELL CENTERED	-0.01	0.01	0.42
School percent Special Education CENTERED	-0.04	0.03	0.14
Blocking variables			
District A block 3	-0.97	0.40	0.01
District A block 2	-0.48	0.50	0.34
District B block 5	-0.33	0.41	0.43
District B block 7	-1.00	0.54	0.06
District B block 8	-0.64	0.41	0.12
District B block 6	-2.02	0.57	<0.001
Interactions with district			
Centered district indicator	-0.40	0.34	0.23
District interaction: Baseline score	0.19	0.18	0.31
District interaction: Teacher is female	-0.12	0.50	0.81
District interaction: Teacher race is black	0.41	0.50	0.41
District interaction: Teacher has only partial certification	-0.28	0.56	0.62
District interaction: School percent FRPL	0.02	0.03	0.38
District interaction: School report card rating – on one scale	-0.14	0.30	0.65
District interaction: School percent non-white	-0.02	0.03	0.39
District interaction: School percent ELL	<0.001	0.02	0.91
District interaction: School percent Special Education	-0.05	0.06	0.42
Constant	0.59	0.19	<0.001
Random effects			
School	<0.001		
Teacher	0.69		
n			
School	69		
Teacher	88		

**Exhibit D-65. Impact of the NTC Model on Communicating with Students –
Cohort 2, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.08	0.27	0.76
Baseline score	0.20	0.13	0.11
Teacher variables			
Teacher is female CENTERED	-0.33	0.30	0.27
Teacher race is black CENTERED	-0.39	0.31	0.20
Teacher has only partial certification CENTERED	0.29	0.41	0.49
School variables			
School percent FRPL CENTERED	0.01	0.01	0.49
School report card rating – on one scale CENTERED	0.07	0.17	0.68
School percent non-white CENTERED	-0.02	0.02	0.32
School percent ELL CENTERED	<0.001	0.01	0.82
School percent Special Education CENTERED	<0.001	0.03	0.96
Blocking variables			
District A block 3	0.25	0.42	0.55
District A block 2	0.11	0.44	0.80
District B block 5	-1.28	0.56	0.02
District B block 7	-0.64	0.75	0.39
District B block 8	-1.39	0.61	0.02
District B block 6	-1.64	0.76	0.03
Interactions with district			
Centered district indicator	-1.68	0.45	<0.001
District interaction: Baseline score	0.35	0.25	0.16
District interaction: Teacher is female	1.04	0.59	0.08
District interaction: Teacher race is black	0.51	0.61	0.40
District interaction: Teacher has only partial certification	-0.30	0.80	0.71
District interaction: School percent FRPL	<0.001	0.03	0.94
District interaction: School report card rating – on one scale	-0.23	0.34	0.49
District interaction: School percent non-white	0.01	0.03	0.69
District interaction: School percent ELL	-0.01	0.03	0.83
District interaction: School percent Special Education	0.05	0.07	0.49
Constant	0.83	0.25	<0.001
Random effects			
School	<0.001		
Teacher	0.91		
n			
School	66		
Teacher	87		

**Exhibit D-66. Impact of the NTC Model on Using Questioning and Discussion Techniques –
Cohort 2, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.28	0.28	0.32
Baseline score	-0.05	0.13	0.68
Teacher variables			
Teacher is female CENTERED	0.11	0.26	0.67
Teacher race is black CENTERED	0.04	0.29	0.88
Teacher has only partial certification CENTERED	0.03	0.39	0.94
School variables			
School percent FRPL CENTERED	<0.001	0.01	0.78
School report card rating – on one scale CENTERED	0.07	0.17	0.65
School percent non-white CENTERED	-0.01	0.01	0.58
School percent ELL CENTERED	<0.001	0.01	0.80
School percent Special Education CENTERED	-0.05	0.03	0.10
Blocking variables			
District A block 3	-0.23	0.44	0.60
District A block 2	0.39	0.46	0.40
District B block 5	-0.54	0.54	0.32
District B block 7	-0.99	0.70	0.16
District B block 8	-0.78	0.58	0.18
District B block 6	-1.70	0.74	0.02
Interactions with district			
Centered district indicator	-1.38	0.44	<0.001
District interaction: Baseline score	0.11	0.25	0.65
District interaction: Teacher is female	0.10	0.52	0.85
District interaction: Teacher race is black	0.45	0.58	0.43
District interaction: Teacher has only partial certification	-0.61	0.76	0.42
District interaction: School percent FRPL	-0.01	0.03	0.58
District interaction: School report card rating – on one scale	0.09	0.34	0.79
District interaction: School percent non-white	0.02	0.03	0.59
District interaction: School percent ELL	0.02	0.03	0.46
District interaction: School percent Special Education	0.05	0.06	0.48
Constant	0.97	0.25	<0.001
Random effects			
School	0.43		
Teacher	0.41		
n			
School	69		
Teacher	88		

**Exhibit D-67. Impact of the NTC Model on Engaging Students in Learning –
Cohort 2, Both RCT Districts Combined**

		Standard	
	Estimate	Error	p-value
Treatment status	0.43	0.28	0.12
Baseline score	-0.06	0.10	0.54
Teacher variables			
Teacher is female CENTERED	0.34	0.30	0.25
Teacher race is black CENTERED	-0.28	0.26	0.29
Teacher has only partial certification CENTERED	-0.25	0.26	0.34
School variables			
School percent FRPL CENTERED	-0.02	0.01	0.06
School report card rating – on one scale CENTERED	-0.07	0.16	0.64
School percent non-white CENTERED	0.01	0.01	0.34
School percent ELL CENTERED	0.01	0.01	0.17
School percent Special Education CENTERED	-0.04	0.03	0.17
Blocking variables			
District A block 3	-0.20	0.37	0.59
District A block 2	0.18	0.69	0.80
District B block 5	-0.26	0.40	0.52
District B block 7	-1.29	0.43	<0.001
District B block 8	-1.17	0.34	<0.001
District B block 6	-1.96	0.50	<0.001
Interactions with district			
Centered district indicator	-0.90	0.34	0.01
District interaction: Baseline score	0.16	0.20	0.44
District interaction: Teacher is female	0.29	0.59	0.63
District interaction: Teacher race is black	1.02	0.54	0.06
District interaction: Teacher has only partial certification	0.10	0.55	0.86
District interaction: School percent FRPL	-0.01	0.02	0.75
District interaction: School report card rating – on one scale	-0.30	0.29	0.30
District interaction: School percent non-white	-0.04	0.03	0.14
District interaction: School percent ELL	0.05	0.02	0.01
District interaction: School percent Special Education	0.02	0.06	0.73
Constant	0.83	0.20	<0.001
Random effects			
School	<0.001		
Teacher	0.66		
n			
School	68		
Teacher	87		

**Exhibit D-68. Impact of the NTC Model on Using Assessment in Instruction –
Cohort 2, Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.18	0.30	0.54
Baseline score	0.05	0.11	0.68
Teacher variables			
Teacher is female CENTERED	0.28	0.26	0.29
Teacher race is black CENTERED	-0.42	0.32	0.19
Teacher has only partial certification CENTERED	-0.37	0.30	0.22
School variables			
School percent FRPL CENTERED	<0.001	0.01	0.97
School report card rating – on one scale CENTERED	0.05	0.14	0.74
School percent non-white CENTERED	0.01	0.01	0.51
School percent ELL CENTERED	0.01	0.01	0.52
School percent Special Education CENTERED	-0.03	0.03	0.30
Blocking variables			
District A block 3	-0.69	0.40	0.09
District A block 2	0.43	0.71	0.54
District B block 5	-0.42	0.36	0.24
District B block 7	-1.97	0.70	<0.001
District B block 8	-0.83	0.42	0.05
District B block 6	-3.06	0.71	<0.001
Interactions with district			
Centered district indicator	-1.18	0.32	<0.001
District interaction: Baseline score	0.34	0.24	0.15
District interaction: Teacher is female	-0.28	0.52	0.60
District interaction: Teacher race is black	0.92	0.66	0.16
District interaction: Teacher has only partial certification	0.11	0.61	0.86
District interaction: School percent FRPL	0.03	0.03	0.22
District interaction: School report card rating – on one scale	0.15	0.26	0.58
District interaction: School percent non-white	-0.05	0.03	0.07
District interaction: School percent ELL	0.03	0.03	0.23
District interaction: School percent Special Education	0.02	0.06	0.66
Constant	0.85	0.19	<0.001
Random effects			
School	<0.001		
Teacher	0.84		
n			
School	69		
Teacher	88		

Cohort 2, District A

Exhibit D-69. Impact of the NTC Model on Creating an Environment of Respect and Rapport – Cohort 2, District A

	Estimate	Standard Error	p-value
Treatment status	-0.19	0.35	0.58
Baseline score	0.07	0.15	0.64
Teacher variables			
Teacher is female CENTERED	-0.48	0.45	0.29
Teacher race is black CENTERED	-0.21	0.35	0.54
Teacher has only partial certification CENTERED	-0.15	0.32	0.63
School variables			
School percent FRPL CENTERED	-0.03	0.02	0.04
School report card rating – on one scale CENTERED	-0.24	0.24	0.31
School percent non-white CENTERED	0.03	0.02	0.22
School percent ELL CENTERED	0.02	0.01	0.27
School percent Special Education CENTERED	-0.09	0.04	0.05
Blocking variables			
District A block 3	-0.05	0.34	0.89
District A block 2	0.04	0.50	0.93
Constant	0.19	0.21	0.37
Random effects			
School	<0.001		
Teacher	0.80		
n			
School	33		
Teacher	44		

**Exhibit D-70. Impact of the NTC Model on Establishing a Culture for Learning –
Cohort 2, District A**

	Estimate	Standard Error	p-value
Treatment status	-0.27	0.50	0.59
Baseline score	0.06	0.14	0.68
Teacher variables			
Teacher is female CENTERED	0.34	0.50	0.51
Teacher race is black CENTERED	0.67	0.40	0.09
Teacher has only partial certification CENTERED	0.07	0.39	0.86
School variables			
School percent FRPL CENTERED	-0.02	0.02	0.22
School report card rating – on one scale CENTERED	-0.30	0.29	0.30
School percent non-white CENTERED	-0.01	0.03	0.78
School percent ELL CENTERED	-0.01	0.02	0.53
School percent Special Education CENTERED	-0.12	0.05	0.02
Blocking variables			
District A block 3	-0.06	0.56	0.91
District A block 2	0.31	0.58	0.60
Constant	0.17	0.32	0.60
Random effects			
School	0.49		
Teacher	0.78		
n			
School	33		
Teacher	44		

Exhibit D-71. Impact of the NTC Model on Managing Classroom Procedures – Cohort 2, District A

		Standard	
	Estimate	Error	p-value
Treatment status	-0.37	0.34	0.28
Baseline score	0.20	0.13	0.12
Teacher variables			
Teacher is female CENTERED	0.27	0.24	0.26
Teacher race is black CENTERED	-0.36	0.27	0.19
Teacher has only partial certification CENTERED	-0.43	0.29	0.13
School variables			
School percent FRPL CENTERED	-0.04	0.02	0.02
School report card rating – on one scale CENTERED	-0.10	0.18	0.57
School percent non-white CENTERED	0.06	0.02	<0.001
School percent ELL CENTERED	0.02	0.02	0.24
School percent Special Education CENTERED	-0.08	0.04	0.05
Blocking variables			
District A block 3	-0.12	0.37	0.74
District A block 2	-0.25	0.48	0.61
Constant	1.48	0.23	<0.001
Random effects			
School	<0.001		
Teacher	0.42		
n			
School	32		
Teacher	37		

Exhibit D-72. Impact of the NTC Model on Managing Student Behavior – Cohort 2, District A

	Estimate	Standard Error	p-value
Treatment status	-0.19	0.31	0.54
Baseline score	0.13	0.13	0.32
Teacher variables			
Teacher is female CENTERED	0.13	0.37	0.73
Teacher race is black CENTERED	0.04	0.35	0.91
Teacher has only partial certification CENTERED	0.03	0.28	0.93
School variables			
School percent FRPL CENTERED	-0.01	0.02	0.58
School report card rating – on one scale CENTERED	-0.01	0.24	0.98
School percent non-white CENTERED	0.02	0.02	0.27
School percent ELL CENTERED	-0.02	0.02	0.31
School percent Special Education CENTERED	-0.07	0.05	0.16
Blocking variables			
District A block 3	-1.00	0.36	0.01
District A block 2	-0.47	0.53	0.37
Constant	0.55	0.24	0.02
Random effects			
School	<0.001		
Teacher	0.74		
n			
School	33		
Teacher	44		

Exhibit D-73. Impact of the NTC Model on Communicating with Students – Cohort 2, District A

	Estimate	Standard Error	p-value
Treatment status	0.11	0.38	0.77
Baseline score	0.33	0.18	0.06
Teacher variables			
Teacher is female CENTERED	0.25	0.38	0.52
Teacher race is black CENTERED	-0.10	0.36	0.78
Teacher has only partial certification CENTERED	0.11	0.41	0.80
School variables			
School percent FRPL CENTERED	0.01	0.02	0.67
School report card rating – on one scale CENTERED	-0.06	0.29	0.85
School percent non-white CENTERED	-0.01	0.03	0.71
School percent ELL CENTERED	<0.001	0.02	0.90
School percent Special Education CENTERED	0.03	0.06	0.56
Blocking variables			
District A block 3	0.30	0.46	0.51
District A block 2	0.19	0.48	0.69
Constant	-0.04	0.30	0.89
Random effects			
School	0.18		
Teacher	0.84		
n			
School	33		
Teacher	43		

**Exhibit D-74. Impact of the NTC Model on Using Questioning and Discussion Techniques –
Cohort 2, District A**

	Estimate	Standard Error	p-value
Treatment status	0.01	0.45	0.99
Baseline score	-0.07	0.17	0.67
Teacher variables			
Teacher is female CENTERED	-0.31	0.24	0.21
Teacher race is black CENTERED	0.31	0.22	0.17
Teacher has only partial certification CENTERED	<0.001	0.27	0.99
School variables			
School percent FRPL CENTERED	<0.001	0.02	0.89
School report card rating – on one scale CENTERED	0.11	0.31	0.74
School percent non-white CENTERED	<0.001	0.03	0.99
School percent ELL CENTERED	0.01	0.03	0.71
School percent Special Education CENTERED	-0.03	0.05	0.58
Blocking variables			
District A block 3	-0.33	0.56	0.56
District A block 2	0.24	0.57	0.68
Constant	0.38	0.35	0.27
Random effects			
School	1.21		
Teacher	0.11		
n			
School	33		
Teacher	44		

Exhibit D-75. Impact of the NTC Model on Engaging Students in Learning – Cohort 2, District A

	Estimate	Standard Error	p-value
Treatment status	0.34	0.38	0.37
Baseline score	-0.03	0.18	0.87
Teacher variables			
Teacher is female CENTERED	0.41	0.38	0.29
Teacher race is black CENTERED	0.16	0.36	0.66
Teacher has only partial certification CENTERED	-0.43	0.39	0.26
School variables			
School percent FRPL CENTERED	-0.03	0.02	0.21
School report card rating – on one scale CENTERED	-0.39	0.29	0.19
School percent non-white CENTERED	-0.01	0.03	0.79
School percent ELL CENTERED	0.04	0.02	0.08
School percent Special Education CENTERED	0.02	0.06	0.74
Blocking variables			
District A block 3	-0.15	0.46	0.74
District A block 2	0.15	0.49	0.77
Constant	0.50	0.30	0.09
Random effects			
School	0.35		
Teacher	0.62		
n			
School	32		
Teacher	43		

Exhibit D-76. Impact of the NTC Model on Using Assessment in Instruction – Cohort 2, District A

	Estimate	Standard Error	p-value
Treatment status	0.33	0.45	0.46
Baseline score	0.22	0.16	0.16
Teacher variables			
Teacher is female CENTERED	0.14	0.39	0.72
Teacher race is black CENTERED	0.01	0.40	0.97
Teacher has only partial certification CENTERED	-0.33	0.38	0.39
School variables			
School percent FRPL CENTERED	0.02	0.02	0.43
School report card rating – on one scale CENTERED	0.13	0.25	0.61
School percent non-white CENTERED	-0.02	0.02	0.42
School percent ELL CENTERED	0.03	0.02	0.24
School percent Special Education CENTERED	-0.02	0.05	0.70
Blocking variables			
District A block 3	-0.70	0.44	0.12
District A block 2	0.46	0.72	0.52
Constant	0.20	0.26	0.44
Random effects			
School	<0.001		
Teacher	0.93		
n			
School	33		
Teacher	43		

Cohort 2, District B

Exhibit D-77. Impact of the NTC Model on Creating an Environment of Respect and Rapport – Cohort 2, District B

	Estimate	Standard Error	p-value
Treatment status	0.07	0.36	0.85
Baseline score	-0.16	0.19	0.40
Teacher variables			
Teacher is female CENTERED	-0.08	0.47	0.86
Teacher race is black CENTERED	-0.52	0.46	0.26
Teacher has only partial certification CENTERED	0.45	0.56	0.43
School variables			
School percent FRPL CENTERED	-0.01	0.01	0.65
School report card rating – on one scale CENTERED	0.11	0.15	0.47
School percent non-white CENTERED	0.01	0.01	0.54
School percent ELL CENTERED	<0.001	0.01	0.74
School percent Special Education CENTERED	-0.08	0.04	0.09
Blocking variable			
District B block 5	<0.001	0.30	0.99
District B block 7	-0.39	0.62	0.53
District B block 8	0.16	0.41	0.69
District B block 6	-1.07	0.83	0.19
Constant	0.33	0.32	0.30
Random effects			
School	<0.001		
Teacher	0.66		
n			
School	36		
Teacher	44		

**Exhibit D-78. Impact of the NTC Model on Establishing a Culture for Learning –
Cohort 2, District B**

	Estimate	Standard Error	p-value
Treatment status	-0.04	0.37	0.92
Baseline score	0.07	0.10	0.51
Teacher variables			
Teacher is female CENTERED	-0.87	0.34	0.01
Teacher race is black CENTERED	-1.02	0.28	<0.001
Teacher has only partial certification CENTERED	-0.27	0.61	0.66
School variables			
School percent FRPL CENTERED	<0.001	0.01	0.97
School report card rating – on one scale CENTERED	0.08	0.15	0.62
School percent non-white CENTERED	<0.001	0.02	0.97
School percent ELL CENTERED	<0.001	0.01	0.95
School percent Special Education CENTERED	0.02	0.03	0.48
Blocking variables			
District B block 5	-1.14	0.62	0.07
District B block 7	-0.77	0.60	0.20
District B block 8	-1.24	0.56	0.03
District B block 6	-1.50	0.65	0.02
Constant	1.50	0.40	<0.001
Random effects			
School	<0.001		
Teacher	0.52		
n			
School	36		
Teacher	44		

Exhibit D-79. Impact of the NTC Model on Managing Classroom Procedures – Cohort 2, District B

	Estimate	Standard Error	p-value
Treatment status	0.60	0.25	0.02
Baseline score	0.08	0.21	0.69
Teacher variables			
Teacher is female CENTERED	0.84	0.55	0.12
Teacher race is black CENTERED	0.07	0.35	0.84
Teacher has only partial certification CENTERED	-0.35	0.58	0.55
School variables			
School percent FRPL CENTERED	-0.01	0.01	0.11
School report card rating – on one scale CENTERED	0.19	0.12	0.10
School percent non-white CENTERED	0.02	0.01	0.13
School percent ELL CENTERED	<0.001	0.01	0.79
School percent Special Education CENTERED	-0.05	0.02	0.04
Blocking variables			
District B block 5	-0.50	0.31	0.10
District B block 7	-1.06	0.53	0.04
District B block 8	-0.45	0.41	0.27
District B block 6	-1.24	0.59	0.03
Constant	0.60	0.28	0.03
Random effects			
School	<0.001		
Teacher	0.47		
n			
School	36		
Teacher	43		

Exhibit D-80. Impact of the NTC Model on Managing Student Behavior – Cohort 2, District B

	Estimate	Standard Error	p-value
Treatment status	0.84	0.34	0.01
Baseline score	-0.01	0.13	0.93
Teacher variables			
Teacher is female CENTERED	0.21	0.32	0.50
Teacher race is black CENTERED	-0.30	0.35	0.39
Teacher has only partial certification CENTERED	-0.01	0.40	0.98
School variables			
School percent FRPL CENTERED	-0.03	0.01	0.01
School report card rating – on one scale CENTERED	0.16	0.12	0.16
School percent non-white CENTERED	0.05	0.02	<0.001
School percent ELL CENTERED	-0.01	0.01	0.29
School percent Special Education CENTERED	-0.05	0.03	0.06
Blocking variables			
District B block 5	-0.35	0.39	0.36
District B block 7	-1.45	0.46	<0.001
District B block 8	-0.98	0.36	0.01
District B block 6	-2.52	0.50	<0.001
Constant	0.45	0.28	0.11
Random effects			
School	<0.001		
Teacher	0.58		
n			
School	36		
Teacher	44		

Exhibit D-81. Impact of the NTC Model on Communicating with Students – Cohort 2, District B

	Estimate	Standard Error	p-value
Treatment status	-0.10	0.41	0.81
Baseline score	0.02	0.17	0.89
Teacher variables			
Teacher is female CENTERED	-0.84	0.44	0.05
Teacher race is black CENTERED	-0.67	0.47	0.15
Teacher has only partial certification CENTERED	0.48	0.67	0.47
School variables			
School percent FRPL CENTERED	0.01	0.02	0.56
School report card rating – on one scale CENTERED	0.16	0.18	0.38
School percent non-white CENTERED	-0.02	0.02	0.21
School percent ELL CENTERED	0.01	0.01	0.64
School percent Special Education CENTERED	-0.02	0.04	0.70
Blocking variables			
District B block 5	-1.28	0.53	0.02
District B block 7	-0.52	0.73	0.48
District B block 8	-1.28	0.60	0.03
District B block 6	-1.49	0.76	0.05
Constant	1.75	0.40	<0.001
Random effects			
School	<0.001		
Teacher	0.80		
n			
School	36		
Teacher	44		

**Exhibit D-82. Impact of the NTC Model on Using Questioning and Discussion Techniques –
Cohort 2, District B**

	Estimate	Standard Error	p-value
Treatment status	0.35	0.26	0.17
Baseline score	-0.06	0.11	0.57
Teacher variables			
Teacher is female CENTERED	0.15	0.40	0.70
Teacher race is black CENTERED	-0.32	0.37	0.39
Teacher has only partial certification CENTERED	0.18	0.61	0.77
School variables			
School percent FRPL CENTERED	<0.001	0.01	0.80
School report card rating – on one scale CENTERED	-0.02	0.15	0.89
School percent non-white CENTERED	-0.01	0.01	0.55
School percent ELL CENTERED	-0.01	0.01	0.37
School percent Special Education CENTERED	-0.07	0.02	<0.001
Blocking variables			
District B block 5	-0.58	0.40	0.15
District B block 7	-1.04	0.56	0.06
District B block 8	-0.79	0.40	0.05
District B block 6	-1.97	0.52	<0.001
Constant	1.65	0.37	<0.001
Random effects			
School	<0.001		
Teacher	0.49		
n			
School	36		
Teacher	44		

Exhibit D-83. Impact of the NTC Model on Engaging Students in Learning – Cohort 2, District B

	Estimate	Standard Error	p-value
Treatment status	0.27	0.25	0.27
Baseline score	-0.15	0.13	0.26
Teacher variables			
Teacher is female CENTERED	0.24	0.36	0.51
Teacher race is black CENTERED	-0.79	0.29	0.01
Teacher has only partial certification CENTERED	-0.27	0.45	0.54
School variables			
School percent FRPL CENTERED	-0.01	0.01	0.12
School report card rating – on one scale CENTERED	0.09	0.11	0.44
School percent non-white CENTERED	0.03	0.01	0.01
School percent ELL CENTERED	-0.01	0.01	0.15
School percent Special Education CENTERED	-0.05	0.02	0.04
Blocking variables			
District B block 5	-0.29	0.41	0.48
District B block 7	-1.14	0.43	0.01
District B block 8	-1.08	0.34	<0.001
District B block 6	-1.85	0.50	<0.001
Constant	1.36	0.26	<0.001
Random effects			
School	<0.001		
Teacher	0.37		
n			
School	36		
Teacher	44		

Exhibit D-84. Impact of the NTC Model on Using Assessment in Instruction – Cohort 2, District B

	Estimate	Standard Error	p-value
Treatment status	-0.07	0.37	0.85
Baseline score	-0.12	0.18	0.53
Teacher variables			
Teacher is female CENTERED	0.42	0.37	0.26
Teacher race is black CENTERED	-0.87	0.51	0.09
Teacher has only partial certification CENTERED	-0.34	0.54	0.53
School variables			
School percent FRPL CENTERED	-0.01	0.01	0.33
School report card rating – on one scale CENTERED	-0.06	0.13	0.62
School percent non-white CENTERED	0.03	0.02	0.07
School percent ELL CENTERED	-0.01	0.01	0.66
School percent Special Education CENTERED	-0.03	0.04	0.39
Blocking variables			
District B block 5	-0.37	0.35	0.30
District B block 7	-1.77	0.77	0.02
District B block 8	-0.66	0.48	0.17
District B block 6	-2.78	0.81	<0.001
Constant	1.54	0.28	<0.001
Random effects			
School	<0.001		
Teacher	0.75		
n			
School	36		
Teacher	44		

APPENDIX E. TEACHER RETENTION IMPACT ANALYSIS AND MODEL RESULTS

Appendix E provides details on the analysis of the impact of the NTC induction model on teacher retention.

Analysis

Sample

The teacher retention analysis included all eligible study teachers with no restrictions. This sample is shown in Exhibit C-3.

Data and Measures

Teacher retention in the district was measured using human resources (HR) data provided by the districts. Teachers were counted as “retained” if they were still employed by the district 2 years after the start of the study. Teachers in Cohort 1 were retained if they were still employed by the district as of October 1, 2015. Teachers in Cohort 2 were retained if they were still employed by the district as of October 1, 2016. Districts were not able to provide consistent data on the job description of the study participants. Therefore, we cannot say definitively whether study teachers were retained in instructional positions, or in other types of positions within the district. Therefore, this analysis measured the impact of the NTC model on retention of teachers in any position within the district. Districts and program staff were able to provide information on reductions-in-force (RIFs) or layoffs due to budget cuts. This analysis was designed to exclude these teachers, as the induction program cannot impact attrition due to layoffs and beginning teachers tend to be disproportionately affected by layoffs. However, no budget-based RIFs or layoffs were identified in the data in either year.

Attrition

Because the districts were able to provide HR data for all study teachers, they were all included in the retention analysis, with zero attrition.

Baseline equivalence

We are unable to provide data on the baseline equivalence of the teacher retention samples on the same metric as the outcome because all study teachers were newly hired at the beginning of the study (i.e., they could not be retained in the district before they were hired). What Works Clearinghouse (WWC) guidance on other potential measures to use for baseline equivalence on this outcome was not yet available at the time this study was designed, and we do not have any baseline data that WWC categorizes as acceptable for this outcome. However, as there was zero attrition in the analytic sample and schools were randomly assigned to treatment and control, baseline equivalence is not required.

Statistical Analysis

To analyze the impact of the NTC model on teacher practice, we estimated the following two-level logistic regression model, with teachers nested within schools:

Level 1 (Teachers): $y_{jk}^* = \beta_{0k} + \beta_1 Treatment_{jk} + \beta_{2k} Z_{jk} + \beta_{3k} Coh_{jk} + \beta_{4k} Z_{jk} Coh_{jk} + r_{jk}$

Level 2 (Schools): $\beta_{0k} = \gamma_{00} + \gamma_{01} W_k + \gamma_{02} Dist_k + \gamma_{03} W_k Dist_k + e_k$

$$\beta_{1k} = \gamma_{10}$$

$$\beta_{2k} = \gamma_{20} + \gamma_{21} Dist_k$$

$$\beta_{3k} = \gamma_{30} + \gamma_{31} W_k + \gamma_{32} Dist_k + \gamma_{33} W_k Dist_k$$

$$\beta_{4k} = \gamma_{40} + \gamma_{41} Dist_k$$

Mixed:

$$y_{jk} = \gamma_{00} + \gamma_{10}Treatment_{jk} + \gamma_{01}W_k + \gamma_{20}Z_{jk} + \gamma_{02}Dist_k + \gamma_{30}Coh_{jk} + \gamma_{03}W_kDist_k \\ + \gamma_{21}Z_{jk}Dist_k + \gamma_{31}W_kCoh_{jk} + \gamma_{40}Z_{jk}Coh_{jk} + \gamma_{32}Dist_kCoh_{jk} \\ + \gamma_{33}W_kCoh_{jk}Dist_k + \gamma_{41}Z_{jk}Coh_{jk}Dist_k + r_{jk} + e_k$$

In this model, y_{jk}^* represents the underlying latent probability that a teacher will be retained. The γ_{10} coefficient estimates the impact of NTC on teacher retention among new teachers. Z_{jk} represents a vector of teacher-level controls, W_k represents a vector of school-level controls, Coh_{jk} is a cohort indicator, and $Dist_k$ is a district indicator. All variables were centered to the analysis sample. The model also included interactions with district, cohort, and a district-by-cohort three-way interaction for each teacher and school control variable. We controlled for school- and teacher-level covariates to account for pre-existing differences between treatment and control that may be associated with teacher retention, and to increase the precision of our estimates. Although treatment was randomly assigned at the school level, there may have been remaining differences between the groups due to chance.

We included cohort and district effects as well as interactions between the covariates and these effects for three reasons. First, we expected that the two districts would have different contexts and therefore the NTC intervention may have a different impact in each district. For that reason, we planned to run separate models for each district as well as a combined model for both districts. Second, we expected that the implementation of the program may improve over time, such that the impact in the two cohorts would be different. For that reason, we planned to run separate models for each cohort. Third, in both districts the context changed substantially over the 3 years of the study, which we expected may affect the outcomes of the two cohorts differently. This strengthened our rationale for including a cohort effect.

Results

The overall proportion of teachers retained in their districts was similar among the treatment and control teachers (79.2 versus 78.7 percent). However, treatment teachers were retained at slightly higher rates in Cohort 1 (83.4 percent versus 76.9 percent) and slightly lower rates in Cohort 2 (75.7 percent versus 81.0 percent) overall (Exhibit -E-1).

Exhibit E-1. Descriptive Differences in Teacher Retention Between Treatment and Control

		Retention into Y3			n Teachers		n Schools	
		Control	Treatment	Difference	Control	Treatment	Control	Treatment
Overall		78.7 (41.0)	79.2 (40.6)	0.5	287	342	119	108
Cohorts Combined	District A	85.8 (35.0)	82.4 (38.2)	-3.4	148	193	44	43
	District B	71.2 (45.4)	75.2 (43.3)	4.0	139	149	75	65

Exhibit E-1. Descriptive Differences in Teacher Retention Between Treatment and Control (concluded)

		Retention into Y3			<i>n</i> Teachers		<i>n</i> Schools	
		Control	Treatment	Difference	Control	Treatment	Control	Treatment
Cohort 1	Overall	76.9 (42.3)	83.4 (37.3)	6.5	156	157	69	69 ⁹
	District A	85.2 (35.7)	87.6 (33.1)	2.4	88	89	35	32
	District B	66.2 (47.7)	77.9 (41.8)	11.7	68	68	34	37
Cohort 2	Overall	81.0 (39.4)	75.7 (43.0)	-5.3	131	185	82	86
	District A	86.7 (34.3)	77.9 (41.7)	-8.8	60	104	26	38
	District B	76.1 (43.0)	72.8 (44.8)	-3.3	71	81	56	48

None of the differences in retention rates between treatment and control teachers was statistically significant when estimated using a multi-level logistic regression and accounting for differences between teachers and schools. However, the direction of the coefficients remains the same as in the descriptive analysis (Exhibit E-2). Full model tables are included below.

Exhibit E-2. The Impact of the NTC Model on Teacher Retention in the District

		Estimate	Standard Error	<i>p</i> -Value	<i>n</i> Teachers	<i>n</i> Schools
Cohorts Combined	Overall	0.04	0.24	0.88	629	227
	District A	-0.30	0.35	0.39	341	87
	District B	0.34	0.39	0.39	288	140
Cohort 1	Overall	0.86	0.43	0.05	313	138
	District A	0.78	0.57	0.18	177	67
	District B	1.04	0.69	0.13	136	71
Cohort 2	Overall	-0.46	0.32	0.15	316	168
	District A	-1.00	0.50	0.05	164	64
	District B	-0.16	0.51	0.76	152	104

Sensitivity to Teacher Joiners

WWC standards address the issue of participants joining the study sample after randomization. These “joiners” may bias the study findings if they are substantially different from the participants who were present at the time of randomization, particularly if they chose to join the school because of the presence of the NTC intervention or study. In this study we could identify when teachers joined the study schools.

Under WWC standards version 3.0, joiners are divided into “late” and “early” joiners. Teachers who were hired or began their employment at a study school shortly after the randomization can be categorized as early joiners. We may reasonably assume early joiners did not choose to work at the

⁹ Note that schools may be in either Cohort 1 or Cohort 2 or both (if they hired first year teachers in both 2013-14 and 2014-15). Therefore, the number of Cohort 1 schools and the number of Cohort 2 schools does not add up to the total number of schools.

school as a result of the NTC intervention or study, which was put into place only days or weeks before. Teachers who were hired or began their employment at the school substantially after randomization are categorized as late joiners. In this study, because all teachers who were hired after October 1, 2013, were excluded from Cohort 1, and all teachers who were hired after October 1, 2014, were excluded from Cohort 2, late joiners include only Cohort 2 teachers who joined existing study schools in the 2014–15 school year. These teachers may have known about the induction program in place in the school and chosen to apply and take their position as a result.

To test the sensitivity of the teacher retention results to the inclusion of joiners, we fit models similar to those estimated for the main treatment effect reported above, but excluding Cohort 2 teachers who joined the study schools that were randomized in 2013–14 under Cohort 1. These models also exclude the district- and cohort-level interaction terms, as the sample is reduced to the point of making estimating these fully interacted models impossible.

We found that excluding late joiner teachers slightly increased the estimate of the impact of the NTC model on teacher retention, but the difference remained statistically insignificant (Exhibit E-3). When late joiners are excluded, approximately 81 percent of treatment teachers were retained into their third year of teaching, compared to 79 percent of control teachers.

Exhibit E-3. Impact of the NTC Model on Teacher Retention after Two Years of Induction Support, Excluding Late Joiner Teachers

	Estimate	Standard Error	<i>p</i> -value	<i>n</i> schools	<i>n</i> teachers
Original estimate	0.04	0.24	0.88	227	629
Excluding late joiner teachers	0.25	0.30	0.40	227	445

* $p < 0.05$

Full Model Tables¹⁰

Exhibit E-4. Impact of the NTC Model on Teacher Retention, Both Cohorts and RCT Districts Combined

	Estimate	Standard Error	p-value
Treatment status	0.038	0.242	0.877
Teacher controls			
Teacher highest degree is Bachelor's CENTERED	0.034	0.272	0.902
Teacher is only partially certified CENTERED	3.237	347.4	0.993
Teacher is female CENTERED	0.098	0.296	0.741
Teacher race is white CENTERED	-0.457	0.257	0.076
School controls			
School percent ELL CENTERED	-0.023	0.015	0.131
School percent IEP CENTERED	-0.079	0.036	0.030
School percent FRPL CENTERED	0.002	0.016	0.913
School percent non-white CENTERED	0.006	0.019	0.740
School report card rating CENTERED	0.040	0.160	0.804
Blocking variables			
District A block 1	0.385	0.843	0.648
District A block 2	-0.589	0.635	0.354
District A block 4	-0.882	1.004	0.379
District A block 5	0.084	0.923	0.927
District B block 9	-0.593	0.970	0.541
District B block 7	-0.504	0.793	0.525
District B block 8	-0.057	0.731	0.938
District B block 2	0.728	0.568	0.200
District B block 6	8.813	1029	0.993
District A block 6	-0.113	0.473	0.811
District interactions			
Centered district indicator	-1.031	185.3	0.996
District interaction: Teacher highest degree is Bachelor's	1.061	0.548	0.053
District interaction: Teacher is only partially certified	-8.671	694.9	0.990
District interaction: Teacher is female	0.405	0.592	0.494
District interaction: Teacher race is white	0.156	0.523	0.765
District interaction: School percent ELL	-0.043	0.031	0.161
District interaction: School percent IEP	-0.062	0.073	0.400
District interaction: School percent FRPL	-0.007	0.032	0.834
District interaction: School percent non-white	-0.011	0.038	0.773
District interaction: School report card rating	-0.340	0.319	0.286
Cohort interactions			
Centered cohort indicator	2.692	185.3	0.988
Cohort interaction: Teacher highest degree is Bachelor's	0.582	0.545	0.286
Cohort interaction: Teacher is only partially certified	9.378	694.9	0.989
Cohort interaction: Teacher is female	0.914	0.590	0.121
Cohort interaction: Teacher race is white	0.391	0.519	0.451
Cohort interaction: School percent ELL	0.037	0.030	0.217
Cohort interaction: School percent IEP	0.131	0.072	0.069
Cohort interaction: School percent FRPL	0.023	0.031	0.468

¹⁰ Blocking variable names are being kept consistent across model results, which resulted in some of the blocking variables appearing out of order in Appendix B exhibits.

**Exhibit E-4. Impact of the NTC Model on Teacher Retention, Both Cohorts and RCT Districts Combined
(concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: School percent non-white	-0.023	0.037	0.527
Cohort interaction: School report card rating	0.306	0.311	0.326
Cohort interaction: District A block 1	-3.787	1.653	0.022
Cohort interaction: District A block 2	-0.374	1.245	0.764
Cohort interaction: District A block 4	-0.874	1.940	0.652
Cohort interaction: District A block 5	-1.583	1.771	0.371
Cohort interaction: District B block 9	-1.379	1.869	0.461
Cohort interaction: District B block 7	1.415	1.579	0.370
Cohort interaction: District B block 8	1.108	1.428	0.438
Cohort interaction: District B block 2	-1.292	1.117	0.248
Cohort interaction: District B block 6	-17.43	2058	0.993
Cohort interaction: District A block 6	0.507	0.926	0.584
District-by-cohort interactions			
District by cohort interaction	-3.185	370.6	0.993
District by cohort interaction: Teacher highest degree is Bachelor's	-0.432	1.090	0.692
District by cohort interaction: Teacher is only partially certified	-17.45	1390	0.990
District by cohort interaction: Teacher is female	0.440	1.182	0.710
District by cohort interaction: Teacher race is white	-0.751	1.023	0.463
District by cohort interaction: School percent ELL	-0.010	0.060	0.871
District by cohort interaction: School percent IEP	0.218	0.145	0.134
District by cohort interaction: School percent FRPL	0.046	0.062	0.460
District by cohort interaction: School percent non-white	0.043	0.073	0.561
District by cohort interaction: School report card rating	1.542	0.624	0.013
Constant	2.630	92.65	0.977
Random effects			
School	0.175	0.312	0.574
n			
Schools	227		
Teachers	629		

**Exhibit E-5. Impact of the NTC Model on Teacher Retention, Both Cohorts Combined –
District A**

	Estimate	Standard Error	p-value
Treatment status	-0.297	0.348	0.394
Teacher controls			
Teacher highest degree is Bachelor's CENTERED	0.587	0.405	0.147
Teacher is only partially certified CENTERED	-1.040	0.367	0.005
Teacher is female CENTERED	0.244	0.396	0.538
Teacher race is white CENTERED	-0.348	0.363	0.339
School controls			
School percent ELL CENTERED	-0.046	0.027	0.092
School percent IEP CENTERED	-0.109	0.063	0.086
School percent FRPL CENTERED	-0.003	0.025	0.905
School percent non-white CENTERED	0.003	0.028	0.906
School report card rating CENTERED	-0.095	0.276	0.731
Blocking variables			
District A block 1	0.624	0.750	0.405
District A block 2	-0.372	0.539	0.490
District A block 4	-0.509	0.876	0.561
District A block 5	0.267	0.776	0.731
Cohort interactions			
Centered cohort indicator	0.752	1.018	0.460
Cohort interaction: Teacher highest degree is Bachelor's	0.391	0.806	0.628
Cohort interaction: Teacher is only partially certified	0.743	0.735	0.312
Cohort interaction: Teacher is female	1.151	0.797	0.149
Cohort interaction: Teacher race is white	-0.032	0.726	0.965
Cohort interaction: School percent ELL	0.023	0.055	0.672
Cohort interaction: School percent IEP	0.239	0.127	0.059
Cohort interaction: School percent FRPL	0.048	0.049	0.325
Cohort interaction: School percent non-white	-0.010	0.054	0.851
Cohort interaction: School report card rating	1.119	0.549	0.042
Cohort interaction: District A block 1	-3.472	1.503	0.021
Cohort interaction: District A block 2	-0.227	1.085	0.834
Cohort interaction: District A block 4	-0.438	1.747	0.802
Cohort interaction: District A block 5	-1.405	1.560	0.368
Constant	2.035	0.550	0.000
Random effects			
School	<0.001		
n			
Schools	87		
Teachers	341		

**Exhibit E-6. Impact of the NTC Model on Teacher Retention, Both Cohorts Combined –
District B**

	Estimate	Standard Error	p-value
Treatment status	0.335	0.393	0.394
Teacher Controls			
Teacher highest degree is Bachelor's CENTERED	-0.552	0.408	0.176
Teacher is only partially certified CENTERED	9.042	4430	0.998
Teacher is female CENTERED	-0.118	0.472	0.803
Teacher race is white CENTERED	-0.642	0.424	0.130
School controls			
School percent ELL CENTERED	-0.006	0.015	0.693
School percent IEP CENTERED	-0.064	0.035	0.072
School percent FRPL CENTERED	0.011	0.023	0.628
School percent non-white CENTERED	0.005	0.029	0.853
School report card rating CENTERED	0.221	0.160	0.166
Blocking variables			
District B block 9	-0.665	0.959	0.488
District B block 7	-0.292	0.903	0.746
District B block 8	0.768	0.669	0.251
District B block 2	10.87	1216	0.993
District B block 6	-0.168	0.576	0.771
Cohort interactions			
Centered cohort indicator	5.170	2362	0.998
Cohort interaction: Teacher highest degree is Bachelor's	0.749	0.808	0.354
Cohort interaction: Teacher is only partially certified	21.97	8859	0.998
Cohort interaction: Teacher is female	0.815	0.937	0.385
Cohort interaction: Teacher race is white	0.980	0.814	0.228
Cohort interaction: School percent ELL	0.049	0.029	0.087
Cohort interaction: School percent IEP	0.035	0.064	0.580
Cohort interaction: School percent FRPL	-0.009	0.042	0.834
Cohort interaction: School percent non-white	-0.042	0.051	0.411
Cohort interaction: School report card rating	-0.527	0.301	0.080
Cohort interaction: District B block 9	1.835	1.883	0.330
Cohort interaction: District B block 7	1.039	1.652	0.529
Cohort interaction: District B block 8	-1.527	1.255	0.224
Cohort interaction: District B block 2	-21.48	2433	0.993
Cohort interaction: District B block 6	0.448	1.071	0.676
Constant	3.620	1181	0.998
Random Effects			
School	1.024		
n			
Schools	140		
Teachers	288		

**Exhibit E-7. Impact of the NTC Model on Teacher Retention, Cohort 1 Teachers Only –
Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	0.862	0.429	0.045
Teacher controls			
Teacher highest degree is Bachelor's CENTERED	-0.345	0.438	0.431
Teacher is only partially certified CENTERED	-1.586	0.615	0.010
Teacher is female CENTERED	-0.180	0.487	0.712
Teacher race is white CENTERED	-0.610	0.420	0.147
School controls			
School percent ELL CENTERED	-0.055	0.027	0.042
School percent IEP CENTERED	-0.181	0.073	0.012
School percent FRPL CENTERED	<0.001	0.028	0.990
School percent non-white CENTERED	0.005	0.034	0.873
School report card rating CENTERED	-0.083	0.276	0.764
Blocking variables			
District A block 1	1.773	1.346	0.188
District A block 2	-0.484	0.977	0.620
District A block 4	-0.379	1.558	0.808
District A block 5	1.333	1.509	0.377
District B block 9	-1.384	1.339	0.301
District B block 7	-0.848	1.161	0.465
District B block 8	1.416	1.056	0.180
District B block 2	15.78	780.7	0.984
District B block 6	-0.526	0.796	0.509
District A block 6	0.395	1.696	0.816
Interactions with district			
Centered district indicator	0.372	1.128	0.741
District interaction: Teacher highest degree is Bachelor's	1.035	0.877	0.238
District interaction: Teacher is only partially certified	0.207	1.193	0.862
District interaction: Teacher is female	0.643	0.976	0.510
District interaction: Teacher race is white	0.741	0.826	0.370
District interaction: School percent ELL	-0.062	0.053	0.246
District interaction: School percent IEP	-0.214	0.139	0.124
District interaction: School percent FRPL	0.010	0.055	0.850
District interaction: School percent non-white	-0.092	0.071	0.191
District interaction: School report card rating	-1.133	0.554	0.041
Constant	0.902	0.603	0.135
Random effects			
School	0.795		
n			
Schools	138		
Teachers	313		

Exhibit E-8. Impact of the NTC Model on Teacher Retention, Cohort 1 Teachers Only – District A

	Estimate	Standard Error	p-value
Treatment status	0.776	0.574	0.176
Teacher controls			
Teacher highest degree is Bachelor's CENTERED	0.173	0.596	0.772
Teacher is only partially certified CENTERED	-1.446	0.559	0.010
Teacher is female CENTERED	0.156	0.570	0.784
Teacher race is white CENTERED	-0.201	0.543	0.711
School controls			
School percent ELL CENTERED	-0.079	0.045	0.079
School percent IEP CENTERED	-0.263	0.114	0.021
School percent FRPL CENTERED	-0.001	0.039	0.974
School percent non-white CENTERED	-0.030	0.044	0.500
School report card rating CENTERED	-0.658	0.445	0.139
Blocking variables			
District A block 1	1.544	1.087	0.156
District A block 2	-0.489	0.831	0.556
District A block 4	-0.389	1.336	0.771
District A block 5	1.201	1.223	0.326
Constant	1.118	0.796	0.160
Random effects			
School	0.273		
n			
Schools	67		
Teachers	177		

Exhibit E-9. Impact of the NTC Model on Teacher Retention, Cohort 1 Teachers Only – District B

	Estimate	Standard Error	p-value
Treatment status	1.042	0.689	0.130
Teacher controls			
Teacher highest degree is Bachelor's CENTERED	-0.875	0.684	0.201
Teacher is only partially certified CENTERED	-1.976	1.221	0.106
Teacher is female CENTERED	-0.507	0.827	0.540
Teacher race is white CENTERED	-1.197	0.728	0.100
School controls			
School percent ELL CENTERED	-0.029	0.026	0.257
School percent IEP CENTERED	-0.090	0.066	0.174
School percent FRPL CENTERED	-0.006	0.042	0.883
School percent non-white CENTERED	0.060	0.058	0.304
School report card rating CENTERED	0.551	0.293	0.060
Blocking variables			
District B block 9	-1.613	1.579	0.307
District B block 7	-0.968	1.386	0.485
District B block 8	1.561	1.224	0.202
District B block 2	17.122	1098.496	0.988
District B block 6	-0.572	0.944	0.544
Constant	0.750	0.744	0.313
Random effects			
School	1.77		
n			
Schools	71		
Teachers	136		

**Exhibit E-10. Impact of the NTC Model on Teacher Retention, Cohort 2 Teachers Only –
Both RCT Districts Combined**

	Estimate	Standard Error	p-value
Treatment status	-0.464	0.324	0.152
Teacher controls			
Teacher highest degree is Bachelor's CENTERED	0.378	0.331	0.254
Teacher is only partially certified CENTERED	9.012	1441.436	0.995
Teacher is female CENTERED	0.483	0.355	0.173
Teacher race is white CENTERED	-0.420	0.334	0.209
School controls			
School percent ELL CENTERED	-0.000	0.017	0.988
School percent IEP CENTERED	-0.008	0.028	0.776
School percent FRPL CENTERED	0.009	0.017	0.591
School percent non-white CENTERED	0.002	0.020	0.932
School report card rating CENTERED	0.249	0.189	0.188
Blocking variables			
District A block 1	-1.579	1.016	0.120
District A block 2	-1.122	0.883	0.204
District A block 4	-1.457	1.226	0.235
District A block 5	-1.282	1.129	0.256
District B block 9	-1.309	1.206	0.278
District B block 7	0.277	1.043	0.791
District B block 8	0.795	1.045	0.447
District B block 2	0.190	0.591	0.747
District B block 6	0.128	1.284	0.920
District A block 6	0.308	0.639	0.630
District interactions			
District indicator	-2.877	768.768	0.997
District interaction: Teacher highest degree is Bachelor's	0.888	0.669	0.184
District interaction: Teacher is only partially certified	-19.599	2882.873	0.995
District interaction: Teacher is female	0.499	0.707	0.480
District interaction: Teacher race is white	-0.329	0.684	0.631
District interaction: School percent ELL	-0.046	0.035	0.194
District interaction: School percent IEP	0.040	0.056	0.471
District interaction: School percent FRPL	0.017	0.034	0.617
District interaction: School percent non-white	0.017	0.041	0.671
District interaction: School report card rating	0.532	0.378	0.159
Constant	4.622	384.382	0.990
Random effects			
School	0.011		
n			
Schools	168		
Teachers	316		

Exhibit E-11. Impact of the NTC Model on Teacher Retention, Cohort 2 Teachers Only – District A

	Estimate	Standard Error	p-value
Treatment status	-0.995	0.501	0.047
Teacher controls			
Teacher highest degree is Bachelor's CENTERED	0.901	0.495	0.069
Teacher is only partially certified CENTERED	-0.714	0.489	0.144
Teacher is female CENTERED	0.760	0.540	0.159
Teacher race is white CENTERED	-0.550	0.459	0.231
School controls	-0.032	0.033	0.325
School percent ELL CENTERED	0.012	0.055	0.827
School percent IEP CENTERED	0.016	0.028	0.555
School percent FRPL CENTERED	0.012	0.032	0.717
School percent non-white CENTERED	0.608	0.356	0.088
School report card rating CENTERED			
Blocking variables	-1.107	0.902	0.220
District A block 1	-0.703	0.763	0.356
District A block 2	-0.708	1.101	0.520
District A block 4	-0.982	1.034	0.342
District A block 5	3.047	0.832	0.000
Constant	-0.995	0.501	0.047
Random effects			
School	<0.001		
n			
Schools	64		
Teachers	164		

Exhibit E-12. Impact of the NTC Model on Teacher Retention, Cohort 2 Teachers Only – District B

	Estimate	Standard Error	p-value
Treatment status	-0.156	0.510	0.759
Teacher controls			
Teacher highest degree is Bachelor's CENTERED	-0.191	0.527	0.716
Teacher is only partially certified CENTERED	14.233	1189.575	0.990
Teacher is female CENTERED	0.258	0.527	0.624
Teacher race is white CENTERED	-0.231	0.553	0.677
School controls			
School percent ELL CENTERED	0.024	0.018	0.179
School percent IEP CENTERED	-0.041	0.038	0.279
School percent FRPL CENTERED	0.006	0.024	0.804
School percent non-white CENTERED	-0.015	0.029	0.614
School report card rating CENTERED	-0.049	0.195	0.801
Blocking variables	0.324	1.269	0.799
District B block 9	0.675	1.256	0.591
District B block 7	0.133	0.715	0.852
District B block 8	0.049	1.448	0.973
District B block 2	0.208	0.770	0.787
District B block 6	4.831	317.220	0.988
Constant			
Random effects			
School	0.943		
n			
Schools	104		
Teachers	152		

APPENDIX F. STUDENT ACHIEVEMENT MODEL RESULTS FOR RCT DISTRICTS

Appendix F provides the details on the analysis of the impact of the NTC induction model on student achievement

Sample

The student achievement analysis sample included all beginning teachers in treatment and control schools who taught reading/ELA and/or mathematics in grades 4 through 8. Special education teachers who taught reading/ELA and/or mathematics and who could be linked to students in district data sets were included in the analysis, along with regular education teachers.¹¹ As in the teacher outcomes analysis, the student outcomes combine teachers and students across cohorts.

As shown in Exhibit F-1, the students in this analysis were predominantly African American (37 to 45 percent) and Hispanic (35 to 49 percent). They were low-income, with close to 90 percent qualifying for free or reduced-price lunch. The teachers of these students were similar to the study sample, predominantly white and female.

Exhibit F-1. Descriptive Statistics for Students in the Achievement Analysis

	Mathematics		Reading/ELA	
	Treatment	Control	Treatment	Control
Student Demographics				
Grade 4	24%	16%	14%	15%
Grade 5	28%	19%	21%	19%
Grade 6	10%	24%	33%	19%
Grade 7	14%	26%	18%	29%
Grade 8	24%	16%	13%	18%
Student Race - African American	37%	45%	48%	39%
Student Race – Hispanic	49%	40%	35%	46%
Student Race – White	6%	9%	8%	9%
Student Race – Asian	5%	5%	7%	4%
Student Race - Native American	0%	0%	0%	0%
Free/reduced-price meals	90%	87%	89%	86%
Special education/IEP	19%	19%	18%	18%
English learner	16%	12%	12%	11%
Student is Female	48%	49%	48%	49%

¹¹ In District A, 2 percent of the teachers in the reading sample and 0 percent of teachers in the mathematics sample were listed as belonging to a special education department. In District B, 16 percent of teachers in the reading sample were listed as special education teachers, as were 17 percent of teachers in the mathematics sample.

Exhibit F-1. Descriptive Statistics for Students in the Achievement Analysis (concluded)

	Mathematics		Reading/ELA	
	Treatment	Control	Treatment	Control
Teacher Demographics				
Teacher Race - African American	18%	23%	30%	34%
Teacher Race – Hispanic	32%	10%	11%	14%
Teacher Race – White	38%	62%	48%	49%
Teacher Race – Asian	2%	1%	8%	1%
Teacher Race - Native American	4%	0%	2%	0%
Teacher Race – Other	6%	1%	1%	1%
Highest degree – Associates	0%	0%	0%	0%
Highest degree – Bachelors	77%	60%	78%	55%
Highest degree – Masters	23%	40%	21%	45%
Not fully certified	9%	25%	24%	12%
Teacher is Female	72%	77%	82%	88%
School Demographics				
Percent free/reduced-price meals	84%	84%	84%	81%
Report Card Rating ¹²	2.2	2.2	2.6	2.2
Percent non-white	92%	90%	91%	88%
Percent English learners	29%	17%	22%	20%
Percent IEP	12%	12%	12%	11%
Blocking variables¹³				
District A block 1	7%	7%	12%	6%
District A block 3	0%	0%	0%	4%
District A block 2	10%	20%	20%	18%
District A block 4	9%	3%	4%	2%
District A block 5	5%	16%	10%	10%
District A block 6	1%	2%	0%	1%
District B block 9	0%	3%	0%	1%
District B block 1	28%	27%	30%	23%
District B block 5	2%	5%	2%	6%
District B block 7	24%	4%	8%	11%
District B block 8	0%	2%	0%	1%
District B block 2	13%	7%	13%	13%
District B block 6	0%	0%	0%	0%
District B block 10	0%	2%	0%	1%
District B block 3	0%	4%	0%	4%
District B block 4	0%	0%	0%	0%
<i>n</i> Students	2,348	2,719	3,434	3,216

¹² 0 = F in BCPS and 3 in CPS; 1 = D in BCPS; 2 = C in BCPS & 2 in CPS; 3 = B in BCPS; 4 = A in BCPS and 1 in CPS.

¹³ Blocking variable names are being kept consistent across model results, which resulted in some of the blocking variables appearing out of order.

Data and Measures

Student achievement was measured using the state standardized test in each district. BCPS used the Florida Comprehensive Assessment Test (FCAT) throughout the study period. CPS, however, began transitioning from the Illinois State Achievement Test (ISAT) to the Partnership for Assessment of Readiness for College and Careers (PARCC) assessments during the 2014–15 school year, the second year of the intervention. The transition was marked by controversy, and not all schools required students to take the PARCC assessment. In the meantime, students were required to take the Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) assessment. Therefore, the study used the ISAT as the outcome and baseline measure in the first year of the intervention (2013–14) and the MAP as the outcome and baseline measure in subsequent years.

Exhibit F-2 shows the mean and standard deviation of the mathematics test scores in each year, combining cohorts. The number of students for each scale score mean varied because the scores were used for different analyses—the 2012–13 scores were used as baseline for Cohort 1, so the mean included only one cohort of students. The 2013–14 scores were used as outcomes for Cohort 1 Year 1 and baseline for Cohort 2 Year 1, so the mean included two cohorts of students. Similarly, the 2014–15 scores were used as outcomes for Cohort 1 Year 2 and Cohort 2 Year 1, as well as baseline for Cohort 2 Year 2, so the mean included two cohorts of students. Finally, the 2015–16 scores were used as outcomes for Cohort 2 Year 2, so the mean included only one cohort of students. Likewise, the number of students in Exhibit F-2 does not align with the number in Exhibit F-4 because they combine cohorts differently. Exhibit F-4 combines Cohort 1 in Year 2 with Cohort 2 in Year 2, with Year 2 of each cohort being the outcome year for the 2-year NTC mentoring program.

Exhibit F-2 shows that the overall standardized scores had a mean near zero with a standard deviation near one. The slight deviation from these exact numbers was due to students who were used in the standardization calculation and then dropped from the analysis because of missing covariates. Exhibit F-2 also indicates that the scale scores in both district increased slightly most years. The achievement analysis accounted for these historical changes by standardizing to the district and year, and including district and cohort controls in the models.

Exhibit F-2. Mathematics Test Descriptives by Study Year

	Treatment			Control		
	Mean	SD	<i>n</i>	Mean	SD	<i>n</i>
Standardized Scores						
Outcome	0.01	0.97	5082	-0.02	0.91	5062
Pretest	0.01	0.97	5082	0.01	1.00	5062
Scale scores						
FCAT 2012–13	208.1	24.9	357	214.4	22.3	752
FCAT 2013–14	210.4	24.4	1264	220.0	22.6	1812
FCAT 2014–15	310.4	26.1	1524	317.3	23.0	1250
FCAT 2015–16	309.6	22.6	551	309.1	20.2	191
ISAT 2012–13 ¹⁴	219.9	32.2	597	212.6	28.5	604
ISAT 2013–14	232.5	28.9	597	225.8	27.0	604
MAP 2012–13	206.6	17.2	592	202.1	15.3	597
MAP 2013–14	214.0	17.0	2548	211.5	17.9	2335
MAP 2014–15	220.8	17.6	2601	219.4	18.1	2455
MAP 2015–16	223.9	15.8	646	229.3	18.3	709

Exhibit F-3 shows the mean and standard deviation of the reading/ELA test scores in each study year, combining cohorts. As in mathematics, the overall reading/ELA standardized scores had a mean near zero with a standard deviation near one. Exhibit F-3 also indicates that the reading/ELA scale scores in both districts increased slightly each year. The achievement analysis accounted for these historical changes by standardizing to the district and year, and including district and cohort controls in the models.

¹⁴ The models used ISAT scores for Cohort 1 Year 1 (2013–14) and MAP scores for all other years. The ISAT was not administered in 2014–15 or 2015–16, so it was not included in the Year 2 descriptives.

Exhibit F-3. Reading/ELA Test Descriptives by Study Year

	Treatment			Control		
	Mean	SD	<i>n</i>	Mean	SD	<i>n</i>
Standardized Scores						
Outcome	0.02	0.98	7380	-0.03	0.99	5418
Pretest	0.01	1.00	7380	-0.04	0.97	5418
Scale scores						
FCAT 2012–13	212.0	23.3	851	210.6	19.6	993
FCAT 2013–14	212.1	24.8	2421	212.9	20.1	1619
FCAT 2014–15	306.4	24.8	2697	308.3	21.5	909
FCAT 2015–16	306.7	22	1127	309.5	20.5	283
ISAT 2012–13 ¹⁵	217.3	32.6	817	220.5	31.5	1032
ISAT 2013–14	226.9	29.5	817	226.4	29.2	1032
MAP 2012–13	203.0	19.5	814	204.1	18.1	1010
MAP 2013–14	207.1	19.9	2875	206.9	17.5	2894
MAP 2014–15	210.3	16.4	3015	209.9	17.5	2484
MAP 2015–16	214.5	15.7	951	216.2	18.2	596

Attrition

In the student achievement analysis, schools attrited from the sample when all teachers who teach tested subjects within the school attrited, i.e., they did not have students included in the achievement analysis. Exhibit F-4 and Exhibit F-5 display the number of treatment and control schools, teachers and students by analysis year and subject, and the school-, teacher-, and student-level attrition by condition in each district and overall. WWC standards for attrition take into account both overall attrition and the difference in attrition between treatment and control groups.

For analysis of the impact of NTC on mathematics achievement after 2 years of induction for teachers (2014–15 for Cohort 1 and 2015–16 for Cohort 2), overall school-level attrition was 11 percent, with differential attrition of 6 percentage points. Within the non-attrited schools, overall teacher-level attrition was 2 percent, with differential attrition of 2 percentage points. Finally, within non-attrited teachers, overall student-level attrition was 11 percent with differential attrition of 2 percentage points. This analysis meets the WWC attrition standards at each level. The same is true within both districts.

For analysis of the impact of NTC on reading/ELA achievement after 2 years of mentoring, overall school-level attrition was 10 percent, with differential attrition of 4 percentage points. Within the non-attrited schools, overall teacher-level attrition was 3 percent, with differential attrition of 2 percentage points. Finally, within non-attrited teachers, overall student-level attrition was 14 percent with differential attrition of 2 percentage points. This analysis meets the WWC attrition standards at each level. The same is true in both districts.

¹⁵ The models used ISAT scores for Cohort 1 Year 1 (2013–14) and MAP scores for all other years. The ISAT was not administered in 2014–15 or 2015–16, so it was not included in the Year 2 descriptives.

Exhibit F-4. Attrition for Analysis of the Impact of NTC on Student Achievement in Mathematics after Two Years of Induction for Teachers

		Schools				Mathematics Teachers				Students			
		Treat	Control	Differential	Overall	Treat	Control	Differential	Overall	Treat	Control	Differential	Overall
District A	Eligible Included in the mathematics model	29	12			33	16			1223	863		
		22	9			32	16			1067	726		
	Percent attrited	24%	25%	1%	24%	3%	0%	3%	2%	13%	16%	3%	14%
	Attrition standard	Met Standard				Met Standard				Met Standard			
District B	Eligible Included in the mathematics model	30	26			39	44			1725	1784		
		29	26			39	42			1584	1595		
	Percent attrited	3%	0%	3%	2%	0%	5%	5%	2%	8%	11%	2%	9%
	Attrition standard	Met Standard				Met Standard				Met Standard			
Overall	Eligible Included in the mathematics model	59	38			72	60			2948	2647		
		51	35			71	58			2651	2321		
	Percent attrited	14%	8%	6%	11%	1%	3%	2%	2%	10%	12%	2%	11%
	Attrition standard	Met Standard				Met Standard				Met Standard			

Exhibit F-5. Attrition for Analysis of the Impact of NTC on Student Achievement in Reading/ELA after Two Years of Induction for Teachers

		Schools				Reading / ELA Teachers				Students			
		Treat	Control	Differential	Overall	Treat	Control	Differential	Overall	Treat	Control	Differential	Overall
District A	Eligible Included in the mathematics model	29	15			37	20			2418	789		
		23	12			35	20			1922	636		
	Percent attrited	21%	20%	1%	20%	5%	0%	5%	4%	21%	19%	1%	20%
	Attrition standard	Met Standard				Met Standard				Met Standard			
District B	Eligible Included in the mathematics model	31	35			45	51			2209	1733		
		30	34			44	50			2023	1566		
	Percent attrited	3%	3%	0%	3%	2%	2%	0%	2%	8%	10%	1%	9%
	Attrition standard	Met Standard				Met Standard				Met Standard			
Overall	Eligible Included in the mathematics model	60	50			82	71			4627	2522		
		53	46			79	70			3945	2202		
	Percent attrited	12%	8%	4%	10%	4%	1%	2%	3%	15%	13%	2%	14%
	Attrition standard	Met Standard				Met Standard				Met Standard			

Baseline Equivalence

Although differential attrition between treatment and control groups was not evident overall, we examined baseline equivalence in the achievement test scores of the students included in each analysis (Exhibit F-6). Baseline equivalence was measured using the same model as was used to measure outcomes (discussed below), applied only to the baseline measures. Where attrition exceeds acceptable standards, an outcome analysis can still Meet Criteria with Reservations if the baseline difference between treatment and control is less than 0.05 standard deviation, or if the difference is between 0.05 and 0.25 standard deviation and a baseline measure is included in the model. In addition to meeting attrition standards, the student achievement analyses also meet baseline equivalence standards in each cohort and district, with the exception of District A in Cohort 2.

Exhibit F-6. Baseline Difference Between Treatment and Control Analysis Samples

Cohort	District	Subject	Treatment Mean (SD)	Comparison Mean (SD)	Unstandardized Difference (Pooled SD)	Standardized Difference	n Students	
							Treatment	Comparison
Cohorts combined	Both Districts Combined	ELA	0.03 (0.97)	-0.06 (0.98)	0.09 (0.98)	0.10	3945	2202
		Math	0.08 (0.98)	0.06 (1.00)	0.01 (0.99)	0.01	2651	2321
	District A	ELA	0.00 (0.98)	-0.17 (0.91)	0.16 (0.97)	0.17	1922	636
		Math	0.08 (1.02)	0.20 (0.97)	-0.11 (1.00)	-0.11	1067	726
	District B	ELA	0.06 (0.96)	-0.02 (1.01)	0.08 (0.98)	0.09	2023	1566
		Math	0.07 (0.95)	0.00 (1.01)	0.07 (0.98)	0.07	1584	1595
Cohort 1	Both Districts Combined	ELA	0.04 (0.99)	-0.05 (1.01)	0.09 (1.00)	0.09	1867	1323
		Math	0.08 (0.94)	0.02 (1.02)	0.06 (0.98)	0.06	1385	1421
	District A	ELA	-0.01 (0.96)	-0.06 (1.05)	0.05 (1.00)	0.05	1072	970
		Math	0.04 (0.95)	-0.18 (1.01)	0.21 (0.99)	0.21	938	886
	District B	ELA	0.12 (1.02)	-0.02 (0.91)	0.14 (0.99)	0.14	795	353
		Math	0.16 (0.92)	0.33 (0.96)	-0.17 (0.94)	-0.18	447	535
Cohort 2	Both Districts Combined	ELA	0.02 (0.95)	-0.09 (0.94)	0.11 (0.95)	0.11	2078	879
		Math	0.08 (1.02)	0.13 (0.97)	-0.06 (1.00)	-0.06	1266	900
	District A	ELA	-0.09 (0.94)	-0.35 (0.88)	0.25 (0.94)	0.27 [†]	1127	283
		Math	0.02 (1.08)	-0.18 (0.92)	0.20 (1.05)	0.19	620	191
	District B	ELA	0.14 (0.95)	0.03 (0.94)	0.11 (0.95)	0.12	951	596
		Math	0.12 (0.95)	0.22 (0.97)	-0.09 (0.96)	-0.10	646	709

[†] Standardized difference is greater than 0.25 standard deviations. Propensity score weight will be applied.

Statistical Analysis

To analyze the impact of the NTC model on student achievement, we estimated the following three-level model, with students nested within teachers nested within schools:

Level 1 (Students): $y_{ijk} = \pi_{0jk} + \pi_{1jk}X_{ijk} + e_{ijk}$

Level 2 (Teachers): $\pi_{0jk} = \beta_{00k} + \beta_{01k}Z_{jk} + \beta_{02k}Coh_{jk} + \beta_{03k}Z_{jk}Coh_{jk} + r_{0jk}$

$$\pi_{1jk} = \beta_{10k} + \beta_{11k}Coh_{jk}$$

Level 3 (Schools): $\beta_{00k} = \gamma_{000} + \gamma_{001}W_k + \gamma_{002}Dist_k + \gamma_{003}W_kDist_k + r_k$

$$\beta_{01k} = \gamma_{010} + \gamma_{011}Dist_k$$

$$\beta_{02k} = \gamma_{020} + \gamma_{021}W_k + \gamma_{022}Dist_k + \gamma_{023}W_kDist_k$$

$$\beta_{03k} = \gamma_{030} + \gamma_{031}Dist_k$$

$$\beta_{10k} = \gamma_{100} + \gamma_{101}Dist_k$$

$$\beta_{11k} = \gamma_{110} + \gamma_{111}Dist_k$$

Mixed:

$$\begin{aligned} y_{ijk} = & \gamma_{000} + \gamma_{001}W_k + \gamma_{010}Z_{jk} + \gamma_{100}X_{ijk} + \gamma_{002}Dist_k + \gamma_{020}Coh_{jk} + \gamma_{003}W_kDist_k \\ & + \gamma_{021}W_kCoh_{jk} + \gamma_{011}Z_{jk}Dist_k + \gamma_{030}Z_{jk}Coh_{jk} + \gamma_{101}X_{ijk}Dist_k \\ & + \gamma_{110}X_{ijk}Coh_{jk} + \gamma_{022}Coh_{jk}Dist_k + \gamma_{023}W_kCoh_{jk}Dist_k + \gamma_{031}Z_{jk}Coh_{jk}Dist_k \\ & + \gamma_{111}X_{ijk}Coh_{jk}Dist_k + r_{0jk} + u_{00k} + e_{ijk} \end{aligned}$$

In this model, y_{ijk} represents the student's score on the reading/ELA or mathematics state assessment, standardized using the mean and standard deviation of all students with study teachers provided by the districts. X_{ijk} is a vector of student-level control variables (including prior achievement on the same assessment), Z_{jk} is a vector of teacher control variables, W_k is a vector of school control variables (including the blocking variables used in randomization), Coh_{jk} is a cohort indicator, and $Dist_k$ is a district indicator. All variables were centered to the analysis sample. The model included interactions with district, cohort, and a district-by-cohort three-way interaction for each student, teacher, and school control variable. These control variables and interactions were included for the same theoretical reasons as the teacher analyses. Additionally, as we examined the relationships between each of the covariates and student achievement, we found that the relationships varied by both district and cohort. A model combining districts and cohorts, without interaction terms, would constrain the relationship between covariates and achievement to be the same across districts and cohorts. This constraint may reduce precision and bias the estimates. This examination empirically supported our theoretical reasons for including the interactions.

As with the teacher outcomes analyses, this analysis was an intent-to-treat analysis at the teacher level, in which teachers and their associated students were included in their original assigned condition, regardless of cross-over, noncompliance, or level of treatment received. Both RCT districts were able to provide student data linked to study teachers for all study teachers who started the school year working in the district. In other words, if a teacher left the district during the school year, his or her students were still linked and we obtained their achievement test scores. As discussed in the attrition section above, this reduced attrition from the analysis and maintained the intent-to-treat design within each year.

Results

The NTC model had a significant impact on student achievement in both reading/ELA and mathematics after 2 years of mentoring. This impact was also evident in mathematics in District B, in mathematics with both districts combined in Cohort 1, and in mathematics in District B in Cohort 1. No other district- and cohort-specific analyses found statistically significant impacts (Exhibit F-7).

Exhibit F-7. Impact of the NTC Model on Student Achievement after Two Years of Induction for Teachers, Overall and by Cohort and District

Cohort	District	Subject	Impact	n Students		n Teachers		n Schools	
				Treatment	Comparison	Treatment	Comparison	Treatment	Comparison
Cohorts Combined	Both Districts Combined	ELA	0.09*	3945	2202	79	70	53	46
		Math	0.15**	2651	2321	71	58	51	35
	District A	ELA	0.05	1922	636	35	20	23	12
		Math	-0.01	1067	726	32	16	22	9
Cohort 1	District B	ELA	0.10	2023	1566	44	50	30	34
		Math	0.19**	1584	1595	39	42	29	26
	Both Districts Combined	ELA	0.11	1867	1323	39	47	30	29
		Math	0.16*	1385	1421	35	41	28	24
	District A	ELA	0.08	795	353	18	14	14	9
		Math	-0.12	447	535	15	12	11	8
	District B	ELA	0.13	1072	970	21	33	16	20
		Math	0.23*	938	886	20	29	17	16
Cohort 2	Both Districts Combined	ELA	<0.001	2078	879	40	23	34	20
		Math	0.07	1266	900	36	17	34	13
	District A	ELA	-0.02†	1127	283	17	6	14	4
		Math	0.20	620	191	17	4	16	2
	District B	ELA	0.03	951	596	23	17	20	16
		Math	0.05	646	709	19	13	18	11

*** $p < 0.001$

** $p < 0.01$

* $p < 0.05$

†Propensity score weight applied to account for inequivalent baseline

Exhibit F-8. Unadjusted Outcome Means for Student Achievement after Two Years of Induction for Teachers, Overall and by District and Cohort

Cohort	District	Subject	Unadjusted Treatment Mean (SD)	Unadjusted Comparison Mean (SD)	<i>n</i> Students	
					Treatment	Comparison
Cohorts Combined	Both Districts Combined	ELA	0.03 (0.96)	-0.03 (1.00)	3945	2202
		Math	0.10 (0.96)	-0.01 (0.90)	2651	2321
	District A	ELA	-0.01 (0.95)	-0.06 (0.95)	1922	636
		Math	0.09 (1.00)	0.02 (0.74)	1067	726
	District B	ELA	0.07 (0.96)	-0.02 (1.02)	2023	1566
		Math	0.11 (0.94)	-0.02 (0.97)	1584	1595
	Cohort 1	ELA	0.03 (0.97)	-0.05 (1.04)	1867	1323
		Math	0.10 (0.91)	-0.09 (0.90)	1385	1421
Cohort 1	District A	ELA	0.00 (0.97)	-0.10 (1.05)	1072	970
		Math	0.10 (0.95)	-0.19 (0.99)	938	886
	District B	ELA	0.07 (0.97)	0.08 (1.01)	795	353
		Math	0.11 (0.83)	0.09 (0.69)	447	535
	Cohort 2	ELA	0.03 (0.95)	-0.01 (0.93)	2078	879
		Math	0.11 (1.01)	0.12 (0.90)	1266	900
	District A	ELA	-0.06 (0.93)	-0.25 (0.84)	1127	283
		Math	0.08 (1.11)	-0.19 (0.83)	620	191
Cohort 2	District B	ELA	0.15 (0.95)	0.10 (0.95)	951	596
		Math	0.13 (0.91)	0.20 (0.90)	646	709

Sensitivity to Teacher Joiners

As in the teacher retention analysis, to test the sensitivity of the student achievement results to the inclusion of joiners, we fit models identical to those estimated for the main treatment effect reported above, but excluding Cohort 2 teachers who joined in in 2014–15 the study schools that were randomized in 2013–14 under Cohort 1. In this case we were able to retain the cohort- and district-level interaction terms, due to the higher degrees of freedom in the model.

We found that excluding late joiner teachers slightly increased the estimates of the impact of the NTC model after 1 year of induction support in both reading and mathematics (Exhibit F-9). The impact of the NTC model on ELA achievement remained positive and statistically significant when late joiner teachers were excluded, with a slight increase in the estimate (from 0.09 to 0.10 standard deviation). Likewise, the impact on mathematics achievement remained positive and statistically significant when late joiner teachers were excluded, with a slight increase in the estimate (from 0.15 to 0.16 standard deviation).

Exhibit F-9. Impact of the NTC Model on Achievement Scores after Two Years of Induction Support, Excluding Late Joiner Teachers

		Impact	<i>n</i> schools	<i>n</i> teachers	<i>n</i> students
ELA	Original estimate	0.09*	99	149	6,147
	Excluding late joiner teachers	0.10*	89	119	4,571
Mathematics	Original estimate	0.15*	86	129	4,972
	Excluding late joiner teachers	0.16*	78	105	3,981

* $p < 0.05$

Full Model Tables

Both Cohorts and RCT Districts Combined

Exhibit F-10. Impact After Two Years of Mentoring – Both Cohorts and RCT Districts Combined, Mathematics

	Estimate	Standard Error	p-value
Treatment indicator	0.145	0.05	0.003
Student variables			
Prior math score	0.706	0.011	<0.001
Grade 5	0.109	0.046	0.018
Grade 6	0.024	0.054	0.656
Grade 7	0.040	0.062	0.52
Grade 8	0.132	0.065	0.042
Student race is black CENTERED	-0.232	0.036	<0.001
Student Hispanic CENTERED	-0.107	0.028	<0.001
Student FRPL CENTERED	-0.132	0.034	<0.001
Student special education CENTERED	-0.188	0.030	<0.001
Student limited English CENTERED	-0.074	0.029	0.011
Student female CENTERED	0.008	0.018	0.668
Teacher variables			
Teacher black CENTERED	-0.011	0.098	0.912
Teacher Bachelor's degree CENTERED	0.089	0.053	0.093
Teacher partially certified CENTERED	0.328	0.154	0.033
Teacher female CENTERED	0.171	0.068	0.012
School variables			
School percent FRPL CENTERED	0.001	0.004	0.807
School Report Card Rating CENTERED	-0.009	0.025	0.725
School percent minority CENTERED	-0.001	0.005	0.768
School percent ELL CENTERED	-0.003	0.002	0.203
School percent IEP CENTERED	0.002	0.008	0.771
Blocking variables			
District A block 1	0.375	0.135	0.005
District A block 2	<0.001	0.148	0.999
District B block 5	0.042	0.067	0.524
District B block 8	0.016	0.088	0.856
Interactions with district			
Centered district indicator	-0.347	0.17	0.041
District interaction: Math pretest	-0.155	0.019	<0.001
District interaction: Grade 5	0.39	0.123	0.002
District interaction: Grade 6	0.243	0.144	0.091
District interaction: Grade 7	0.545	0.167	0.001
District interaction: Grade 8	0.272	0.172	0.114
District interaction: Student black	0.089	0.052	0.087

**Exhibit F-10. Impact After Two Years of Mentoring – Both Cohorts and RCT Districts Combined,
Mathematics (continued)**

	Estimate	Standard Error	p-value
District interaction: Student Hispanic	0.055	0.049	0.263
District interaction: Student FRPL	0.064	0.05	0.197
District interaction: Student SPED	0.02	0.048	0.680
District interaction: Student LEP	0.014	0.061	0.823
District interaction: Student female	-0.034	0.03	0.255
District interaction: Teacher black	-0.227	0.145	0.118
District interaction: Teacher Bachelor's degree	-0.549	0.112	<0.001
District interaction: Teacher certification	-0.300	0.174	0.084
District interaction: Teacher female	0.056	0.127	0.661
District interaction: School percent FRPL	0.005	0.008	0.543
District interaction: School rating	0.172	0.076	0.024
District interaction: School percent non-white	0.004	0.010	0.726
District interaction: School percent student LEP	0.010	0.005	0.060
District interaction: School percent SPED	-0.030	0.017	0.086
Interactions with cohort			
Centered cohort indicator	0.031	0.118	0.793
Cohort interaction: Math pretest	0.033	0.023	0.147
Cohort interaction: Grade 5	-0.067	0.090	0.454
Cohort interaction: Grade 6	-0.013	0.103	0.898
Cohort interaction: Grade 7	1.042	0.314	0.001
Cohort interaction: Grade 8	0.513	0.123	<0.001
Cohort interaction: Student black	0.236	0.073	0.001
Cohort interaction: Student Hispanic	-0.073	0.057	0.198
Cohort interaction: Student FRPL	0.031	0.068	0.649
Cohort interaction: Student SPED	-0.054	0.060	0.364
Cohort interaction: Student LEP	0.132	0.058	0.024
Cohort interaction: Student female	0.007	0.036	0.847
Cohort interaction: Teacher black	-0.096	0.196	0.626
Cohort interaction: Teacher Bachelor's degree	0.120	0.107	0.264
Cohort interaction: Teacher certification	0.211	0.304	0.487
Cohort interaction: Teacher female	0.090	0.131	0.492
Cohort interaction: School percent FRPL	-0.006	0.007	0.458
Cohort interaction: School rating	-0.076	0.046	0.097
Cohort interaction: School percent non-white	0.004	0.010	0.713
Cohort interaction: School percent student LEP	0.011	0.004	0.008
Cohort interaction: School percent SPED	-0.002	0.015	0.910
Cohort interaction: District A block 1	0.572	0.238	0.016
Cohort interaction: District A block 2	0.007	0.258	0.978
Cohort interaction: District B block 5	-0.120	0.123	0.329
Cohort interaction: District B block 8	-0.210	0.184	0.254

**Exhibit F-10. Impact After Two Years of Mentoring – Both Cohorts and RCT Districts Combined,
Mathematics (concluded)**

	Estimate	Standard Error	p-value
District-by-Cohort interaction			
District by cohort interaction	-0.746	0.326	0.022
District by cohort interaction: Math pretest	0.060	0.038	0.115
District by cohort interaction: Grade 5	0.897	0.246	<0.001
District by cohort interaction: Grade 6	0.593	0.292	0.042
District by cohort interaction: Grade 8	0.509	0.336	0.13
District by cohort interaction: Student black	-0.141	0.104	0.175
District by cohort interaction: Student Hispanic	0.179	0.098	0.068
District by cohort interaction: Student FRPL	0.124	0.100	0.214
District by cohort interaction: Student SPED	0.161	0.095	0.092
District by cohort interaction: Student LEP	-0.448	0.123	<0.001
District by cohort interaction: Student female	0.029	0.060	0.628
District by cohort interaction: Teacher black	-0.669	0.288	0.020
District by cohort interaction: Teacher Bachelor's degree	-1.162	0.238	<0.001
District by cohort interaction: Teacher certification	0.316	0.348	0.364
District by cohort interaction: Teacher female	0.317	0.249	0.203
District by cohort interaction: School percent FRPL	0.043	0.016	0.008
District by cohort interaction: School rating	0.295	0.130	0.023
District by cohort interaction: School percent non-white	-0.027	0.020	0.187
District by cohort interaction: School percent student LEP	0.010	0.009	0.247
District by cohort interaction: School percent SPED	-0.010	0.032	0.761
Constant	-0.123	0.066	0.062
Random effects			
School	0.017		
Teacher	0.011		
Student	0.236		
n			
Schools	86		
Teachers	129		
Students	4972		

**Exhibit F-11. Impact After Two Years of Mentoring –
Both Cohorts and RCT Districts Combined, Reading/ELA**

	Estimate	Standard Error	p-value
Treatment indicator	0.088	0.039	0.026
Student variables			
Prior reading score	0.709	0.013	<0.001
Grade 5	-0.021	0.036	0.560
Grade 6	0.037	0.038	0.320
Grade 7	-0.037	0.049	0.446
Grade 8	-0.020	0.049	0.692
Student race is black CENTERED	-0.114	0.038	0.003
Student Hispanic CENTERED	-0.006	0.031	0.839
Student FRPL CENTERED	-0.090	0.033	0.007
Student special education CENTERED	-0.290	0.028	<0.001
Student limited English CENTERED	-0.117	0.031	<0.001
Student female CENTERED	0.036	0.018	0.047
Teacher variables			
Teacher black CENTERED	0.129	0.067	0.055
Teacher Bachelor's degree CENTERED	-0.021	0.039	0.582
Teacher partially certified CENTERED	0.013	0.120	0.915
Teacher female CENTERED	0.071	0.050	0.155
School variables			
School percent FRPL CENTERED	-0.006	0.003	0.022
School Report Card Rating CENTERED	0.007	0.022	0.746
School percent minority CENTERED	0.007	0.004	0.066
School percent ELL CENTERED	-0.005	0.002	0.011
School percent IEP CENTERED	-0.001	0.006	0.849
Blocking Variables			
District A block 1	0.139	0.105	0.188
District A block 2	-0.057	0.109	0.602
District B block 5	0.001	0.064	0.982
District B block 8	0.181	0.070	0.009
Interactions with district			
District indicator	-0.280	0.141	0.047
District interaction: Reading pretest	-0.026	0.018	0.153
District interaction: Grade 5	0.358	0.104	0.001
District interaction: Grade 6	0.163	0.105	0.120
District interaction: Grade 7	0.415	0.125	0.001
District interaction: Grade 8	0.287	0.135	0.033
District interaction: Student black	-0.013	0.053	0.803
District interaction: Student Hispanic	-0.070	0.050	0.161
District interaction: Student FRPL	0.024	0.049	0.631
District interaction: Student SPED	0.015	0.044	0.728

**Exhibit F-11. Impact After Two Years of Mentoring –
Both Cohorts and RCT Districts Combined, Reading/ELA (continued)**

	Estimate	Standard Error	p-value
District interaction: Student LEP	-0.037	0.056	0.504
District interaction: Student female	0.026	0.028	0.356
District interaction: Teacher black	-0.041	0.083	0.621
District interaction: Teacher Bachelor's degree	-0.002	0.059	0.967
District interaction: Teacher certification	-0.041	0.134	0.759
District interaction: Teacher female	-0.037	0.147	0.802
District interaction: School percent FRPL	-0.003	0.007	0.671
District interaction: School rating	0.012	0.058	0.843
District interaction: School percent non-white	-0.002	0.009	0.829
District interaction: School percent student LEP	0.013	0.005	0.005
District interaction: School percent SPED	-0.023	0.014	0.099
Interactions with cohort			
Cohort indicator	0.189	0.083	0.022
Cohort interaction: Reading pretest	-0.065	0.025	0.011
Cohort interaction: Grade 5	-0.078	0.067	0.244
Cohort interaction: Grade 6	0.047	0.068	0.490
Cohort interaction: Grade 7	0.744	0.224	0.001
Cohort interaction: Grade 8	0.027	0.091	0.763
Cohort interaction: Student black	-0.050	0.072	0.488
Cohort interaction: Student Hispanic	0.043	0.049	0.384
Cohort interaction: Student FRPL	-0.040	0.065	0.538
Cohort interaction: Student SPED	-0.200	0.056	<0.001
Cohort interaction: Student LEP	0.076	0.061	0.216
Cohort interaction: Student female	0.017	0.036	0.639
Cohort interaction: Teacher black	0.576	0.134	<0.001
Cohort interaction: Teacher Bachelor's degree	0.052	0.083	0.531
Cohort interaction: Teacher certification	0.028	0.235	0.905
Cohort interaction: Teacher female	0.092	0.091	0.316
Cohort interaction: School percent FRPL	-0.007	0.005	0.196
Cohort interaction: School rating	-0.056	0.034	0.097
Cohort interaction: School percent non-white	-0.001	0.007	0.854
Cohort interaction: School percent student LEP	0.005	0.003	0.097
Cohort interaction: School percent SPED	<0.001	0.009	0.995
Cohort interaction: District A block 1	-0.121	0.172	0.481
Cohort interaction: District A block 2	-0.337	0.182	0.065
Cohort interaction: District B block 5	-0.149	0.100	0.135
Cohort interaction: District B block 8	-0.209	0.147	0.155
District-by-cohort interactions			
District-by-cohort interaction	-0.733	0.272	0.007
District-by-cohort interaction: Reading pretest	0.012	0.037	0.739

**Exhibit F-11. Impact After Two Years of Mentoring –
Both Cohorts and RCT Districts Combined, Reading/ELA (concluded)**

	Estimate	Standard Error	p-value
District-by-cohort interaction: Grade 5	0.789	0.206	<0.001
District-by-cohort interaction: Grade 6	0.850	0.209	<0.001
District-by-cohort interaction: Grade 8	0.666	0.264	0.012
District-by-cohort interaction: Student black	0.052	0.084	0.538
District-by-cohort interaction: Student Hispanic	0.124	0.097	0.197
District-by-cohort interaction: Student FRPL	0.167	0.087	0.055
District-by-cohort interaction: Student SPED	0.012	0.037	0.739
District-by-cohort interaction: Student LEP	-0.079	0.111	0.476
District-by-cohort interaction: Student female	-0.102	0.056	0.067
District-by-cohort interaction: Teacher black	-0.814	0.169	<0.001
District-by-cohort interaction: Teacher Bachelor's degree	-0.090	0.124	0.466
District-by-cohort interaction: Teacher certification	0.155	0.258	0.548
District-by-cohort interaction: Teacher female	-0.115	0.306	0.707
District-by-cohort interaction: School percent FRPL	-0.001	0.014	0.938
District-by-cohort interaction: School rating	0.049	0.092	0.594
District-by-cohort interaction: School percent non-white	0.009	0.017	0.613
District-by-cohort interaction: School percent student LEP	-0.001	0.006	0.884
District-by-cohort interaction: School percent SPED	0.005	0.022	0.820
Constant	-0.022	0.052	0.675
Random effects			
School	0.021		
Teacher	<0.001		
Student	0.274		
n			
School	99		
Teacher	149		

Both Cohorts Combined, District A

Exhibit F-12. Impact After Two Years of Mentoring – District A Both Cohorts Combined, Mathematics

	Estimate	Standard Error	p-value
Treatment indicator	-0.010	0.066	0.880
Student variables			
Prior math score	0.556	0.014	<0.001
Grade 5	0.458	0.103	<0.001
Grade 6	0.221	0.115	0.054
Grade 7	0.472	0.124	<0.001
Grade 8	0.301	0.125	0.016
Student race is black CENTERED	-0.148	0.035	<0.001
Student Hispanic CENTERED	-0.055	0.037	0.137
Student FRPL CENTERED	-0.074	0.034	0.029
Student special education CENTERED	-0.165	0.035	<0.001
Student limited English CENTERED	-0.061	0.050	0.223
Student female CENTERED	-0.028	0.022	0.211
Teacher variables			
Teacher black CENTERED	-0.230	0.072	0.001
Teacher Bachelor's degree CENTERED	-0.432	0.069	<0.001
Teacher partially certified CENTERED	0.039	0.055	0.484
Teacher female CENTERED	0.184	0.065	0.005
School variables			
School percent FRPL CENTERED	0.003	0.005	0.572
School Report Card Rating CENTERED	0.159	0.043	<0.001
School percent minority CENTERED	0.003	0.006	0.593
School percent ELL CENTERED	0.009	0.003	0.001
School percent IEP CENTERED	-0.025	0.010	0.009
Blocking variables			
District A block 1	0.317	0.080	<0.001
District A block 2	0.028	0.082	0.733
Interactions with cohort			
Cohort indicator	-0.598	0.245	0.015
Cohort interaction: Math pretest	0.099	0.028	<0.001
Cohort interaction: Grade 5	0.764	0.206	<0.001
Cohort interaction: Grade 6	0.471	0.238	0.048
Cohort interaction: Grade 7	0.601	0.263	0.022
Cohort interaction: Grade 8	0.759	0.256	0.003
Cohort interaction: Student black	0.087	0.070	0.211
Cohort interaction: Student Hispanic	0.106	0.075	0.154
Cohort interaction: Student FRPL	0.154	0.068	0.024
Cohort interaction: Student SPED	0.119	0.069	0.085
Cohort interaction: Student LEP	-0.306	0.100	0.002

**Exhibit F-12. Impact After Two Years of Mentoring – District A Both Cohorts Combined, Mathematics
(concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: Student female	0.032	0.044	0.477
Cohort interaction: Teacher black	-0.784	0.144	<0.001
Cohort interaction: Teacher Bachelor's degree	-0.976	0.143	<0.001
Cohort interaction: Teacher certification	0.377	0.133	0.005
Cohort interaction: Teacher female	0.335	0.136	0.014
Cohort interaction: School percent FRPL	0.029	0.009	0.002
Cohort interaction: School rating	0.198	0.085	0.020
Cohort interaction: School percent non-white	-0.009	0.012	0.468
Cohort interaction: School percent student LEP	0.013	0.006	0.023
Cohort interaction: School percent SPED	-0.010	0.019	0.583
Cohort interaction: District A block 1	0.674	0.162	<0.001
Cohort interaction: District A block 2	0.219	0.184	0.233
Constant	-0.284	0.135	0.035
Random effects			
School	<0.001		
Teacher	0.005		
Student	0.206		
n			
Schools	31		
Teachers	48		
Students	1793		

Exhibit F-13. Impact After Two Years of Mentoring – District A Both Cohorts Combined, Reading/ELA

	Estimate	Standard Error	p-value
Treatment indicator	0.045	0.056	0.424
Student variables			
Prior reading score	0.684	0.014	<0.001
Grade 5	0.330	0.101	0.001
Grade 6	0.197	0.101	0.052
Grade 7	0.398	0.110	<0.001
Grade 8	0.309	0.122	0.011
Student race is black CENTERED	-0.126	0.040	0.002
Student Hispanic CENTERED	-0.079	0.043	0.063
Student FRPL CENTERED	-0.069	0.038	0.071
Student special education CENTERED	-0.273	0.035	<0.001
Student limited English CENTERED	-0.149	0.049	0.002
Student female CENTERED	0.060	0.023	0.008
Teacher variables			
Teacher black CENTERED	0.075	0.045	0.096
Teacher Bachelor's degree CENTERED	-0.024	0.042	0.569
Teacher partially certified CENTERED	-0.017	0.048	0.723
Teacher female CENTERED	0.090	0.118	0.445
School constant			
School percent FRPL CENTERED	-0.007	0.005	0.151
School Report Card Rating CENTERED	0.029	0.041	0.483
School percent minority CENTERED	0.002	0.006	0.739
School percent ELL CENTERED	0.010	0.003	0.001
School percent IEP CENTERED	-0.026	0.010	0.006
Blocking Variables			
District A block 1	0.150	0.078	0.054
District A block 2	-0.043	0.080	0.593
Interactions with cohort			
Cohort indicator	-0.569	0.244	0.019
Cohort interaction: Reading pretest	-0.057	0.028	0.041
Cohort interaction: Grade 5	0.704	0.201	<0.001
Cohort interaction: Grade 6	0.846	0.203	<0.001
Cohort interaction: Grade 7	0.777	0.216	<0.001
Cohort interaction: Grade 8	0.778	0.242	0.001
Cohort interaction: Student black	-0.008	0.079	0.916
Cohort interaction: Student Hispanic	0.047	0.085	0.584
Cohort interaction: Student FRPL	0.089	0.076	0.244
Cohort interaction: Student SPED	-0.037	0.070	0.601
Cohort interaction: Student LEP	-0.003	0.098	0.973
Cohort interaction: Student female	-0.085	0.045	0.059
Cohort interaction: Teacher black	-0.256	0.095	0.007

**Exhibit F-13. Impact After Two Years of Mentoring – District A Both Cohorts Combined, Reading/ELA
(concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: Teacher Bachelor's degree	-0.021	0.086	0.806
Cohort interaction: Teacher certification	0.131	0.093	0.160
Cohort interaction: Teacher female	0.042	0.247	0.864
Cohort interaction: School percent FRPL	-0.004	0.010	0.688
Cohort interaction: School rating	<0.001	0.076	0.999
Cohort interaction: School percent non-white	0.005	0.012	0.664
Cohort interaction: School percent student LEP	0.002	0.005	0.741
Cohort interaction: School percent SPED	0.004	0.016	0.815
Cohort interaction: District A block 1	-0.058	0.146	0.692
Cohort interaction: District A block 2	-0.340	0.149	0.023
Constant	-0.292	0.126	0.021
Random effects			
School	0.007		
Teacher	<0.001		
Student	0.307		
n			
Schools	35		
Teachers	55		
Students	2558		

Both Cohorts Combined, District B

Exhibit F-14. Impact After Two Years of Mentoring – District B Both Cohorts Combined, Mathematics

	Estimate	Standard Error	p-value
Treatment indicator	0.188	0.068	0.006
Student variables			
Prior math score	0.706	0.012	<0.001
Grade 5	0.120	0.050	0.017
Grade 6	0.029	0.059	0.627
Grade 7	0.028	0.067	0.672
Grade 8	0.158	0.071	0.027
Student race is black CENTERED	-0.231	0.038	<0.001
Student Hispanic CENTERED	-0.106	0.030	<0.001
Student FRPL CENTERED	-0.132	0.035	<0.001
Student special education CENTERED	-0.187	0.031	<0.001
Student limited English CENTERED	-0.074	0.030	0.014
Student female CENTERED	0.007	0.019	0.699
Teacher variables			
Teacher black CENTERED	-0.010	0.115	0.932
Teacher Bachelor's degree CENTERED	0.096	0.062	0.123
Teacher partially certified CENTERED	0.344	0.179	0.055
Teacher female CENTERED	0.183	0.080	0.022
School variables			
School percent FRPL CENTERED	0.002	0.004	0.734
School Report Card Rating CENTERED	-0.011	0.029	0.718
School percent minority CENTERED	-0.002	0.006	0.734
School percent ELL CENTERED	-0.003	0.003	0.223
School percent IEP CENTERED	0.001	0.009	0.900
Blocking variables			
District B block 5	0.038	0.078	0.630
District B block 8	<0.001	0.103	0.999
Interactions with cohort			
Cohort indicator	0.006	0.137	0.962
Cohort interaction: Math pretest	0.035	0.024	0.135
Cohort interaction: Grade 5	-0.073	0.097	0.457
Cohort interaction: Grade 6	-0.003	0.111	0.977
Cohort interaction: Grade 8	0.585	0.134	<0.001
Cohort interaction: Student black	0.237	0.076	0.002
Cohort interaction: Student Hispanic	-0.074	0.059	0.212
Cohort interaction: Student FRPL	0.029	0.070	0.680
Cohort interaction: Student SPED	-0.056	0.063	0.370
Cohort interaction: Student LEP	0.135	0.060	0.025
Cohort interaction: Student female	0.007	0.037	0.856

**Exhibit F-14. Impact After Two Years of Mentoring – District B Both Cohorts Combined, Mathematics
(concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: Teacher black	-0.097	0.231	0.673
Cohort interaction: Teacher Bachelor's degree	0.133	0.125	0.289
Cohort interaction: Teacher certification	0.218	0.354	0.538
Cohort interaction: Teacher female	0.102	0.155	0.509
Cohort interaction: School percent FRPL	-0.006	0.009	0.499
Cohort interaction: School rating	-0.085	0.054	0.115
Cohort interaction: School percent non-white	0.004	0.012	0.709
Cohort interaction: School percent student LEP	0.012	0.005	0.017
Cohort interaction: School percent SPED	-0.004	0.017	0.814
Cohort interaction: District B block 5	-0.098	0.146	0.502
Cohort interaction: District B block 8	-0.220	0.216	0.309
Constant	-0.146	0.077	0.060
Random effects			
School	0.025		
Teacher	0.018		
Student	0.253		
n			
Schools	55		
Teachers	81		
Student	3179		

Exhibit F-15. Impact After Two Years of Mentoring – District B Both Cohorts Combined, Reading/ELA

	Estimate	Standard Error	p-value
Treatment indicator	0.099	0.057	0.082
Student variables			
Prior reading score	0.707	0.012	<0.001
Grade 5	-0.032	0.036	0.381
Grade 6	0.027	0.037	0.465
Grade 7	-0.056	0.049	0.250
Grade 8	-0.034	0.048	0.486
Student race is black CENTERED	-0.112	0.037	0.002
Student Hispanic CENTERED	-0.006	0.030	0.833
Student FRPL CENTERED	-0.090	0.032	0.005
Student special education CENTERED	-0.293	0.027	<0.001
Student limited English CENTERED	-0.120	0.029	<0.001
Student female CENTERED	0.036	0.017	0.035
Teacher variables			
Teacher black CENTERED	0.121	0.072	0.091
Teacher Bachelor's degree CENTERED	-0.015	0.040	0.707
Teacher partially certified CENTERED	0.008	0.131	0.952
Teacher female CENTERED	0.073	0.053	0.167
School variables			
School percent FRPL CENTERED	-0.006	0.003	0.047
School Report Card Rating CENTERED	0.008	0.025	0.736
School percent minority CENTERED	0.007	0.004	0.109
School percent ELL CENTERED	-0.005	0.002	0.023
School percent IEP CENTERED	-0.002	0.007	0.760
Blocking variables			
District B block 5	-0.004	0.072	0.959
District B block 8	0.190	0.077	0.014
Interactions with cohort			
Cohort Indicator	0.191	0.088	0.030
Cohort interaction: Reading pretest	-0.065	0.024	0.008
Cohort interaction: Grade 5	-0.085	0.067	0.206
Cohort interaction: Grade 6	0.044	0.067	0.507
Cohort interaction: Grade 7	<0.001	<0.001	
Cohort interaction: Grade 8	0.013	0.088	0.886
Cohort interaction: Student black	-0.063	0.073	0.389
Cohort interaction: Student Hispanic	0.040	0.060	0.505
Cohort interaction: Student FRPL	-0.042	0.063	0.508
Cohort interaction: Student SPED	-0.200	0.054	<0.001
Cohort interaction: Student LEP	0.078	0.059	0.181
Cohort interaction: Student female	0.016	0.034	0.645
Cohort interaction: Teacher black	0.606	0.145	<0.001

**Exhibit F-15. Impact After Two Years of Mentoring – District B Both Cohorts Combined, Reading/ELA
(concluded)**

	Estimate	Standard Error	p-value
Cohort interaction: Teacher Bachelor's degree	0.055	0.087	0.529
Cohort interaction: Teacher certification	0.060	0.256	0.813
Cohort interaction: Teacher female	0.103	0.096	0.284
Cohort interaction: School percent FRPL	-0.007	0.006	0.232
Cohort interaction: School rating	-0.058	0.036	0.110
Cohort interaction: School percent non-white	-0.001	0.008	0.850
Cohort interaction: School percent student LEP	0.005	0.003	0.121
Cohort interaction: School percent SPED	-0.002	0.010	0.813
Cohort interaction: District B block 5	-0.138	0.105	0.191
Cohort interaction: District B block 8	-0.222	0.164	0.176
Constant	-0.020	0.057	0.724
Random effects			
School	0.030		
Teacher	<0.001		
Student	0.250		
n			
Schools	64		
Teachers	94		
Students	3589		

Cohort 1, Both RCT Districts Combined

Exhibit F-16. Impact After Two Years of Mentoring – Cohort 1, Both RCT Districts Combined, Mathematics

	Estimate	Standard Error	p-value
Treatment indicator	0.158	0.067	0.019
Student variables			
Prior math score	0.688	0.015	<0.001
Grade 5	0.154	0.045	0.001
Grade 6	0.036	0.059	0.543
Grade 7	0.047	0.070	0.502
Grade 8	-0.123	0.072	0.085
Student race is black CENTERED	-0.351	0.057	<0.001
Student Hispanic CENTERED	-0.068	0.044	0.122
Student FRPL CENTERED	-0.145	0.053	0.007
Student special education CENTERED	-0.161	0.035	<0.001
Student limited English CENTERED	-0.141	0.036	<0.001
Student female CENTERED	0.004	0.025	0.884
Teacher variables			
Teacher black CENTERED	0.060	0.123	0.627
Teacher Bachelor's degree CENTERED	0.032	0.074	0.661
Teacher partially certified CENTERED	0.254	0.149	0.088
Teacher female CENTERED	0.128	0.103	0.211
School variables			
School percent FRPL CENTERED	0.004	0.005	0.478
School Report Card Rating CENTERED	0.039	0.032	0.224
School percent minority CENTERED	-0.003	0.007	0.654
School percent ELL CENTERED	-0.009	0.003	0.012
School percent IEP CENTERED	0.003	0.008	0.696
Blocking variables			
District B block 5	0.082	0.101	0.416
District B block 8	0.090	0.129	0.484
District A block 1	0.081	0.196	0.681
District A block 2	0.022	0.218	0.920
Interactions with district			
District Indicator	0.037	0.253	0.885
District interaction: Math pretest	-0.185	0.027	<0.001
District interaction: Grade 5	-0.067	0.187	0.723
District interaction: Grade 6	-0.069	0.194	0.722
District interaction: Grade 7	-0.012	0.228	0.956
District interaction: Grade 8	-0.003	0.234	0.991

**Exhibit F-16. Impact After Two Years of Mentoring – Cohort 1, Both RCT Districts Combined,
Mathematics (concluded)**

	Estimate	Standard Error	p-value
District interaction: Student black	0.162	0.073	0.027
District interaction: Student Hispanic	-0.038	0.068	0.578
District interaction: Student FRPL	0.001	0.073	0.986
District interaction: Student SPED	-0.060	0.066	0.365
District interaction: Student LEP	0.240	0.086	0.005
District interaction: Student female	-0.048	0.042	0.254
District interaction: Teacher black	0.049	0.216	0.820
District interaction: Teacher Bachelor's degree	-0.016	0.192	0.932
District interaction: Teacher certification	-0.513	0.198	0.009
District interaction: Teacher female	-0.111	0.148	0.453
District interaction: School percent FRPL	-0.016	0.008	0.040
District interaction: School rating	0.034	0.094	0.718
District interaction: School percent non-white	0.018	0.014	0.201
District interaction: School percent student LEP	0.006	0.008	0.463
District interaction: School percent SPED	-0.021	0.029	0.472
Constant	-0.118	0.089	0.185
Random effects			
School	0.023		
Teacher	0.015		
Student	0.262		
n			
Schools	52		
Teachers	76		
Students	2806		

**Exhibit F-17. Impact After Two Years of Mentoring – Cohort 1, Both RCT Districts Combined,
Reading/ELA**

	Estimate	Standard Error	p-value
Treatment indicator	0.111	0.061	0.070
Student variables			
Prior reading score	0.737	0.017	<0.001
Grade 5	-0.007	0.042	0.863
Grade 6	-0.036	0.047	0.448
Grade 7	-0.151	0.063	0.017
Grade 8	-0.068	0.054	0.208
Student race is black CENTERED	-0.073	0.055	0.188
Student Hispanic - CENTERED	-0.021	0.049	0.663
Student FRPL CENTERED	-0.069	0.055	0.207
Student special education CENTERED	-0.196	0.038	<0.001
Student limited English CENTERED	-0.163	0.039	<0.001
Student female CENTERED	0.029	0.024	0.235
Teacher variables			
Teacher black CENTERED	-0.073	0.089	0.409
Teacher Bachelor's degree CENTERED	-0.036	0.052	0.487
Teacher partially certified CENTERED	0.017	0.086	0.846
Teacher female CENTERED	0.052	0.078	0.503
School variables			
School percent FRPL CENTERED	-0.004	0.005	0.412
School Report Card Rating CENTERED	0.035	0.030	0.242
School percent minority CENTERED	0.007	0.007	0.288
School percent ELL CENTERED	-0.006	0.003	0.053
School percent IEP CENTERED	<0.001	0.007	0.952
Blocking variables			
District B block 5	0.065	0.094	0.489
District B block 8	0.283	0.106	0.007
District A block 1	0.255	0.163	0.117
District A block 2	0.175	0.173	0.310
Interactions with district			
District indicator	-0.032	0.021	0.126
District interaction: Reading pretest	-0.031	0.026	0.240
District interaction: Grade 5	-0.006	0.177	0.975
District interaction: Grade 6	-0.260	0.176	0.138
District interaction: Grade 7	0.101	0.191	0.596
District interaction: Grade 8	-0.059	0.196	0.763
District interaction: Student black	-0.057	0.081	0.482
District interaction: Student Hispanic	-0.076	0.081	0.347
District interaction: Student FRPL	-0.035	0.077	0.650
District interaction: Student SPED	-0.064	0.066	0.336

**Exhibit F-17. Impact After Two Years of Mentoring – Cohort 1, Both RCT Districts Combined,
Reading/ELA (concluded)**

	Estimate	Standard Error	p-value
District interaction: Student LEP	0.005	0.090	0.959
District interaction: Student female	0.078	0.040	0.053
District interaction: Teacher black	0.282	0.126	0.025
District interaction: Teacher partially certified	0.029	0.082	0.722
District interaction: Teacher female	-0.159	0.119	0.181
District interaction: School percent FRPL	-0.050	0.134	0.712
District interaction: School rating	-0.002	0.007	0.839
District interaction: School percent non-white	-0.041	0.080	0.611
District interaction: School percent student LEP	-0.007	0.010	0.509
District interaction: School percent SPED	0.013	0.007	0.050
Constant	-0.086	0.071	0.225
Random effect			
School	0.030		
Teacher	<0.001		
Student	0.283		
n			
Schools	59		
Teachers	86		
Students	3190		

Cohort 1, District A

Exhibit F-18. Impact After Two Years of Mentoring – District A Cohort 1, Mathematics

	Estimate	Standard Error	p-value
Treatment indicator	-0.119	0.063	0.058
Student variables			
Prior math score	0.512	0.020	<0.001
Student in 5th grade	0.094	0.149	0.530
Student in 6th grade	0.034	0.142	0.810
Student in 7th grade	0.323	0.167	0.053
Student in 8th grade	0.001	0.161	0.994
Student race is black CENTERED	-0.131	0.033	<0.001
Student FRPL CENTERED	-0.168	0.045	<0.001
Student special education CENTERED	-0.227	0.051	<0.001
Student limited English CENTERED	0.065	0.070	0.349
Student female CENTERED	-0.045	0.031	0.147
Teacher variables			
Teacher black CENTERED	0.267	0.092	0.004
Teacher Bachelor's degree CENTERED	0.153	0.094	0.103
Teacher partially certified CENTERED	-0.026	0.080	0.745
Teacher female CENTERED	0.014	0.042	0.742
Teacher variables			
School percent FRPL CENTERED	-0.012	0.002	<0.001
School Report Card Rating CENTERED	0.030	0.037	0.414
School percent minority CENTERED	<0.001	0.006	0.983
School percent ELL CENTERED	0.007	0.004	0.061
School percent IEP CENTERED	-0.024	0.013	0.070
Blocking variables			
District A block 1	-0.116	0.095	0.221
District A block 2	-0.157	0.099	0.111
Constant	0.038	0.165	0.819
Random effects			
School	<0.001		
Teacher	<0.001		
Student	0.223		
n			
Schools	19		
Teachers	27		
Students	982		

Exhibit F-19. Impact After Two Years of Mentoring – District A Cohort 1, Reading/ELA

	Estimate	Standard Error	p-value
Treatment indicator	0.079	0.086	0.360
Student variables			
Prior reading score	0.710	0.021	<0.001
Student in 5th grade	-0.025	0.172	0.884
Student in 6th grade	-0.270	0.167	0.105
Student in 7th grade	-0.024	0.175	0.890
Student in 8th grade	-0.109	0.185	0.556
Student race is black CENTERED	-0.059	0.043	0.167
Student FRPL CENTERED	-0.118	0.056	0.034
Student special education CENTERED	-0.262	0.056	<0.001
Student limited English CENTERED	-0.166	0.084	0.048
Student female CENTERED	0.104	0.034	0.002
Teacher variables			
Teacher black CENTERED	0.215	0.083	0.010
Teacher Bachelor's degree CENTERED	-0.014	0.064	0.822
Teacher partially certified CENTERED	-0.124	0.074	0.093
Teacher female CENTERED	0.039	0.091	0.671
School variables			
School percent FRPL CENTERED	-0.006	0.004	0.150
School Report Card Rating CENTERED	0.005	0.057	0.923
School percent minority CENTERED	<0.001	0.006	0.992
School percent ELL CENTERED	0.009	0.005	0.058
School percent IEP CENTERED	-0.031	0.016	0.043
Blocking variables			
District A block 1	0.242	0.131	0.065
District A block 2	0.188	0.133	0.159
Constant	-0.049	0.211	0.816
Random effects			
School	0.011		
Teacher	<0.001		
Student	0.305		
n			
Schools	23		
Teachers	32		
Students	1148		

Cohort 1, District B

Exhibit F-20. Impact After Two Years of Mentoring – District B Cohort 1, Mathematics

	Estimate	Standard Error	p-value
Treatment indicator	0.227	0.096	0.017
Student variables			
Prior math score	0.688	0.016	<0.001
Student in 5th grade	0.162	0.047	0.001
Student in 6th grade	0.037	0.062	0.551
Student in 7th grade	0.041	0.075	0.580
Student in 8th grade	-0.128	0.076	0.091
Student race is black CENTERED	-0.303	0.050	<0.001
Student FRPL CENTERED	-0.161	0.054	0.003
Student special education CENTERED	-0.160	0.036	<0.001
Student limited English CENTERED	-0.145	0.038	<0.001
Student female CENTERED	0.004	0.026	0.885
Teacher variables			
Teacher black CENTERED	0.060	0.136	0.658
Teacher Bachelor's degree CENTERED	0.048	0.081	0.558
Teacher partially certified CENTERED	0.269	0.162	0.096
Teacher female CENTERED	0.120	0.114	0.295
School variables			
School percent FRPL CENTERED	0.005	0.007	0.462
School Report Card Rating CENTERED	0.047	0.039	0.221
School percent minority CENTERED	-0.004	0.009	0.624
School percent ELL CENTERED	-0.010	0.004	0.016
School percent IEP CENTERED	0.004	0.010	0.702
Blocking variables			
District B block 5	0.079	0.124	0.523
District B block 8	0.092	0.153	0.547
Constant	-0.147	0.106	0.163
Random effects			
School	0.042		
Teacher	0.017		
Student	0.284		
n			
Schools	33		
Teachers	49		
Students	1824		

Exhibit F-21. Impact After Two Years of Mentoring – District B Cohort 1, Reading/ELA

	Estimate	Standard Error	p-value
Treatment indicator	0.131	0.092	0.152
Student variables			
Prior reading score	0.735	0.017	<0.001
Student in 5th grade	-0.017	0.042	0.690
Student in 6th grade	-0.049	0.048	0.303
Student in 7th grade	-0.176	0.064	0.006
Student in 8th grade	-0.073	0.053	0.170
Student race is black CENTERED	-0.053	0.045	0.236
Student FRPL CENTERED	-0.074	0.052	0.158
Student special education CENTERED	-0.199	0.037	<0.001
Student limited English CENTERED	-0.167	0.038	<0.001
Student female CENTERED	0.030	0.024	0.207
Teacher variables			
Teacher black CENTERED	-0.088	0.094	0.345
Teacher Bachelor's degree CENTERED	-0.025	0.052	0.628
Teacher partially certified CENTERED	-0.006	0.086	0.946
Teacher female CENTERED	0.054	0.083	0.515
School variables			
School percent FRPL CENTERED	-0.004	0.006	0.472
School Report Card Rating CENTERED	0.036	0.036	0.312
School percent minority CENTERED	0.008	0.008	0.363
School percent ELL CENTERED	-0.006	0.003	0.086
School percent IEP CENTERED	0.001	0.008	0.922
Blocking variables			
District B block 5	0.063	0.113	0.578
District B block 8	0.307	0.124	0.014
Constant	-0.096	0.083	0.247
Random effects			
School	0.047		
Teacher	<0.001		
Student	0.27		
n			
Schools	36		
Teachers	54		
Students	2042		

Cohort 2, Both Districts Combined

Exhibit F-22. Impact After Two Years of Mentoring – Cohort 2, Both RCT Districts Combined, Mathematics

	Estimate	Standard Error	p-value
Treatment indicator	0.071	0.074	0.335
Student variables			
Prior math score	0.724	0.016	<0.001
Grade 5	0.061	0.071	0.393
Grade 6	-0.013	0.083	0.872
Grade 7	-0.119	0.190	0.531
Grade 8	0.332	0.097	0.001
Student race is black CENTERED	-0.117	0.045	0.010
Student Hispanic - CENTERED	-0.148	0.036	<0.001
Student FRPL CENTERED	-0.116	0.041	0.005
Student special education CENTERED	-0.209	0.045	<0.001
Student limited English CENTERED	-0.008	0.043	0.845
Student female CENTERED	0.013	0.025	0.605
Teacher variables			
Teacher black CENTERED	-0.090	0.118	0.443
Teacher Bachelor's degree CENTERED	0.103	0.073	0.156
Teacher partially certified CENTERED	0.274	0.209	0.189
Teacher female CENTERED	0.179	0.076	0.018
School variables			
School percent FRPL CENTERED	-0.001	0.004	0.781
School Report Card Rating CENTERED	-0.060	0.029	0.039
School percent minority CENTERED	-0.002	0.006	0.786
School percent ELL CENTERED	0.002	0.002	0.421
School percent IEP CENTERED	0.003	0.012	0.824
Blocking variables			
District B block 5	0.055	0.076	0.469
District B block 8	-0.081	0.106	0.444
District A block 1	0.640	0.144	<0.001
District A block 2	0.096	0.153	0.530
Interactions with district			
District indicator	-0.635	0.205	0.002
District interaction: Math pretest	-0.124	0.026	<0.001
District interaction: Grade 5	0.821	0.159	<0.001
District interaction: Grade 6	0.532	0.204	0.009
District interaction: Grade 7	1.058	0.274	<0.001
District interaction: Grade 8	0.437	0.223	0.050

**Exhibit F-22. Impact After Two Years of Mentoring – Cohort 2, Both RCT Districts Combined,
Mathematics (concluded)**

	Estimate	Standard Error	p-value
District interaction: Student black	0.018	0.072	0.801
District interaction: Student Hispanic	0.147	0.068	0.031
District interaction: Student FRPL	0.123	0.066	0.063
District interaction: Student SPED	0.098	0.066	0.136
District interaction: Student LEP	-0.208	0.085	0.015
District interaction: Student female	-0.022	0.041	0.590
District interaction: Teacher black	-0.511	0.164	0.002
District interaction: Teacher Bachelor's degree	-1.023	0.130	<0.001
District interaction: Teacher certification	-0.031	0.229	0.893
District interaction: Teacher female	0.220	0.159	0.165
District interaction: School percent FRPL	0.020	0.011	0.062
District interaction: School rating	0.326	0.090	<0.001
District interaction: School percent non-white	-0.003	0.012	0.789
District interaction: School percent student LEP	0.015	0.005	0.004
District interaction: School percent SPED	-0.037	0.016	0.023
Constant	-0.069	0.089	0.439
Random effect			
School	<0.001		
Teacher	0.011		
Student	0.203	0.006	0
n			
Schools	47		
Teachers	53		
Students	2166		

**Exhibit F-23. Impact After Two Years of Mentoring – Cohort 2, Both RCT Districts Combined,
Reading/ELA**

	Estimate	Standard Error	p-value
Treatment indicator	0.004	0.043	0.923
Student variables			
Prior reading score	0.685	0.019	<0.001
Grade 5	0.096	0.057	0.090
Grade 6	0.142	0.063	0.024
Grade 7	0.108	0.074	0.145
Grade 8	0.141	0.087	0.105
Student race is black CENTERED	-0.124	0.052	0.017
Student Hispanic - CENTERED	0.012	0.038	0.754
Student FRPL CENTERED	-0.107	0.038	0.005
Student special education CENTERED	-0.378	0.040	<0.001
Student limited English CENTERED	-0.073	0.046	0.117
Student female CENTERED	0.048	0.027	0.074
Teacher variables			
Teacher black CENTERED	0.398	0.081	<0.001
Teacher Bachelor's degree CENTERED	-0.054	0.052	0.296
Teacher partially certified CENTERED	-0.049	0.170	0.771
Teacher female CENTERED	0.112	0.054	0.037
School variables			
School percent FRPL CENTERED	-0.009	0.002	<0.001
School Report Card Rating CENTERED	-0.017	0.024	0.469
School percent minority CENTERED	0.007	0.003	0.005
School percent ELL CENTERED	-0.002	0.002	0.114
School percent IEP CENTERED	0.011	0.007	0.120
Blocking variables			
District B block 5	-0.064	0.065	0.324
District B block 8	0.053	0.084	0.529
District A block 1	0.145	0.083	0.081
District A block 2	-0.136	0.090	0.131
Interactions with district			
District indicator	-0.608	0.151	<0.001
District interaction: Reading pretest	-0.030	0.025	0.240
District interaction: Grade 5	0.597	0.116	<0.001
District interaction: Grade 6	0.499	0.128	<0.001
District interaction: Grade 7	0.746	0.143	<0.001
District interaction: Grade 8	0.577	0.170	0.001
District interaction: Student black	-0.002	0.071	0.980
District interaction: Student Hispanic	-0.067	0.063	0.287
District interaction: Student FRPL	0.082	0.061	0.178

**Exhibit F-23. Impact After Two Years of Mentoring – Cohort 2, Both RCT Districts Combined,
Reading/ELA (concluded)**

	Estimate	Standard Error	p-value
District interaction: Student SPED	0.086	0.056	0.125
District interaction: Student LEP	-0.082	0.066	0.217
District interaction: Student female	-0.029	0.039	0.451
District interaction: Teacher black	-0.487	0.101	<0.001
District interaction: Teacher Bachelor's degree	0.025	0.079	0.748
District interaction: Teacher certification	0.059	0.181	0.746
District interaction: Teacher female	-0.012	0.209	0.954
District interaction: School percent FRPL	0.001	0.008	0.885
District interaction: School rating	0.034	0.063	0.584
District interaction: School percent non-white	-0.004	0.010	0.662
District interaction: School percent student LEP	0.014	0.003	<0.001
District interaction: School percent SPED	-0.033	0.012	0.007
Constant	0.026	0.068	0.703
Random effects			
School	<0.001		
Teacher	0.004		
Student	0.263		
n			
Schools	54		
Teachers	63		
Students	2957		

Cohort 2, District A

Exhibit F-24. Impact After Two Years of Mentoring – District A Cohort 2, Mathematics

	Estimate	Standard Error	p-value
Treatment indicator	0.200	0.136	0.142
Student variables			
Prior math score	0.600	0.020	<0.001
Student in 5th grade	0.898	0.138	<0.001
Student in 6th grade	0.590	0.193	0.002
Student in 7th grade	0.953	0.201	<0.001
Student in 8th grade	0.766	0.190	<0.001
Student race is black CENTERED	-0.098	0.041	0.016
Student FRPL CENTERED	0.006	0.049	0.904
Student special education CENTERED	-0.111	0.046	0.015
Student limited English CENTERED	-0.221	0.071	0.002
Student female CENTERED	-0.007	0.032	0.818
Teacher variables			
Teacher black CENTERED	-0.615	0.099	<0.001
Teacher Bachelor's degree CENTERED	-0.896	0.095	<0.001
Teacher partially certified CENTERED	0.315	0.104	0.002
Teacher female CENTERED	0.478	0.146	0.001
School variables			
School percent FRPL CENTERED	0.022	0.010	0.020
School Report Card Rating CENTERED	0.293	0.076	<0.001
School percent minority CENTERED	-0.007	0.010	0.473
School percent ELL CENTERED	0.017	0.004	<0.001
School percent IEP CENTERED	-0.043	0.014	0.002
Blocking variables			
District A block 1	0.640	0.124	<0.001
District A block 2	0.013	0.151	0.932
Constant	-0.809	0.220	<0.001
Random effect			
School	<0.001		
Teacher	0.007		
Student	0.189		
n			
Schools	18		
Teachers	21		
Students	811		

Exhibit F-25. Impact After Two Years of Mentoring – District A Cohort 2, Reading/ELA

	Estimate	Standard Error	p-value
Treatment indicator	-0.015	0.069	0.826
Student variables			
Prior reading score	0.657	0.019	<0.001
Student in 5th grade	0.653	0.108	<0.001
Student in 6th grade	0.563	0.120	<0.001
Student in 7th grade	0.804	0.119	<0.001
Student in 8th grade	0.704	0.154	<0.001
Student race is black CENTERED	-0.097	0.038	0.011
Student FRPL CENTERED	-0.035	0.051	0.491
Student special education CENTERED	-0.291	0.042	<0.001
Student limited English CENTERED	-0.146	0.051	0.004
Student female CENTERED	0.017	0.030	0.568
Teacher variables			
Teacher black CENTERED	-0.080	0.045	0.078
Teacher Bachelor's degree CENTERED	-0.036	0.049	0.461
Teacher partially certified CENTERED	0.020	0.046	0.659
Teacher female CENTERED	0.166	0.177	0.347
School variables			
School percent FRPL CENTERED	-0.006	0.007	0.344
School Report Card Rating CENTERED	0.025	0.052	0.636
School percent minority CENTERED	0.001	0.007	0.880
School percent ELL CENTERED	0.011	0.002	<0.001
School percent IEP CENTERED	-0.024	0.008	0.004
Blocking variables			
District A block 1	0.156	0.070	0.026
District A block 2	-0.165	0.073	0.024
Constant	-0.538	0.135	<0.001
Random effect			
School	<0.001		
Teacher	<0.001		
Student	0.309		
n			
Schools	18		
Teachers	23		
Students	1410		

Cohort 2, District B

Exhibit F-26. Impact After Two Years of Mentoring – District B Cohort 2, Mathematics

	Estimate	Standard Error	p-value
Treatment indicator	0.045	0.093	0.629
Student variables			
Prior math score	0.732	0.017	<0.001
Student in 5th grade	0.084	0.078	0.278
Student in 6th grade	0.006	0.090	0.944
Student in 7th grade	-0.128	0.200	0.523
Student in 8th grade	0.390	0.106	<0.001
Student race is black CENTERED	-0.017	0.039	0.661
Student FRPL CENTERED	-0.144	0.042	0.001
Student special education CENTERED	-0.213	0.047	<0.001
Student limited English CENTERED	-0.002	0.044	0.961
Student female CENTERED	0.014	0.026	0.593
Teacher variables			
Teacher black CENTERED	-0.088	0.134	0.513
Teacher Bachelor's degree CENTERED	0.089	0.083	0.284
Teacher partially certified CENTERED	0.289	0.233	0.215
Teacher female CENTERED	0.178	0.087	0.041
School variables			
School percent FRPL CENTERED	-0.002	0.005	0.690
School Report Card Rating CENTERED	-0.057	0.034	0.089
School percent minority CENTERED	-0.001	0.006	0.884
School percent ELL CENTERED	0.003	0.003	0.361
School percent IEP CENTERED	0.007	0.014	0.608
Blocking variables			
District B block 5	0.025	0.087	0.773
District B block 8	-0.119	0.121	0.325
Constant	-0.069	0.101	0.496
Random effects			
School	<0.001		
Teacher	0.016		
Student	0.214		
n			
Schools	29		
Teachers	32		
Students	1355		

Exhibit F-27. Impact After Two Years of Mentoring – District B Cohort 2, Reading/ELA

	Estimate	Standard Error	p-value
Treatment indicator	0.033	0.056	0.559
Student variables			
Prior reading score	0.678	0.017	<0.001
Student in 5th grade	0.087	0.058	0.133
Student in 6th grade	0.139	0.059	0.019
Student in 7th grade	0.124	0.072	0.087
Student in 8th grade	0.137	0.083	0.101
Student race is black CENTERED	-0.142	0.043	0.001
Student FRPL CENTERED	-0.108	0.034	0.001
Student special education CENTERED	-0.383	0.038	<0.001
Student limited English CENTERED	-0.074	0.043	0.087
Student female CENTERED	0.046	0.024	0.057
Teacher variables			
Teacher black CENTERED	0.418	0.094	<0.001
Teacher Bachelor's degree CENTERED	-0.049	0.057	0.386
Teacher partially certified CENTERED	-0.039	0.178	0.825
Teacher female CENTERED	0.130	0.058	0.025
School variables			
School percent FRPL CENTERED	-0.009	0.002	<0.001
School Report Card Rating CENTERED	-0.023	0.027	0.398
School percent minority CENTERED	0.007	0.003	0.015
School percent ELL CENTERED	-0.003	0.002	0.132
School percent IEP CENTERED	0.006	0.008	0.483
Blocking variables			
District B block 5	-0.076	0.071	0.284
District B block 8	0.040	0.093	0.665
Constant	0.015	0.072	0.830
Random effects			
School	0.007		
Teacher	0.001		
Student	0.221		
n			
Schools	36		
Teachers	40		
Students	1547		

APPENDIX G. SENSITIVITY TESTS FOR RCT RESULTS

Following the student achievement analyses, we carried out several tests of the sensitivity of our results to different specifications of the model. The first of these tests (late joiners) was reported in Appendix F. This section is designed to address four additional questions:

1. Is the estimated impact of the NTC model on student achievement different for elementary school students than for middle school students?
2. Is the relationship between the NTC induction model and student achievement affected by students taking mathematics or reading/ELA classes with more than one study teacher?
3. Does the relationship between the NTC induction model and student achievement vary by school characteristics?
4. Are measures of the frequency, duration, or quality of the mentoring experience associated with higher student achievement?

These analyses used multi-level regressions of the same structure as the achievement analysis reported in Appendix A, with students nested within teachers nested within schools, controls for student variables, teacher variables, school variables, and blocking variables, and interactions with cohort and district indicators.

Is the estimated impact of the NTC model on student achievement different for elementary school students than for middle school students?

The student achievement impact estimates combined students in grades 4 through 8, with controls for grade level included in the models. However, previous studies have found that achievement gains often vary substantially by grade level (e.g., Lipsey, et. al, 2012).¹⁶ Therefore, we examined the impact of the NTC model on student achievement for elementary and middle school grades separately and tested whether the impact varied.¹⁷

The overall impact of the NTC induction model after 2 years of support in ELA was statistically significant in elementary school, but not in middle school. However, the difference in the size of these impacts was not statistically significant. The overall impact in mathematics was statistically significant in both middle school and elementary school, and the difference in impact by grade level was not statistically significant (Exhibit G-1). This result indicates that the impact of the NTC model on student achievement does not vary significantly by school level.

¹⁶ 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*. (NCSE 2013–3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences (IES), U.S. Department of Education. This report is available on the IES website, <http://ies.ed.gov/ncser/>

¹⁷ The sample sizes at each grade level were too small to examine impact estimates for each grade separately.

Exhibit G-1. Estimates of the Impact of NTC on Student Achievement Overall, by School Level, Two Years of Induction Support, RCT Districts

		Treatment	Control	Impact	<i>n</i> Schools	<i>n</i> Teachers	<i>n</i> Students
Reading/ELA	Elementary	0.07	-0.03	0.10*	69	94	2,691
	Middle school	0.04	-0.03	0.08	59	78	3,456
	Difference in impacts			0.02	99	149	6,147
Mathematics	Elementary	0.03	-0.12	0.15*	60	86	2,258
	Middle school	0.15	0.01	0.14*	45	59	2,714
	Difference in impacts			0.01	86	129	4,972

* $p < 0.05$

Is the relationship between the NTC induction model and student achievement affected by students taking mathematics or ELA classes with more than one study teacher?

Some students in both BCPS and CPS took more than one mathematics or ELA class in the same year, usually with two or more different teachers. If these students appeared in the classroom for more than one study teacher, they were included in the estimation for both teachers. However, this approach double-counted some students. In the mathematics analysis, 462 students appeared in more than one teacher's classroom. In the ELA analysis, 1,716 students appeared in more than one teacher's classroom. In this analysis we tested whether excluding these duplicate students affects the results.

Excluding duplicate students slightly decreased the size of the estimates of the impact of the NTC model after 2 years of mentoring in mathematics and ELA (Exhibit G-2). The impact of the NTC model on mathematics achievement remained positive and statistically significant when duplicate students were excluded, with a slight decrease in the estimate (from 0.15 to 0.11 standard deviation). The impact of the NTC model on ELA achievement was reduced by only 0.01 SD, but the resulting coefficient was not statistically significant. This sensitivity analyses drastically reduced the sample (from 6,147 students to 4,431 students), which affected our power to detect an impact.

Exhibit G-2. Estimates of the Impact of the NTC on Student Achievement Overall, Excluding Duplicate Students, Two Years of Induction Support, RCT Districts

		Impact	<i>n</i> Schools	<i>n</i> Teachers	<i>n</i> Students
ELA	Original estimate	0.09*	99	149	6,147
	Excluding duplicate students	0.08	87	127	4,431
Mathematics	Original estimate	0.15**	86	129	4,972
	Excluding duplicate students	0.11*	86	125	4,510

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

Does the relationship between the NTC induction model and student achievement vary by school characteristics?

After 2 years of mentoring, there were positive impacts in both reading and mathematics. However, these main effects may mask a variation in the impact of the model depending on school characteristics. For example, the NTC model may have a stronger impact in schools with a greater proportion of minority students. Therefore, we tested interactions between school characteristics and student achievement outcomes.

The overall impact of NTC induction on student achievement after 2 years of induction support did not significantly vary by:

- School report card rating
- Percent English learners
- Percent free or reduced-price lunch (FRPL)
- Percent minority
- Percent individualized education program (special education)

However, the impact of the NTC induction model on student achievement after 2 years of induction support did vary by school characteristics in District A. In schools with an above-average proportion of students receiving FRPL,¹⁸ the impact of NTC on student achievement in mathematics was not statistically significant. However, in schools with a *below*-average proportion of students receiving FRPL, the impact was statistically significant and negative. The difference between the impact in the two types of schools was statistically significant. (Exhibit G-3).

Conversely, in schools in District A with a below-average proportion of students with individualized education programs (IEPs), the impact of NTC on student achievement in mathematics was not statistically significant. However, in schools with an *above*-average proportion of students with IEPs, the impact was statistically significant and negative. The difference between the impact in the two types of schools was statistically significant (Exhibit G-3).

We conducted similar analyses for District B but there were no statistically significant results.

¹⁸ District averages at the time of randomization were used.

Exhibit G-3. Estimates of the Impact of NTC on Student Achievement in District A, by School Demographics, Two Years of Induction Support

		Treatment	Control	Impact	<i>n</i> schools	<i>n</i> teachers	<i>n</i> students
Percent FRPL	Above average	-0.41	-0.42	0.01	22	36	1,358
	Below average	-0.13	0.24	-0.38*	9	12	435
	Difference in impacts			0.39*	31	48	1,793
Percent IEP	Above average	-0.39	-0.05	-0.34*	13	17	486
	Below average	-0.19	-0.26	0.07	18	31	1,307
	Difference in impacts			-0.41*	31	48	1,793

* $p < 0.05$

Are measures of the frequency, duration, or quality of the mentoring experience associated with higher student achievement?

The estimates of the impact of the NTC model on student achievement compared the entire treatment group to the control group. However, implementation of the NTC model varied by treatment teacher, and some aspects of the model were also experienced by control teachers. Therefore, we used teacher survey and Learning Zone data to examine correlations between frequency, duration, and quality of mentoring and student achievement. Because teachers were not randomly assigned to different levels of mentoring, this analysis is *strictly correlational* and cannot be used to infer a causal relationship between these aspects of mentoring and student achievement.

Survey measures were available on both treatment and control teachers, but were self-reported:

- Presence of a mentor (Yes/No)
- Frequency of meeting with a mentor (At least weekly vs. Less than weekly)
- Duration of meetings with mentor (An hour or more vs. Less than an hour)
- Frequency of high-leverage mentoring activities¹⁹ (Continuous)
- Value of mentoring activities to the teacher²⁰ (Continuous)

Learning Zone measures were available only for treatment teachers:

- Average number of minutes of mentoring per month (Continuous)
- Percent of in-person mentoring sessions where a Formative Assessment and Support System (FAS) tool was used (Continuous)

¹⁹ The items that make up this scale are listed in Appendix G.

²⁰ The items that make up this scale are listed in Appendix G.

All teachers (both treatment and control) who had students in the achievement analysis also reported having a mentor. With no variation, we could not test the relationship between having a mentor and teachers' student achievement.

There was a significant positive relationship between student mathematics achievement and (1) the self-reported frequency of teachers' meetings with mentors, (2) the self-reported length of mentoring meetings, and (3) the self-reported frequency of high-leverage mentoring activities. Teachers who reported meeting at least weekly with their mentors had students with mathematics scores 0.13 standard deviation higher than those who met with their mentors less often ($p < 0.01$). Teachers who reported meeting with their mentors for an hour or more had students with mathematics test scores 0.14 standard deviation higher than those who met with their mentors for a shorter time period. Finally, an increase of one standard deviation in high-leverage mentoring activities was associated with an increase of 0.11 standard deviation in mathematics achievement (Exhibit G-4).

In reading/ELA, there was a statistically significant positive association between student achievement and the length of mentoring meetings. Teachers who reported meeting with their mentors for an hour or more had students with reading/ELA test scores 0.17 standard deviation higher than those who met with their mentors for a shorter time period. There was no statistically significant relationship between student achievement in reading/ELA and the frequency of mentoring meetings or the frequency of high leverage mentoring activities (Exhibit G-4).

There was no statistically significant relationship detected between student achievement in either reading/ELA or mathematics and:

- Value of mentoring activities to the teacher
- Learning Zone measures of meeting length (within treatment only)
- Percent of in-person mentoring sessions where a Formative Assessment and Support System (FAS) tool was used (within treatment only)

Exhibit G-4. Estimates of the Relationship Between Survey Responses and Student Achievement, Cohort 1 Year 2 (2014–15) and Cohort 2 Year 2 (2015–16), RCT Districts

		Mathematics	ELA
Frequency of meetings (treatment and control combined)	Less than weekly	-0.13	0.03
	At least weekly	0.00	0.06
	Difference	0.13*	0.03
	<i>n</i> schools	64	72
	<i>n</i> teachers	89	97
	<i>n</i> students	3,592	4,255
Length of meetings (treatment and control combined)	Less than one hour	-0.15	-0.09
	An hour or more	-0.01	0.08
	Difference	0.14*	0.17*
	<i>n</i> schools	62	68
	<i>n</i> teachers	84	91
	<i>n</i> students	3,380	4,050
Frequency of high-leverage mentoring activities (treatment and control combined)	Average frequency of activities	-0.23	0.00
	One SD above average	-0.12	0.02
	Difference	0.11*	0.02
	<i>n</i> schools	63	71
	<i>n</i> teachers	87	96
	<i>n</i> students	3,538	4,206

* $p < 0.05$

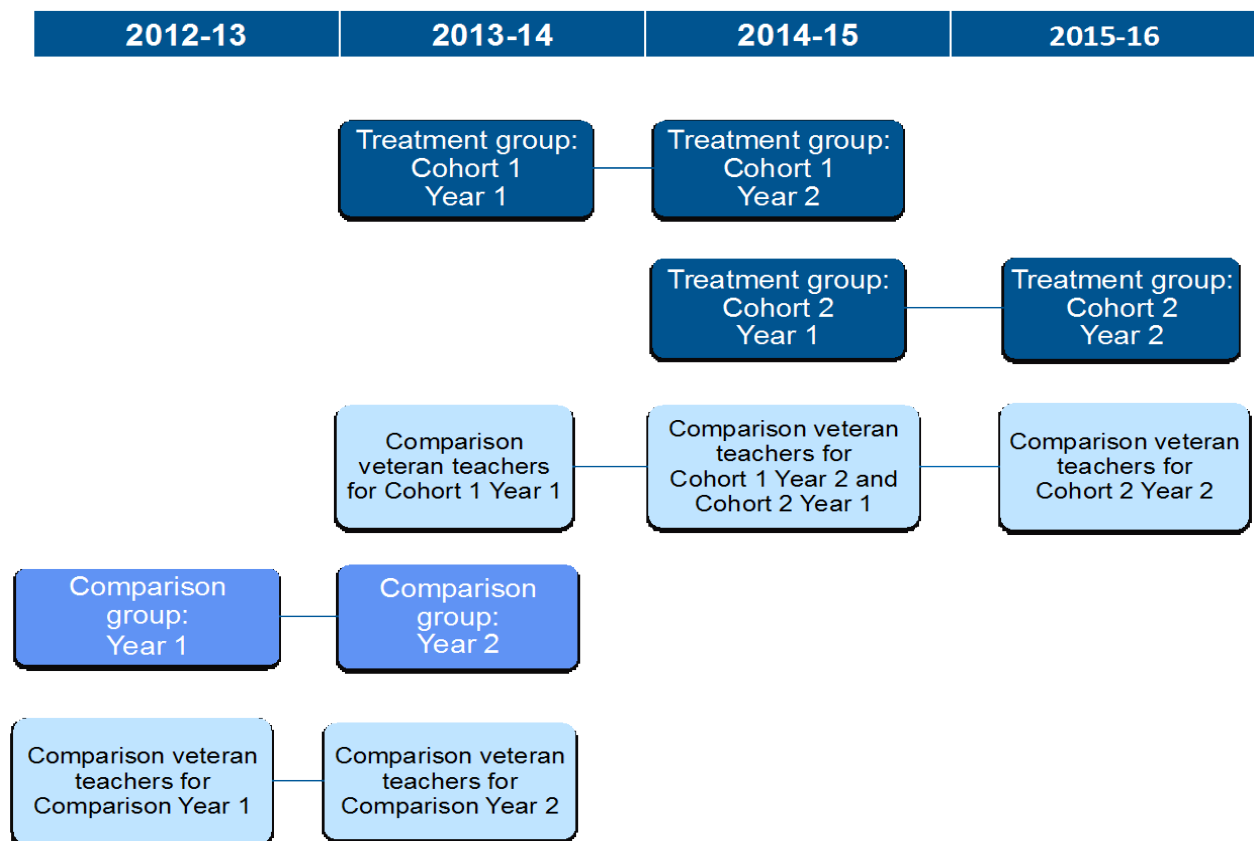
APPENDIX H. QED STUDY METHODS AND STUDENT ACHIEVEMENT MODEL RESULTS

This appendix presents the methods and results for student achievement analysis using a difference-in-differences design in GWAEA.

Difference-in-Differences Design

Because all new teachers in the GWAEA participating districts began receiving NTC mentoring in 2013–14, we could not use random assignment to study impact in GWAEA. Instead, we applied a quasi-experimental difference-in-differences approach to estimate the impact of NTC mentoring on student achievement. That is, we compared new teachers in GWAEA participating districts who began teaching in 2012–13 (comparison group) and did not receive NTC mentoring to new teachers in the same districts who began teaching in 2013–14 (treatment group: Cohort 1) and in 2014–15 (treatment group: Cohort 2), when NTC mentoring was offered to all new teachers, adjusting for differences between veteran teachers in the same school and year as each of these two cohorts of new teachers. Exhibit H-1 presents the difference-in-differences design.

Exhibit H-1. Schematic of Treatment and Control Teachers, Years 1 and 2



Baseline Equivalence

Exhibits H-2 through 4 detail the baseline equivalence among students of the four groups of teachers for the difference-in-differences analyses for Cohort 1 Year 1, Cohort 2 Year 1, and Cohort 2 Year 2.

Exhibit H-2. Cohort 1 Year 1 Baseline Student Test Scores, by Groups of Teachers

	Cohort 1 Teachers (Year 1 Treated, 2013-14)	Comparison Teachers (Year 1, 2012-13)	Comparison Veteran Teachers for Cohort 1 Teachers (2013-14)	Comparison Veteran Teachers for Comparison Teachers (2012-13)
ELA				
Mean	-0.13	0.15	0.00	0.11
SD	1.01	1.06	1.01	1.03
<i>n</i> students	204	217	773	892
Mathematics				
Mean	-0.28	-0.08	0.06	-0.02
SD	1.08	0.95	1.04	0.98
<i>n</i> students	214	178	708	398
ELA (weighted)				
Mean	-0.13	-0.04	-0.17	-0.08
SD	1.01	1.04	1.00	1.02
<i>n</i> students	204	217	773	892
Mathematics (weighted)				
Mean	-0.28	-0.30	-0.20	-0.25
SD	1.08	0.98	1.04	0.98
<i>n</i> students	214	178	708	398

Exhibit H-3. Cohort 2 Year 1 Baseline Student Test Scores, by Groups of Teachers

	Cohort 2 Teachers (Year 1 Treated, 2014-15)	Comparison Teachers (Year 1 2012-13)	Comparison Veteran Teachers for Cohort 2 Teachers (2014-15)	Comparison Veteran Teachers for Comparison Teachers (2012-13)
ELA				
Mean	-0.08	0.15	0.05	0.11
SD	1.00	1.06	1.01	1.03
<i>n</i> students	506	217	2238	892
Mathematics				
Mean	0.11	-0.08	0.02	-0.02
SD	1.08	0.95	1.01	0.98
<i>n</i> students	358	178	1815	398

Exhibit H-4. Cohort 2 Year 2 Baseline Student Test Scores, by Groups of Teachers

	Cohort 2 Teachers (Year 2 Treated, 2015-16)	Comparison Teachers (Year 2 2013-14)	Comparison Veteran Teachers for Cohort 1 Teachers (2015-16)	Comparison Veteran Teachers for Comparison Teachers (2013-14)
ELA				
Mean	0.06	0.10	0.05	0.01
SD	0.93	0.89	1.03	1.08
<i>n</i> students	424	143	928	642
Mathematics				
Mean	0.01	-0.05	0.09	-0.12
SD	1.12	0.93	1.01	0.98
<i>n</i> students	360	126	2086	210

Distribution of Outcomes

Exhibits H-5 through H-8 provide the means and standard deviations of standardized outcome scores for each of the four groups of teachers included in each of the four impact analyses: Cohort 1 Year 1, Cohort 1 Year 2, Cohort 2 Year 1, and Cohort 2 Year 2.

Exhibit H-5. Cohort 1 Year 1 Outcome Student Test Scores, by Groups of Teachers

	Cohort 1 Teachers (Year 1 Treated, 2013-14)	Comparison Teachers (Year 1, 2012-13)	Comparison Veteran Teachers for Cohort 1 Teachers (2013-14)	Comparison Veteran Teachers for Comparison Teachers (2012-13)
ELA				
Mean	-0.13	0.19	0.02	0.05
SD	1.04	1.08	1.06	1.01
<i>n</i> students	204	217	773	892
Mathematics				
Mean	-0.27	-0.14	0.05	-0.14
SD	1.03	0.94	1.04	0.93
<i>n</i> students	214	178	708	398

Exhibit H-6. Cohort 1 Year 2 Outcome Student Test Scores, by Groups of Teachers

	Cohort 1 Teachers (Year 2 Treated, 2014-15)	Comparison Teachers (Year 2 2013-14)	Comparison Veteran Teachers for Cohort 1 Teachers (2014-15)	Comparison Veteran Teachers for Comparison Teachers (2013-14)
ELA				
Mean	-0.17	0.14	-0.19	0.08
SD	1.03	0.93	1.02	0.99
<i>n</i> students	340	194	1189	1305
Mathematics				
Mean	-0.21	0.19	0.01	0.11
SD	0.97	0.97	1.01	0.96
<i>n</i> students	533	299	1822	1140

Exhibit H-7. Cohort 2 Year 1 Outcome Student Test Scores, by Groups of Teachers

	Cohort 2 Teachers (Year 1, Treated, 2014-15)	Comparison Teachers (Year 1, 2012-13)	Comparison Veteran Teachers for Cohort 2 Teachers (2014-15)	Comparison Veteran Teachers for Comparison Teachers (2012-13)
ELA				
Mean	-0.06	0.19	0.08	0.05
SD	0.97	1.08	0.99	1.01
<i>n</i> students	506	217	2238	892
Math				
Mean	0.14	-0.14	0.08	-0.14
SD	1.06	0.94	1.00	0.93
<i>n</i> students	358	178	1815	398

Exhibit H-8. Cohort 2 Year 2 Outcome Student Test Scores, by Groups of Teachers

	Cohort 2 Teachers (Year 2 Treated, 2015-16)	Comparison Teachers (Year 2, 2013-14)	Comparison Veteran Teachers for Cohort 2 Teachers (2015-16)	Comparison Veteran Teachers for Comparison Teachers (2013-14)
ELA				
Mean	0.01	0.15	0.01	0.00
SD	0.99	0.91	1.01	1.04
<i>n</i> students	424	143	928	642
Math				
Mean	-0.04	0.00	0.07	-0.01
SD	1.08	0.88	1.00	0.93
<i>n</i> students	360	126	2086	210

Statistical Analysis

To analyze the impact of the NTC model on student achievement, we estimated a three-level difference-in-differences model for first and second year impact of Cohort 1 and Cohort 2 NTC teachers respectively, with students nested within teachers nested within schools. Because treatment and comparison beginning teachers may be in the same schools, NTC impact was estimated at the teacher level. The model is:

$$y_{ijk} = \beta_0 + \beta_1 + (\textit{Beginning teacher}_{jk}) + \beta_2(\textit{Treatment Year}_{jk}) + \beta_3(\textit{Beginning teacher} * \textit{Treatment Year})_{jk} + \beta_m(\textit{mth student covariate}_{ijk}) + \beta_k(\textit{kth teacher covariate}_{jk}) + \beta_l(\textit{lth school covariate}_k) + r_{jk} + u_k$$

In this model, y_{ijk} represents the student's score on the reading/ELA or mathematics state assessment. *Beginning teacher* is an indicator for whether the student's teacher is a beginning teacher, with the reference being veteran teachers. The coefficient β_1 estimates the effect of having a beginning teacher, as compared to a veteran teacher, for comparison beginning teachers. The variable *Treatment Year* is an indicator of whether the student achievement outcome is taken in the first/second year of teaching for treatment teachers versus the first/second year of teaching for control teachers. Its coefficient, β_2 , estimates the impact of time: the difference in achievement between the year of treatment group achievement versus the year of the control group achievement. The interaction between these two variables provides the difference-in-differences estimate. Therefore, β_3 is the estimated impact of NTC on achievement for students with beginning teachers served by NTC.

Each model controls for student prior achievement, student background characteristics (grade level, race, free or reduced-price lunch [FRPL], special education status, limited English proficiency, and gender), teacher background characteristics (race, degree, gender, and whether the teacher is a special education teacher), school characteristics (school enrollment, percentage FRPL, percentage minority, percentage English language learners [ELL], percentage with Individualized Education Plans [IEP]), percentage passing reading or math tests, and an alternative school indicator.

The analyses were restricted to students with non-missing pretest and non-missing posttest assessment data. Less than 5% of the sample was missing data on other covariates, and we used list-wise deletion to handle missing values on those covariates.

Full Model Results

Exhibits H-9 through H-16 are the full model tables for the student achievement impact analysis. Although many of these variables are not statistically significant, we include them in all models to increase the precision with which we can estimate the treatment effect and to minimize omitted variable bias.

**Exhibit H-9. First-Year Impact on Student ELA Achievement,
Cohort 1 Teachers, QED Site**

ELA Standardized Score Model	Coefficient		SE
New teacher	0.080		0.108
Cohort 1 year	0.019		0.080
Cohort 1 year*New teacher	-0.123		0.135
Prior reading score	0.735	***	0.034
Tested in winter	-0.479	**	0.161
Tested in fall	-0.207		0.155
School demographics			
School enrollment	-0.000	**	0.000
% FRPL	-0.001		0.005
% Black and Hispanic students	-0.012		0.007
% IEP students	0.007		0.032
% EL students	-0.004		0.007
% students passing reading test	0.002		0.009
Alternative school	-0.222		1.067
Teacher background			
Teacher female	-0.030		0.097
Teacher Black or Hispanic	0.043		0.110
Teacher master's degree	0.068		0.075
Special education teacher	-0.127	*	0.056
Student demographics			
Female	0.083	**	0.025
Black	-0.109		0.066
Hispanic	-0.112		0.102
Asian	0.106		0.203
Free or reduce-d price lunch	-0.166	**	0.056
Special education	-0.253	***	0.062
English learner	-0.198	*	0.095
Grade 5	0.293		0.152
Grade 6	0.153		0.136
Grade 7	0.173		0.148
Grade 8	0.186		0.142
Constant	0.020		0.069

**Exhibit H-9. First-Year Impact on Student ELA Achievement,
Cohort 1 Teachers, QED Site (concluded)**

ELA Standardized Score Model	Coefficient	SE
Random effects		
School random effect	0.000	0.000
Teacher random effect	0.000	0.000
Student random effect	0.313	0.010
<i>n</i>		
Schools	20	
Teachers	80	
Students	2086	

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

**Exhibit H-10. First-Year Impact on Student Mathematics Achievement,
Cohort 1 Teachers, QED Site**

Math Standardized Score Model	Coefficient		SE
New teacher	-0.009		0.103
Cohort 1 year	-0.077		0.064
Cohort 1 year*New teacher	-0.067		0.142
Prior reading score	0.737	***	0.035
Tested in winter	0.261	**	0.089
Tested in fall	0.863	***	0.084
School demographics			
School enrollment	0.001	***	0.000
% FRPL	0.013	***	0.002
% Black and Hispanic students	0.012		0.007
% IEP students	-0.047	***	0.008
% EL students	-0.020	***	0.004
% students passing reading test	0.013	***	0.003
Alternative school	1.178	***	0.285
Teacher background			
Teacher female	-0.145	*	0.071
Teacher Black or Hispanic	0.054		0.182
Teacher master's degree	0.146	*	0.068
Special education teacher	0.108		0.115
Student demographics			
Female	-0.010		0.034
Black	-0.231	***	0.064
Hispanic	0.075	*	0.034
Asian	0.042		0.090
Free or reduced-price lunch	-0.097	**	0.037
Special education	-0.246	***	0.066
English learner	-0.084		0.105
Grade 5	0.134		0.079
Grade 6	-0.070		0.106
Grade 7	-0.114		0.080
Grade 8	0.135		0.073
Constant	-0.014		0.040
Random effects			
School random effect	0.000		.
Teacher random effect	0.000		0.000
Student random effect	0.263		0.031
<i>n</i>			
Schools	18		
Teachers	79		
Students	1498		

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

**Exhibit H-11. Second-Year Impact on Student ELA Achievement,
Cohort 1 Teachers,QED Site**

ELA Standardized Score Model	Coefficient		SE
New teacher	0.120	*	0.054
Cohort 1 year	-0.085	*	0.043
Cohort 1 year*New teacher	-0.078		0.065
Prior reading score	0.740	***	0.012
Tested in winter	-0.389	**	0.142
Tested in fall	0.005		0.058
Tested fall to spring	0.114	*	0.046
School demographics			
School enrollment	0.000	**	0.000
% FRPL	-0.003		0.002
% Black and Hispanic students	0.005		0.004
% IEP students	-0.006		0.007
% EL students	0.000		0.004
% students passing reading test	0.008		0.004
Alternative school	0.554		0.495
Teacher background			
Teacher female	-0.008		0.028
Teacher Black or Hispanic	0.000		0.000
Teacher master's degree	0.032		0.029
Special education teacher	-0.075		0.076
Student demographics			
Female	-0.007		0.025
Black	-0.088	**	0.033
Hispanic	-0.111	*	0.050
Asian	0.028		0.069
Free or reduced-price lunch	-0.142	***	0.024
Special education	-0.202	***	0.035
English learner	0.075		0.063
Grade 5	0.035		0.049
Grade 6	0.129		0.068
Grade 7	0.259	**	0.084
Grade 8	0.232	**	0.073
Constant	-0.023		0.025
Random effects			
School random effect	0.000		0.000
Teacher random effect	0.002		0.001
Student random effect	0.314		0.008
<i>n</i>			
Schools	23		
Teachers	121		
Students	3028		

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

**Exhibit H-12. Second-Year Impact on Student Mathematics Achievement,
Cohort 1 Teachers, QED Site**

Math Standardized Score Model	Coefficient	SE
New teacher	-0.037	0.034
Cohort 1 year	-0.092	0.055
Cohort 1 year*New teacher	-0.011	0.042
Prior reading score	0.776 ***	0.013
Tested in winter	0.003	0.087
Tested in fall	0.126	0.101
Tested fall to spring	0.863 ***	0.084
School demographics		
School enrollment	0.000 *	0.000
% FRPL	0.005	0.002
% Black and Hispanic students	-0.024 ***	0.006
% IEP students	0.001	0.005
% EL students	0.016 ***	0.004
% students passing reading test	0.002	0.005
Alternative school	0.501	0.378
Teacher background		
Teacher female	0.012	0.030
Teacher Black or Hispanic	0.000	0.000
Teacher master's degree	-0.028	0.024
Special education teacher	-0.095	0.091
Student demographics		
Female	-0.046	0.028
Black	-0.038	0.038
Hispanic	0.028	0.059
Asian	0.197	0.101
Free or reduced-price lunch	-0.069 *	0.034
Special education	-0.143 *	0.069
English learner	-0.149	0.113
Grade 5	-0.103	0.065
Grade 6	0.031	0.066
Grade 7	-0.027	0.054
Grade 8	-0.055	0.050
Constant	0.087	0.039
Random effects		
School random effect	0.000	0.000
Teacher random effect	0.000	0.000
Student random effect	0.275	0.124
<i>n</i>		
Schools	22	
Teachers	125	
Students	3794	

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

**Exhibit H-13. First-Year Impact on Student ELA Achievement,
Cohort 2 Teachers, QED Site**

ELA Standardized Score Model	Coefficient		SE
New teacher	0.052		0.065
Cohort 1 year	0.063		0.046
Cohort 1 year*New teacher	-0.081		0.073
Prior reading score	0.752	***	0.010
Tested in winter	-0.088		0.055
Tested in fall	0.058		0.057
School demographics			
School enrollment	-0.000		0.000
% FRPL	0.001		0.003
% Black and Hispanic students	-0.009	*	0.004
% IEP students	0.015	*	0.008
% EL students	0.015		0.012
% students passing reading test	0.002		0.003
Alternative school	-0.188		0.156
Teacher background			
Teacher female	-0.005		0.035
Teacher Black or Hispanic	-0.015		0.088
Teacher master's degree	0.060		0.034
Special education teacher	-0.224	**	0.079
Student demographics			
Female	0.007		0.031
Black	-0.064	*	0.031
Hispanic	0.050		0.045
Asian	0.052		0.064
Free or reduced-price lunch	-0.139	***	0.021
Special education	-0.203	***	0.032
English learner	-0.181	*	0.079
Grade 5	-0.013		0.052
Grade 6	0.054		0.050
Grade 7	0.036		0.056
Grade 8	0.027		0.062
Constant	0.018		0.037
Random effects			
School random effect	0.000		0.000
Teacher random effect	0.008		0.002
Student random effect	0.297		0.008
<i>n</i>			
Schools	24		
Teachers	144		
Students	3853		

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

**Exhibit H-14. First-Year Impact on Student Mathematics Achievement,
Cohort 2 Teachers, QED Site**

Math Standardized Score Model	Coefficient	SE
New teacher	-0.019	0.079
Cohort 1 year	0.171 *	0.067
Cohort 1 year*New teacher	-0.025	0.090
Prior reading score	0.795 ***	0.011
Tested in winter	0.105	0.081
Tested in fall	0.322 ***	0.084
School demographics		
School enrollment	0.000	0.000
% FRPL	-0.002	0.004
% Black and Hispanic students	-0.002	0.006
% IEP students	-0.007	0.010
% EL students	0.004	0.019
% students passing reading test	-0.014 **	0.005
Alternative school	0.075	0.154
Teacher background		
Teacher female	0.014	0.042
Teacher Black or Hispanic	-0.080	0.154
Teacher master's degree	-0.033	0.045
Special education teacher	-0.148	0.086
Student demographics		
Female	0.022	0.038
Black	-0.045	0.031
Hispanic	0.075	0.049
Asian	0.125	0.066
Free or reduced-price lunch	-0.116 ***	0.022
Special education	-0.174 ***	0.035
English learner	-0.109	0.106
Grade 5	0.190 **	0.062
Grade 6	0.134 *	0.068
Grade 7	0.152 *	0.072
Grade 8	0.121	0.077
Constant	-0.078	0.058
Random effects		
School random effect	0.000	0.000
Teacher random effect	0.013	0.004
Student random effect	0.236	0.007
<i>n</i>		
Schools	23	
Teachers	109	
Students	2749	

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

**Exhibit H-15. Second-Year Impact on Student ELA Achievement,
Cohort 2 Teachers, QED Site**

ELA Standardized Score Model	Coefficient	SE
New teacher	0.083	0.061
Cohort 1 year	-0.035	0.055
Cohort 1 year*New teacher	-0.092	0.068
Prior reading score	0.758 ***	0.014
Tested in winter	-0.060	0.069
Tested in fall	-0.215	0.112
School demographics		
School enrollment	-0.000	0.000
% FRPL	-0.004	0.003
% Black and Hispanic students	0.003	0.003
% IEP students	0.039 *	0.017
% EL students	-0.013	0.015
% students passing reading test	0.012	0.007
Alternative school	0.000	0.000
Teacher background		
Teacher female	-0.024	0.035
Teacher Black or Hispanic	0.000	0.000
Teacher master's degree	0.043	0.034
Special education teacher	0.002	0.102
Student demographics		
Female	0.033	0.033
Black	-0.159 ***	0.045
Hispanic	-0.034	0.068
Asian	0.091	0.092
Free or reduced-price lunch	-0.106 ***	0.029
Special education	-0.243 ***	0.042
English learner	-0.022	0.107
Grade 5	-0.012	0.048
Grade 6	-0.110	0.076
Grade 7	-0.305 ***	0.076
Grade 8	-0.206 **	0.074
Constant	0.034	0.037
Random effects		
School random effect	0.000	0.000
Teacher random effect	0.000	0.002
Student random effect	0.314	0.010
<i>n</i>		
Schools	17	
Teachers	79	
Students	2137	

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

**Exhibit H-16. Second-Year Impact on Student Mathematics Achievement,
Cohort 2 Teachers, QED Site**

Math Standardized Score Model	Coefficient		SE
New teacher	-0.071		0.100
Cohort 1 year	-0.213	*	0.096
Cohort 1 year*New teacher	-0.020		0.112
Prior reading score	0.788	***	0.011
Tested in winter	0.106		0.159
Tested in fall	0.438	**	0.168
School demographics			
School enrollment	0.001	**	0.000
% FRPL	0.032	***	0.006
% Black and Hispanic students	-0.032	**	0.011
% IEP students	-0.053	**	0.017
% EL students	0.079	**	0.024
% students passing reading test	0.016		0.009
Alternative school	0.000		0.000
Teacher background			
Teacher female	0.041		0.051
Teacher Black or Hispanic	0.000		0.000
Teacher master's degree	0.016		0.044
Special education teacher	-0.013		0.097
Student demographics			
Female	-0.066		0.049
Black	-0.142	***	0.033
Hispanic	-0.019		0.050
Asian	0.106		0.059
Free or reduced-price lunch	-0.082	***	0.023
Special education	-0.106	**	0.034
English learner	-0.075		0.080
Grade 5	0.035		0.063
Grade 6	0.277	**	0.086
Grade 7	0.174		0.111
Grade 8	-0.014		0.111
Constant	0.253	**	0.087
Random effects			
School random effect	0.000		.
Teacher random effect	0.014		0.004
Student random effect	0.226		0.006
<i>n</i>			
Schools	14		
Teachers	77		
Students	2782		

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

SRI Education

SRI Education, a division of SRI International, is tackling the most complex issues in education to identify trends, understand outcomes, and guide policy and practice. We work with federal and state agencies, school districts, foundations, nonprofit organizations, and businesses to provide research-based solutions to challenges posed by rapid social, technological and economic change. SRI International is a nonprofit research institute whose innovations have created new industries, extraordinary marketplace value, and lasting benefits to society.

Silicon Valley

(SRI International headquarters)
333 Ravenswood Avenue
Menlo Park, CA 94025
+1.650.859.2000

education@sri.com

Washington, DC Metro Area

1100 Wilson Boulevard, Suite
2800 Arlington, VA 22209
+1.703.524.2053

www.sri.com/education