NCEE 2010-4027

U.S. DEPARTMENT OF EDUCATION

Impacts of Comprehensive Teacher Induction

Final Results from a Randomized Controlled Study

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

This page is intentionally left blank.

Impacts of Comprehensive Teacher Induction

Final Results from a Randomized Controlled Study

June 2010

Steven Glazerman
Eric Isenberg
Sarah Dolfin
Martha Bleeker
Amy Johnson
Mary Grider
Matthew Jacobus
Mathematica Policy Research

Melanie Ali
Project Officer
Institute of Education Sciences

NCEE 2010-4027
U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

June 2010

The report was prepared for the Institute of Education Sciences under Contract No. ED-04-CO-0112/001. The project officer is Melanie Ali in the National Center for Education Evaluation and Regional Assistance.

IES evaluation reports present objective information on the conditions of implementation and impacts of the programs being evaluated. IES evaluation reports do not include conclusions or recommendations or views with regard to actions policymakers or practitioners should take in light of the findings in the reports.

To order copies of this report,

- Write to ED Pubs, Education Publications Center, U.S. Department of Education, P.O. Box 1398, Jessup, MD 20794-1398.
- Call in your request toll free to 1-877-4ED-Pubs. If 877 service is not yet available in your area, call 800-872-5327 (800-USA-LEARN). Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY) should call 800-437-0833.
- Fax your request to 301-470-1244.
- Order online at www.edpubs.org.

This report also is available on the IES website at http://ies.ed.gov/ncee.

Upon request, this report is available in alternate formats such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

ACKNOWLEDGMENTS

This study would not have been possible without the support and cooperation we received from many people. The technical working group for the study provided useful insight into and reactions to earlier presentations of methods and findings. The group included Carol Bartell, Mark Berends, Larry Hedges, Hamilton Lankford, Rebecca Maynard, Sandra Odell, Jeffrey Smith, and Todd Stinebrickner. We wish we could name the many dedicated individuals in the 17 participating school districts who provided us with data and cooperated with the study in many ways. These people were extraordinarily generous with their time and expertise.

At WestEd, Edward (Ted) Britton and Cathy Ringstaff did an excellent job coordinating the field visits and other work needed to monitor implementation of the induction programs. Staff from the two induction programs—Educational Testing Service and the New Teacher Center at the University of California, Santa Cruz—were extremely cooperative with study demands and generously shared all facets of their programs with us.

At Mathematica, Pat Nemeth oversaw a complex multi-wave data collection effort as the survey director, ably assisted by Linda Mendenko, Nakis Evgeniou, and Joe Baker. Phyllis Schulman and Susan Golden played critical roles in achieving high response rates to the surveys. Julieta Lugo-Gil, Megan Hague Angus, Andrea Mraz Esposito, Kristin Hallgren, Samina Sattar, Kathy Sonnenfeld, and Ama Takyi-Laryea coordinated with districts at several stages. Dominic Harris and Norma Altshuler provided expert programming assistance. Korantema Kaleem provided project management. Roberto Agodini carefully reviewed drafts of the report and provided helpful comments. The report was edited by Laura Bernstein and was produced by Donna Dorsey and Lisa Walls.

This study incorporates data on individual teachers' college entrance examination scores provided to Mathematica by the College Board and by ACT.

This page is intentionally left blank.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST[1]

The research team for this evaluation consists of a prime contractor, Mathematica Policy Research of Princeton, NJ, and one subcontractor, WestEd of San Francisco, CA. Neither of these organizations nor their key staff members have financial interests that could be affected by findings from the evaluation of the two comprehensive induction programs considered in this report. No one on the Technical Working Group, convened by the research team to provide advice and guidance, has financial interests that could be affected by findings from the evaluation.

[1] Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose other professional work may not be entirely independent of or separable from the tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

This page is intentionally left blank.

CONTENTS

This page intentionally left blank.

TABLES

Appendix Tables

FIGURES

Appendix Figures

EXECUTIVE SUMMARY

High teacher turnover and inadequate teacher preparation represent challenges for education policymakers. High turnover among teachers in urban school districts can hurt student achievement by exposing more students to inexperienced teachers (Darling-Hammond 2000); disrupt schools; and impose a high cost on districts that must recruit, hire, and train replacement teachers (Ingersoll and Smith 2003; King and Newmann 2000). Even teachers who persist may struggle with pedagogy or classroom management if they are not adequately supported early in their careers (Kauffman et al. 2002).

To support beginning teachers, most districts offer some form of teacher induction or mentoring, but they often provide a limited set of services in response to an unfunded state mandate and with modest local resources (Berry et al. 2002; Smith and Ingersoll 2004). We refer to this usual level of induction support as informal or low-intensity teacher induction, which may include pairing each new teacher with another full-time teacher without providing training, supplemental materials, or release time for the induction to occur.

One policy option in response to the problems of high turnover and inadequate preparation is to support teachers with a formal, more comprehensive induction program during their initial years in the classroom. Support that is intensive, structured, and sequentially delivered is sometimes referred to as "comprehensive" induction. It is often delivered through experienced, trained full-time mentors and may also include a combination of school and district orientation sessions, special in-service training (professional development), classroom observations, and constructive feedback through formative assessment.

In 2004, the U.S. Department of Education's Institute of Education Sciences contracted with Mathematica Policy Research to conduct a large-scale evaluation of comprehensive teacher induction. The purpose of the study was to determine whether augmenting the set of services districts usually provide to support beginning teachers with a more comprehensive program improves teacher and student outcomes. This is the study's third and final report on the program's impacts.

To evaluate the impact of comprehensive teacher induction relative to the usual induction support, we conducted a randomized experiment in a set of districts that were not already implementing comprehensive induction. We assigned 418 elementary schools in 17 urban districts by lottery to either (1) a treatment group whose beginning teachers were offered comprehensive teacher induction or (2) a control group whose beginning teachers received the district's usual, less comprehensive or intensive induction services. Random assignment ensures that any systematic differences in outcomes between the treatment and control group can be attributed to comprehensive induction.

The comprehensive services were provided by either the Educational Testing Service (ETS) of Princeton, New Jersey or the New Teacher Center at the University of California, Santa Cruz (NTC), depending on the district's preference. These program providers implemented their respective comprehensive programs in each school and district to which they were assigned. The providers began by helping districts select the mentors. Beginning teachers in treatment schools were then assigned to a full-time mentor with a 12 to 1 ratio. Mentors received ongoing training and a curriculum of materials to support the teachers' development. Beginning teachers were offered

monthly professional development sessions, opportunities to observe veteran teachers, and an end-of-year colloquium.

In 10 of the 17 districts, the services were offered to treatment schools for one year only ("one-year districts"). In the remaining 7 districts, services were offered to treatment schools for two years ("two-year districts"). Because the two sets of districts implemented different versions of the treatment and they were not randomly chosen to implement one or two years of comprehensive induction, we present most results separately for one- and two-year districts.

The research team collected survey and administrative data for four years after the initial random assignment in summer 2005. Teacher surveys were used to measure the effect of comprehensive induction on support services that teachers reported receiving and the impact it had on workforce outcomes (teacher attitudes, teacher retention, and composition of the teacher workforce) for the full sample of 1,009 teachers. We conducted classroom observations to measure the impact on teaching practices in the first year for the subsample of approximately 700 teachers who were teaching literacy skills. District-provided data on student test scores were used to measure the impact on test scores for the subsample of approximately 200 teachers in grades and subjects that had both an end of year test and a test of prior achievement from the previous year.

Key findings:

- **During the comprehensive induction program, treatment teachers received more support than control teachers.**[2] For example, in the first year they were more likely to have a mentor assigned to them (90 versus 72 percent in one-year districts and 96 versus 79 percent in two-year districts), spent more time with a mentor (85 versus 68 minutes per week on average in one-year districts and 108 versus 82 minutes per week on average in two-year districts), and participated in more activities such as observing other teachers (68 versus 39 percent in one-year districts and 72 versus 47 percent in two-year districts), as reported in the spring. The pattern of statistically significant differences favoring the treatment group continued in the second year for the districts where comprehensive services were offered over two years. However, treatment teachers in districts receiving one year of comprehensive induction received less support in their second year on all these dimensions than control teachers in the same districts. In the third and fourth years of teaching, treatment teachers received levels of support that were similar to their control group counterparts, whether we considered one-year or two-year districts.

[2] Unless stated, all comparisons in the executive summary and in the report are statistically significant at the 0.05 level using a two-sided hypothesis test. Statistical significance means that the observed differences are not likely due to chance.

- **The extra induction support for treatment teachers did not translate into impacts on classroom practices in the first year.** We observed teachers giving a literacy lesson in the spring of their first year and found no impacts on teachers' implementation of the literacy lesson, content of the literacy lesson, or classroom culture.

- **For teachers who received one year of comprehensive induction, there was no impact on student achievement.** In each of the first three years of teachers' careers, students of treatment teachers receiving one year of comprehensive induction support performed no better on average than students of the corresponding control teachers.

- **For teachers who received two years of comprehensive induction, there was no impact on student achievement in the first two years. In the third year, there was a positive and statistically significant impact on student achievement.**

  - In the third year, in districts and grades in which students' test scores from the current and prior year are available, students of treatment teachers outperformed students of the corresponding control teachers on average. These impacts are equivalent to effect sizes of 0.11 in reading and 0.20 in math, which is enough to move the average student from the 50th percentile up 4 percentile points in reading and 8 percentile points in math.

  - These results are based on the subset of data for which students' test scores from the current and prior year are available. If the analyses are conducted without requiring test scores from the prior year, we do not find an impact on math or reading scores. This alternative approach nearly doubles the available sample of study teachers but the lack of data on students' prior achievement results in a less precise estimate. This means that we are less likely to detect a true impact if it exists, despite the larger sample size.

- **Neither exposure to one year nor exposure to two years of comprehensive induction had a positive impact on retention or other teacher workforce outcomes**:

  - Treatment teachers did not report being more satisfied or feeling more prepared to teach than control teachers at any of the six time points over the four school years in which we collected data.

  - There was no impact on teacher retention over the first four years of the teachers' careers. This was true of retention in the original school, the original school district, and the teaching profession.

  - We found no evidence that comprehensive induction improved the composition of the teacher workforce through selective retention. There were no statistically significant positive differences between treatment teachers retained in the district and control teachers retained in the district in teacher characteristics such as college entrance exam scores, college selectivity, and advanced degrees; nor in performance measures such as first year classroom observation scores or third year student test scores.

Study Design

Participating Districts: 17 school districts in 13 states participated in the study. In addition to willingness to participate in the evaluation, districts had to meet criteria for size, poverty, and need for induction: at least 570 teachers in elementary schools, at least 10 elementary schools with 50 percent or more students eligible for free or reduced-price meals under the federal National School Lunch Program, and no existing comprehensive induction program offered in study schools. To be in the study, districts had to have schools with no full-time mentors and an expenditure on induction of less than $1,000 per beginning teacher. Although the districts did not form a statistically representative sample of the nation, they were drawn from states with a variety of regulatory, administrative, and demographic contexts. Treatment schools in each district worked with one of the two providers of comprehensive induction—ETS or NTC—based primarily on district preferences.

Participating Schools: 418 elementary schools participated in the study. Together with participating districts, we selected schools for the study that had eligible beginning teachers and that were not already implementing a comprehensive induction program.

Participating Teachers: 1,009 teachers participated in the study. Within each study school, all eligible teachers participated if they were new to the profession, taught in grades K-6, and were not already receiving induction support from a teacher preparation or certification program.

Random Assignment: Within each district, we randomly assigned schools to either the treatment group, in which case teachers at the school were offered comprehensive teacher induction, or to the control group, in which case teachers at the school took part in the district's usual set of induction services. Assigning entire schools helps ensure that teachers in the control group are not receiving the benefits of services offered to the treatment group.

Years of Treatment: The treatment included one year of comprehensive induction services for 10 districts ("one-year districts") and two years of such services for the other 7 districts ("two-year districts."). We selected a set of districts to receive a second year of the treatment based on such factors as whether the mentors were available for a second year. Dividing the sample in this way does not allow for and should not be used to make direct comparisons between districts that received a different number of years of treatment. Impacts are presented separately for one-year and two-year districts.

Outcomes: We examined impacts on classroom outcomes—evidence of best teaching practices through classroom observations and effects on student test scores—as well as workforce outcomes, including teacher satisfaction and preparedness, the rate of teacher retention, and the composition of the teacher workforce.

Model-based Approach: To estimate impacts, we used regression methods, which compare the treatment and control groups, controlling for the confounding effects of any chance differences in a range of student, teacher, or school characteristics, such as grade level, students' eligibility for free or reduced price lunch, or teacher certification. The regression model is also used to account for the fact that teachers or students are clustered within schools.

Data

Teacher Surveys: A teacher survey designed to measure the background of the study teachers, receipt of induction services and alternative support services, teacher attitudes, and mobility patterns was administered six times over four years. Response rates on teacher surveys ranged from 88 to 97 percent for the treatment group and 83 to 94 percent for the control group.

Classroom Observations: In the first year of the study, trained classroom observers used a rubric designed to quantify evidence of best teaching practice in literacy lessons in 639 classrooms. The observations focused on teachers responsible for English or language arts.

Student Achievement Data: Districts provided student test scores for regularly administered reading and math tests as well as associated student background data for each of the first three years of the study. For each year, the district provided annual test scores from the current year (posttest) and the prior year (pretest). Of the 1,009 teachers who began in the study in the 2005-2006 school year, districts provided valid student test score data for teachers in the most recent year of the study, the 2007-2008 school year. The other teachers were either no longer teaching in the district, teaching grades or subjects that were not tested, or teaching in one of the two districts that did not provide test score data for the 2007-2008 school year.

Induction Support for Beginning Teachers

To select a comprehensive induction program and program provider for the study, we issued a Request for Proposals (RFP) in 2004. The RFP specified that the induction program include several components that earlier research and professional wisdom gleaned from practice had suggested were important features of successful teacher induction programs (Alliance for Excellent Education 2004; Ingersoll and Smith 2004; Smith and Ingersoll 2004; Kelly 2004; Serpell and Bozeman 2000). A group of outside expert reviewers ranked the proposals and selected ETS and NTC as the providers whose programs most closely met the study's specified requirements. The two programs were roughly comparable in structure and included the required components:

- Carefully selected and trained full-time mentors
- A curriculum of intensive and structured support for beginning teachers that includes an orientation, professional development opportunities, and weekly meetings with mentors
- A focus on instruction, with opportunities for novice teachers to observe experienced teachers
- Formative assessment tools that permit evaluation of practice on an ongoing basis and require observations and constructive feedback
- Outreach to district and school-based administrators to educate them about program goals and to garner their systemic support for the program

Mathematica contracted with both providers to deliver comprehensive induction services to the districts in the study, with nine of the districts assigned to ETS, and the remaining eight to NTC. Staff from WestEd, a subcontractor for Mathematica, served on the implementation team and was charged with monitoring the implementation of the comprehensive induction services to help providers ensure fidelity to the core service model as well as identifying and helping address any implementation challenges that arose.

Both the ETS and NTC programs are based on a curriculum expected to promote effective teaching. The ETS program defines effective teaching in terms of 22 components organized into four domains of professional practice. The components are aligned with the Interstate New Teacher Assessment and Support Consortium (INTASC 1992) principles. The NTC induction model defines effective teaching in terms of six Professional Teaching Standards. Each standard, or domain, is broken into a succession of more discretely defined categories of teaching behaviors.

The curriculum that formed the foundation of both programs included a number of activities. Mentors were asked to meet weekly with treatment teachers for approximately two hours. Conversation was expected to center around the induction programs' teacher learning activities, but mentors also exercised professional judgment in selecting additional activities to meet beginning teachers' needs, including observing instruction or providing a demonstration lesson; reviewing lesson plans, instructional materials, or student work; or interacting with students to gain an additional perspective on teachers' instructional practices. Treatment teachers were provided monthly professional development sessions to complement their interactions with mentors, and the ETS districts also offered monthly study groups—mentor-facilitated peer support meetings for treatment teachers during which beginning teachers met monthly to discuss their local needs and practices. Treatment teachers also observed veteran teachers once or twice during the year. At the end of each school year, treatment teachers in both ETS and NTC districts participated in a colloquium celebrating the year's successes and teachers' professional growth.

The providers adapted the curricula of the second year of their usual induction programs for the second year of induction services in the two-year districts. While programs provided induction activities to these districts' treatment teachers during the second year that were similar to those in the first year, the content was designed to reflect the growth of mentors and beginning teachers and the evolution of their circumstances and needs. In two-year districts served by ETS, mentors led Teacher Learning Communities, an adaptation of the first year's study groups that included specific content for each session and a formal structure for teachers to try out approaches to instruction. During second year professional development sessions in the two-year districts served by NTC, mentors elaborated on standardized topics and designed activities to reflect local needs.

At the heart of the comprehensive induction services was the support provided by a full-time mentor trained by the program providers. The goal of the study was to assign each mentor to 12 beginning teachers. At the outset of the study, the program providers sought mentor candidates with a minimum of five years of teaching experience in elementary school, recognition as an exemplary teacher, and experience in providing professional development or mentoring other teachers (particularly beginning teachers). During Years 1 and 2, the providers brought their respective mentors together for 8-12 days of training spread across 3 to 4 sessions during the summer and school year. Trainings previewed the content of upcoming professional development sessions and gradually introduced processes of mentor/mentee work in such areas as reflecting on instructional practices and analyzing student work. All induction activities were voluntary for beginning teachers.

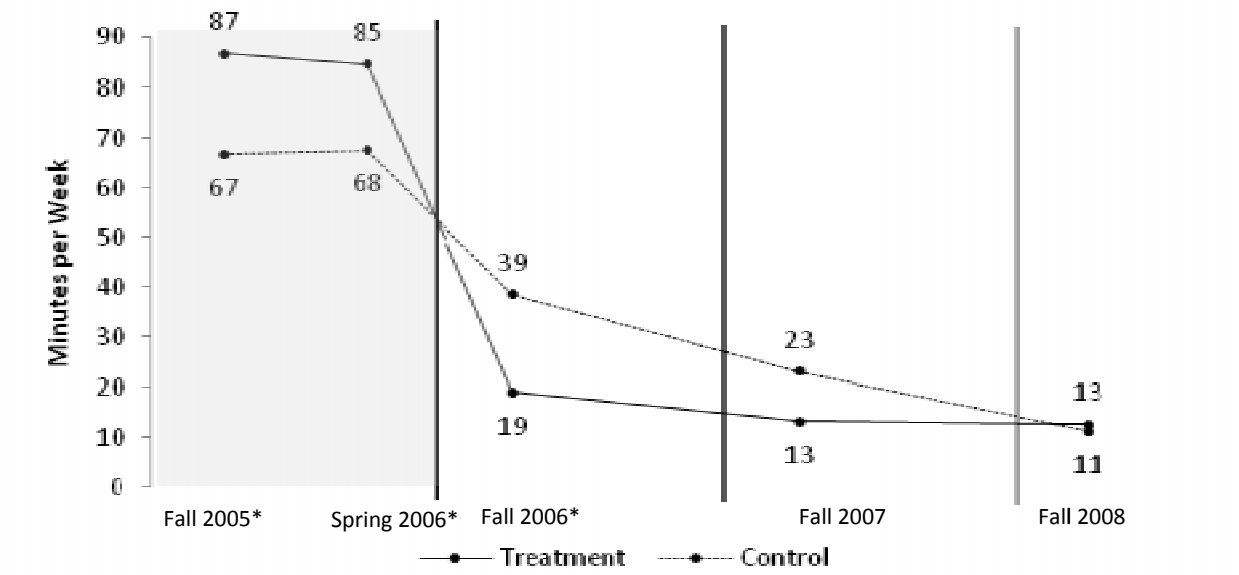
We found that assignment to the treatment group changed the pattern of induction services reported by beginning teachers. Figures ES.1 and ES.2 summarize the timing and intensity of the program's core support—mentoring—measured in minutes per week spent with mentors. Figure ES.1 presents data for one-year districts and Figure ES.2 presents data for two-year districts. The figures illustrate statistically significant differences between the treatment and control groups in weekly time spent with mentors during the intervention. Meeting time falls significantly for both treatment and control groups after the end of the intervention and treatment-control differences become statistically insignificant except for a significant negative difference in one-year districts in fall 2006. Mentoring is just one measure of induction support. We examined dozens of other measures and found similar patterns of support over time:

- **In one-year districts, both treatment and control teachers reported receiving substantial induction support. However, treatment teachers received more and different support than control teachers during the comprehensive induction program (their first year of teaching).** For instance, relative to control teachers, treatment teachers were more likely to have an assigned mentor (90 versus 70 percent in fall 2005, p-value 0.000; 90 versus 72 percent in spring 2006, p-value 0.000) and spent more time per week meeting with their mentors (87 versus 67 minutes in fall 2005, p-value 0.007; 85 versus 68 minutes in spring 2006, p-value 0.039); these differences were all statistically significant.

- **In two-year districts, treatment and control teachers reported receiving substantial induction support as well. However, similar to the findings in one-year districts, treatment teachers received more and different support than control teachers during the comprehensive induction program (their first two years of teaching).** For instance, relative to control teachers, treatment teachers were more likely to have an assigned mentor (between fall 2005 and spring 2007, the percent of treatment teachers ranged from 80 to 96 and the percent of control teachers ranged from 34 to 79; p-values all 0.000) and spent more time per week meeting with their mentors (between fall 2005 and spring 2007, time spent by treatment teachers ranged from 79 to 124 minutes and the time spent by control teachers ranged from 41 to 82 minutes; p-values ranged from 0.001 to 0.087); these differences were statistically significant with the exception of meeting time in spring 2006.

- **In their second year, immediately following the end of the comprehensive induction program, treatment teachers in one-year districts received less and different induction support than control teachers.** For measures such as the percentage of teachers with an assigned mentor and time spent meeting with mentors per week, this reflects a significant drop in support among control teachers and an even larger significant drop in support among treatment teachers. A survey of teachers in one-year districts conducted in fall 2006 showed that there were statistically significant differences favoring the control teachers in several areas: for instance, treatment teachers were less likely than control teachers to have an assigned mentor (20 percent of treatment teachers versus 29 percent of control teachers, p-value 0.017) and spent less time per week meeting with their mentors (19 minutes for treatment teachers versus 39 minutes for control teachers, p-value 0.002). No statistically significant differences favoring the treatment teachers were found.

Figure ES.1. Treatment-Control Differences in Total Minutes Spent in Mentoring per Week: One-Year Districts

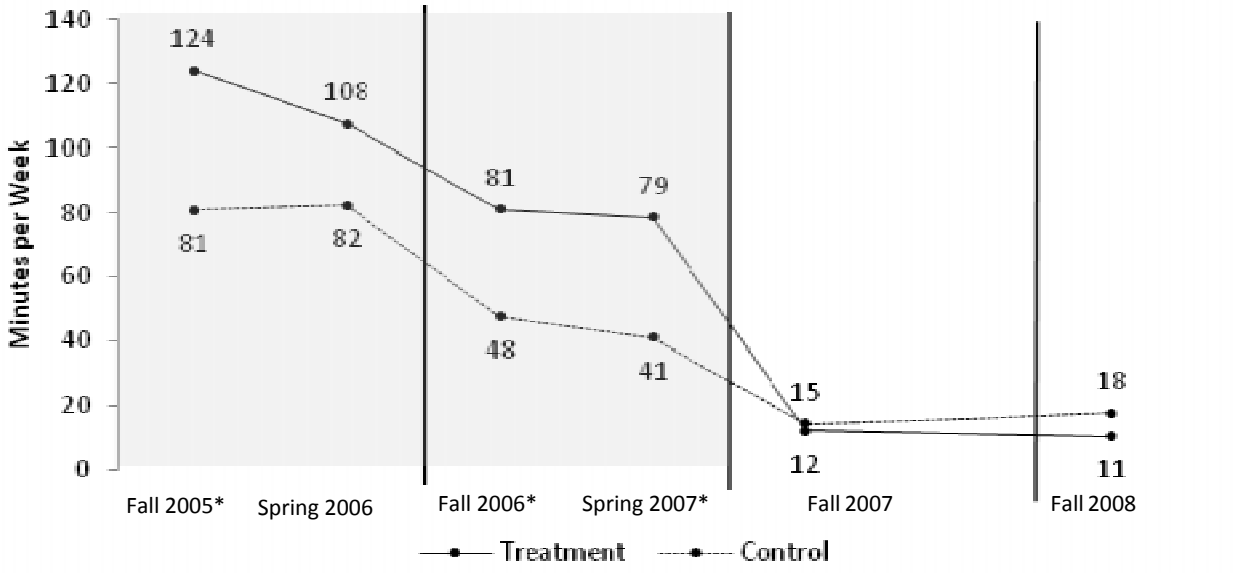


Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: N = 503 teachers in fall 2005, 499 teachers in spring 2006, 472 teachers in fall 2006, 426 teachers in fall 2007, and 398 teachers in fall 2008.

* Treatment-control difference is significantly different from zero at the 0.05 level.

Figure ES.2. Treatment-Control Differences in Total Minutes Spent in Mentoring per Week: Two-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: N = 395 teachers in fall 2005, 386 teachers in spring 2006, 360 teachers in fall 2006, 372 teachers in spring 2007, 326 teachers in fall 2007, and 321 teachers in fall 2008.

* Treatment-control difference is significantly different from zero at the 0.05 level.

- **In the third and fourth years of teaching, after the intervention ended for all districts, treatment and control teachers received similar levels of support.** In both one- and two-year districts, there were statistically significant differences in fewer than 7 percent of the 134 measures we surveyed.

Impacts on the Classroom

To measure the effect of comprehensive induction in the classroom, we compared these outcomes for teachers in the treatment and control groups: (1) the use of best practices in teaching a literacy lesson and (2) standardized test scores for teachers' students. For the classroom practices analysis, we focused on those teachers responsible for English language arts or literacy classes (698 teachers). For test score analyses, we focused on teachers in tested grades and subjects (about 200 teachers per year). Results pertaining to literacy instruction do not necessarily apply to teachers of other subjects. Similarly, results for teachers in tested grades do not necessarily apply to teachers of other grades or subjects.

Classroom Practices. We sent trained observers into treatment and control classrooms to administer the Diagnostic Classroom Observation (DCO) in spring 2006 (year 1 of the study). We observed literacy (or reading/language arts) lessons in 639 classrooms. Based on a set of 16 indicators, observers scored teachers on a five-point scale, ranging from "no evidence" to "extensive evidence" of effective teaching practice. We produced summary scores by averaging the indicators within each of three domains.

After controlling for teacher and school characteristics, we observed no statistically significant differences between treatment and control teachers' performance on the three domains measured by the DCO: implementation of a literacy lesson, content of a literacy lesson, or classroom culture.

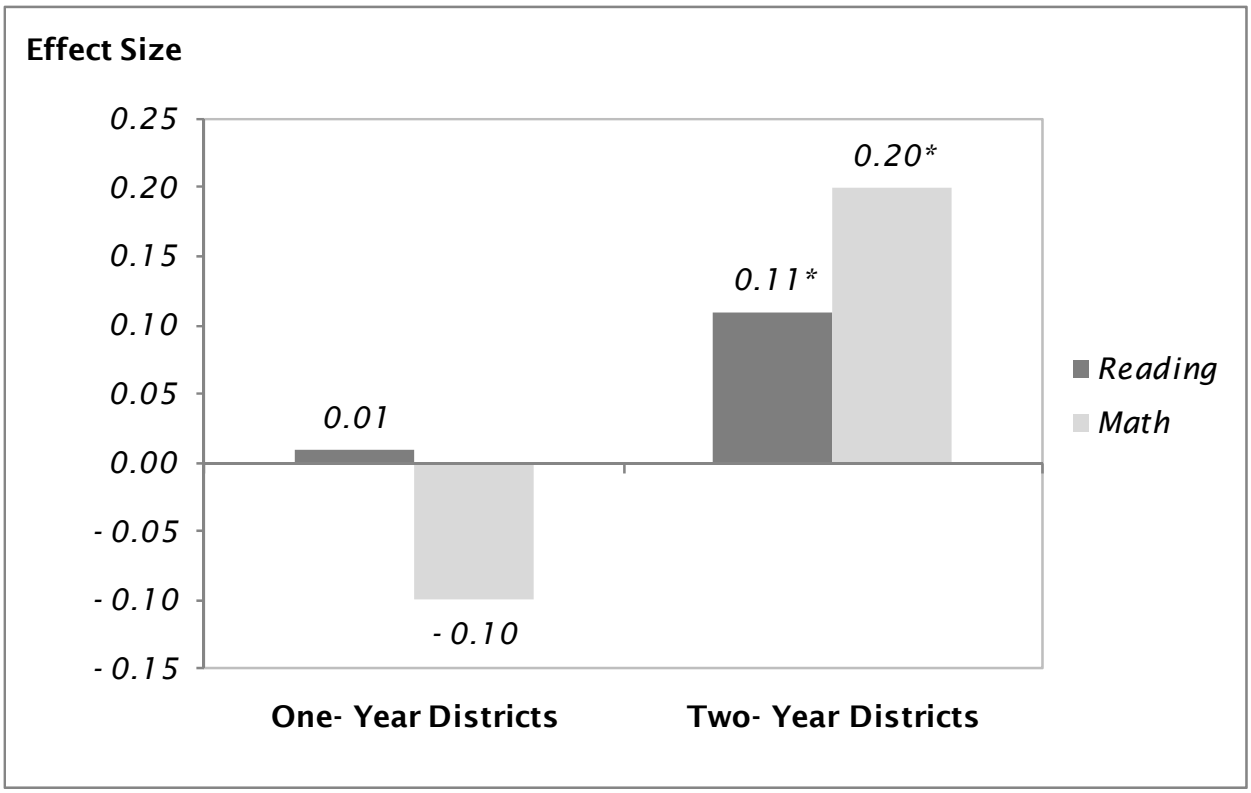
Student Achievement. To measure impacts on student achievement we compared the test scores for students of treatment teachers at the end of the year (posttest) to those of control teachers, accounting for any preexisting differences in prior achievement (pretest) and background characteristics of students and teachers. We present results in this report for the teachers' third year, the 2007–2008 school year.[3] This was the second year after the treatment ended in one-year districts, and the first year after the treatment ended in two-year districts.

For one-year districts, the impacts on math and reading scores in the study's third year were not significantly different from zero. For two-year districts, the impacts on math and reading scores were both positive and statistically significant. The results for the two-year districts, presented in Figure ES.3, show that comprehensive induction led to an increase in test scores of 11 percent of a standard deviation in reading, which is enough to move the average student from the 50th percentile up 4 percentile points, and an increase of 20 percent of a standard deviation in math scores, enough to move the average student up 8 percentile points.

[3] Results for the first year showed no overall impacts in either math or reading, as documented in an earlier report (Glazerman et al. 2008). In the second year, when we looked separately at one- and two-year districts, we continued to find no overall impacts in either subject (Isenberg et al. 2009).

As specified in the study design (Glazerman et al. 2005), the eligible sample for the test score analysis was limited to teachers in tested grades and subjects. Because our design also aimed to account for preexisting differences in student achievement, we included only students who took both a pretest and posttest, thereby excluding the lowest tested grade. For example, in many districts, the lowest tested grade was grade 3. If we expand the sample of teachers and grades by not requiring test scores from the prior year (approximately doubling the number of teachers included in the analyses), we do not find an impact on math or reading scores in either one-year or two-year districts. However, the lack of data on students' prior achievement produces a less precise estimate. This means that with the expanded sample we are less likely to detect a true impact if it exists, despite the larger sample size.

Figure ES.3. Impacts on Test Scores, Year 3 (grades with current and prior year tests)



Source: Mathematica analysis using data from the 2006–2007 and 2007–2008 school years provided by participating school districts; Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data are regression adjusted and account for clustering of students within schools. N = 99 teachers and 1,690 students in reading and 95 teachers and 1,629 students in math in one-year districts; 74 teachers and 1,347 students in reading and 68 teachers and 1,198 students in math in two-year districts.

*Treatment-control difference is significantly different from zero at the 0.05 level.

Impacts on the Teaching Workforce

To measure the effect of comprehensive induction on the teacher workforce, we examined the impacts of comprehensive induction on (1) teachers' attitudes that relate to career decisions, including their satisfaction with teaching and their feelings of preparedness to deal with different aspects of their jobs; (2) teachers' mobility; and (3) the mix of teachers who decide to stay in the district.

Teacher Attitudes. Using items from the induction activities surveys, we measured teachers' feelings of satisfaction in 19 areas on a four-point scale ranging from "very dissatisfied" to "very satisfied" and teachers' feelings of preparedness in 13 areas on a four-point scale from "not at all prepared" to "very well prepared." Factor analysis suggested that teacher satisfaction and teacher preparedness could be grouped into three categories each: satisfaction with (1) school, (2) class, and (3) career; and preparedness to (1) instruct, (2) work with students, and (3) work with others.

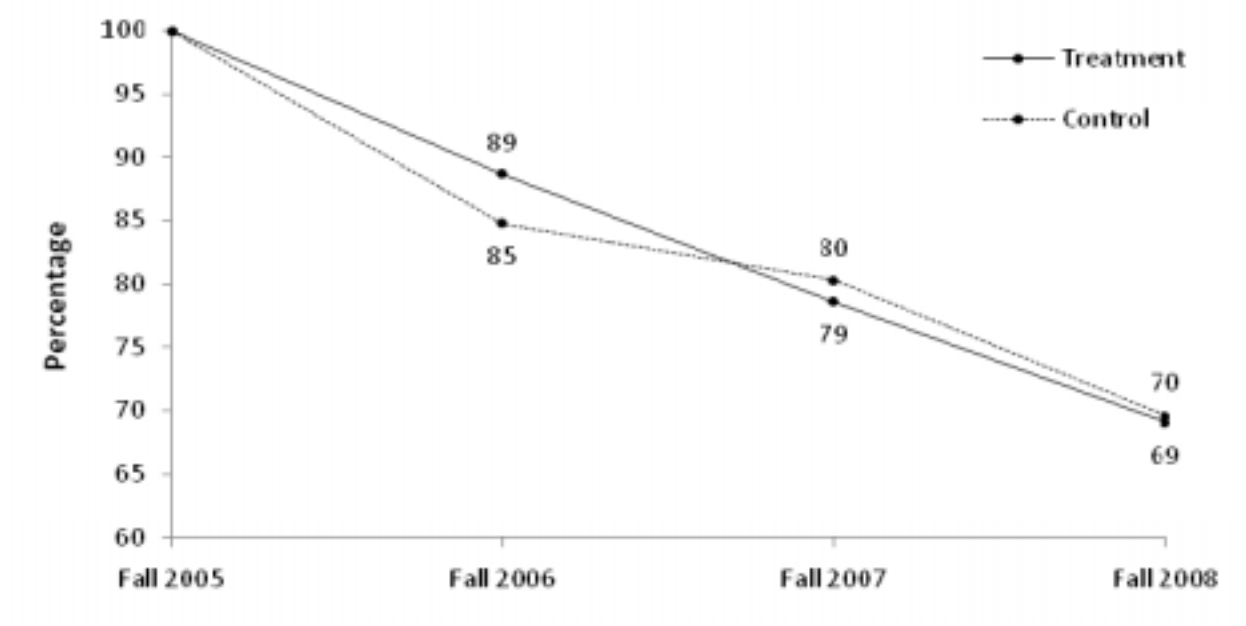
Comprehensive induction did not make teachers feel more satisfied or prepared. Teachers from the treatment and control groups reported feelings of satisfaction and preparedness that differed by 0.1 or less on the four-point scales at all points in time at which we measured these attitudes. These results are robust to alternate ways of aggregating the data, including rescaling the responses as binary variables and considering each original area of satisfaction or preparedness individually rather than combined into larger categories.

Teacher Mobility. Comprehensive induction did not make beginning teachers more likely to stay in their schools, their districts, or the profession. To measure teacher mobility, we surveyed teachers annually to learn whether they were still teaching, and if so, where they were teaching. By the end of the study period, 69 percent of teachers in one-year districts and 63 percent of teachers in two-year districts were still teaching in their original district. Figures ES.4 and ES.5 illustrate the lack of statistically significant treatment-control differences. They show a set of survival curves that plot the percentage of teachers retained in their original district in each year of the study for the one-year districts (Figure ES.4) and two-year districts (Figure ES.5) separately by treatment status.

The treatment group is represented by a solid line and the control group is represented by a dashed line. The differences were statistically insignificant at each time point in both one-year and two-year districts. When we conducted similar analyses of retention in the profession and the original school, we also found no significant treatment-control differences. For example, in terms of retention in the profession, 87 percent of teachers in one-year districts returned for a fourth year of teaching, with no significant difference between treatment and control teachers; 85 percent of teachers in two-year districts returned for a fourth year of teaching with the treatment-control difference being statistically insignificant.

When we examined in more detail where the movers and leavers went, we still did not find significant differences between treatment and control group mobility patterns in one-year or two-year districts. For example, statistically similar percentages of treatment and control teachers (about 47 percent) stayed in the same school in one-year districts, whereas 15 percent moved within the district, 10 percent moved to a new district (including charter schools), and 5 percent moved to private schools. We examined the reasons teachers gave for moving to a teaching position outside their original school or for leaving the profession and again found no significant treatment-control differences.

Figure ES.4. Survival Curve for One-Year Districts: Percentage Remaining in the District

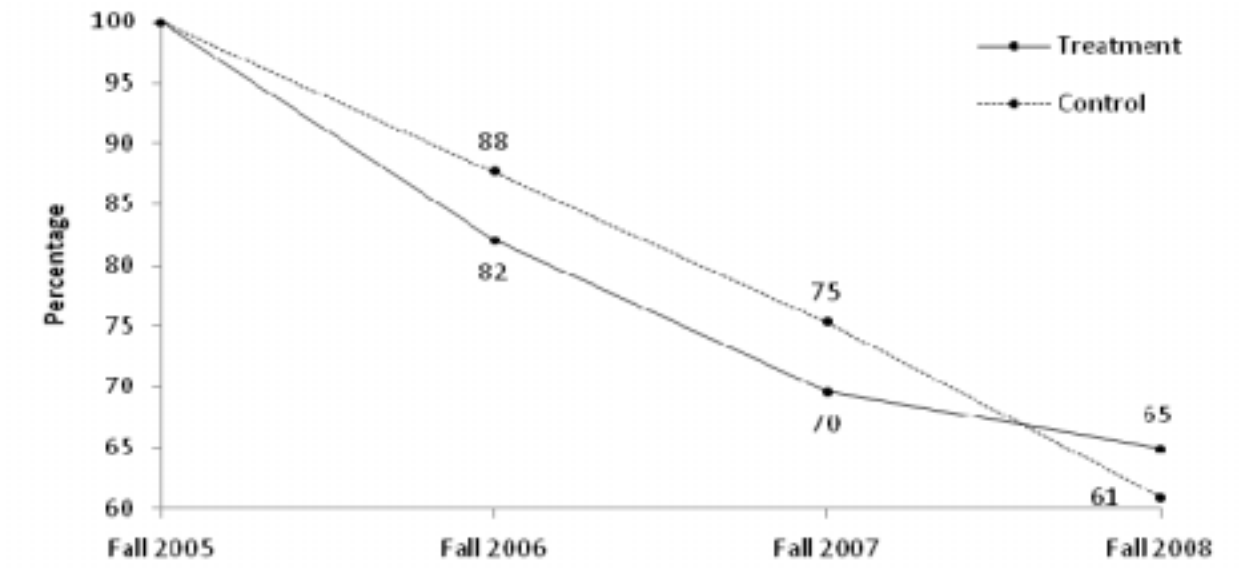


Source: Mathematica First, Second, and Third Teacher Mobility Surveys administered in fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. N = 561 teachers in fall 2005, 500 teachers in fall 2006, 476 teachers in fall 2007, and 417 teachers in fall 2008.

Treatment-control differences are not significantly different from zero at the 0.05 level.

Figure ES.5. Survival Curve for Two-Year Districts: Percentage Remaining in the District



Source: Mathematica First, Second, and Third Teacher Mobility Surveys administered in fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. N = 448 teachers in fall 2005, 382 teachers in fall 2006, 364 teachers in fall 2007, and 345 teachers in fall 2008.

Treatment-control differences are not significantly different from zero at the 0.05 level.

We conducted sensitivity analyses to confirm the findings of no impacts on retention rates. To address concerns about potential bias introduced by which teachers respond to the surveys, we used a variety of alternative methods for defining mobility, defining the eligible sample, and estimating the impacts, and continued to find no impact on mobility for all plausible assumptions regarding survey nonrespondents.

Composition of the Workforce. We investigated the impacts of comprehensive induction on the composition of the teaching force in the district to understand whether comprehensive induction raised the quality of teaching by encouraging the weakest teachers to leave or lowered it by discouraging the strongest ones from staying. To test this hypothesis, we used measures of teachers' professional qualifications, classroom practice ratings from their first year, and student test data from the third year of teaching.

We found similar levels of professional qualifications for treatment and control teachers who remained in their original district in their fourth year of teaching ("stayers"). Restricting the sample to stayers, we compared the average values of several teacher characteristics for treatment teachers to control teachers. The top panel of Table ES.1 shows results for one-year districts; Table ES.2 shows results for two-year districts. For each teacher characteristic, we found no statistically significant difference between treatment stayers and control stayers.

Similarly, when we measured teacher quality using classroom performance measures, comprehensive induction did not improve composition of the teacher workforce. The bottom panels of Tables ES.1 and ES.2 focus on performance measures. Restricting the sample to the stayers, treatment teachers did not exhibit stronger evidence than control teachers of effective classroom practices during the first year of the study. As shown in the bottom panel of Table ES.1, which refers to one-year districts, stayers in the control group outperformed stayers in the treatment group in raising students' math test scores by a statistically significant margin in year 3. There was no significant difference in reading test scores. Results for two-year districts are shown in Table ES.2. Unlike the full group of teachers included in the analysis in the third year of the study, among stayers returning for a fourth year in their original district there was no significant difference between treatment and control teachers in reading or math.

Association Between Levels of Induction Support and Outcomes

To complement the experimental analysis, which was based on the random assignment of teachers to treatment and control groups, we conducted two correlational analyses. Correlational analyses should be interpreted with caution because they are not causal. Unlike with the randomized experiment that used variation in treatment status, the variation in induction services that we explore here can be caused by confounding factors that also explain teachers' attitudes, workforce attachment, and effectiveness in the classroom.

The first correlational analysis tests whether there is a relationship between the study outcomes and the level or intensity of induction services more generally. We exploit the natural variation in induction support that occurred across teachers within and between experimental groups to determine whether there is a relationship between the level of induction support and the study outcomes. We use the same sample (treatment and control) and the same regression methods as the experimental analyses, but instead of assignment to treatment status as the key explanatory variable, we used four measures of induction support based on the number of years the teacher had an assigned mentor and indices of the breadth, intensity, and instructional focus of induction services constructed from the survey data on induction activities.

Table ES.1. Characteristics of District Stayers After Three Years, by Treatment Status (Percentages Except Where Noted): One-Year Districts

Teacher Characteristic	Treatment Stayers	Control Stayers	Difference	P-value
Background Characteristic				
College entrance exam score (SAT combined score or equivalent)	1040	1013	27	0.325
Attended highly selective college	27.5	27.2	0.3	0.954
Major or minor in education	78.7	80.9	-2.1	0.665
Student teaching experience (weeks)	15.8	15.4	0.4	0.772
Highest degree Is master's or doctorate	22.4	28.2	-5.8	0.311
Entered the profession through traditional four-year program	67.6	58.9	8.7	0.171
Certified (regular or probationary)	94.7	94.7	0.0	0.999
Career changer	14.5	12.8	1.7	0.682
Sample Size (Teachers)	148	139		
Sample Size (Schools)	88	84		
Year 1 Classroom Observation Score (on 1 to 5 scale)				
Content of a literacy lesson	2.3	2.6	-0.3*	0.024
Implementation of a literacy lesson	2.6	2.8	-0.2	0.151
Classroom culture	3.0	3.1	-0.1	0.607
Sample Size (Teachers)	100	94		
Sample Size (Schools)	71	65		
Year 3 Test Scores (standard deviation units)				
Reading	-0.27	-0.28	0.02	0.764
Math	-0.26	-0.11	-0.15*	0.008
Sample Size (Teachers)	25	36		
Sample Size (Schools)	23	29		

Source: Mathematica analysis using data from the College Board and ACT, Inc.; data from the 2006-2007 and 2007-2008 school years provided by participating school districts; Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted to account for the study design. The analysis of college entrance exam scores relied on a smaller sample (84 treatment and 86 control teachers and 61 treatment and 62 control schools). The analysis of Year 3 Test Scores relied on a different sample for reading (26 treatment and 34 control teachers and 24 treatment and 27 control schools) and math (per table values).

*Significantly different from zero at the 0.05 level.

Table ES.2. Characteristics of District Stayers After Three Years, by Treatment Status (Percentages Except Where Noted): Two-Year Districts

Teacher Characteristic	Treatment Stayers	Control Stayers	Difference	P-value
Background Characteristic				
College entrance exam scores (SAT combined score or equivalent)	905	935	-30	0.330
Attended highly selective college	23.7	21.4	2.3	0.703
Major or minor in education	67.8	66.6	1.2	0.874
Student teaching experience (weeks)	12.3	12.3	0.1	0.975
Highest degree Is master's or doctorate	16.7	10.2	6.5	0.196
Entered the profession through traditional four-year program	61.1	66.4	-5.4	0.443
Certified (regular or probationary)	95.8	92.9	2.9	0.366
Career changer	17.1	11.7	5.4	0.292
Sample Size (Teachers)	124	93		
Sample Size (Schools)	67	52		
Year 1 Classroom Observation Score (on 1 to 5 scale)				
Content of a literacy lesson	2.4	2.4	0.0	0.690
Implementation of a literacy lesson	2.7	2.6	0.1	0.583
Classroom culture	3.1	3.1	0.1	0.624
Sample Size (Teachers)	87	62		
Sample Size (Schools)	50	41		
Year 3 Test Scores (standard deviation units)				
Reading	-0.23	-0.27	0.05	0.302
Math	-0.12	-0.24	0.11	0.054
Sample Size (Teachers)	31	16		
Sample Size (Schools)	21	14		

Source: Mathematica analysis using data from the College Board and ACT, Inc.; data from the 2006-2007 and 2007-2008 school years provided by participating school districts; Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted to account for the study design. The analysis of college entrance exam scores relied on a smaller sample (56 treatment and 47 control teachers and 40 treatment and 35 control schools). The analysis of Year 3 Test Scores relied on a different sample for reading (33 treatment and 17 control teachers and 24 treatment and 15 control schools) and math (per table values).

None of the differences is statistically significant at the 0.05 level.

The relationships between induction support and student achievement were mixed: statistically significant (positive) and statistically insignificant. For math, there were both positive, statistically significant associations and statistically insignificant associations. For reading, there were only statistically insignificant associations.

Beginning teachers who received more induction support reported being more satisfied, on average, than those who received less. Induction intensity and instructional focus stood out as the two aspects of support that were positively related to teacher attitudes. The relationship of induction services to teachers' reported feelings of preparedness exhibited a similar pattern but with only one statistically significant relationship (induction intensity). These feelings of satisfaction and preparedness did not translate into better retention. None of the four measures of beginning teacher support was related to retention in the district or in the profession.

In the second correlational analysis, we examined whether better outcomes are associated with matching between the mentor and mentee on two dimensions, race/ethnicity and grade. We conducted this analysis using only the treatment group, which is the part of the sample for which we have detailed information on mentor background.

Beginning teachers who had the same race/ethnicity as their mentor or taught the same grade as had their mentor had lower rates of retention in the district and in the profession than those who did not have such a match. This contradicts the hypothesis that better matching would produce better outcomes. When we examined the other two outcomes, teacher attitudes and student achievement, we found no evidence of a statistically significant relationship with a mentor match.

I. INTRODUCTION AND BACKGROUND

Policymakers and researchers have recently been concerned about shortages of highly qualified teachers in hard-to-staff school districts (Howard 2003; Ng 2003), particularly in urban areas (Murphy et al. 2003). These concerns have generated debate about how to attract new teachers (Levin and Quinn 2003), although some researchers have argued that the shortages may have less to do with the difficulties of attracting new teachers than with retaining them (Ingersoll 2001). A frequently cited statistic from national data on teacher mobility suggests that 24 percent of beginning teachers leave the classroom by the end of their second year and 46 percent leave by the end of their fifth year (Ingersoll 2003).

High teacher turnover can have negative consequences. It can hurt student achievement by exposing more students to inexperienced teachers (Darling-Hammond 2000). It can also impose a high cost on districts that must recruit, hire, and train replacement teachers, and it can disrupt schools (Ingersoll and Smith 2003; King and Newmann 2000; Alliance for Excellent Education 2004).

Even teachers who manage to persist can find themselves struggling if they are not adequately supported early in their careers, especially if they were not adequately prepared for the challenges of the classroom. The hardest-to-staff schools tend to have classroom conditions that challenge even the best-trained teacher candidates. Teachers who start their careers in these settings may face challenges in pedagogy or classroom management for which they were not fully prepared (Kauffman et al. 2002).

One policy option in response to the problems of high turnover and inadequate preparation is to support teachers with a formal, comprehensive induction program during their initial years in the classroom. Such a program might include a combination of school and district orientation sessions, special in-service training (professional development), mentoring by an experienced teacher, classroom observation, and constructive feedback through formative assessment. To support beginning teachers, most districts offer some form of teacher induction or mentoring, but they often provide a limited set of services in response to an unfunded state mandate and with modest local resources (Berry et al. 2002; Smith and Ingersoll 2004). An example of informal or low-intensity teacher induction includes pairing each new teacher with another full-time teacher without providing any training, supplemental materials, or release time for the induction to occur. In short, although teacher induction is common, induction that is intensive, structured, and sequentially delivered in response to teachers' emerging pedagogical needs is not common. Throughout this report, we refer to the more formal, structured programs as "comprehensive" induction.

One reason that school districts do not offer more support to new teachers could be that comprehensive teacher induction is expensive. Costs for induction programs, as estimated in recent literature, range from $1,660 to $6,605 per teacher per year (Villar and Strong 2007; Alliance for Excellent Education 2004).[4] Moreover, there is little empirical evidence on whether investing

[4] These reports note costs for five programs—four are two-year programs and one is a one-year program. The data sources include state, district, county, and local data. The period to which the data pertains is 2003-2004 for three programs and unspecified for the other two. Several other studies of the costs of teacher turnover present estimates of induction or teacher-training costs, but these measures are expressed in terms of costs per vacancy. Without additional

additional resources in a more comprehensive, and hence more expensive, induction program would help districts attract, develop, and retain beginning teachers.

According to several research reviews (Ingersoll and Kralik 2004; Totterdell et al. 2004; Lopez et al. 2004; Borman and Dowling 2008), studies of teacher induction to date have been neither conclusive nor rigorous. Research based on federal statistics (for example, Shen 1997; Smith and Ingersoll 2004; Henke et al. 2000; Alt and Henke 2007) can provide a useful, nationally representative perspective on the issue, but it is limited in the extent to which it can capture the intensity of induction supports and in the range of outcomes that can be examined. Research at the local level (for example, Fuller 2003; Youngs 2002; Rockoff 2008; Youngs 2007; Wechsler et al. 2010) has yielded more detailed descriptions of teacher supports but, like the national studies, has relied on non-experimental approaches that do not necessarily provide unbiased estimates of the causal impacts of interest: the retention rate for participants or test scores of participants' students compared to what they would have been in the absence of the program. Researchers in this tradition have attempted to address selection bias to varying degrees. Some researchers have reported retention rates for program participants absent a comparison group or have simply referred to the overall state retention rate as a benchmark (Odell and Ferraro 1992; Tushnet et al. 2002).

Congressional interest in formal teacher induction has grown, despite the lack of evidence. The No Child Left Behind Act of 2001 (NCLB), which reauthorized the Elementary and Secondary Education Act of 1965 (ESEA), emphasizes the importance of teacher quality in student improvement. Title II, Part A, of ESEA—the Improving Teacher Quality State Grants program—provides nearly $3 billion per year to states to train, recruit, and prepare high-quality teachers. The implementation of teacher induction programs is one allowable use of these funds. Current discussions on the reauthorization of NCLB argue for a continued focus on supporting teachers through professional development opportunities and teacher-mentoring programs, with a call to fund "proven models" to meet these objectives. In addition, the Higher Education Opportunity Act of 2008 authorizes grants that include teacher induction or mentoring programs for new teachers. These initiatives demonstrate federal interest in a policy response grounded in providing induction support as a core means to improve teacher quality. They also, however, stress the need to conduct rigorous research to determine whether efforts to implement comprehensive teacher induction programs produce a measurable impact on teacher retention and other positive outcomes for teachers and students.

A. Research Questions and Study Design Overview

To provide Congress and state and local education agencies with the scientific evidence that will support sound decisions about teacher induction, the National Center for Education Evaluation and Regional Assistance within the U.S. Department of Education's (ED) Institute of Education Sciences (IES) contracted with Mathematica Policy Research to conduct the Evaluation of the Impact of Teacher Induction Programs. The study examines whether augmenting the set of services districts usually provide to support beginning teachers with a more comprehensive program

(continued)
information on the number of vacancies, this measure does not provide sufficient information to be helpful to districts considering the adoption of an induction program. See National Commission on Teaching and America's Future (2007), Barnes et al. (2007), Milanowski and Odden (2007), and Fuller (2000).

improves teacher retention rates and other positive teacher and student outcomes. More specifically, the analysis is designed to address the following research questions:

1. What is the effect of comprehensive teacher induction on the types and intensity of induction services teachers receive, relative to the types and intensity of services they receive from districts' current induction programs?
2. What impacts does comprehensive induction have in the classroom? Specifically, what are the impacts on:

   a. teachers' classroom practices?
   b. student achievement?

3. What impacts does comprehensive induction have on the teaching workforce? Specifically, what are the impacts on:

   a. teacher attitudes (satisfaction and feelings of preparedness)?
   b. teacher retention, including retention in the district and in the profession?
   c. composition of the teaching force?

As part of this study, we issued a request for proposals in 2004 to identify a promising comprehensive teacher induction program. Among the proposals received in response to our request, two described highly similar programs operated by different providers; each program earned the highest rating from an expert review committee. The providers were Educational Testing Service (ETS), Princeton, New Jersey, and the New Teacher Center (NTC) at the University of California–Santa Cruz. Mathematica contracted with both providers to deliver one year of the services that we characterize as comprehensive. Of the 17 districts participating in the study, ETS operated in 9 districts, and NTC operated in 8 districts.

The study used an experimental design in which we randomly assigned a selected group of elementary schools within each of the 17 participating districts either to a treatment group, which received comprehensive teacher induction from either ETS or NTC (depending on the district) or to a control group, which took part in the district's usual teacher induction program if one existed. We assigned 418 elementary schools with 1,009 eligible beginning teachers across the 17 urban districts with at least ten high-poverty elementary schools. Although the districts selected for the study did not form a statistically representative sample of the nation, they were drawn from 13 states with a variety of regulatory, administrative, and demographic contexts. The study includes elementary schools only.

After the first year of intervention services was delivered in treatment schools, IES decided to expand the treatment to include a second year of services for a subsample of the districts, in effect creating two studies: one for districts that received one year of services, and the other for districts that received two years. The teachers assigned to treatment in the one-year districts started in fall 2005 and received induction services in the 2005–2006 school year; the teachers assigned to treatment in two-year districts also started in fall 2005 but received services in the 2005–2006 and 2006–2007 school years.

We selected the districts to receive a second year of the treatment based on factors such as whether the mentors who had been trained within the district by ETS or NTC were available for a second year and whether the group of districts selected for a second year would include approximately one-half of the total number of teachers participating in the evaluation. Dividing the sample in this way does not allow for and should not be used to make direct comparisons between

the districts that received one year of treatment and those that received two years of treatment, but instead it allows us to investigate the effectiveness of one-year programs separately from that of two-year programs. Seven districts (four for ETS and three for NTC) continued the program to a second year.

This report presents findings from all three years of the study but emphasizes findings from the third year of the study. Except where noted, we present findings separately for the set of 10 districts that received one year of treatment and the other set of 7 districts that received two years of treatment.

Researchers from WestEd, a subcontractor to Mathematica, monitored the implementation of the comprehensive induction services. WestEd staff played a critical role by providing regular, on-site oversight of the implementation to help ensure that it was faithful to the core service model and to identify and help address any implementation challenges that arose.

B. Previous Findings from the Study

Two interim reports from this study (Glazerman et al. 2008; Isenberg et al. 2009) showed that teachers assigned to the treatment group reported more induction support than control teachers while treatment services were being offered, but they also showed that the additional support did not translate into positive impacts on key outcomes after either of the first two years of the study.[5] The first year report showed that the offer of comprehensive induction support amounted to a greater likelihood of having a mentor formally assigned to beginning teachers (93 versus 75 percent) in the teacher's first year, more time spent in meetings with the mentor (95 versus 74 minutes per week), and greater frequency of receiving assistance in all 10 induction activities asked about for the week preceding the spring 2006 survey (such as suggestions to improve practice and help with state and district standards) and in all 22 areas asked about for the three months preceding the spring survey (including classroom management, reviewing student work, and communicating with parents).

The first year report found no positive impacts on classroom practices, student achievement, teacher retention, or the composition of the district's teaching workforce. Nor did the first interim report find any evidence of positive impacts on teachers' satisfaction or feelings of preparedness (Glazerman et al. 2008). The followup after the second year continued to find treatment-control contrasts in induction supports favoring the treatment group in two-year districts, but revealed no impacts on student achievement, teacher attitudes, or teacher mobility in either one-year or two-year districts (Isenberg et al. 2009).

The current report summarizes all of the findings for the study through the final followup, which took place after the study teachers' third year.

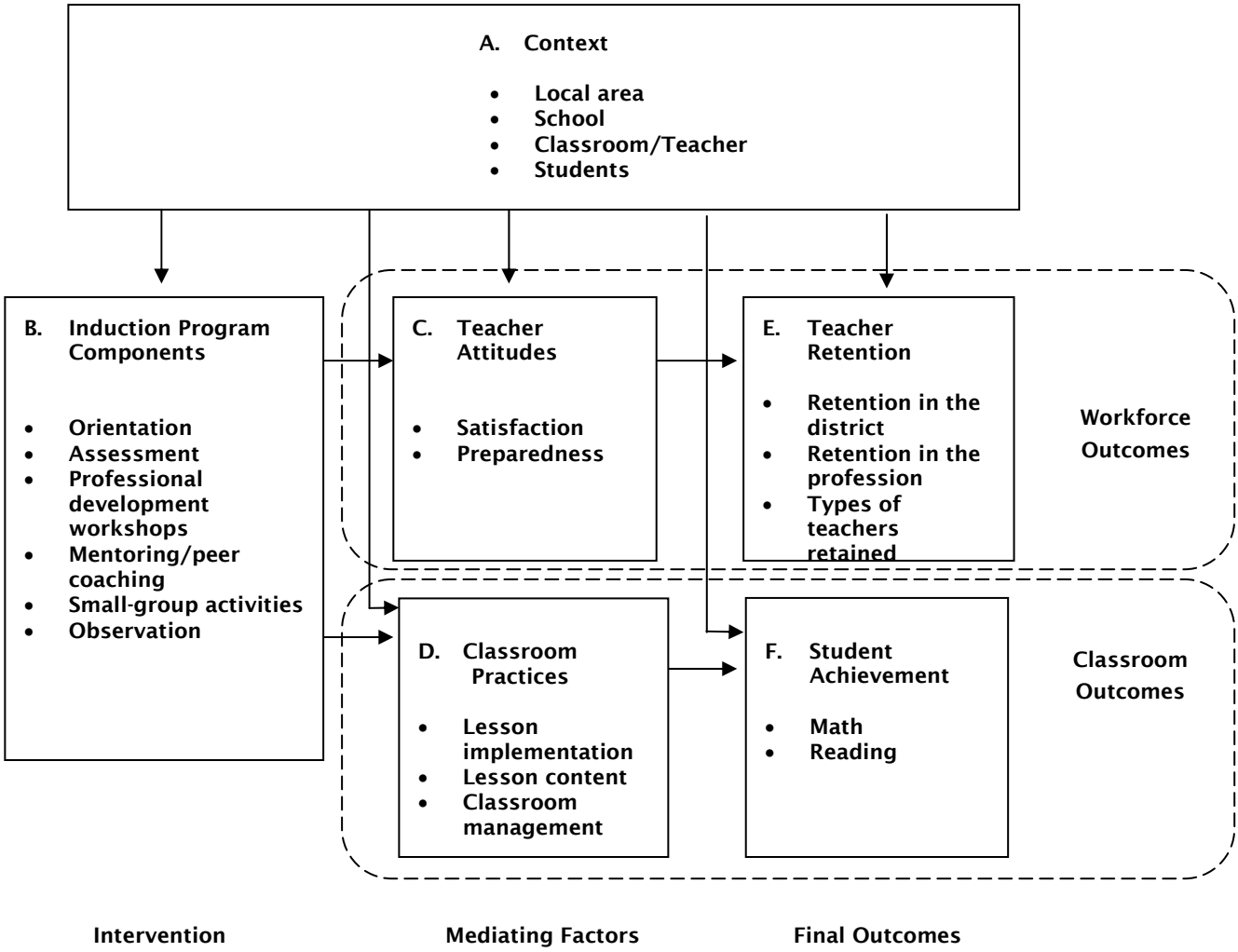
C. Conceptual Background for the Study

To answer the research questions, we began by identifying the pathways through which teacher induction programs could lead to teacher and student outcomes. Figure I.1 illustrates these pathways

[5] All comparisons discussed in this report are statistically significant at the 0.05 level unless otherwise stated.

and highlights some of the contextual factors that are useful to consider when planning and interpreting these analyses. More specifically, the figure shows how induction program components, contextual factors, and other mediating factors might affect teacher and student outcomes.

Figure I.1. Effects of Teacher Induction on Teacher and Student Outcomes: Conceptual Framework



Context. Context is important. The structure and functioning of an induction program are likely to be influenced by the characteristics of the local area, the school, the beginning teacher's classroom, the teacher, and her students (Figure I.1, Box A). Teacher and student outcomes may be directly affected, for example, by neighborhood demographics, the degree of administrative and financial support for beginning teachers, the percentage of a classroom's students with special needs or special education status, and teachers' employment histories. For this reason, the study examines variation *within* each of 17 districts by using district as a stratum within which random assignment is conducted (see Chapter II).

Induction Program. Induction programs may include a variety of possible components (Figure I.1, Box B). There is no one-size-fits-all model of teacher induction: different programs emphasize different approaches. For instance, programs may stress to a greater or lesser degree components such as orientation, assessment, professional development workshops, mentoring/peer coaching, small-group activities, and classroom observation. Presumably, the more intense the

emphasis on a given component, the larger the effect it will have on outcomes. But even the intensity with which a program implements a given component may vary in terms of quality, duration, and frequency. In this study, teachers in the treatment group received a specially selected comprehensive program of induction supports.

Counterfactual Condition. This study does not compare comprehensive induction to no induction or support for beginning teachers. Rather, it addresses the policy-relevant question: What would happen if school districts that were not already implementing comprehensive programs were to begin doing so? The school districts in the study that represent this state of the world (the counterfactual condition) were carefully selected to be the ones that might consider adopting comprehensive programs in the future. They may already have an existing set of informal arrangements for supporting beginning teachers, but the expectation is that a new, comprehensive program would expand these supports and possibly change the means by which support is provide and the focus of that support. Thus in Figure I.1 we hypothesize that the breadth, intensity, and nature of induction services (Box B) will differ on average for treatment and control classrooms.

Outcomes. Induction may benefit school districts in two ways: by strengthening the teacher workforce through reducing attrition and/or improving the composition of the workforce (Figure I.1, Box E) and by enabling teachers to improve student academic outcomes (Figure I.1, Box F). Induction may affect mediating factors that help explain changes in these final outcomes. For instance, two possible precursors to teacher mobility are dissatisfaction and the feeling of being unprepared, both of which can presumably be mitigated with more intensive induction support (Figure I.1, Box C). In addition, students' academic outcomes may improve through the mediating factor of improved classroom practices (Figure I.1, Box D).

D. Organization and Content of This Report

The conceptual framework presented above guided the organization of this report. After presenting the methods (Chapter II) and data (Chapter III), the report outlines the induction program components under study and the services that treatment and control teachers report receiving, in Chapter IV. Next, we present estimates of the effect of the treatment by examining impacts on classroom practices and student test scores (Chapter V) and effects on the teaching workforce by examining impacts on teacher attitudes and mobility behavior (Chapter VI). The final chapter presents correlational analyses that relate measures of different aspects of induction to key outcomes (Chapter VII) as a way to add context to the experimental findings.

II. STUDY DESIGN AND METHODS

The centerpiece of the design for the teacher induction evaluation is the use of random assignment to construct a group of teachers who were exposed to comprehensive teacher induction services (treatment) and an equivalent group who were exposed to the induction services normally offered by the districts (control). This chapter documents the study design, discusses the methods for selecting districts, schools, and teachers for inclusion in the study, and describes the data analysis methods.

The sample selection process described in this chapter is summarized in Figure II.1. Although we undertook a purposeful selection of districts and schools, the schools, once selected to be in the study, were then randomly assigned within each district to a treatment or control group. This ensures that the resulting impact estimates are internally valid. The description of the district and school selection process below is meant to help readers understand the population to which the findings generalize.

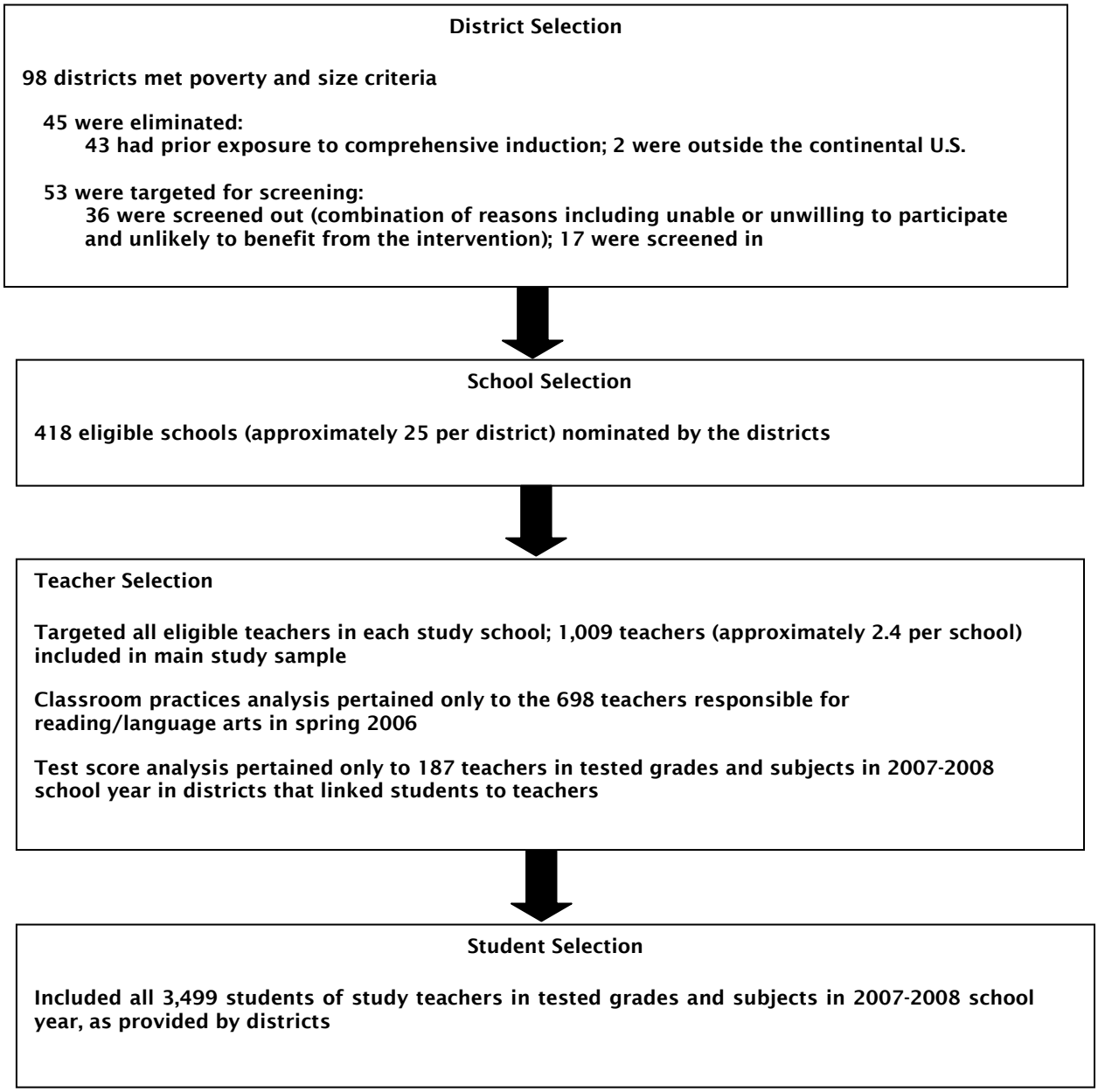
A. Selection of Districts

We sought a group of approximately 16 to 20 school districts that were not already providing comprehensive teacher induction in all the schools that needed it, but would be candidates for future adoption of such a program. The initial list of targeted districts was selected according to size and poverty levels in order to guarantee a sufficiently large sample for statistical precision while including hard-to-staff schools. We first used data from the National Center for Education Statistics' Common Core of Data (CCD) 2004-2005 to identify all school districts in the United States with at least 570 teachers in elementary schools and at least 10 elementary schools with 50 percent of students eligible for free or reduced-price meals under the federal government's National School Lunch Program (NSLP). We developed these size and poverty targets in consultation with the Institute of Education Sciences (IES), based on an earlier feasibility analysis (Glazerman et al. 2005). Nationally, 98 districts were determined to meet these targets.

We narrowed the list of districts through a screening and recruitment process. Mathematica subcontracted with the Penn Center for Educational Leadership (CEL) at the University of Pennsylvania to conduct a series of screening interviews with state and district officials to determine each district's suitability for inclusion in the study. Beginning with the list of 98 districts, Mathematica and CEL eliminated 2 districts that were outside the continental United States and 43 that had previous exposure to teacher induction programs of similar intensity and comprehensiveness to the ones selected for the study. Most of those districts were in California, Louisiana, Ohio, or Texas, but we also eliminated districts in other states that reported hiring staff to provide mentoring services full time, offering stipends of more than $1,000 per mentor (for one-on-one mentoring), or budgeting an equivalent of $1,000 or more per beginning teacher for induction services.

We eliminated another 36 districts that refused to participate, had no interest in implementing an induction program, or did not believe that they could benefit from the intervention being offered. Many such districts were in the process of reducing their teaching force and therefore did not want to introduce interventions to promote retention.

Figure II.1. Sample Selection Flow Chart



At the end of the screening and recruiting process, we had a final sample of 17 school districts in 13 states. By selecting districts that both met our criteria and had leaders who agreed to be in the study, we identified those most likely to need and implement comprehensive teacher induction in the future. These districts, with some combination of rising enrollments, high teacher turnover, and a limited supply of new teachers, are the most promising candidates for teacher induction and hence for this study.

Each district was assigned to one of the two providers of treatment services, either Educational Testing Service (ETS) or New Teacher Center (NTC), based primarily on district preferences. The preference-based method of assigning districts to providers does not allow for and should not be used to make direct comparisons of one provider to the other. Observed differences in impacts between ETS and NTC districts may be due to the programs or to the set of districts each provider worked with; those effects cannot be separated. The process of selecting program providers does

not affect the internal validity of the impact estimates, which are computed within district and district type.

Similarly, the decision of which districts would receive a second year of intervention was preference based. We selected the districts to receive a second year of the treatment based on convenience and feasibility. We ensured a balance of ETS and NTC districts in the two-year group. The non-random selection of districts means that they may differ in unobserved ways beyond having one or two years of treatment. Therefore, we avoid direct comparisons of one-year to two-year districts just as we avoided comparing ETS to NTC districts. Again, this method of district selection does not affect the impact estimates themselves.

Table II.1 shows the characteristics of districts included in the study. The districts served low-income students, with more than 40 percent of students in each district qualifying for the NSLP. The study included districts with a majority of their students being African American (7 of the 17 districts), Hispanic (2 of 17), and white (3 of 17), and 5 diverse districts without a racial/ethnic majority. The districts were all urban; 9 of 17 districts enrolled more than 50,000 students, and 11 of 17 included more than 50 elementary schools. The districts were in three of the nation's four Census regions: Northeast, Midwest, and South.

Table II.1 also shows the characteristics of one-year and two-year districts. Seven of the one-year districts and two of the two-year districts had more than 50,000 students. Two of the seven 2-year districts and none of the one-year districts served a student population that was majority (greater than 50 percent) Hispanic. All four of the study districts in the Midwest region were selected to implement the treatment for one year. Districts in the Northeast and South were part of one-year and two-year groups. Throughout most of this report, we present findings for the one-year and two-year districts separately.

B. Selection of Schools and Teachers

Within each district, either all or a subset of elementary schools was selected for the study. Large districts exercised some discretion over the subset of schools considered for the study. Otherwise, we selected all schools with eligible teachers and then selected all the teachers within those schools who met the following eligibility criteria:

- **Elementary Grade.** Teachers in K–6 were considered elementary. We excluded teachers of part-day prekindergarten classes. We focused on elementary rather than secondary schools because we needed a large number of schools per district to ensure feasibility of the study design.
- **New to the Profession.** We encountered 58 teachers who reported more than two years of teaching experience in some capacity, even if the district did not recognize such experience. They were included if (1) the district considered such teachers as new from the perspective of eligibility for beginning teacher induction services and (2) the method for identifying teachers for the study was applied consistently to all schools within each district.

Table II.1. Characteristics of Districts in Teacher Induction Sample by Length of Induction Program

	Number of Districts			Percentage
District Characteristics	One-Year	Two-Year	All	All
Demographics				
Low Income (Percentage Eligible for NSLP)				
<65	4	3	7	41.2
65–70	2	0	2	11.8
70–75	2	1	3	17.6
75–80	2	3	5	29.4
>80	0	0	0	0.0
Race/Ethnicity				
Majority African American	4	3	7	41.2
Majority Hispanic	0	2	2	11.8
Majority white	3	0	3	17.6
No single majority group	3	2	5	29.4
Region				
Northeast	2	2	4	23.5
Midwest	4	0	4	23.5
West	0	0	0	0.0
South	4	5	9	52.9
District Size				
Student Enrollment				
5,000–24,999	1	0	1	5.9
25,000–49,999	2	5	7	41.2
50,000–100,000	4	1	5	29.4
More than 100,000	3	1	4	23.5
Number of Elementary Schools				
Fewer than 50	3	3	6	35.3
50–100	2	3	5	29.4
More than 100	5	1	6	35.3
Study Sample				
Number of Mentors				
2	7	4	11	64.7
3	2	2	4	23.5
4	1	0	1	5.9
5	0	1	1	5.9
Number of Sample Teachers				
25–49	6	2	8	47.1
50–74	2	4	6	35.3
75–100	2	0	2	11.8
More than 100	0	1	1	5.9
Sample Size (Districts)	10	7	17	100.0

Source: Mathematica analysis using the Common Core of Data 2004-2005 from the National Center for Education Statistics; Mathematica teacher induction survey management system.

NSLP = National School Lunch Program.

- **Not Already Receiving Support.** Some alternative teacher preparation or certification programs continue to support teachers during their first year of teaching. Although teachers receiving such support were rare in study schools, we excluded them from the study in order to prevent duplication of induction services. We did, however, include teachers in alternative certification programs who were not receiving induction services from their programs.

We ultimately included 418 elementary schools in the study across the 17 districts. Tables II.2 and II.3 show the percentages of schools in one- and two-year districts serving low-income and minority students, as well as the grade configurations of the schools. Most of the schools in both types of districts employed one, two, or three eligible beginning teachers.

Table II.2. School Characteristics in One-Year Districts by Treatment Status (Percentages)

School Characteristic	All Schools	Treatment	Control	Difference	P-value
Percent Eligible for NSLP					0.592
<50%	8.5	9.3	7.8	1.5	
50–75%	23.7	21.0	26.4	-5.4	
75–100%	67.8	69.7	65.8	3.9	
Race/Ethnicity					0.863
Majority African American	43.8	43.3	44.3	-1.0	
Majority Hispanic	13.9	15.7	12.1	3.6	
Majority white	23.4	22.1	24.6	-2.5	
Other/mixed	18.9	18.9	19.0	-0.1	
Grade Configuration					0.907
Pre-K to 5 or K to 5	64.4	65.5	63.4	2.1	
Pre-K to 8 or K to 8	26.4	26.1	26.7	-0.7	
Other	9.2	8.4	9.9	-1.5	
Number of Sample Teachers					0.270
1	41.6	39.3	43.8	-4.5	
2	23.3	23.8	22.8	1.0	
3	20.4	23.0	17.8	5.2	
4	6.1	8.2	4.1	4.1	
More than 4	8.6	5.7	11.6	-5.8	
Sample Size (Schools)	252	124	128		

Source: Mathematica analysis using the Common Core of Data 2004-2005 from the National Center for Education Statistics.

Note: Data are weighted to account for the study design. Significance tests for categorical variables are design-adjusted F-tests of the difference in distributions. None of the differences is statistically significant at the 0.05 level.

NSLP = National School Lunch Program.

Table II.3. School Characteristics in Two-Year Districts by Treatment Status (Percentages)

School Characteristic	All Schools	Treatment	Control	Difference	P-value
Percent Eligible for NSLP					0.365
<50%	8.7	11.1	6.2	4.9	
50–75%	19.3	15.4	23.4	-8.0	
75–100%	72.0	73.5	70.4	3.1	
Race/Ethnicity					0.383
Majority African American	44.7	44.6	44.8	-0.2	
Majority Hispanic	33.8	37.8	29.6	8.2	
Majority white	6.7	7.2	6.2	1.1	
Other/mixed	14.8	10.3	19.4	-9.1	
Grade Configuration					0.662
Pre-K to 5 or K to 5	81.3	84.0	78.6	5.4	
Pre-K to 8 or K to 8	11.5	9.5	13.6	-4.1	
Other	7.2	6.5	7.8	-1.3	
Number of Sample Teachers					0.695
1	32.1	29.9	34.3	-4.4	
2	24.9	27.8	22.0	5.9	
3	14.7	17.3	12.1	5.2	
4	12.5	11.4	13.7	-2.3	
More than 4	15.7	13.5	17.9	-4.4	
Sample Size (Schools)	166	86	80		

Source: Mathematica analysis using the Common Core of Data 2004-2005 from the National Center for Education Statistics.

Note: Data are weighted to account for the study design. Significance tests for categorical variables are design-adjusted F-tests of the difference in distributions.

None of the differences is statistically significant at the 0.05 level.

NSLP = National School Lunch Program.

C. Random Assignment of Schools to Treatment

The defining feature of the study is the random assignment of schools to a treatment group that received the comprehensive induction services or to a control group that received the prevailing induction services provided by the district. Given the large sample, we can attribute the differences in average outcomes between the two groups to the addition of comprehensive induction services, ruling out all other confounding factors.

1. Method of Random Assignment

Eligible teachers in a school were either all exposed or all not exposed to treatment, a method known as *cluster random assignment*. Cluster random assignment was necessary because varying the types of induction services available in the same school building could result in contamination between services. For example, a mentor might feel uncomfortable being told not to provide any assistance to the colleague of one of his or her beginning teachers if the colleague was struggling with a problem. Furthermore, the presence of a mentor in the building could affect how existing supports and other resources are distributed among faculty at that school. Therefore, we assigned all eligible teachers within a school to treatment or control status based on the school in which they were expected to teach at the point of random assignment (baseline).

To increase statistical precision, we used block random assignment, with school districts as blocks. In other words, we conducted random assignment of schools within districts to ensure that each district was represented equally in both groups and that treatment status was not confounded with the school district. Block random assignment took into account the considerable variation among districts in the policies, student populations, and environments that could affect the study's outcomes.

Within districts, we used an efficient randomization technique called *constrained minimization*. For each district, we listed all admissible allocations of schools to treatment and control groups, and we randomly selected one allocation, with each allocation having an equal probability of selection. The admissible allocations were those that achieved an appropriate degree of balance between the treatment and control groups in terms of the overall number of eligible teachers and teaching assignment (grade level). Because the admissible allocations were defined independently of treatment status, every school and every teacher had a 50 percent probability of assignment to the treatment group. Glazerman et al. (2005) provide details on this random assignment method.

2. Treatment-Control Balance at Baseline

Random assignment produced groups that were equivalent on a wide variety of measures. Tables II.2 to II.11 describe the sample of schools and teachers along the dimensions measured, presenting the average characteristics separately by treatment status. The treatment and control schools exhibited similar percentages of low-income students and minority students, as shown in Tables II.2 and II.3.

While teachers were randomized indirectly, via their schools, the treatment and control teachers were similar in terms of demographic characteristics. Tables II.4 and II.5 present demographic characteristics of the study teachers by treatment group from one-year and two-year districts, respectively. Of 532 teachers in one-year districts responding to the baseline survey, similar percentages of treatment and control group members were white (74 and 77 percent, respectively), female (86 and 88 percent), under age 25 (51 and 49 percent), married (47 and 45 percent), and had no children at home (74 and 75 percent). Of 421 teachers in two-year districts responding to the baseline survey, similar percentages of treatment and control group members were white (43 and 44 percent, respectively), female (89 and 91 percent), under age 25 (48 and 47 percent), married (43 percent for both groups), and had no children at home (66 and 63 percent).

Treatment and control teachers were similar in terms of most professional characteristics, with some exceptions. Tables II.6 through II.9 describe the professional backgrounds of teachers for the one-year and two-year districts, respectively. In one-year districts, similar percentages of treatment and control teachers had advanced degrees, earned bachelor's degrees from highly selective colleges[6], had an education degree, and entered the profession with no student teaching (Table II.6). There was a statistically significant difference in how the teachers in one-year districts entered the profession, with a higher percentage of treatment teachers coming from a traditional four-year program (62 versus 56 percent) and a lower percentage of treatment teachers entering through an alternative preparation program (13 versus 22 percent). There was also a statistically significant

[6] A "highly selective" college or university is one that is rated as "most competitive," "highly competitive," or "very competitive" by the 2003 edition of the *Barron's Profile of American Colleges*.

difference in the type of teaching certificate held, with a higher percentage of treatment teachers holding a regular certificate (70 versus 60 percent) and a lower percentage of treatment teachers holding a probationary certificate (23 versus 36 percent). For those teachers who gave us permission to obtain their SAT or ACT score and for whom scores were available, we found no statistically significant differences in scores between the treatment and control teachers (Table II.7). With two-year districts, none of the differences between treatment and control teachers in professional background characteristics were statistically significant (Tables II.8 and II.9).

Table II.4. Teacher Demographic Characteristics by Treatment Status (Percentages): One-Year Districts

Teacher Characteristics	All Teachers	Treatment	Control	Difference	P-value
Gender					0.519
Male	12.6	13.6	11.6	2.0	
Female	87.4	86.4	88.4	-2.0	
Race/Ethnicity					0.585
White, non-Hispanic	75.5	74.1	77.0	-2.9	
African American, non-Hispanic	14.0	15.1	13.0	2.1	
Hispanic	5.5	4.8	6.2	-1.4	
Other/mixed/unknown	5.0	6.0	3.9	2.2	
Age (Years)[a]					0.902
20–25	49.8	50.5	49.1	1.4	
26–29	19.5	18.2	20.8	-2.6	
30–39	18.9	19.6	18.2	1.4	
40 or older	11.8	11.7	11.9	-0.1	
Marital Status					0.685
Married or living with a partner	45.7	46.6	44.6	2.0	
Single, separated, divorced, or widowed	54.3	53.4	55.4	-2.0	
Children Living in the Home					0.713
None	74.5	73.9	75.1	-1.2	
One or more children younger than age 5	10.4	11.5	9.3	2.2	
One or more children, none younger than age 5	15.1	14.6	15.6	-1.0	
Sample Size (Teachers)	532	267	265		

Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data are weighted to account for the study design. Significance tests for categorical variables are design-adjusted F-tests of the difference in distributions.

[a]Age of teacher is measured as of December 31, 2005, during the school year in which the study began.

None of the differences is statistically significant at the 0.05 level.

Table II.5. Teacher Demographic Characteristics by Treatment Status (Percentages): Two-Year Districts

Teacher Characteristics	All Teachers	Treatment	Control	Difference	P-value
Gender					0.604
Male	10.1	10.9	9.3	1.6	
Female	89.9	89.1	90.7	-1.6	
Race/Ethnicity					0.382
White, non-Hispanic	43.5	42.8	44.3	-1.5	
African American, non-Hispanic	25.5	29.5	21.4	8.1	
Hispanic	27.1	23.5	31.0	-7.5	
Other/mixed/unknown	3.8	4.3	3.3	0.9	
Age (Years)[a]					0.388
20–25	47.4	47.5	47.3	0.2	
26–29	20.0	20.9	19.0	1.8	
30–39	21.3	18.2	24.5	-6.3	
40 or older	11.4	13.5	9.2	4.3	
Marital Status					0.910
Married or living with a partner	43.1	43.4	42.8	0.6	
Single, separated, divorced, or widowed	56.9	56.6	57.2	-0.6	
Children Living in the Home					0.807
None	64.5	65.7	63.4	2.3	
One or more children younger than age 5	19.7	19.8	19.7	0.1	
One or more children, none younger than age 5	15.7	14.6	16.9	-2.3	
Sample Size (Teachers)	421	222	199		

Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data are weighted to account for the study design. Significance tests for categorical variables are design-adjusted F-tests of the difference in distributions.

[a]Age of teacher is measured as of December 31, 2005, during the school year in which the study began.

None of the differences is statistically significant at the 0.05 level.

Table II.6. Teacher Professional Background by Treatment Status (Percentages): One-Year Districts

Teacher Characteristics	All Teachers	Treatment	Control	Difference	P-value
Has Master's or Doctoral Degree	26.3	24.0	28.7	-4.7	0.289
Earned a Bachelor's Degree from a Highly Selective College	30.6	31.2	30.0	1.2	0.790
Earned a Degree with Education-Related Major or Minor	77.7	76.8	78.5	-1.7	0.680
How Entered the Profession					0.048*
Traditional program (four-year)	59.1	62.4	55.7	6.7	
Traditional program (post-baccalaureate)	22.6	22.9	22.4	0.5	
Teach for America	0.7	1.5	0.0	1.5	
Other alternative preparation program or unknown	17.5	13.3	21.9	-8.6	
Career Changer	13.3	12.9	13.9	-1.0	0.731
Teaching Certificate					0.009*
Regular	64.8	69.8	59.5	10.3	
Probationary	29.4	23.3	36.0	-12.6	
Emergency/waiver/other	5.7	6.8	4.5	2.3	
Weeks of Student Teaching					0.277
Zero	13.7	12.0	15.5	-3.5	
1–12	20.0	19.3	20.7	-1.5	
13–16	38.2	36.8	39.6	-2.7	
17 or more	28.2	31.9	24.2	7.7	
Sample Size (Teachers)	532	267	265		

Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data are weighted to account for the study design. Significance tests for categorical variables are design-adjusted F-tests of the difference in distributions.

*Significantly different from zero at the 0.05 level.

Table II.7. Teacher College Entrance Exams by Treatment Status: One-Year Districts

Teacher Characteristics	All Teachers	Treatment	Control	Difference	P-value
College Entrance Exam Scores (Percentages)					0.109
Did not take exam	8.9	8.3	9.5	-1.2	
Did not consent to obtain scores	19.3	16.6	22.2	-5.6	
Scores not found	10.6	13.6	7.5	6.2	
Scores reported	61.2	61.4	60.8	0.6	
SAT Combined Score (or ACT Equivalent)	1030	1033	1028	5	0.789
Sample Size (All Teachers)	561	275	286		
Sample Size (Teachers with Usable ACT or SAT Scores)	327	164	163		

Source: Mathematica analysis using data from the College Board and ACT, Inc.

Note: ACT scores were converted to SAT score equivalents using concordance tables in Dorans et al. (1997). Significance tests for categorical variables are design-adjusted F-tests of the difference in distributions.

None of the differences is statistically significant at the 0.05 level.

Statistically significant differences were found between treatment and control groups in teachers' assignments. For both the one-year and two-year districts, a smaller percentage of control than treatment teachers said that they were responsible for reading outcomes (86 percent of control teachers versus 92 percent of treatment teachers in the one-year districts, and 78 percent of control teachers versus 90 percent of treatment teachers in the two-year districts, as shown in Tables II.10 and II.11). The control group in the two-year districts contained a higher percentage of subject teachers than did the treatment group (12 versus 3 percent). Subject teachers included those who taught a single core subject such as math or science, as well as those who taught subjects such as art and music. This could mean that the process for identifying eligible teachers worked differently in the treatment and control schools, although non-classroom (including special subject) teachers were automatically excluded from the student test score analyses. The special subject teachers were included in the analysis of induction services received, teacher attitudes, and retention in order to measure outcomes for all teachers to whom districts would normally provide comprehensive induction services. The findings were robust to the inclusion or exclusion of special subject teachers.

3. Integrity of the Random Assignment Design

A randomized trial is the strongest evaluation design for identifying causal relationships, but even randomized experiments are subject to threats that can undercut a researcher's ability to draw inferences about the effectiveness of the intervention. We examined two typical threats to random assignment studies—noncompliance and attrition (study dropouts)—and found that these issues were not sufficiently serious to undermine the integrity of the study's findings.

Table II.8. Teacher Professional Background by Treatment Status (Percentages): Two-Year Districts

Teacher Characteristics	All Teachers	Treatment	Control	Difference	P-value
Has Master's or Doctoral Degree	15.9	16.2	15.7	0.5	0.915
Earned a Bachelor's Degree from a Highly Selective College	28.8	30.0	27.5	2.6	0.565
Earned a Degree with Education-Related Major or Minor	64.6	63.6	65.7	-2.1	0.689
How Entered the Profession					0.395
Traditional program (four-year)	61.5	59.3	63.7	-4.4	
Traditional program (post-baccalaureate)	9.2	7.8	10.6	-2.7	
Teach for America	6.2	5.7	6.6	-0.8	
Other alternative preparation program/unknown	23.2	27.1	19.1	8.0	
Career Changer	14.9	15.9	13.9	2.0	0.597
Teaching Certificate					0.892
Regular	50.4	49.5	51.3	-1.7	
Probationary	41.9	42.1	41.7	0.4	
Emergency/waiver/other	7.7	8.4	7.1	1.3	
Weeks of Student Teaching					0.445
Zero	28.5	30.6	26.2	4.4	
1–12	18.3	16.2	20.5	-4.2	
13–16	34.6	36.8	32.3	4.5	
17 or more	18.6	16.3	21.0	-4.7	
Sample Size (Teachers)	421	222	199		

Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data are weighted to account for the study design. Significance tests for categorical variables are design-adjusted F-tests of the difference in distributions.

None of the differences is statistically significant at the 0.05 level.

Table II.9. Teacher College Entrance Exams by Treatment Status: Two-Year Districts

Teacher Characteristics	All Teachers	Treatment	Control	Difference	P-value
College Entrance Exam Scores (Percentages)					0.891
Did not take exam	14.3	13.0	15.6	-2.6	
Did not consent to obtain scores	22.7	23.4	22.0	1.5	
Scores not found	11.6	12.3	10.9	1.5	
Scores reported	51.4	51.2	51.6	-0.3	
SAT Combined Score (or ACT Equivalent)	975	961	990	30	0.287
Sample Size (All Teachers)	448	231	217		
Sample Size (Teachers with usable ACT or SAT Scores)	221	117	104		

Source: Mathematica analysis using data from the College Board and ACT, Inc.

Note: ACT scores were converted to SAT score equivalents using concordance tables in Dorans et al. (1997). Significance tests for categorical variables are design-adjusted F-tests of the difference in distributions.

None of the differences is statistically significant at the 0.05 level.

Noncompliance. Noncompliance with treatment assignment—a concern in randomized experiments in which subjects in the control group receive treatment services or subjects in the treatment group fail to take up treatment (Angrist et al. 1996)—was not a serious problem in the teacher induction study. We put several safeguards in place to document teachers' compliance with treatment assignment and districts' cooperation with program implementation. First, an induction activities survey, administered twice during the implementation year, allowed us to measure the induction services each sample member received. Second, researchers from WestEd, a subcontractor to Mathematica, monitored implementation of the comprehensive induction services and fidelity to the induction model by collecting information on attendance at program activities and watching for services that might have been extended to teachers in schools not randomly assigned to the treatment group. Third, we monitored program mentor interactions via program logs and teacher mobility using field reports that were filed in a tracking system to complement the survey data on teacher mobility. Collectively, these data sources yielded a complete picture of service receipt.

The main form of noncompliance—"crossover" resulting from control group members' receipt of treatment—was not a problem. We designed the study to avoid contamination within the school and found limited mobility between school types (control to treatment or vice versa) during the school year. We identified fewer than three teachers out of more than 1,000 who transferred from a control to a treatment school and received services.

Table II.10. Teaching Assignments by Treatment Status (Percentages): One-Year Districts

Teacher Characteristics	All Teachers	Treatment	Control	Difference	P-value
Grade Level					0.151
Kindergarten	13.6	12.7	14.6	-1.8	
Grade 1	15.2	14.2	16.2	-1.9	
Grade 2	14.4	16.9	11.8	5.0	
Grade 3	13.2	15.3	10.9	4.4	
Grade 4	12.9	14.5	11.1	3.4	
Grade 5	10.0	8.4	11.6	-3.2	
Multiple, other	20.8	17.9	23.8	-5.9	
Responsible for Reading Outcomes	89.3	92.2	86.2	6.0*	0.034
Responsible for Mathematics Outcomes	91.0	93.0	88.9	4.1	0.110
Subject Specialty[a]					
Teaches only one grade level	82.0	85.3	78.5	6.7	0.104
Specialist: bilingual, ESL, or ELL	[b]	[b]	[b]		
Specialist: special education	7.5	5.7	9.4	-3.7	0.142
Specialist: core academic or other subject (e.g., reading, social studies, mathematics, science, computers, foreign language, art, music, gym)	4.9	3.9	6.0	-2.1	0.288
Teaching in Preferred Grade and Subject	79.6	81.6	77.6	4.0	0.138
Sample Size (Teachers)	532	267	265		

Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data are weighted to account for the study design. Significance tests for categorical variables are design-adjusted F-tests of the difference in distributions.

[a]Subject specialty variables are not exhaustive or mutually exclusive. In this table, a "specialist" is someone who does not teach just one grade level.

[b]Exact value suppressed to protect respondent confidentiality.

*Significantly different from zero at the 0.05 level.

ESL = English as a Second Language; ELL = English Language Learner.

Table II.11. Teaching Assignments by Treatment Status (Percentages): Two-Year Districts

Teacher Characteristics	All Teachers	Treatment	Control	Difference	P-value
Grade Level					0.151
Kindergarten	18.3	19.5	17.1	2.4	
Grade 1	14.4	14.4	14.4	0.0	
Grade 2	16.3	17.4	15.1	2.2	
Grade 3	13.6	13.7	13.5	0.2	
Grade 4	9.9	9.8	10.1	-0.3	
Grade 5	7.9	8.9	6.9	2.0	
Multiple, other	20.8	17.9	23.8	-5.9	
Responsible for Reading Outcomes	84.4	90.3	78.2	12.1*	0.003
Responsible for Mathematics Outcomes	83.3	86.4	80.1	6.3	0.092
Subject Specialty[a]					
Teaches only one grade level	82.9	85.4	80.3	5.1	0.209
Specialist: bilingual, ESL, or ELL	1.7	1.7	1.7	0.0	0.995
Specialist: special education	5.3	6.6	4.0	2.6	0.301
Specialist: core academic or other subject (e.g., reading, social studies, mathematics, science, computers, foreign language, art, music, gym)	7.5	3.4	11.8	-8.4*	0.003
Teaching in Preferred Grade and Subject	78.4	78.7	78.1	0.7	0.876
Sample Size (Teachers)	421	222	199		

Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data are weighted to account for the study design. Significance tests for categorical variables are design-adjusted F-tests of the difference in distributions.

[a]Subject specialty variables are not exhaustive or mutually exclusive. In this table, a "specialist" is someone who does not teach just one grade level.

*Significantly different from zero at the 0.05 level.

ESL = English as a Second Language; ELL = English Language Learner.

The second form of noncompliance—"no-shows" resulting from treatment group members failing to adopt the treatment—did not occur frequently. We did see some treatment group teachers refusing induction services or transferring to schools in which the induction services would not be available (for example, if they left the district). Nine treatment schools representing 12 teachers in one district and 3 teachers in another district refused to implement the comprehensive induction services that were offered. These 15 teachers made up 3 percent of the treatment group. The degree of program dropout is discussed in Chapter IV. All sample members are included in the impact analysis regardless of compliance status and classified according to their school's original treatment assignment.

Nonresponse and Study Attrition. Nonresponse and study attrition, especially differential attrition by treatment status, is another issue that affects the quality of any randomized experiment (or any longitudinal study regardless of design). For this study, response rates were at least 88 percent for the full sample on all major surveys in year 1 of the study, at least 83 percent in year 2, and at least 85 percent in year 3 (see Chapter III, Table III.1), yet we observed differences in

response rates by treatment status that were statistically significant. For example, the control group response rate for the spring 2006 induction activities questionnaire was 83 percent and the corresponding treatment group rate was 93 percent. A concern with differential response rates is that if nonresponse is not random with respect to outcomes, then the degree to which nonresponse affects the average outcomes will differ by treatment status, and the impact estimates—which are differences in mean outcomes for respondents only—will be biased. If, for example, nonrespondents have worse outcomes than do respondents, we would expect the lower response rates for the control group to translate into an upwardly biased estimate of the counterfactual outcome and therefore a downwardly biased estimate of the impact.

To mitigate such an outcome, we constructed nonresponse adjustment weights. Such weights let the respondents within each treatment group who look most like nonrespondents carry a greater weight so that they can stand in for their missing counterparts. We adjusted the weights to account for the variations in design implementation across districts. A full discussion of weights is included in Appendix A. We used these weights in the impact estimation, although the weights did not substantially change the findings.

D. Impact Estimation

The goal of the impact analysis is to estimate the effect of comprehensive teacher induction on a range of teacher outcomes relative to those that would have been observed in the absence of the comprehensive program. To that end, we examined whether student achievement gains, teacher mobility patterns, and other outcomes for teachers randomly assigned to the receipt of comprehensive induction services differed from the outcomes for those we assigned to the receipt of the prevailing induction services offered by the district.

Appendix A details the methods used for estimating the impacts of the comprehensive induction programs, as well as the alternate estimation approaches we used for testing the robustness of the study's findings. We illustrate the effect of alternate approaches by using a benchmark model that imposes the most reasonable set of assumptions and measurement rules and then comparing it to a set of alternatives that implement deviations—one at a time—from that benchmark. For example, the benchmark model specifies a set of variables used as covariates for regression adjustment of the impact estimates. The set of benchmark covariates differs for each outcome.

One virtue of random assignment is its analytic simplicity. The difference between the average outcome for the treatment and control groups is an unbiased estimate of the impact of the treatment on any outcome of interest. A *t*-test of the difference in average outcomes enables the evaluator to assess whether the observed difference could have been attributable to chance or to the program.

In the case of the teacher induction experiment, the hypothesis tests must be constructed in a way that is consistent with the study design. Specifically, we must account for the fact that we randomly assigned schools, rather than individual teachers, to treatment groups. Recognizing that teachers from the same school share the same principal, school culture, building conditions, neighborhood, and other characteristics that might affect teacher outcomes, we cannot treat teachers in the same school as independent observations.

Therefore, we use a model-based approach to estimate program impacts. The statistical model not only allows us to represent the non-independence of observations explicitly, it also allows us to exploit the data on student, teacher, and school background characteristics to increase the precision

of the estimates of treatment effects. The regression model allows us to control for the effects of a range of background characteristics, not just treatment status, on the outcomes of interest. By accounting for the many variables that affect teacher retention, for example, we can reduce the amount of unexplained variation in mobility decisions and thereby increase our confidence in the estimates of treatment effects.

The other advantage of the regression model is its ability to acknowledge the hierarchical structure of the data—for example, the nesting of teachers within schools. Accordingly, the units of analysis can be properly specified, and unbiased estimates of the standard errors used to conduct hypothesis tests can be devised. Although the study defines outcomes at the teacher level, we performed random assignment at the school level; hence, the regression model must account for the clustering of teachers within schools. Appendix A describes the statistical methods in more detail.

Impact findings are presented in two ways in this report. First, we present them as differences between the (regression-adjusted) means or percentages for the treatment and control groups. Second, for continuous outcome variables, we present the impact as an effect size, defined as the fraction of a standard deviation of the outcome variable. Effect sizes are a common metric used to compare findings across studies that rely on different measurement instruments. Effect sizes are computed as the impact divided by the standard deviation of the outcome variable. The standard deviation is computed using the full sample (treatment and control groups).

E. Interpreting Impact Estimates When There Are Multiple Comparisons

To interpret the impact estimates, this report relies on conventional notions of statistical significance. To determine if an impact estimate represents a true effect of the treatment or just a chance difference between the treatment and control groups, we conduct a statistical hypothesis test. The effect is deemed statistically significant if the probability of observing a difference (the "p-value") in the absence of a true impact is less than 5 percent. In other words, there is a 5 percent chance of "Type I error," declaring a finding to be statistically significant when the treatment was not responsible for the effect.

Using these rules, the probability of committing a Type I error is always 5 percent for any one test, but as the number of tests increases, the chance of committing at least one such error rises, leading to what is known as the multiple comparison problem—the risk of ignoring a large number of nonsignificant results and regarding one or two statistically significant results as true impacts.

There are many solutions to this problem, but we discuss two here. One solution, which we followed in this report, is to note the number of non-significant findings when reporting on significant findings, so the reader has the appropriate context. For example, it would be inappropriate to suppress non-significant findings from a table without at least noting that the additional tests were conducted.

Another set of solutions includes formalized approaches to controlling the *family-wise* Type I error rate, which is the probability of making a single Type I error in a group of hypothesis tests, or that try to control the False Discovery Rate (FDR), which is the percentage of tests that result in a Type I error. One such formalized approach that we considered for this report is an FDR control procedure developed by Benjamini and Hochberg (1995). The method calls for rank-ordering the

tests by their p-value from lowest to highest and determining a cutoff p-value above which all of the findings are deemed statistically insignificant, even if their individual p-values may fall below 0.05.[7]

This report used the first approach of contextualizing the findings and did not present any adjustments based on the Benjamini-Hochberg (BH) method because such adjustments were unnecessary or inappropriate (see Isenberg et al. 2009 for a fuller discussion). There was only one impact estimate in this report where the method could have been appropriately applied and a different test result would have been reached. In Table VI.1 a significant negative impact on the prior literacy lesson content score of teachers who remained in the treatment group versus those in the control group would not be regarded as significant after applying a BH adjustment. In the text, we discuss the non-significant findings and do not draw conclusions based on this finding out of context.

[7] This cutoff is determined to be the last test in the list, rank-ordered from lowest to highest p-value, for which the test's p-value is less than 0.05*(i/m), where i is the rank and m is the number of tests being conducted.

III. DATA

In accordance with the conceptual framework presented in Chapter I, we collected detailed data on teacher induction services, outcomes, and contextual factors. The data collection effort was most intense during the 2005–2006 school year, while the comprehensive induction programs were being implemented in the treatment schools in all districts, and continued for an additional three years. At the start of the programs, we surveyed mentors on their background characteristics and reviewed program documents from the Educational Testing Service (ETS) and the New Teacher Center (NTC). We administered a background teacher survey in fall 2005, at which time we also requested teachers' permission to obtain their college entrance exam scores (SAT or ACT). Surveys of teacher induction activities were administered to both treatment and control teachers during all four years of the study (2005–2006 through 2008–2009). For the study's core outcomes, we observed classrooms in spring 2006, collected the districts' student records data following the 2005–2006, 2006–2007, and 2007–2008 school years, and conducted teacher mobility surveys in fall 2006, 2007, and 2008 to learn about teacher retention. Figure III.1 shows a timeline for the data collection activities.

This report presents findings pertaining to all four years of data collection, both for the set of districts that received one year of treatment and for those that received two years of treatment. Response rates and brief descriptions of each data collection activity are provided below. Copies of the survey instruments may be found in Glazerman et al. (2005). At the end of this chapter, we present flow diagrams that explain how we used the data we collected to derive our analysis samples from the pool of teachers we originally identified as eligible for the study. Figure III.2 shows a flow diagram for one-year districts, and Figure III.3 shows a similar diagram for two-year districts.

A. Mentor Survey

As part of the treatment intervention, ETS and NTC worked with district staff to hire 44 mentors who would deliver the intervention services, offering support and guidance to help beginning teachers use evidence from their own practice to recognize and implement effective instruction. The mentor hiring and duties are described in Chapter IV.

During the ETS and NTC mentor-training sessions in fall 2005, we surveyed all 44 mentors on their previous mentoring experience, professional background, and basic demographic characteristics. All of these factors may influence the effect of mentor training on the mentor's practice and, in turn, the effect of mentoring on outcomes for beginning teachers. The survey was a self-administered, paper-and-pencil questionnaire.

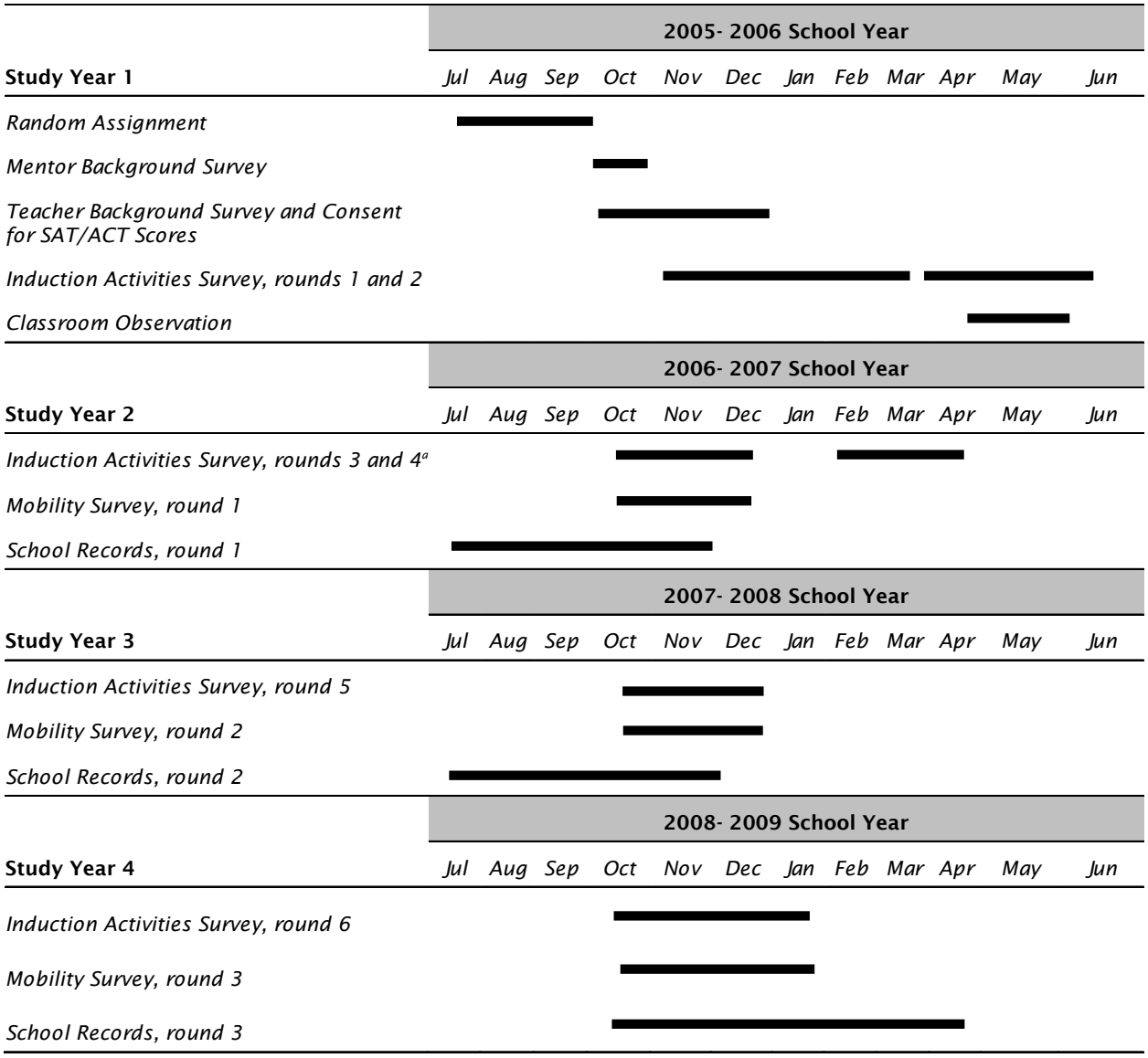
B. Beginning Teacher Surveys

1. Teacher Background Survey

Starting in October 2005, we administered a baseline survey to the treatment and control teachers to gather detailed information about their professional backgrounds, current teaching assignments, and demographic characteristics. The survey addressed teachers' professional credentials, participation in teacher preparation programs, perceptions of the teaching profession, and personal background characteristics, many of which (marital status, spouse's occupation and relocation history, number of young children, and salary at the start of the first year) are hypothesized to affect career decisions and hence retention. We mailed the surveys to all sample members at their schools and followed up by telephone and in person. Although most surveys were

returned in late 2005, we continued to follow up with sample members throughout the school year in order to achieve a final response rate of 94 percent (92 percent of control group teachers and 97 percent of treatment group teachers).

Figure III.1. Timing of Data Collection



[a] In spring 2007, the Induction Activities Survey was administered only to teachers in the 7 two-year districts.

One component of this background survey was a consent form asking teachers to permit the research team to obtain their college entrance exam scores, either SAT or ACT. These scores, which we received from 52 percent of teachers, provide an objective measure of a teacher's cognitive ability before they received any special preparation to enter the profession. Such a measure is useful as a potential correlate for teacher effectiveness or a description of the types of teachers who choose to stay in or leave the teaching profession (Ferguson and Ladd 1996; Greenwald et al. 1996).

2. Induction Activities Survey

It was important to understand the differences in the services delivered by the comprehensive and prevailing programs, and to investigate teachers' participation in induction activities after treatment ended. Treatment teachers in one-year and two-year districts were offered the same usual services as control teachers following the conclusion of the intervention. Our post-intervention data can show whether the intervention induced future changes in treatment teachers' usage of these services beyond what it would have been in the absence of the intervention. To that end, we administered a survey of teacher induction activities to both treatment and control teachers twice during the 2005–2006 school year, and again in fall 2006, fall 2007, and fall 2008.[8] Teachers in the seven districts that received two years of comprehensive teacher induction were surveyed an additional time during spring 2007 to gather more in-depth information about the induction activities in which they participated. Given that the nature of induction activities may change often during the school year, the administration of multiple surveys reduced any difficulties teachers might have had in recalling the activities over the course of the study, allowing us to detect changes over time in the types and intensity of services, such as the amount of time spent in mentor meetings or the number of times that administrators observed teachers in the classroom. The current report presents the findings from the induction activities surveys administered at all six time points (fall 2005, spring 2006, fall 2006, spring 2007, fall 2007, and fall 2008).

These surveys included questions applicable to services delivered by both the comprehensive and prevailing programs. The survey asked questions about mentoring from any source, timing and duration of mentor interactions, other induction activities such as classroom observations, professional development workshops, feedback on instructional practices, and the extent to which respondents are satisfied with various aspects of teaching. We mailed the surveys and followed up by telephone and in some cases used field interviewers to complete the survey in person to achieve a high response rate.

3. Teacher Mobility Survey

We sent mobility surveys to all teachers in fall 2006, fall 2007, and fall 2008 to track their career progress—whether they returned to teaching and, if so, whether they returned to the same school or district. For those who left teaching, we asked about the circumstances, reasons, and timing of the change as well as about their current employment status and plans for returning (if applicable). For example, we asked about job responsibilities and salary for those who had changed jobs. As with the other teacher surveys, the mobility surveys were self-administered, mail questionnaires with telephone and in-person follow-up interviews for those who did not complete the instrument by mail.

[8] The fall 2005 and spring 2006 induction activities surveys were administered over a period that stretched from November to early March and late March to June, respectively. Large shares of the surveys were returned in January and March (28 percent for the first induction activities survey and 48 percent for the second, respectively). One reason for the variation in completion dates is the variation in the start and end dates for the academic calendars among the 17 districts included in the study.

4. Response Rates to Teacher Surveys

Response rates on teacher surveys ranged from 88 to 97 percent for the treatment group and 78 to 92 percent for the control group (Table III.1). Table III.2 shows response rates for different subgroups. Despite overall response rates above 80 percent, the control group response rates persistently fell below those of the treatment group by a margin that was statistically significant. The degree to which the differential rates bias the findings depends on overall levels of nonresponse and the nature of nonresponse. Differences between the sample of respondents to the background survey and the full set of respondents and nonrespondents on observable school characteristics—the only data available for respondents and nonrespondents—are not statistically significant (see Table III.3). This suggests that the sample of teachers for whom we have survey data is similar to the population of eligible respondents that the sample represents.

Table III.1. Response Rates to Teacher Surveys by Treatment Status

Data Collection Instrument	Number of Eligible Respondents	Response Rate (Percentages)		
		Full Sample	Treatment	Control
Mentor Background Survey	44	100.0	100.0	n/a
Teacher Background Survey*	1,009	94.4	96.6	92.2
Induction Activities Survey				
Fall 2005*	1,009	89.0	93.3	84.7
Spring 2006*	1,009	87.7	92.5	82.9
Fall 2006*	1,009	88.7	91.5	85.9
Spring 2007*	447[a]	83.2	87.9	78.2
Fall 2007*	1,009	85.3	90.2	80.2
Fall 2008*	1,009	85.0	89.6	80.2
Teacher Mobility Survey				
Fall 2006*	1,009	88.7	91.5	85.9
Fall 2007*	1,009	85.3	90.2	80.2
Fall 2008*	1,009	85.0	89.6	80.2

Source: Mathematica teacher induction survey management system.

Note: The Induction Activities Survey and Teacher Mobility Survey were administered together in fall 2006, fall 2007, and fall 2008.

[a]**The spring 2007 survey was administered only in the seven districts that received two years of comprehensive teacher induction.**

***Response rates significantly different between treatment and control at the .05 level.**

n/a = not applicable.

Table III.2. Response Rates to Teacher Surveys by Subgroup and Treatment Status

	Response Rate (Percentages)													
	Teacher Background Survey, Fall 2005		Induction Activities, Fall 2005		Induction Activities, Spring 2006		Induction Activities/Mobility Survey, Fall 2006		Induction Activities, Spring 2007		Induction Activities/Mobility Survey, Fall 2007		Induction Activities/Mobility Survey, Fall 2008	
	T	C	T	C	T	C	T	C	T	C	T	C	T	C
District Type (Years of Implementation)														
One year	**97.1**	92.7	94.2	85.7	93.5	84.3	92.7	**87.2**	n/a	**n/a**	89.5	**82.9**	**86.9**	**80.4**
Two year	96.1	**91.7**	92.2	83.4	91.3	81.1	88.8	82.0	87.9	77.9	**90.0**	75.6	91.8	78.8
Grade Level														
K or Pre-K	96.3	97.2	95.0	90.3	92.5	90.3	94.7	91.3	93.2	86.2	91.3	80.6	90.0	79.2
1	98.6	97.2	95.9	94.4	95.9	87.3	95.4	89.7	83.9	81.5	95.9	88.7	91.8	81.7
2	97.6	91.0	95.2	78.2	89.3	76.9	91.0	89.0	92.1	82.9	89.3	76.9	90.5	78.2
3	97.5	94.7	95.1	86.0	96.3	80.7	89.7	84.3	91.2	77.8	86.4	84.2	87.7	87.7
4	96.7	91.7	95.0	88.3	93.3	86.7	91.1	84.5	85.0	78.3	85.0	73.3	85.0	76.7
5	100.0	96.2	95.7	88.5	97.8	90.4	93.0	91.1	83.3	82.4	91.3	84.6	89.1	88.5
Other/ multiple	91.5	84.1	82.9	75.2	85.4	75.2	83.5	72.9	82.5	67.8	89.0	74.3	89.0	73.5
School Type (Percent in Free Lunch Program)														
0–49.9%	100.0	93.1	94.6	72.4	94.6	72.4	94.6	86.2	100.0	66.7	91.9	89.7	100.0	82.8
50–74.9%	95.9	91.4	92.9	84.4	91.8	81.3	90.2	85.6	90.9	77.6	89.8	80.5	87.8	77.3
75–100%	97.1	92.1	94.1	86.8	92.4	84.9	91.8	85.2	86.0	78.2	89.7	78.9	88.6	81.7
Unknown	90.0	96.6	83.3	75.9	93.3	79.3	79.3	79.3	89.3	79.3	86.7	75.9	86.7	65.5

Source: Mathematica teacher induction survey management system; Mathematica Teacher Background Survey administered in fall 2005, Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008, and Mathematica First, Second, and Third Teacher Mobility Surveys administered in fall 2006, fall 2007, and fall 2008 to all study teachers; Mathematica Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: The Mathematica Induction Activities Survey and Teacher Mobility Survey were administered together in fall 2006, fall 2007, and fall 2008. Treatment and control group sample sizes are shown in Appendix Table A.6.

T = Treatment; C = Control; n/a = not applicable.

Table III.3. School Characteristics of Respondents and Nonrespondents

	Respondents Only			
	Background Survey (n=953)	Induction Activities Surveys (n=964)	Mobility Surveys (n=922)	Respondents and Nonrespondents (n=1,009)
Percent Free Lunch in School				
Unknown	5.8	5.6	5.3	5.9
0–49.9%	6.7	6.6	6.9	6.5
50–74.9%	22.1	22.3	22.2	22.4
75–100%	65.4	65.5	65.5	65.2
Percent White in School				
Unknown	0.9	0.9	1.0	0.9
0–49.9%	81.1	81.0	80.6	81.4
50–74.9%	16.7	16.5	16.8	16.3
75–100%	1.6	1.6	1.6	1.5
Percent Black in School				
Unknown	0.9	0.9	1.0	0.9
0–49.9%	59.3	60.0	59.8	59.8
50–74.9%	6.9	6.9	7.3	6.8
75–100%	32.8	32.3	32.0	32.5

Source: Mathematica analysis using the Common Core of Data 2004–2005 from the National Center for Education Statistics.

Note: None of the differences between respondents and the full sample (respondents and non-respondents) are statistically significant at the 0.05 level.

C. Classroom Observations

We observed classrooms of teachers in the treatment and control groups to measure their classroom practices in the areas of reading and literacy.[9] We excluded from this data collection any teachers who were responsible for small classes such as special education resource teachers, taught special populations such as bilingual classes, taught mathematics only, were not first-year teachers, or were no longer teaching in the district. Thus, the eligible sample for the classroom observations (698) was smaller than the full study sample (1,009). Among eligible teachers, we achieved a response rate of 94 percent for the treatment group and 89 percent for the control group (Table III.4).

We applied the eligibility rules uniformly to both the treatment and control groups. Some teachers with prior experience were in the study because their districts insisted, per their normal practice, that induction be offered to teachers who were new to the district. Because school districts chose to provide comprehensive induction services to these individuals, it was important to understand the impact of such services on their subsequent mobility behavior. However, we excluded such teachers from the classroom practices analysis to focus on the true novice teachers, those for whom induction was most likely to have an impact on classroom practices. We classified those who had left the classroom as ineligible for observation instead of "missing" because we

[9] We chose to focus on reading and literacy given the central role of this subject in elementary education.

already planned a separate, detailed analysis to deal with attrition from teaching (the teacher retention/mobility analysis).

Table III.4. Response Status to Classroom Observation and Reasons for Nonresponse

Status/Reason	Number of Teachers	Percentage of Teachers	Percent of Eligible Teachers		
			All Eligibles	Treatment	Control
Eligibles					
Completes	639	63.3	91.6	93.8	89.1
Refusals	59	5.6	8.5	6.2	10.9
Ineligibles					
Does not teach reading in classroom setting	175	17.3	n/a	n/a	n/a
Not teaching	64	6.3	n/a	n/a	n/a
Not beginning teachers/other	72	7.1	n/a	n/a	n/a
Total	1,009	100.0	100.0	100.0	100.0

Source: Mathematica teacher induction survey management system.

n/a = not applicable

The observations focused on pedagogical practices and classroom management. All of the classroom observers were or had been classroom teachers themselves and underwent special training for this study. They visited the study classrooms in late spring 2006 (toward the end of the year), when differences in teacher practices resulting from the comprehensive induction program would most likely be evident. They were blind to the treatment status of the classrooms they observed.

The instrument used to conduct the observations was the Diagnostic Classroom Observation (DCO), formerly known as the Vermont Classroom Observation Tool (VCOT). This classroom observation tool and the methods used to train observers are described in greater detail in Appendix A. We considered many alternative measures of classroom practices but selected the DCO for several reasons. First and foremost, the tool incorporates the most appropriate level of detail on practices that are believed to be part of good instruction. Although some of the alternatives lent themselves to consistent and easy measurement, they tended to focus on activities that could be counted, such as the number of times students raise their hands. In addition, they did not capture complex teacher behaviors, such as whether the teacher makes connections between reading and writing. The DCO measures the teacher practices that current research suggests are essential to good teaching or that have been linked to student achievement growth (Cawelti 2004). Second, the DCO measures instructional practices that closely reflect those recognized by both the ETS and NTC induction programs, particularly literacy instruction. Third, the DCO is simple to complete while in the field. Finally, the DCO is an attractive choice because its developers pair the instrument and written materials with thorough training.[10]

[10] Inter-rater reliability indices from the publisher are not available. In the current study, observers were deemed certified to conduct observations based on a comparison of their 16-item scores to the observations of a "gold standard" panel; following certification, however, inter-rater reliability was not measured in the field.

We observed study teachers once while they were teaching a literacy unit. The observations lasted between one and two hours, with duration dependent on how the district or school structured its class periods. To reduce some of the variability that can occur with literacy classes, trained schedulers asked schools to invite observers into the school during the time when teachers were most likely to teach reading. More detail on the observation procedures can be found in Appendix A.

Observers scored teachers in each of three constructs based on a set of items that are believed to be indicators of good practice: implementation of a lesson, content of a lesson, and classroom culture. The three domains are composed of five, four, and seven items, respectively. Observers rated the extent of evidence of teacher behavior for each item on a five-point scale, showing (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence. For example, for lesson implementation: "The pace of the lesson is appropriate for the developmental level of the students"; for literacy content: "Understanding of content and concepts is taught through *close reading of text* and vocabulary instruction"; and for classroom culture: "Classroom management maximizes learning opportunities." The tool provides observers with examples of specific behaviors to look for in assessing the extent of evidence of teacher practice within each item. We found all items within each of the three literacy constructs to be highly correlated with other items in the construct, based on standardized inter-item reliability coefficients. Psychometric details are presented in Appendix A.

D. Student Records

To gauge how comprehensive induction affected student achievement, we collected student data directly from school districts. The data include student scores, linked to teachers, on standardized tests administered in the spring of each study year (the posttest) and scores for the same students from tests taken in the spring of the prior year (the pretest). For example, in the third year of the study, districts provided pretest scores from spring 2007 and posttest scores from spring 2008.[11] Districts also provided student background data, including race/ethnicity, eligibility for free or reduced-price meals under the National School Lunch Program, English language learner status, disability status, and date of birth (to determine which students were over age for grade).

As shown in Figures III.2 and III.3, some teachers were not eligible for the test score analysis because their districts did not provide pretest and posttest data. State assessment systems under No Child Left Behind typically test students beginning in grade 3, which implies that only teachers in grades 4 and 5 in K–5 elementary schools routinely have students with both posttest and pretest scores. Across one-year and two-year districts and treatment and control groups, of the 1,009 teachers who began in the study in the 2005-2006 school year, districts provided student test score data for 190 teachers in the most recent year of the study, the 2007-2008 school year. The other 819 teachers were either no longer teaching in the district, teaching in non-tested grades or

[11] For three districts that tested at least some students in the fall, we used a fall test as a pretest (at the beginning of the year in the study teacher's classroom) and/or a fall test as a posttest (at the beginning of the next year following enrollment in the study teacher's classroom).

subjects, or teaching in one of two districts that did not provide usable test score data for the teachers' third year of teaching.[12]

Of the 190 teachers for whom we received student test score data, we excluded 27 teachers from the student achievement analysis. A total of 7 teachers were linked to an implausibly high number of students to be a regular classroom teacher, as explained in Appendix A. We believe that these teachers may have been mistakenly linked to students. Another 19 teachers were teaching in grade levels for which a treatment-control comparison could not be made within their district. We deleted one teacher with anomalous test score gains.[13] The requirement that each grade within a district have both treatment and control teachers is to ensure that any peculiar test characteristics, if they exist, are represented in both the treatment and control groups.[14]

The result of these eligibility restrictions is that the analysis sample for conducting test score analysis is smaller than the analysis sample for other outcomes, as anticipated in the design report (Glazerman et al. 2005). Teachers in the student achievement analysis sample in year 3 represented 82 percent of eligible teachers in reading and 84 percent in math. The resulting standard errors of test score impact estimates were in the range of 0.036 to 0.064, meaning that an impact in effect size units of 0.071 to 0.126 would be statistically significant. Although the eligibility restrictions for the test score analysis result in a sample that has a different mix of districts and grades than the larger sample used to analyze teacher retention and other outcomes, and the reasons that some teachers or students are excluded from the analysis sample may be related to test score outcomes, none of these reasons is likely to be related to treatment status. Therefore, we conclude that the estimated impacts on test scores are internally valid. Nonetheless, readers should exercise caution in generalizing the findings because the grades and districts are not a random subset of the full sample.

We made treatment-control comparisons within grades within districts, and then aggregated across grades and districts. Scores were scaled scores, normal curve equivalents, percent correct, or percentile rankings. Within each district-grade combination, we rescaled tests by subtracting the mean score of all students who took that test and dividing by the standard deviation for all test takers. Typically, we used means and standard deviations from a state reference group or a grade-representative norm sample. Further details on aggregation are presented in Appendix A.

[12] All districts provided test score data for the teachers' first two years of teaching, as detailed in Glazerman et al. (2008) and Isenberg et al. (2009), but one of these districts was unable to link teachers to students. We did not collect data from this district for the teachers' third year. Another district refused to provide data for the teachers' third year.

[13] We deleted teachers whose average student gain scores were greater than 1.5 standard deviations above the mean for the reference group (state or norm sample). This resulted in the loss of one classroom, whose students had gains in reading scores that were below the state average, in line with other classrooms in the study, but gains in math scores that would have placed most of them in the 94th percentile or above for the state.

[14] Counts of teachers with valid data pertain to math scores only, for illustration. The corresponding sample sizes for the reading analysis are shown along with those for the math analysis in Figures III.2 and III.3.

Figure III.2. Flow of Teachers Through the Study in One-Year Districts

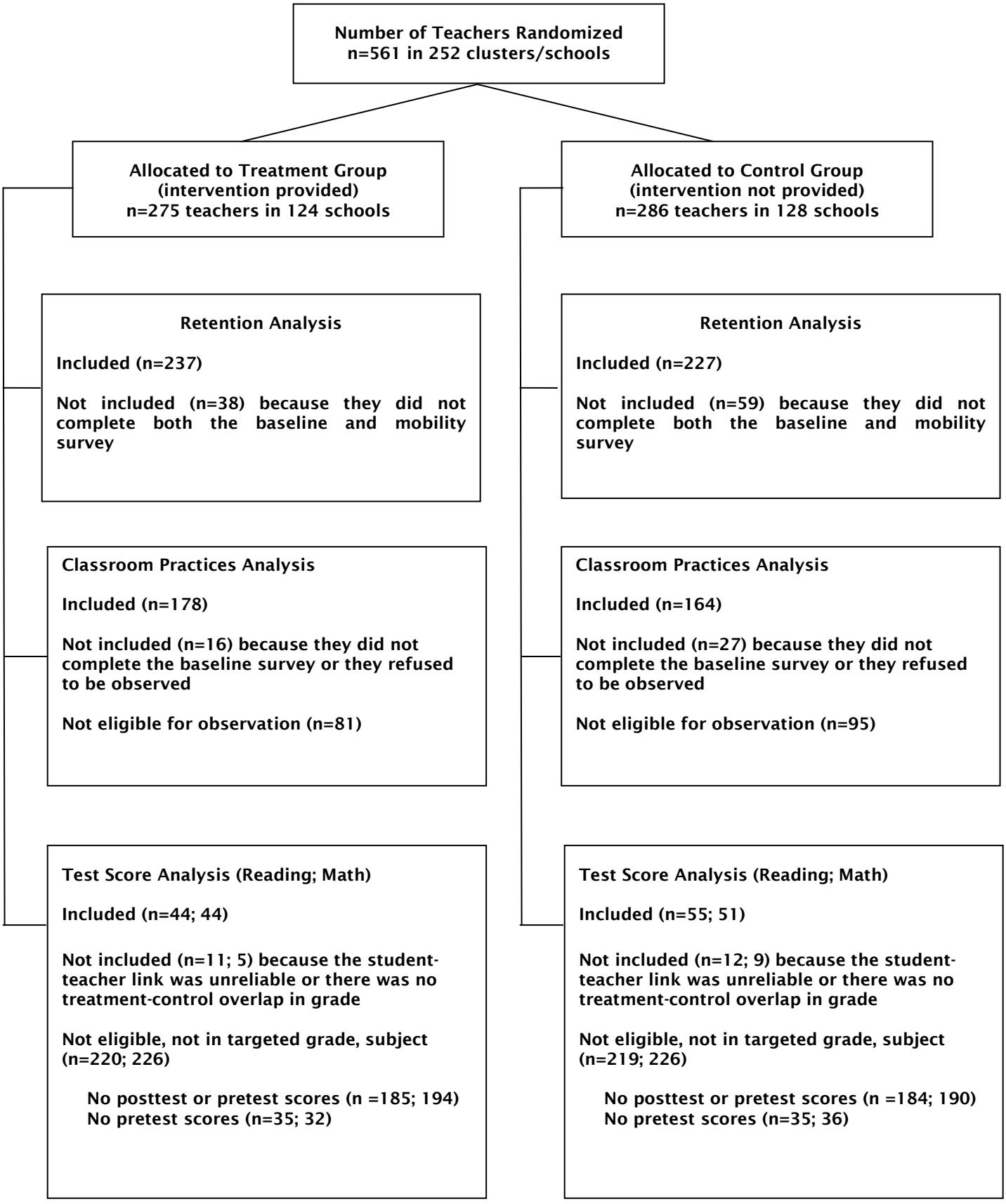
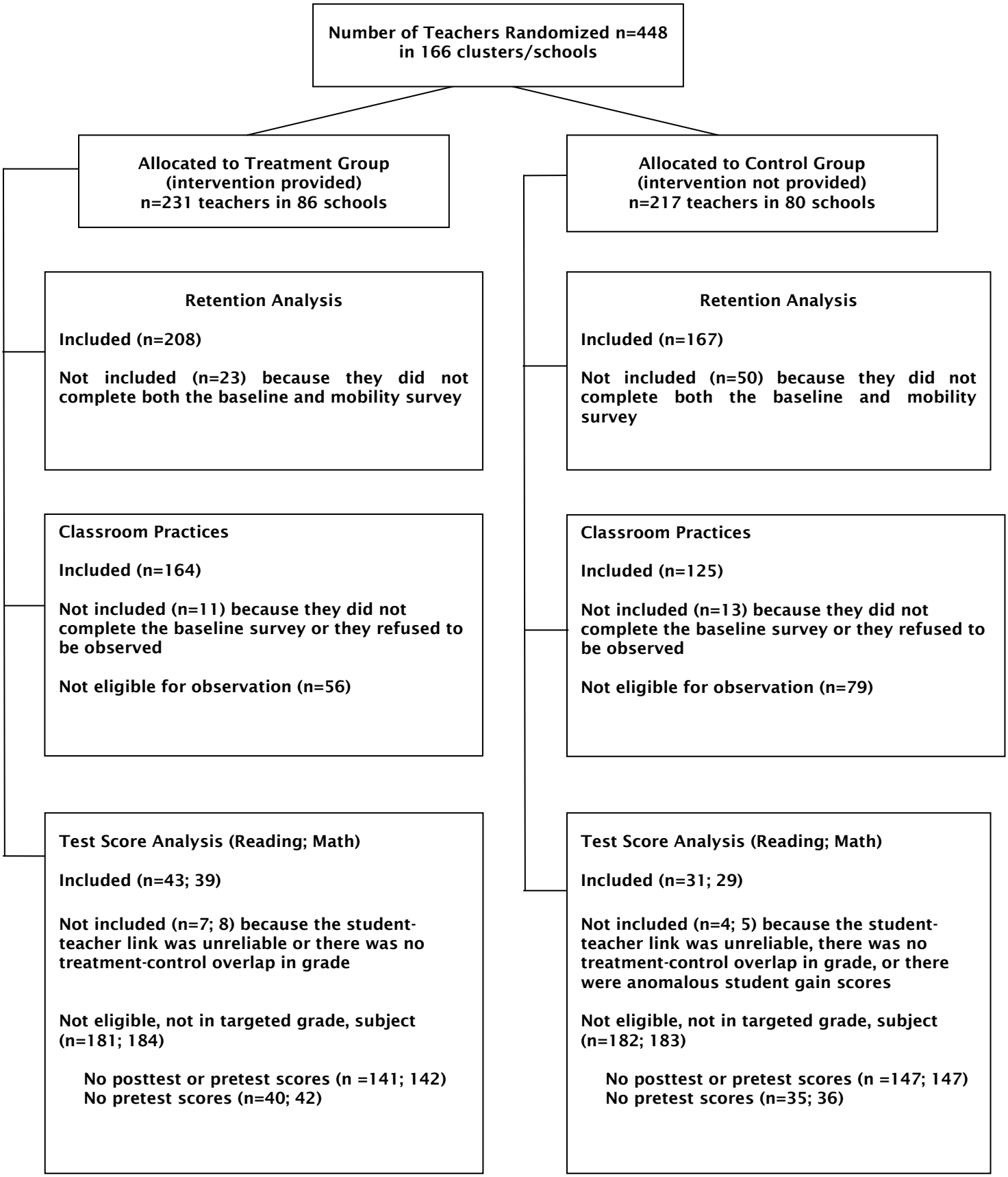


Figure III.3. Flow of Teachers Through the Study in Two-Year Districts



E. Other Supporting Data

To interpret the impact findings, we needed to understand how comprehensive teacher induction programs were delivered and how they compared to the existing array of services. The induction activities surveys were the primary data source, but we gathered supplemental data to enrich the analysis. WestEd staff reviewed materials supplied by the two comprehensive induction program providers (ETS and NTC) to supplement the information we collected through the induction activities surveys. The materials, which provided the basis for the detailed description of program support (see Chapter IV), included documents such as training agenda and materials, curriculum guides, and assessment tools.

IV. PROGRAM IMPLEMENTATION AND INDUCTION SERVICE CONTRAST

To characterize the nature of comprehensive teacher induction and the level of services provided to beginning teachers in the control condition, we measured the types, frequency, and duration of induction activities in both the treatment and control groups from the perspective of the teachers. For the treatment group, we collected additional data on teacher attendance at program events and mentor background characteristics and experience.

This chapter has two parts. The first part describes the intervention provided to the treatment group during the 2005–2006 and 2006–2007 school years. During the 2005–2006 school year, services were provided in all 17 study districts. In 2006–2007, services continued in 7 of the 17 districts.

The second part of the chapter compares the induction experiences of teachers in the treatment group with the experiences of those in the control group, both during and after implementation of the comprehensive induction services in the treatment schools, in both one-year and two-year districts. The gap in services, or service contrast, represents the effect of offering treatment on the type and intensity of induction services received. According to our model of induction services (Figure I.1), the service contrast should be an important precursor to impacts on desirable outcomes such as student test scores and teacher retention.

A. Comprehensive Teacher Induction

To test the hypothesis that a comprehensive teacher induction program would be more effective than the services normally provided to beginning teachers by their schools and districts, we had to identify such a program as well as a provider of program services. Accordingly, we issued a Request for Proposals (RFP) in 2004. The RFP specified that the induction program should include components that earlier research and professional wisdom gleaned from practice had suggested were important features of successful teacher induction programs (Alliance for Excellent Education 2004; Ingersoll and Smith 2004; Smith and Ingersoll 2004; Kelly 2004; Serpell and Bozeman 2000). The components include carefully selected and trained full-time mentors; a curriculum of intensive and structured support for beginning teachers, including orientation, professional development opportunities, and weekly meetings with mentors; a focus on instruction, with opportunities for novice teachers to observe experienced teachers; formative assessment tools that permit evaluation of practice on an ongoing basis and require observations and constructive feedback; and outreach to district- and school-based administrators to educate them about program goals and to garner their systemic support for the program.

A group of outside expert reviewers read and scored the proposals received in response to the RFP. Among those submitted, the ETS and NTC proposals stood out as most closely meeting the study's specified requirements. We selected these programs in order to determine whether the comprehensive induction model is effective in improving classroom practices, student achievement, and teacher retention, rather than whether a particular comprehensive induction program is effective in improving these outcomes. Including two programs increased our ability to generalize findings of the comprehensive induction model relative to including just one program. Furthermore, the expert panel that was convened to select the study's intervention rated both the ETS and NTC programs as high in quality, and the panel agreed that they were similar enough in goals and structure that including both (and pooling impact data across the two programs) would be a fair test of the comprehensive induction model.

The detailed description of the two programs in the following sections is based on information from program documents and data from WestEd's external monitoring of the induction programs' implementation in all districts during 2005–2006 and in the seven districts implementing a second year of induction during 2006–2007. In the first year, WestEd monitors observed all mentor training sessions and webinars (web-based seminars provided by ETS) conducted by the programs and reviewed materials for each event in advance. Monitors interviewed program leaders and staff and received reports from them regularly, weekly at start-up and monthly later in the school year. For each program, the monitors also observed one initial local orientation for beginning teachers, one for administrators, and an end-of-year colloquium for beginning teachers.

WestEd monitors visited each district in the fall; in the spring, they either visited again or conducted semi-structured telephone interviews.[15] Monitors also conducted end-of-year visits, observed a professional development and/or study group session for beginning teachers, observed one weekly mentor meeting, and joined at least one mentor during regular weekly visits with two to four of his or her beginning teachers. During visits and telephone calls, monitors spoke separately with the district coordinator and each mentor to gauge whether districts were receiving all prescribed services from the induction programs; whether the nature and level of effort in districts' implementation were consonant with the programs' intent; whether district coordinators were enabling mentors to fulfill their roles, and whether mentors were carrying out their roles as planned; what local challenges were impeding implementation, if any; and what plans districts and programs had for addressing such challenges.

In the second year of implementation in the seven two-year districts, WestEd reviewed materials and attendance data for each major professional development event and conducted interviews and received reports on a schedule similar to that of the first year. WestEd monitors also made two- or three-day site visits in the first months of the school year to two of the three NTC districts and three of the four ETS districts. During these visits, monitors interviewed district coordinators and mentors and observed professional development events for beginning teachers. Monitors also conducted semi-structured telephone interviews with all district coordinators at the beginning and end of the school year. All but two districts were followed by the same WestEd monitor as in year 1. In these two exceptions, circumstances made it necessary to assign different WestEd monitors, but they had had full monitoring experience with other districts during year 1.

Practitioners and policymakers should be aware that the programs implemented in this study by ETS and NTC were not necessarily the same models that would be delivered outside the study context. First, for study purposes, the objective was consistent implementation of each program, with a high level of fidelity to program design and a quick response to any implementation issues. Second, the providers adapted their programs to ensure that the required components were included in a one-year curriculum to reflect the initial study design. Once it was decided to add a second year, the programs made additional modifications and adaptations to extend the curriculum another year. Finally, the providers adjusted their usual methods of service delivery to meet the requirements of the study in both years. To implement the mentor training, each program organized off-site mentor training sessions, bringing together the mentors from all of the districts in which they were operating, as described later. Outside the study context when there is district-wide implementation

[15] Four of the nine ETS districts (44 percent) and three of the eight NTC districts (38 percent) were visited. The others were interviewed by telephone.

with a larger number of mentors, training typically occurs within the district, rather than off site with mentors from other districts.

1. Administrative Support Structure

To understand the treatment provided by each program, we begin with an overview of the key roles played by designated staff members in implementing the programs (Figure IV.1). Oversight for implementation of the ETS and NTC programs was the responsibility of a designated staff member from the respective organizations.[16] These program leaders directed all activities and provided substantive leadership. They led the adaptation of program materials for use in the study, played integral roles in the design and delivery of mentor trainings, and supported the work of their own program staff and site-based district coordinators. They held monthly staff meetings and stayed in close contact with district coordinators for purposes such as preparing or debriefing the weekly mentor meetings, providing ideas for optimizing mentors' working conditions, monitoring the fidelity of district implementation of induction program content and activities, and fostering productive relationships among various staff members. In year 2, an ETS co-leader left the study and was replaced by one of the mentors, whereas the NTC leader continued in her role.[17]

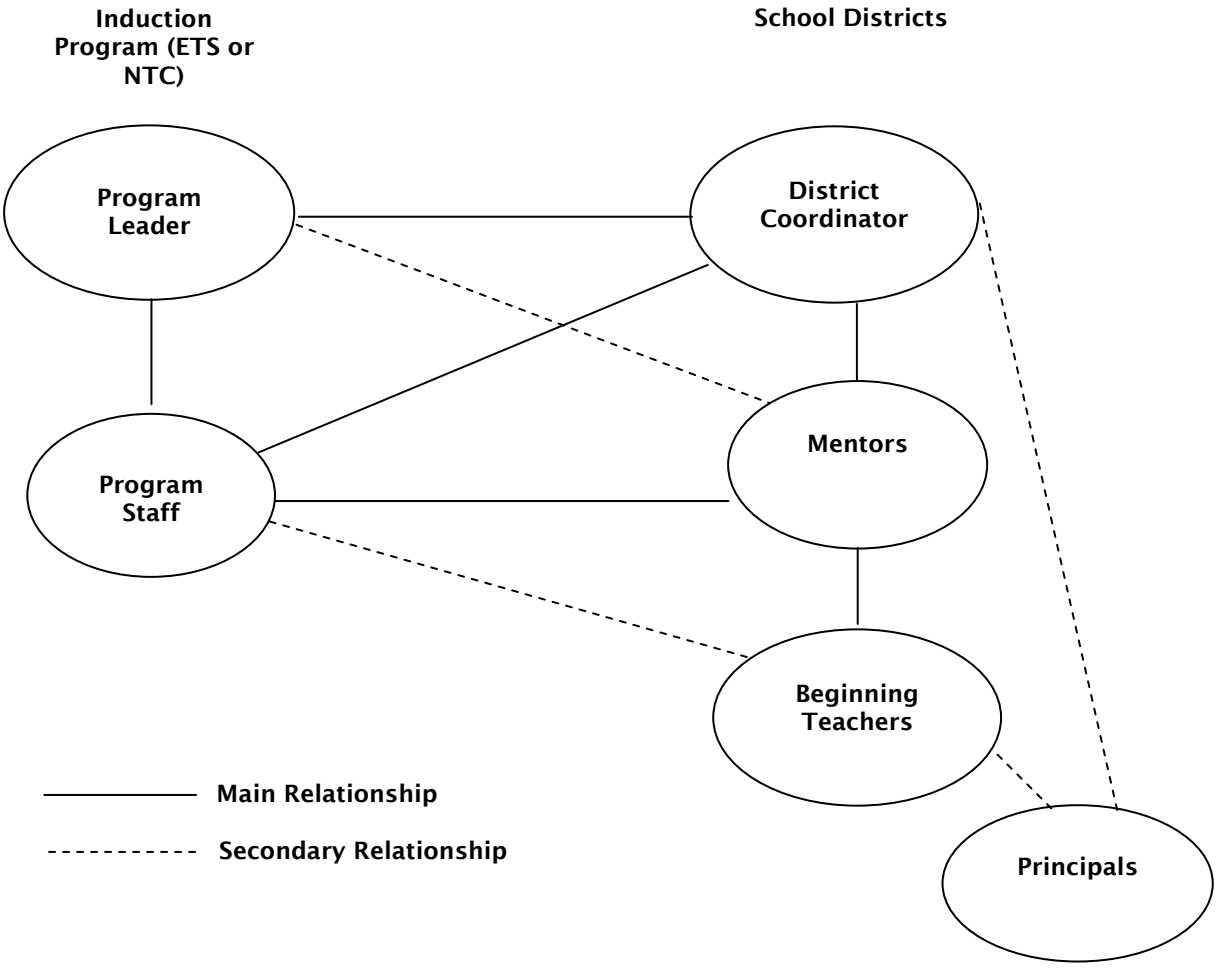
In collaboration with the program leaders, designated ETS and NTC program staff worked with assigned districts to help implement the program consistently across the districts.[18] In the second year, in the seven districts that continued implementation, all program staff had experience in this role from the previous year. Three districts were served by the same person as in year 1; two ETS and two NTC districts were served by a different person in the second year. The program staff made monthly visits to each district, during which they delivered or facilitated a professional development session for beginning teachers, worked with district coordinators on issues related to program implementation, met with the mentors to continue building their skills, and shadowed them on their weekly visits with beginning teachers. While shadowing the mentors, program staff could observe firsthand any needs for program support as related to mentoring skills or the use of program processes and tools. This provided staff with the opportunity to discuss how the program could best address the needs and circumstances of teachers in each setting. Between visits, program staff engaged in regular and frequent communication with mentors and district coordinators to discuss any issues that surfaced and to provide ongoing direction.

[16] In addition, WestEd staff provided external oversight of services provided in order to help address any issues that arose and to keep implementation consistent across all sites.

[17] The ETS co-leader for the study, who had served under the program leader in year 1, left because of personal circumstances. A mentor from year 1 was promoted to serve as co-leader in year 2, and this person also continued to serve as program staff for a district. Whereas the NTC leader continued in this role, this person also served as program staff for one of the districts in year 2.

[18] Each program staff member served one or two districts. Staff members spent between 20 percent and 30 percent of their time serving each district.

Figure IV.1. Structure of Roles in the Induction Program



Districts designated their own staff members to provide local oversight to program implementation. District coordinators worked in departments of human resources or professional development. In year 1, a key function was to help establish district positions for mentors and recruit candidates for these positions, establish procedures for job reporting and evaluation, create functional working conditions for mentors by locating office space, and setting up email and telephone access. They also helped identify beginning teachers to participate in the study, assign teachers to mentors, find appropriate settings for program events and schedule them on the district's master calendar, and address occasional program implementation challenges. In both years of program implementation, district coordinators facilitated mentors' weekly meetings and joined mentors at off-site trainings throughout the year. To reduce the chances that treatment and control groups would share any services or resources, we asked districts to assign coordinators who would not also be involved in the district's own induction activities at the elementary level.

The individuals serving as district coordinators in year 1 continued in that role in year 2; in one district of each program, however, a replacement was named because the original person could not continue due to changes in her main position. The district coordinators worked with the programs at the outset of year 2 to adjust mentors' workloads depending on which beginning teachers stayed or left after year 1, arranged settings for program events, and scheduled them on the district's master calendar. In both years, district coordinators spent 10 to 15 percent of their time on these functions,

with considerably more time early in the year and much less time as the year progressed (about 30 percent and less than 10 percent, respectively, in year 1, and about 20 percent and less than 10 percent, respectively, in year 2).

According to interviews with district coordinators by WestEd monitors, those with more influence in the district were better able to broker the organizational arrangements that needed to be made across district departments and levels. For example, coordinators had to obtain approval for scheduling professional development sessions on the district's master calendar and locate mentor offices or rooms to serve as meeting spaces. Factors that helped coordinators in their role included the support of high-level district administrators, coaching or mentoring experience, and good rapport with program staff. In contrast, smooth program implementation was more difficult when coordinators were less responsive or influential. Given that the coordinator role was an addition to a full set of existing responsibilities, coordinators struggled to carve out the time needed for program implementation.[19]

Principals also played an important role in program implementation. Both ETS and NTC asked principals to encourage and support beginning teachers' participation in induction activities, particularly by permitting them to attend professional development sessions and minimizing conflicts that could impede mentors' efforts to schedule time with them. In both school years, the programs offered an initial orientation for administrators, and NTC held a fall and spring administrator briefing over breakfast.[20] During these events, program leaders and district coordinators sought to gain administrators' support for their beginning teachers' participation in the induction program and for the involvement of the mentor assigned to their school. The orientation events provided brief overviews of beginning teachers' needs for support and development and the induction program's purposes and activities. Both programs strongly cautioned mentors against sharing specific information with principals that could affect the beginning teachers' job evaluations and compromise confidentiality and openness in the mentor/mentee relationship.

Overall, school and district officials evidenced wide variation in the level of principal support, ranging from those who were extremely supportive, actively encouraging teachers to make the most of the induction opportunities, to principals who actively resisted participation and would not permit teachers to be released for program activities.[21] The resistant principals either required beginning teachers to attend school or district events that conflicted with induction program activities or imposed heavy restrictions on when mentors could visit teachers. During year 1, five principals out of the 210 treatment schools in the study fell into this latter category. Such resistance abated over the course of this year and the next in response to the intervention of district coordinators, mentors, and program staff. Induction programs encouraged mentors to visit their beginning teachers' principals at least once a month. When program staff shadowed mentors, they also met briefly with principals who did not strongly support the induction program in order to help convince them of its value.

---

[19] When ETS and NTC are contracted by a district to implement their respective programs, not in the context of a study, district coordinators spend more than 15 percent of their time on program implementation.

[20] In year 2, NTC facilitated mentors taking a presentation role for part of the event to enhance principals' perception of their roles and expertise.

[21] WestEd's monitors gathered this information through interviews with program leaders, district coordinators, and mentors, and through direct observations of participants at the NTC administrator breakfast briefings.

2. Mentors

At the heart of the comprehensive induction services was the support provided by a highly trained, full-time mentor. Mentors were most frequently responsible for 12 beginning teachers (32 percent), although caseloads ranged from 8 to 14 teachers over the course of each year. With mentoring as the largest component of the comprehensive induction programs, mentors necessarily underwent careful selection and training. At the outset of the study, programs worked with each district by providing them with a written job posting, guidance on selection of the Interview Team, and a set of interview questions and rubric for mentor selection. The selection rubric called for individuals with a minimum of five years of teaching experience in elementary school, recognition as an exemplary teacher, and expertise in designing and implementing standards-based instruction. In each district, candidates who met these criteria were interviewed by a committee that included the district coordinator for the study and other participants, such as representatives from human resources, the teacher's union, and professional development; an assistant superintendent for instruction; other experienced mentors; and/or school administrators. Staff from the comprehensive induction programs traveled to the interviews or conducted telephone consultations with the district coordinators to help in the selection of mentors, though districts were responsible for the final decisions. In all but three districts, two or more people applied for each mentor position. One instance of turnover among mentors occurred during the first year of program implementation. Mentors involved in year 1 implementation continued to fill the mentor positions for year 2 of the study. Because some beginning teachers left teaching or the participating districts after year 1, mentor caseloads were adjusted at the beginning of year 2. Whenever possible, beginning teachers were served by the same mentor during years 1 and 2.[22]

Table IV.1 describes the background of the 44 mentors selected to deliver the comprehensive induction services in the study districts. These data are taken from a survey administered to mentors at the outset of program implementation in year 1. All mentors reported at least 5 years of teaching experience, with an average of 17.9 years. Most (86 percent) held a master's degree and 14 percent were certified through the National Board of Professional Teaching Standards. A majority (82 percent) had come to the mentoring role from a position as a classroom teacher, and 46 percent had ever worked in nonteaching positions in education. The average age of these mentors was 43 in 2005 and 51 percent were white, non-Hispanic. Although the mentors were implementing the particular program under study for the first time during the 2005–2006 school year, 77 percent reported having prior mentoring experience—6.2 years on average—and among those, 74 percent had previously attended mentor training. The most commonly reported areas of training addressed classroom management, giving effective feedback, and mentor roles (over 87 percent for each area).

[22] Halfway through year 2, one NTC mentor left the study for a career advancement opportunity; the service loads of remaining mentors in this district were reconfigured to distribute responsibility for the beginning teachers previously assigned to the departing mentor.

Table IV.1. Mentor Characteristics

Characteristics	Percentage
Race/Ethnicity: Percentage White, Non-Hispanic	51.2
Education: Has Master's Degree	86.4
Certified Through National Board of Professional Teaching Standards (NBPTS)	13.6
Teaching Experience	
Last position before mentoring was as a classroom teacher	81.8
Ever worked in nonteaching position(s) within education	45.5
Mentoring Background	
Any mentoring experience	77.3
Any previous mentoring training (if have mentoring experience)	73.5
Areas of Mentor Training (If Received Mentor Training)	
Classroom management	87.5
Giving effective feedback	87.5
Mentor roles	87.5
Coaching strategies	80.0
Lesson planning	79.2
Classroom observations	65.2
Helping adult learners set goals	52.2
Analyzing student work	50.0
Leading study groups	39.1
Coaching in literacy/language	27.5
Coaching in math	20.8

	Average	Range (Min., Max.)
Age in 2005 (Years)	43.0	(28, 61)
Teaching Experience (Years)	17.9	(5, 35)
Experience in Nonteaching Position(s) Within Education (Years)	1.4	(0, 6.8)
Years of Mentoring Experience (If Have Mentoring Experience)	6.2	(1, 30)
Caseload (Number of Beginning Teachers)	11.7	(8, 14)
Sample Size (Mentors)	44	

Source: Mathematica Mentor Survey administered in fall 2005 to all study mentors.

Once mentors were selected for program participation, both ETS and NTC trained their respective mentors during the first year of program implementation in four training sessions each that were extensive, intensive, and focused. Two of the eight trainings were fully attended. One mentor was absent at the six other trainings (a different person in each instance). These absences were due to personal circumstances. Each program brought mentors together for a total of 10 or 12 days (ETS and NTC, respectively), devoting two to three days per session (Figure IV.2). By convening mentors from all of a program's study sites at a single location, trainings provided opportunities for cross-site collaboration designed to enrich learning the programs' curricula and also to foster concrete discussions about how best to address any implementation issues. By holding

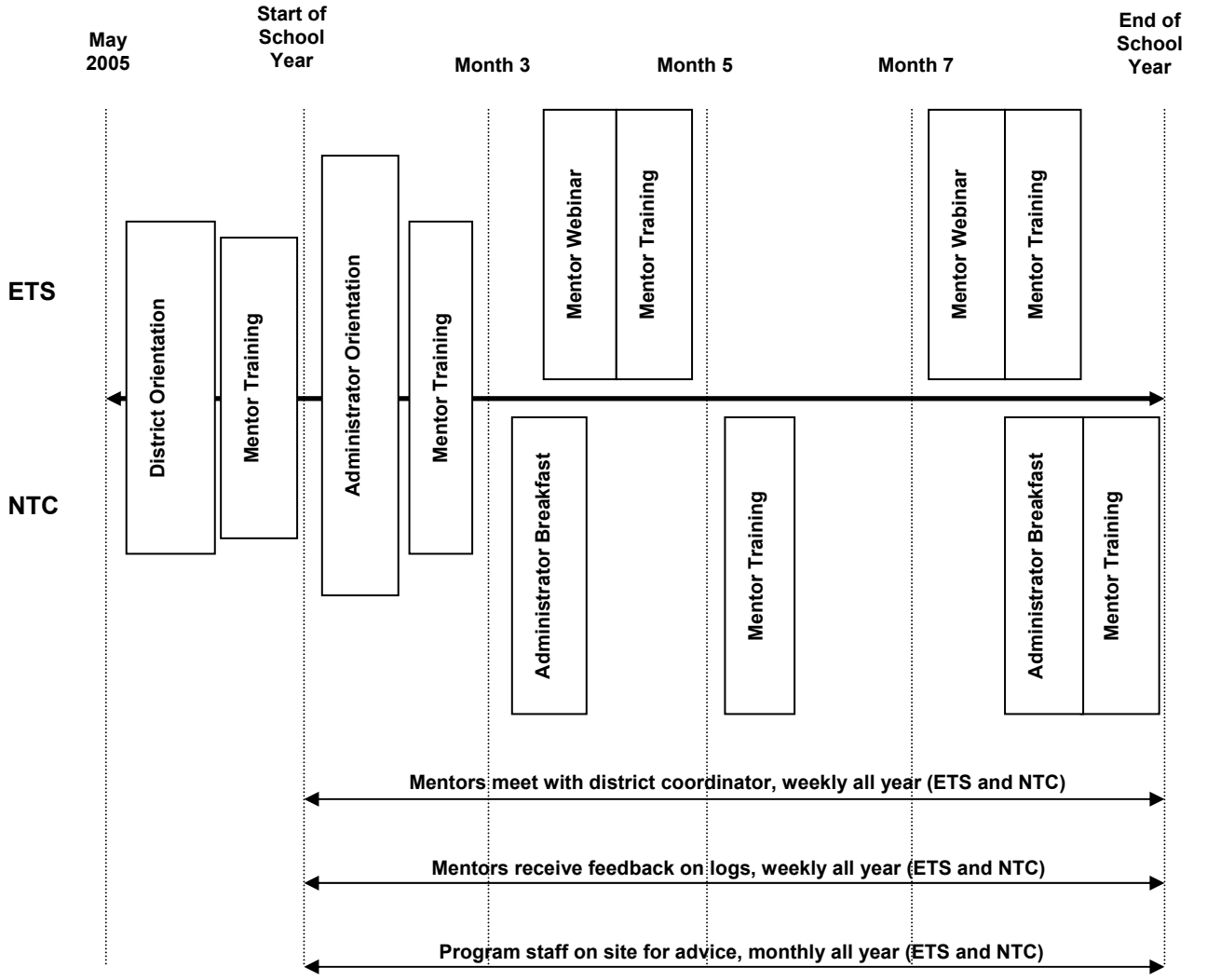
sessions over the course of the 2005–2006 school year, program staff were able to provide training as it was needed. Trainings previewed the content of upcoming professional development sessions and gradually introduced forms and processes of mentor/mentee work. For example, forms and processes for beginning teachers' midyear reflections on their instructional practices and professional development were not introduced to mentors until the second training (fall); ways for beginning teachers to analyze student work in the spring were introduced during the third training (winter); and the fourth training (spring) explored ways of prompting beginning teachers to initiate longer-range goals for their development.

Trainings focused on active learning in two main areas: (1) improving beginning teachers' instruction, including the use of forms and processes to advance it; and (2) mentoring skills for working with beginning teachers, such as using evidence from teachers' instruction rather than presenting opinions, and conversational techniques, such as paraphrasing and asking clarifying questions. Programs also spent some training time on how to address beginning teachers' survival needs and other more general needs, with ETS spending 5 percent of mentors' training time and NTC spending up to 10 percent of training time on this topic.[23]

The programs were also intentionally designed to provide mentors with support and development opportunities throughout the academic year via activities beyond the four formal training sessions. The planned activities involved interaction with program staff, other mentors, and district coordinators. WestEd's monitoring data indicate that when program staff visited their districts each month, they joined the weekly meeting to help mentors become more familiar with program content and tools. The weekly meetings also allowed mentors to exchange ideas on successes and challenges in working with beginning teachers and in gaining the support of building administrators. At the outset of the school year, district coordinators provided substantive advice during weekly mentor meetings and three-quarters of them continued to join mentor meetings throughout the year. Program staff and district coordinators regularly responded to telephone or email inquiries from mentors, and the ETS program held two one-hour webinars for mentors and district coordinators. The fall webinar helped mentors shift from providing the type of general support needed by beginning teachers at the outset of the year to focusing on specific development of teachers' instructional practices. During the spring webinar, coordinators and mentors shared ideas for planning the end-of-year colloquium. (The NTC program did not include webinars but covered these topics during its additional two days of mentor training over the year.)

[23] Examples of survival and more general needs include how to interact with the principal, how to deal with teachers' emotional needs, how to deal with a particularly difficult student, or how to find classroom resources.

Figure IV.2. Comprehensive Induction Program Training for Mentors, District Coordinators, and Administrators: 2005–2006 School Year



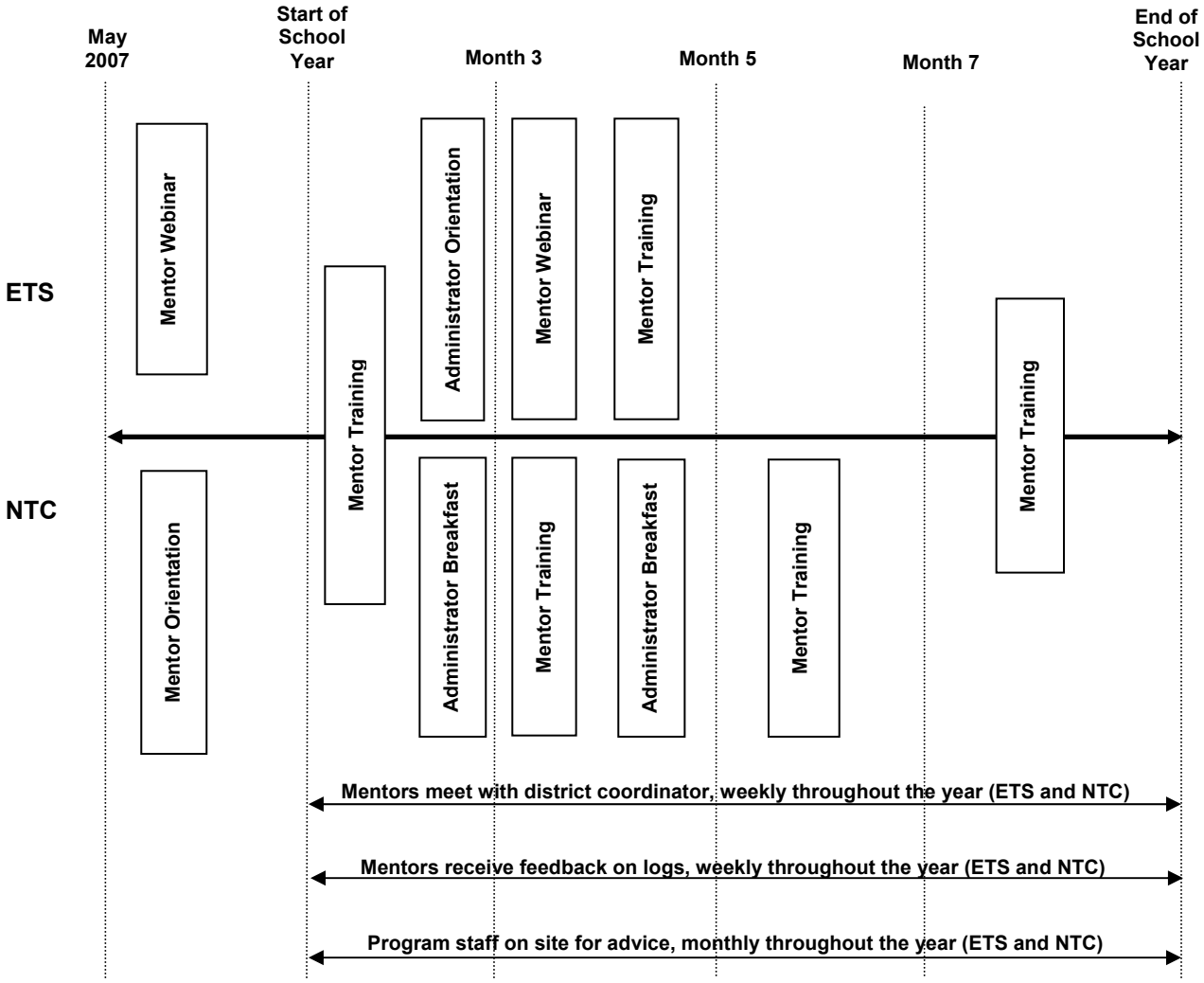
Note: Activities common to both providers are shown on both sides of the horizontal divider between ETS and NTC. The district orientation was offered to district coordinators and administrators from the central office. The administrator orientation was offered to school building administrators.

The program leaders and program staff also reviewed and provided feedback on the logs used by mentors to summarize weekly meetings with teachers. Feedback included discussion about why a beginning teacher was requiring or receiving more or less contact time than average, ideas for addressing beginning teachers' needs, how to use program tools, and how to stay on schedule with program implementation.

During the second year, ETS and NTC continued intensive training of their respective mentors in the seven districts that continued program implementation. Each program brought mentors together for a total of 8 and 10 days over three and four sessions (ETS and NTC, respectively), devoting 1.5 to 2.5 days per session (Figure IV.3). In addition to trainings, NTC held a late summer retreat with its mentors to debrief the first year of program implementation and help with the final strategic planning for the second year. At the outset of the 2006–2007 school year, ETS held a two-hour webinar for initial orientation of its mentors, whereas NTC held an early training session. A

second ETS webinar was held between the first two ETS trainings. For a training later in the year, one of the districts hosted the training.

Figure IV.3. Comprehensive Induction Program Training for Mentors, District Coordinators, and Administrators: 2006–2007 School Year



Note: Activities common to both providers are shown on both sides of the horizontal divider between ETS and NTC. The administrator orientation was offered to school building administrators.

All mentors participated in the trainings, which reflected a focus similar to that of year 1. Given mentors' experience from their training in the first year, activities during the second year included less emphasis on learning mentoring skills. Instead, NTC training paid particular attention to the equitable engagement of diverse students, and part of the spring training was spent having mentors shadow their peers during meetings with beginning teachers. For ETS, the training was expanded to include a focus on the content and conduct of its Teacher Learning Communities, a new component of its professional development activities in year 2, described later in this chapter.

Similar to the support described for year 1 of implementation, the programs were also intentionally designed to provide mentors with support and development opportunities throughout the academic year through activities beyond the four formal training sessions, using the same strategies described earlier for year 1.

3. Program Services and Activities

The comprehensive induction programs both provided structured mentoring following a curriculum designed to improve instructional practice, monthly professional development session, opportunities to observe veteran teachers, and an end-of-year colloquium. Below we describe the services and activities in the first and second years of program implementation.

a. Year 1 Program Services and Activities (2005–2006 School Year)

In the first year of program implementation, mentoring of beginning teachers began during the first week of school whenever possible, following an orientation session during which teachers were introduced to induction program goals and schedules. On average across the districts, half of the mentors were able to visit their beginning teachers before the first day of school to get acquainted and help set up classrooms.[24] Once the school year was underway, mentors tried to visit their beginning teachers at the same time every week, but meetings were rearranged as needed to accommodate circumstances or to accomplish a specific task, such as observing a particular lesson.[25]

All beginning teachers in the treatment group were also expected to participate in monthly professional development (PD) sessions, and the ETS districts offered monthly study groups—mentor-facilitated peer support meetings for beginning teachers. Beginning teachers also observed veteran teachers once or twice during the year. At the end of the school year, beginning teachers participated in a colloquium. Each of these induction activities is described in more detail below.

Mentoring. Both the ETS and NTC programs consist of a yearlong curriculum for beginning teachers that focuses on effective teaching (Table IV.2). The ETS program defines effective teaching in terms of 22 critical components organized into four general domains of professional practice. The components are aligned with the Interstate New Teacher Assessment and Support Consortium (INTASC 1992) principles.[26] The NTC induction model defines effective teaching in terms of six Professional Teaching Standards.[27] Each standard or domain is broken into a succession of more discretely defined categories of teaching behaviors. The mentor's goal is to help beginning teachers use evidence from their own practice to recognize and implement effective instruction as defined by the domains or standards. Both induction programs use a continuum of performance as a means for teachers to establish a benchmark and improve their instructional practice (Table IV.3).

---

[24] The primary obstacle to holding these early meetings was the delay in district staff's identifying the beginning teachers in each school for the study. This challenge was due to operating in a study context; districts may have been able to begin providing mentoring services more quickly in the absence of the study since they could have sent mentors out to schools in which principals could readily identify beginning teachers with whom they would work. Additionally, 12 percent of beginning teachers were hired after the school year began, further contributing to delays in identifying teachers and assigning mentors.

[25] Especially in the early part of the 2005–2006 school year, mentors spent extra time with beginning teachers who were experiencing serious survival or instructional challenges (data on the frequency and duration of these meetings are unavailable). Program staff monitored these situations to ensure that such services did not take time away from focusing on instruction for those teachers who were on track in their development.

[26] The ETS program derives its content from *Enhancing Professional Practice: A Framework for Teaching* (Association for Supervision and Curriculum Development 1996).

[27] The content of the NTC program is based on two documents—*California's Standards for the Teaching Profession* (California Commission on Teacher Credentialing 1997) and *Continuum of Teacher Development* (New Teacher Center 2002).

The first-year curriculum of ETS is organized around seven Pathwise Induction Events**,** each of which is designed to help beginning teachers explore a particular aspect of their practice and become increasingly proficient as an educator. The initial event requires teachers to investigate their school and community and to develop profiles of the students in their class. In two events, mentors observe beginning teachers in the classroom and provide feedback on their practices, planning materials, and students' work. Three events involve a structured series of activities through which teachers explore a certain aspect of their practice as related to (1) establishing a positive classroom environment, (2) designing an instructional experience, and (3) analyzing students' work. Teachers identify a particular practice in each of these areas, implement it, and then reflect on the experience. Each event concludes with the development of an Individual Growth Plan in that respective area. The last event is a colloquium for all beginning teachers in a district during which they conduct a self-assessment.

The centerpiece of the NTC mentoring model is the NTC Formative Assessment System (FAS). FAS involves a series of collaborative processes between the mentor and the beginning teacher that aims to collect and analyze a variety of data focused on teacher practices and student learning. A set of protocols and forms helps structure mentor/teacher interactions, although an individual teacher's needs determine the precise focus and pace. FAS's central tool is a collaborative assessment log that provides the framework for the mentor's and beginning teacher's weekly conversation. The teacher uses the log to record information on recent successes and challenges and specific next steps. FAS focuses on two key areas in a teacher's development: (1) professional goal setting and (2) classroom practices. Professional goal setting involves both setting goals and reflecting on instructional practices in relation to the model's six teaching standards (Table IV.2) and the continuum of performance (Table IV.3). Teachers identify an area of practice as a focus area, develop a plan to achieve particular goals, and then assess their progress. Teachers establish an individual learning plan and conduct a midyear review to assess progress in meeting goals.

Table IV.2. ETS and NTC Content: Four Domains and Six Professional Teaching Standards

ETS Domains of Professional Practice		
Domains	Example, Subcategories of a Domain (*Instruction*)	Example, Details of Subcategory (*Engaging Students in Learning*)
1. Planning and preparation 2. Classroom environment 3. ***Instruction**** 4. Professional responsibilities **See next column for details*	Communicating clearly and accurately Using questioning and discussion techniques ***Engaging students in learning**** Providing feedback to students Demonstrating flexibility and responsiveness **See next column for details*	Representation of content Activities and assignments Grouping of students Instructional materials and resources Structure and pacing
NTC Professional Teaching Standards		
Professional Teaching Standards	Example, Subcategories of a Standard (*Engaging/Supporting All Students in Learning*)	Example, Details of Subcategory (*Promoting Self-Directed, Reflective Learning for All Students*)
1. Planning instruction and designing learning experiences 2. Creating/maintaining effective environments 3. Understanding/organizing subject matter 4. Development as a professional educator 5. ***Engaging/supporting all students in learning**** 6. Assessing student learning **See next column for details*	Connecting prior knowledge, life experiences, and interests with learning goals ***Promoting self-directed, reflective learning for all students**** Using variety of instructional strategies and resources to respond to students' diverse needs Facilitating learning experiences that promote autonomy, interaction, and choice Engaging students in problem solving and critical thinking to make subject matter meaningful **See next column for details*	Motivate students to initiate their own learning and strive for challenging goals Describe their learning processes and progress Explain clear learning goals for students Engage students in examining their work and work of peers Help students develop and use strategies for knowing, reflecting on, and monitoring their learning Help students use strategies for accessing knowledge and information *(Note: Above entries are slightly abbreviated versions of the source document.)*

Source: The ETS program derives its content from *Enhancing Professional Practice: A Framework for Teaching* (Danielson 1996). The content of the NTC program is based on two documents—*California's Standards for the Teaching Profession* (California Commission on Teacher Credentialing 1997) and *Continuum of Teacher Development* (New Teacher Center 2002).

Table IV.3. Example of ETS and NTC Detailed Specifications for Development of Beginning Teachers' Practices

ETS: Domain 3 (Instruction): Engaging Students in Learning: Representation of Content			
Level 1: Unsatisfactory **Representation of content is inappropriate and unclear or uses poor examples and analogies.**	**Level 2: Basic** **Representation of content is inconsistent in quality; some portions are done skillfully, with examples, while others are difficult to follow.**	**Level 3: Proficient** **Representation of content is appropriate and links well with students' knowledge and experience.**	**Level 4: Distinguished** **Representation of content is appropriate and links well with students' knowledge and experiences. Students contribute to representation of content.**

NTC: Standard 5 (Engaging/Supporting All Students in Learning): Promoting Self-Directed, Reflective Learning for All Students				
Level 1: Beginning **Directs student learning experiences and monitors students' progress within a specific lesson. Assistance is provided as requested by students.**	**Level 2: Emerging** **Provides some opportunities for students to monitor their own work and to reflect on progress and process.**	**Level 3: Applying** **Supports students in developing skills needed to monitor their own learning. Students have opportunities to reflect on and discuss progress and process.**	**Level 4: Integrating** **Structures learning activities that enable students to set goals and develop strategies for demonstrating, monitoring, and reflecting on progress and process.**	**Level 5: Innovating** **Facilitates students to initiate learning goals and set criteria for demonstrating and evaluating work. Students reflect on progress/process as a regular part of learning experiences.**

Source: The ETS program derives its content from *Enhancing Professional Practice: A Framework for Teaching* (Danielson 1996). The content of the NTC program is based on two documents—*California's Standards for the Teaching Profession* (California Commission on Teacher Credentialing 1997) and *Continuum of Teacher Development* (New Teacher Center 2002).

Classroom practice focuses on students' learning needs and teachers' instruction. Various FAS tools help mentors and teachers collaboratively develop an understanding of school and community resources as well as student profiles. Additional tools focus on analyzing students' work to permit development of a better understanding of learning needs and how to address them, communicating effectively with parents, and planning lessons. Several tools help the mentor collect data from regular classroom observations of the teacher.

To cover the ETS and NTC program curricula, programs expected mentors to allocate approximately two hours for contact time each week with every beginning teacher in their caseload.[28] Mentors were expected to spend some of that time every week meeting with beginning

[28] Average actual time spent with a mentor in one-year and two-year districts is shown in Tables V.3 and VI.3, respectively. However, these data do not distinguish between time spent with a treatment mentor and time spent with other mentors.

teachers for one-on-one conversation, particularly around the induction programs' teacher learning activities. For the balance of the weekly allotment of time, mentors exercised professional judgment in using a range of strategies for assisting beginning teachers with induction program activities or general beginning teacher needs, for example, observing instruction, reviewing lesson plans and instructional materials, providing a demonstration lesson, reviewing student work, or interacting with students to assist teachers in understanding their students' learning challenges.

Monthly Professional Development Sessions.[29] During the 2005–2006 school year, both ETS and NTC held monthly, two-hour professional development sessions (Figure IV.4),[30] which complemented the interactions between mentors and beginning teachers as described in the seven ETS events and NTC's FAS (Table IV.4). On average, the professional development sessions drew 72 percent and 65 percent of the ETS and NTC beginning teachers, respectively, as shown in Tables IV.5 and IV.6. However, average attendance ranged from almost universal attendance in one district (93 percent) to less than half in another (43 percent).

Study Groups. In the ETS program, the mentors and beginning teachers met monthly in informal study groups. This gave teachers an opportunity to discuss with their mentors how they were progressing in their practice, challenges they faced, and approaches for addressing the challenges. The meetings also enabled teachers to exchange ideas and information related to their teaching practices. The average attendance at ETS monthly study groups was 69 percent, ranging across districts from 63 to 84 percent.

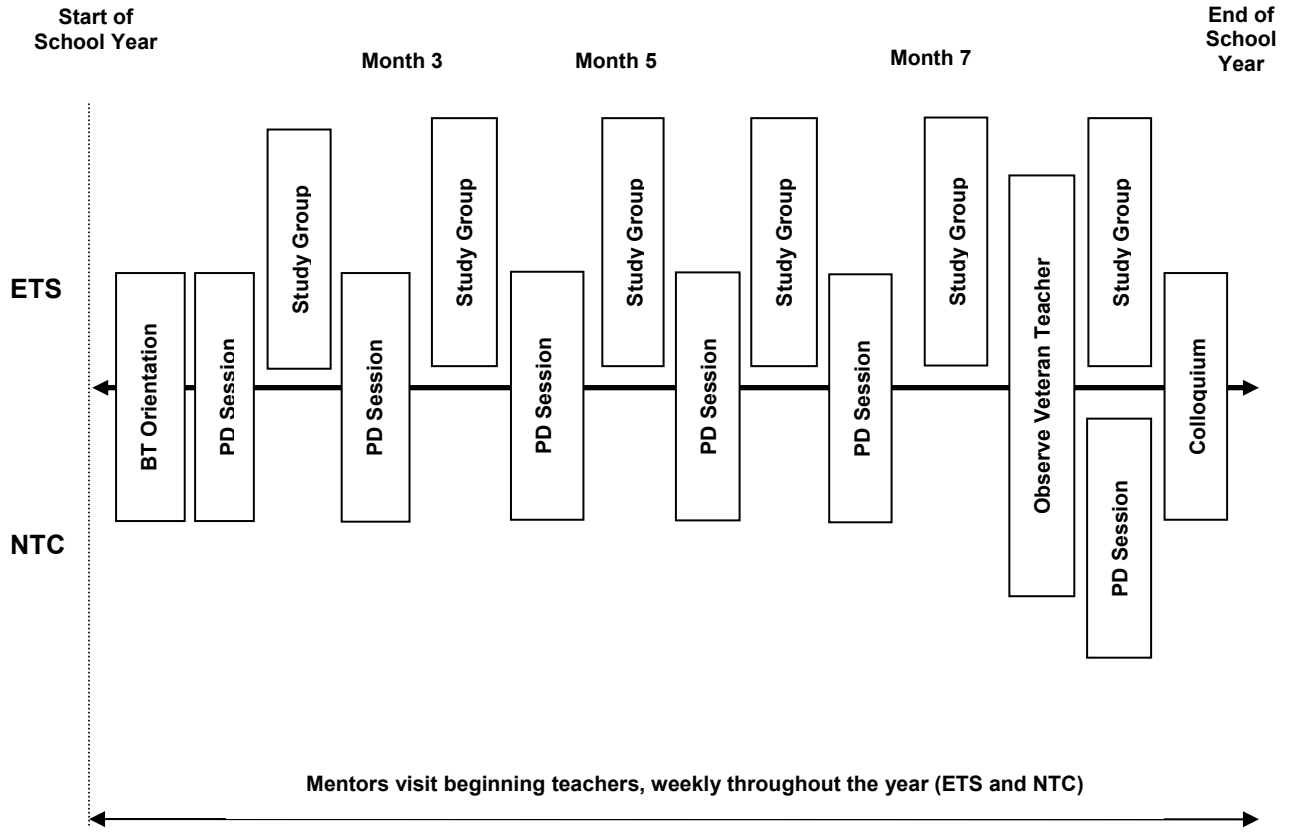
Observation of Veteran Teachers. Mentors arranged one or two formal opportunities for beginning teachers to observe experienced teachers, with an attempt to select observations that would be relevant to the instructional goals of the beginning teachers. They provided advance guidance to beginning teachers on what to observe, as well as methods and forms for attending to the focal instructional practices and recording observations of them. Mentors debriefed the observations with beginning teachers to discuss what they learned.[31]

[29] In five districts, unexpected scheduling conflicts in the master calendar or other district factors (for example, temporary labor disputes) resulted in cancellation of one professional development session with no opportunity to reschedule.

[30] The first NTC session was a full day.

[31] To limit the time burden on teachers, no professional development session was held in the month(s) when the observations were conducted. Programs encouraged mentors to accompany beginning teachers for the observations, but it was challenging for mentors to accomplish this while maintaining their regular weekly travel to multiple schools for a meeting with every beginning teacher in their caseload. Data on the percentage of treatment teachers who observed veteran teachers together with their mentors and who discussed the observations with mentors during debriefings are unavailable.

Figure IV.4. Comprehensive Induction Program Activities for Beginning Teachers: 2005–2006 School Year



Notes: BT = beginning teacher; PD = professional development. Activities common to both providers are shown on both sides of the horizontal divider between ETS and NTC.

Table IV.4. Topics for Monthly Professional Development Sessions, by Program

ETS	NTC
Communication with families	**Effective learning environment (the only full-day session)**
Classroom management	**Engaging all students**
Differentiated instruction for ELL and special needs students	**Assessing all students**
Evidence-centered teaching and assessment	**Planning instruction**
Analyzing and sharing student work	**Understanding and organizing subject matter**
Examining evidence of professional growth by sharing work from induction program activities	**Developing as a professional educator (colloquium)**
Beginning teacher self-assessment and sharing of learning (colloquium)	

Source: The ETS program derives its content from *Enhancing Professional Practice: a Framework for Teaching* (Danielson 1996). The content of the NTC program is based on two documents—*California's Standards for the Teaching Profession* (California Commission on Teacher Credentialing 1997), *Continuum of Teacher Development* (New Teacher Center 2002), and other unpublished materials provided to the study authors by program staff.

Table IV.5. Teacher Attendance at ETS Induction Activities (Percentages): 2005–2006 School Year

		Range of Average Attendance Across Districts		Regularity of Attendance	
Activity	**Average Attendance of BTs**	**High**	**Low**	**Teachers Missing No More Than 1 Session**	**Teachers Missing 3 or More Sessions**
Orientation[a]	**n/a**	**n/a**	**n/a**	**n/a**	**n/a**
Monthly PD sessions (five sessions)[b]	**72**	**92**	**56**	**20**	**29**
Study groups	**69**	**84**	**63**	**25**	**33**
End-of-year colloquia[a]	**87**	**96**	**75**	**n/a**	**n/a**

Source: WestEd attendance logs for activities of treatment teachers in districts receiving the ETS induction program.

[a]Data not available for orientations. Data available from four of nine districts for end-of-year colloquia.

[b]Average of district averages across all five sessions.

BT = beginning teacher; PD = professional development; n/a = not applicable.

N = 259 teachers.

Table IV.6. Teacher Attendance at NTC Induction Activities (Percentages): 2005–2006 School Year

Activity	Average Attendance of BTs	Range of Average Attendance Across Districts		Regularity of Attendance	
		High	Low	Teachers Missing No More Than 1 Session	Teachers Missing 3 or More Sessions
Orientation	51	94	26	n/a	n/a
Monthly PD sessions (six sessions)[a]	65	93	43	23	22
End-of-year colloquia	60	96	46	n/a	n/a

Source: WestEd attendance logs for activities of treatment teachers in districts receiving the NTC induction program.

[a]Average of district averages across all six sessions.

BT = beginning teacher; PD = professional development; n/a = not applicable.

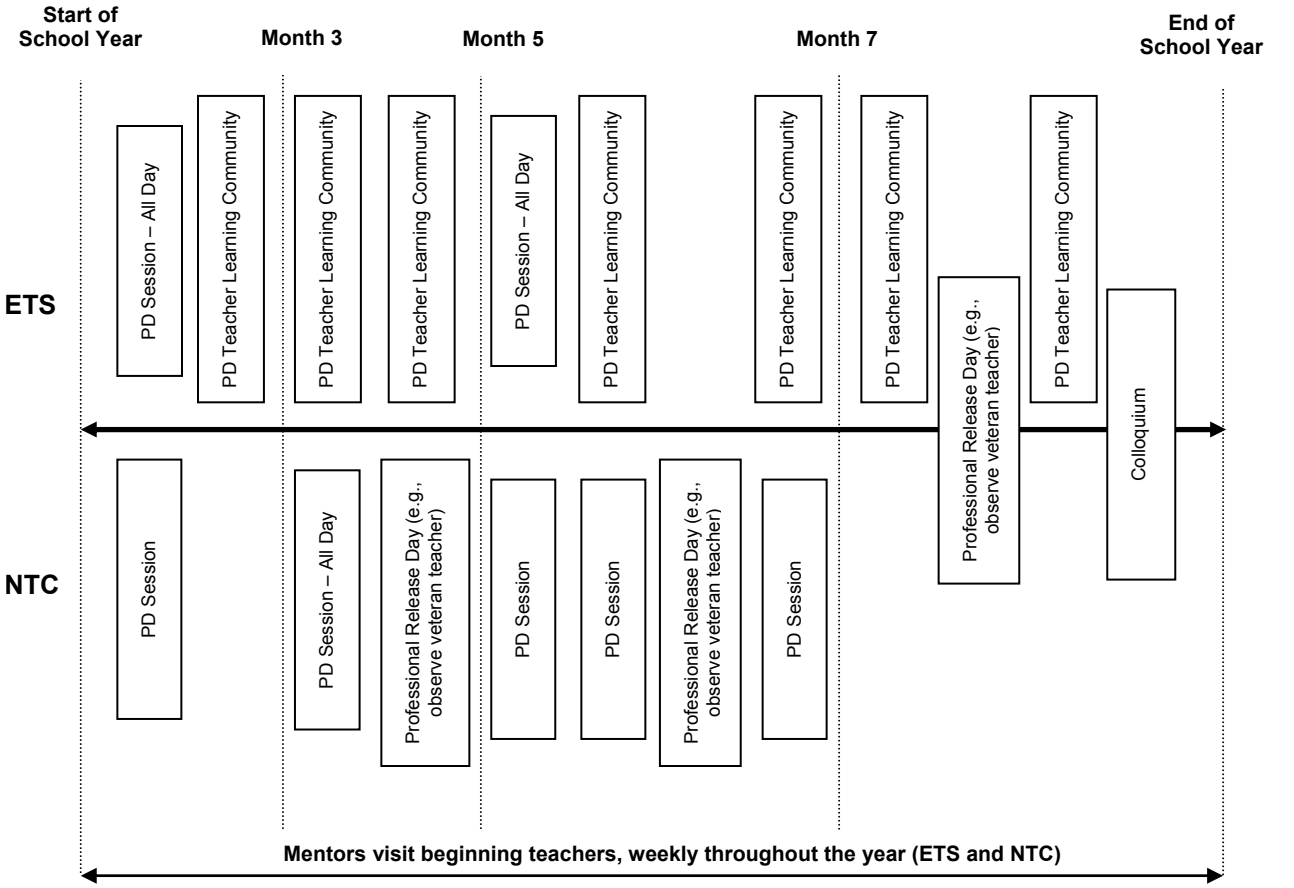
N = 247 teachers.

End-of-Year Colloquium. The two- to three-hour colloquium in each district focused on celebrating the first year's successes and teachers' professional growth. It also encouraged teachers to set goals for improved instruction for the year ahead. Attendance at the end-of-year colloquia was similar to that of other events, with about two-thirds participation across the study (87 percent across ETS districts and 60 percent across NTC districts), but considerably lower and higher levels in some districts (ranging from 46 to 96 percent).

b. Year 2 Program Services and Activities (2006–2007 School Year)

A second year of program implementation was provided in 7 of the original 17 districts. As in year 1, mentoring of beginning teachers (those who were randomly assigned to treatment in year 1 and were now in their second year of teaching) began during the first week of school and continued weekly throughout the year, with a similar structure. In addition to this, all treatment teachers were expected to participate in professional development sessions, as noted in Figure IV.5. The ETS district mentors also held monthly Teaching Learning Community (TLC) meetings with their beginning teachers. In year 1, these meetings were called "study groups" and mentors primarily facilitated general peer support among their beginning teachers. In year 2, the meetings focused more on enhancing particular aspects of instruction. Beginning teachers also had release days to observe veteran teachers or to work with their mentors on other development tasks, just as they had in year 1. Similar to year 1, at the end of this second school year, beginning teachers participated in a colloquium. Each of these induction activities is described in more detail below.

Figure IV.5. Comprehensive Induction Program Activities for Beginning Teachers for 2006–2007 School Year



Note: Activities common to both providers are shown on both sides of the horizontal divider between ETS and NTC.

PD = professional development.

Mentoring. Mentoring in the second year was similar to the support provided in the first year. Programs again expected mentors to allocate approximately two hours of contact each week with every beginning teacher in their caseload and to engage in the same kinds of mentor/novice interactions described for year 1. The framework for ETS mentors was again Pathwise Induction Events, whereas NTC mentors again used the FAS.

Professional Development. The ETS and NTC programs included between 35 and 42 hours of professional development for beginning teachers in year 2.[32] In ETS districts, a total of eight 2-hour sessions were held, as well as two all-day sessions (in months one and four of the school year) and a release day for observation of other teachers. NTC districts held one all-day session in month two or three, five 2-hour sessions throughout the year, and three release days for observation

[32] There was variation within and between districts in the amount of time devoted to any particular session, but the total time allocated in any district fell within this range.

of other teachers or individual work with mentors.[33] As in year 1, session topics continued to be related to the mentors' weekly work with their beginning teachers.

Programs changed the content and conduct of the professional sessions during this second year to reflect the growth of mentors and beginning teachers and the evolution of their circumstances and needs. Although staff of both programs traveled to districts to conduct the all-day sessions, mentors took the lead in carrying out the rest of the professional development sessions. Following the initial program-led sessions, mentors in each NTC district fleshed out details of nationally assigned topics (for example, differentiation in instruction) and designed activities to reflect local needs, in consultation with the program leader and their district coordinator. As in year 1, the NTC sessions used active-learning activities. The ETS Teacher Learning Communities were led by mentors and were an adaptation of the first year's study groups during which beginning teachers met monthly to discuss their individual needs and practices. In year 2, the ETS program provided specific content for each session and a formal structure for taking teachers through a cycle that consisted of (1) illustrating possible approaches for the instruction, (2) having teachers try them out, and (3) debriefing the resulting experience in the next session.

On average, the professional development sessions drew 62 percent and 58 percent of the beginning teachers over the course of the year, ETS and NTC respectively (Table IV.7).[34,35] The attendance at the all-day sessions in both programs generally was higher than at the two-hour sessions that were most often held after school: 75 percent and 79 percent for the first ETS and NTC all-day sessions, and 55 percent for the second ETS all-day session. Thirty-eight percent and 27 percent of teachers (ETS and NTC, respectively) *participated* in 80 percent or more of the sessions. Approximately one-third of teachers *missed* the majority (more than 50 percent) of the sessions (36 percent and 35 percent of ETS and NTC teachers, respectively).

Table IV. 8 lists the topics for the professional development sessions, by program. The topics for the first two NTC sessions—communication with families and equitable instruction and student achievement—were extensions of topics introduced in year 1. NTC selected these topics from an analysis of needs expressed by treatment teachers in an NTC-administered survey in the latter part of the first year. The ETS TLC sessions employed an existing ETS professional development product, *Keeping Learning on Track: Integrating Assessment with Instruction through Teacher Learning Communities*. The content of the product, described in Table IV.8, was introduced in the two all-day professional development sessions; during their monthly TLC meetings, teachers then discussed the topics and whatever experiences they had in applying the practices in their classrooms. ETS staff continually made minor but important adaptations of the product for specific use with beginning teachers in the study; for example, developing more elementary school examples than the standard product contained.

---

[33] In one ETS district, a single professional development session had to be cancelled because of unexpected local scheduling conflicts.

[34] WestEd attendance logs are the source data for discussion of participation of beginning teachers in professional development sessions.

[35] Average attendance ranged widely among the districts from 36 to 71 percent, and 48 to 74 percent (ETS and NTC, respectively).

Table IV.7. ETS and NTC Teacher Attendance: Professional Development Sessions and Colloquia: 2006–2007 School Year

		Range of Average Attendance Across Districts (Percentage)		Regularity of Attendance (Percentage)	
Activity	BT Average Attendance	High	Low	BTs Attending Most Sessions	BTs Missing Most Sessions
Monthly PD Sessions					
ETS (9 sessions)	62	71	36	38 (miss 1–2 of 9)	36 (miss 5+ of 9)
NTC (5 sessions)	58	74	48	27 (miss 1 of 5)	35 (miss 3+ of 5)
End-of-Year Colloquium					
ETS	61	70	29	n/a	n/a
NTC	60	61	58	n/a	n/a

Source: WestEd attendance logs for activities of treatment teachers in districts receiving the induction program.

BT = beginning teacher; PD = professional development; n/a = not applicable.

N = 192 teachers (ETS) and 206 teachers (NTC).

Observation of Veteran Teachers. Mentors arranged formal opportunities for beginning teachers to observe experienced teachers, with an attempt to select observations that would be relevant to the instructional goals of the beginning teachers. Both programs required one observation, but NTC participants also could use another of their three release days for additional observations. ETS and NTC mentors provided similar types of guidance and observation debriefings as in the first year.

End-of-Year Colloquium. As in the first year, the two- to three-hour end-of-year colloquia in each district focused on celebrating the year's successes and teachers' professional growth. They also encouraged teachers to set goals for improved instruction for the next school year. Attendance at the end-of-year colloquia was similar to that of other professional development events (61 percent and 60 percent of teachers, ETS and NTC, respectively), with notably higher and lower levels among individual districts (ranging from 29 to 70 percent).

Table IV.8. Topics for Professional Development Sessions, by Program

ETS	
Expanded examination of framework for teaching	This session is a review of the conceptual framework that shaped the ETS induction program in year 1 (see Table IV.2).
Using evidence to inform practice; norms for teacher learning communities	This session established a focus on teaching (versus providing general peer support). It also set norms for professional and interpersonal behavior during sessions, and a structure and timetable to use in each session.
Using learning intentions to strengthen starts and ends of lessons	This session focused on establishing clear expectations/goals for lessons and an assessment of goal attainment.
Providing formative feedback	This session focused on the range and frequency of written feedback provided on student assignments.
Developing quality hinge questions	This session focused on using optimal questioning strategies to engineer effective classroom discussions, questions, and learning tasks.
Student self- and peer-assessment	This session focused on the value of, and how to establish, clear scoring/grading rubrics.
NTC	
Expanded examination of standards for teaching	This session is a review of the six professional teaching standards.
Strong parental relationships and communication	This session focused on family-teacher conferences, general and specific strategies for communication with families, and ways to enlist and build partnerships with families.
Equitable instruction and student achievement (the only full-day session)	This session focused on recognizing individual student needs, and analyzing student work to identify individual needs.
Differentiated instruction	This session focused on differentiating instruction to meet individual needs, by tailoring instructional materials and varying modes of instruction.
Other topics[a]	These sessions typically delved further into topics begun in prior sessions.

Source: ETS: *Keeping Learning on Track*; NTC: varied proprietary documents from the induction program.

[a]Identified in consultation with NTC staff and inspection of their data from year 1 participant survey.

B. Treatment-Control Differences in Teacher Induction Services

We do not compare comprehensive teacher induction to the absence of any support services for new teachers; rather, we compare comprehensive teacher induction to the prevailing level of induction services in the selected districts. We used the control group to characterize the types and intensity of district and school support that beginning teachers in the study schools would normally receive in the absence of the experimental intervention. The intervention gave treatment teachers the opportunity to receive services through the comprehensive induction programs, but participation was voluntary. By comparing service receipt in the treatment group with that in the control group, we derived estimates of the service contrast, which provides the necessary context for understanding the results of analyses on teacher and student outcomes presented in Chapters V and VI. Estimates of the service contrast were computed using an ordinary least squares regression

model with district and grade fixed effects that accounted for clustering of teachers within schools; weights were applied to adjust for survey nonresponse and the study design. One-year and two-year districts were analyzed separately.

The induction activities surveys that were administered during fall 2005, spring 2006, fall 2006, spring 2007 (for teachers in two-year districts), fall 2007, and fall 2008 were used to characterize the induction services received by the treatment and control teachers during their first four years in the profession. We examined differences in services received by treatment and control teachers both before and after the comprehensive induction services ended. Treatment teachers in one-year and two-year districts were offered the same usual district services as control teachers following the conclusion of the intervention (beginning in year 2 for treatment teachers in one-year districts and year 3 for treatment teachers in two-year districts). The examination of service usage in these later years is important to providing a full picture of the service contrast. Analysis of the services received by control teachers in the later years provides a description of typical district and school induction support following teachers' initial years in the classroom. Moreover, the analysis can show whether the intervention induced future changes in treatment teachers' usage of these services beyond what it would have been in the absence of the intervention.

This section presents differences in service receipt between treatment and control teachers in the following areas: mentor assignments, number and types of mentors, meeting time with mentors, mentor activities, areas of mentor guidance, observation and feedback, professional development activities, and professional development session topic areas. We first discuss results for one-year districts, followed by results for two-year districts. Findings are shown in figures; Appendix B presents tables of treatment and control group means and service contrast estimates for each data collection period.

Overall, results showed statistically significant differences in the amount, type, and content of induction support received by treatment and control teachers. The findings are given below and described in more detail in the remainder of this chapter:[36]

- **In one-year districts, both treatment and control teachers reported receiving substantial induction support. However, treatment teachers, whose participation in the comprehensive induction program was voluntary, received more and different support than control teachers during the comprehensive induction program (their first year of teaching).** Relative to control teachers, treatment teachers were more likely to have an assigned mentor (90 versus 70 percent in fall 2005, p-value 0.000; 90 versus 72 percent in spring 2006, p-value 0.000) and spent more time per week meeting with their mentors (87 versus 67 minutes in fall 2005, p-value 0.007; 85 versus 68 minutes in spring 2006,[37] p-value 0.039); these differences were all statistically significant. There were additional statistically significant differences favoring treatment

[36] Throughout this chapter we present results from a large number of hypothesis tests. When conducting many independent hypothesis tests, a small percentage of results will be statistically significant even if no underlying relationship exists (5 percent, when using an alpha level of 0.05). To guard against this problem of multiple comparisons, we present all findings in context, discussing the non-significance of the tests where we fail to reject the null hypothesis as well as the statistical significance of those that do. See the end of Chapter II for further discussion of this issue.

[37] Teachers who report not having a mentor are given a value of zero for all variables related to time spent with a mentor and supports received from a mentor.

teachers in the time spent being observed by mentors during the last full week of teaching (34 versus 10 minutes in fall 2005, p-value 0.000; 27 versus 7 minutes in spring 2006, p-value 0.000), the frequency of informal feedback during the three months prior to the survey (3.2 versus 2.4 times in fall 2005, p-value 0.000; 2.5 versus 1.9 times in spring 2006, p-value 0.001), whether a teacher worked with a study group of new teachers during the three months prior to the survey (66 versus 34 percent in fall 2005, p-value 0.000; 71 versus 29 percent in spring 2006, p-value 0.000), and whether a teacher observed others teaching in their classrooms during the three months prior to the survey (61 versus 44 percent in fall 2005, p-value 0.000; 68 versus 39 percent in spring 2006, p-value 0.000). Treatment teachers were also more likely to receive mentors' guidance in 9 of 10 areas covered by the survey during the last full week of teaching. (Supporting details may be found in Appendix Tables B.3 and B.8.) No adjustments for multiple comparisons were made in assessing the statistical significance of the results because such adjustments were unnecessary or inappropriate. The analysis examined 132 comparisons during the comprehensive induction program for one-year districts, of which 75 were statistically significant and 6.6 would be significant by chance if the program had no effect.

- **In two-year districts, treatment and control teachers reported receiving substantial induction support as well. However, similar to the findings in one-year districts, treatment teachers received more and different support than control teachers during the comprehensive induction program (their first two years of teaching).** Relative to control teachers, treatment teachers were more likely to have an assigned mentor (between fall 2005 and spring 2007, the percent of treatment teachers ranged from 80 to 96 and the percent of control teachers ranged from 34 to 79; p-values all 0.000) and spent more time per week meeting with their mentors (between fall 2005 and spring 2007, time spent by treatment teachers ranged from 79 to 124 minutes and the time spent by control teachers ranged from 41 to 82 minutes; p-values ranged from 0.087 to 0.001); these differences were statistically significant with the exception of meeting time in spring 2006. There were additional statistically significant differences favoring treatment teachers: the time spent being observed by mentors during the last full week of teaching (between fall 2005 and spring 2007, the time spent by treatment teachers ranged from 19 to 38 minutes and the time spent by control teachers ranged from 7 to 17 minutes; p-values ranged from 0.000 to 0.003), the time spent watching mentors model lessons during the last full week of teaching (between fall 2005 and spring 2007, the time spent by treatment teachers ranged from 10 to 16 minutes and the time spent by control teachers ranged from 4 to 10 minutes; p-values ranged from 0.003 to 0.027), the frequency of informal feedback during the three months prior to the survey (between fall 2005 and spring 2007, the frequency for treatment teachers ranged from 1.9 to 2.8 times and the frequency for control teachers ranged from 1.5 to 2.5 times; p-values ranged from 0.001 to 0.266), and whether a teacher worked with a study group of new teachers during the three months prior to the survey (between fall 2005 and spring 2007, the percent of treatment teachers ranged from 42 to 67 and the percent of control teachers ranged from 14 to 25; p-values all 0.000). Treatment teachers were also more likely to receive mentors' guidance in all 10 areas covered by the survey during the last full week of teaching. (Supporting details may be found in Appendix Tables B.18 and B.23.) No adjustments for multiple comparisons were made in assessing the statistical significance of the results because such adjustments were unnecessary or inappropriate. The analysis examined 264 comparisons during the comprehensive

induction program for one-year districts, of which 70 were statistically significant and 13.2 would be significant by chance if the program had no effect.

- **In their second year, immediately following the end of the comprehensive induction program, treatment teachers in one-year districts received less and different induction support than control teachers.** For measures such as the percentage of teachers with an assigned mentor and time spent meeting with mentors per week, this reflects a significant drop in support among control teachers and an even larger significant drop in support among treatment teachers. A survey of teachers in one-year districts conducted in fall 2006 showed that there were statistically significant differences favoring the control teachers in several areas: treatment teachers were less likely than control teachers to have an assigned mentor (20 percent of treatment teachers versus 29 percent of control teachers, p-value 0.017), spent less time per week meeting with their mentors (19 minutes for treatment teachers versus 39 minutes for control teachers, p-value 0.002), spent less time being observed by mentors during the last full week of teaching (2 minutes by treatment teachers versus 6 minutes by control teachers, p-value 0.021), were less likely to work with a study group of new teachers during the three months prior to the survey (11 percent of treatment teachers versus 21 percent of control teachers, p-value 0.003), and were less likely to receive mentors' guidance in all 10 areas covered by the survey during the last full week of teaching. (Supporting details may be found in Appendix Table B.3.) No statistically significant differences favoring the treatment teachers were found. No adjustments for multiple comparisons were made in assessing the statistical significance of the results because such adjustments were unnecessary or inappropriate. The analysis examined 66 comparisons during fall 2006 for one-year districts, of which 29 were statistically significant and 3.3 would be significant by chance if the program had no effect.

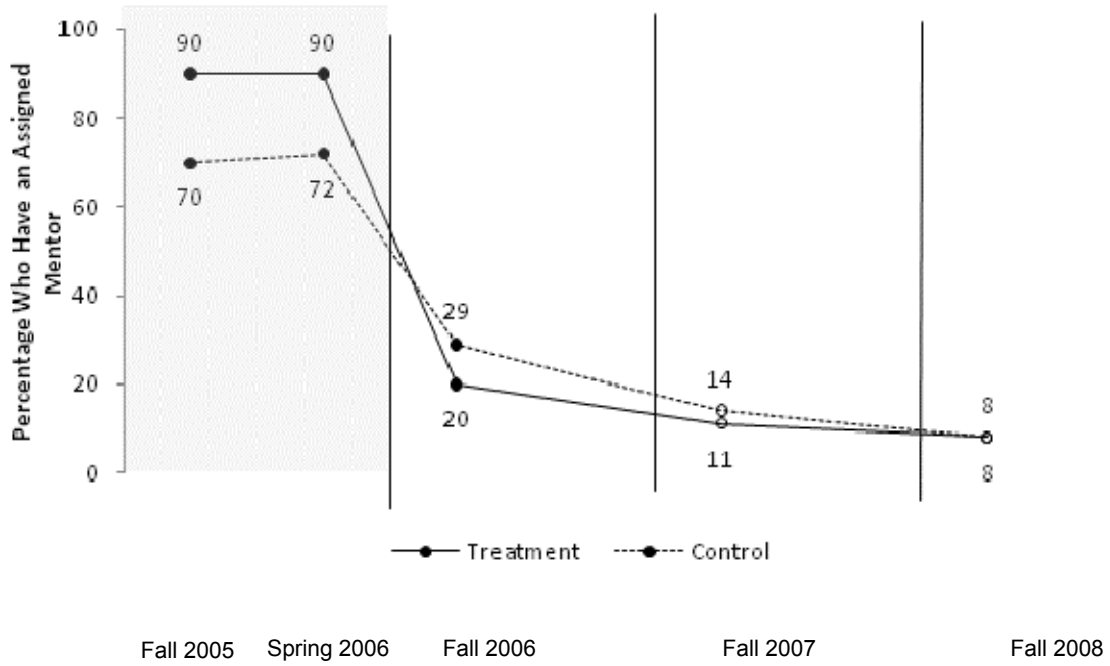
- **In the third and fourth years of teaching, after the intervention ended for all districts, treatment and control teachers received similar levels of support.** In both one- and two-year districts, there were statistically significant differences in fewer than seven percent of the 134 measures we surveyed. No adjustments for multiple comparisons were made in assessing the statistical significance of the results because such adjustments were unnecessary or inappropriate. The analysis examined 132 comparisons during study teachers' third and fourth years of teaching for one-year districts, of which 8 were statistically significant and 6.6 would be significant by chance if the program had no effect. We also examined 132 comparisons during study teachers' third and fourth years of teaching for one-year districts, of which 9 were statistically significant and 6.6 would be significant by chance if the program had no effect.

1. One-Year Districts

a. Mentor Assignments

A majority of both treatment and control teachers reported having a mentor assigned to them during the comprehensive induction program, but treatment teachers were significantly more likely than control teachers to report this support. Figure IV.6 illustrates the levels of support for treatment and control teachers over time. The differences were 90 versus 70 percent in fall 2005 and 90 versus 72 percent in spring 2006. While 10 percent of treatment teachers do not report having a mentor assigned to them, all treatment teachers were assigned a mentor as part of the comprehensive induction program.

Figure IV.6. Treatment-Control Differences in Percent Who Have an Assigned Mentor: One-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. N=503 teachers in fall 2005, 499 teachers in spring 2006, 472 teachers in fall 2006, 426 teachers in fall 2007, and 398 teachers in fall 2008.

Treatment-control differences are significantly different from zero at the 0.05 level except in fall 2007 and fall 2008.

The percent of both treatment and control teachers with an assigned mentor dropped significantly from spring 2006 to fall 2006 (p-values both 0.000), after the comprehensive induction program ended. The drop was larger for treatment teachers; treatment teachers were significantly *less* likely than control teachers to report having an assigned mentor (20 versus 29 percent) during the fall after the comprehensive induction program ended. By fall 2007 and also in fall 2008, differences between treatment and control teachers were insignificant (11 versus 14 percent and 8 versus 8 percent). Differences between treatment and control teachers in the likelihood of having any mentor followed a similar pattern over time.

b. Number and Types of Mentors

Treatment teachers' number of mentors and mentor profiles differed significantly from those of control teachers during the comprehensive induction program. Treatment teachers were significantly more likely than control teachers to report having multiple mentors (25 versus 15 percent in fall 2005 and 22 versus 12 percent in spring 2006), having two mentors assigned to them (19 versus 7 percent in fall 2005 and 18 versus 7 percent in spring 2006), and having a full-time mentor (74 versus 8 percent in fall 2005 and 72 versus 10 percent in spring 2006). Also in this first year, treatment teachers were significantly less likely than control teachers to report having a mentor who was a teacher (25 versus 64 percent in fall 2005 and 25 versus 66 percent in spring 2006).

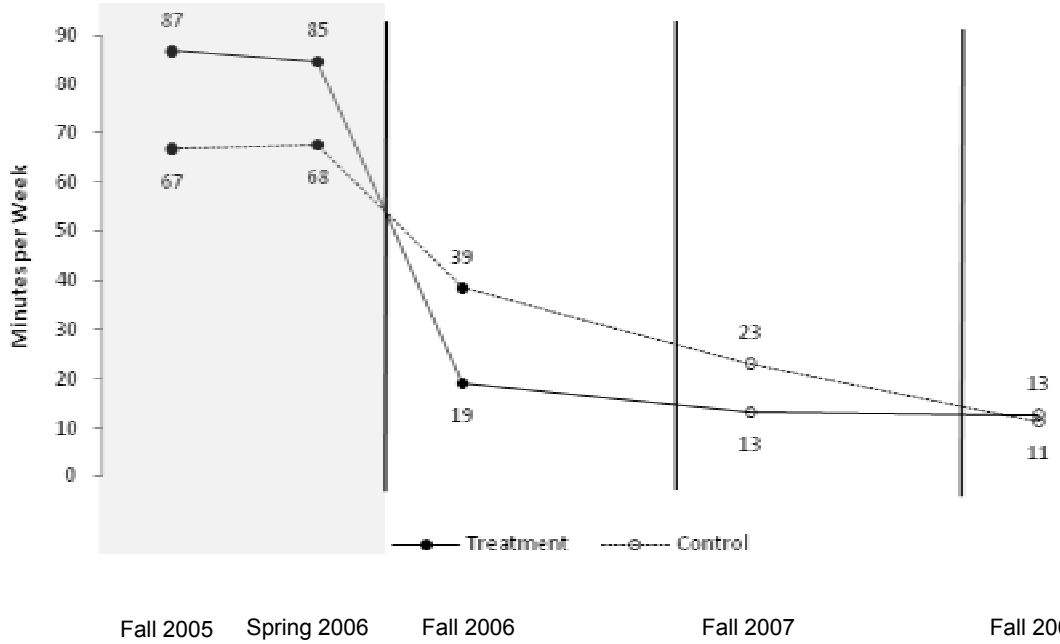
In the fall after the comprehensive induction program ended, treatment teachers were significantly *less* likely than control teachers to report having two assigned mentors (2 versus 6 percent) or having a mentor who was a teacher (21 versus 31 percent). Treatment and control teachers did not differ significantly in having multiple mentors or a full-time mentor. The number and type of treatment and control teachers' mentors were not significantly different in fall 2007 and fall 2008.

c. Meetings with Mentors

Treatment and control teachers both spent more than an hour a week in mentor meetings and activities during the comprehensive induction program, but treatment teachers spent significantly more time than control teachers. Figure IV.7 illustrates average total weekly mentor meeting time for treatment and control teachers over time. Combining usual scheduled time and informal time during the most recent full week of teaching, we found that treatment teachers spent an average of 87 minutes in mentor meetings compared to 67 minutes for control teachers in fall 2005, and 85 versus 68 minutes in spring 2006. Since total meeting time was not reported directly but had to be constructed from reports of the frequency and duration of usual scheduled meetings and the time spent in informal meetings, we could not determine precisely whether treatment teachers met with their study mentors for two hours per week as the ETS and NTC programs expected. The statistically significant differences in meeting time (20 minutes and 17 minutes) were attributable entirely to differences in the duration of the usual scheduled meetings (23 versus 10 minutes in fall 2005 and 23 versus 11 minutes in spring 2006). Differences in total meeting time in fall 2005 and fall 2006 are shown separately by district in Figures B.1–B.2 in Appendix B. Details of the method that generates these results can be found in Appendix A.

Differences in the time spent with full-time mentors and mentors who were teachers reflected differences in the types of mentors that treatment and control teachers reported. Treatment teachers reported spending significantly more time during the most recent week of teaching meeting with full-time mentors than did control teachers (60 versus 4 minutes in fall 2005 and 52 versus 6 minutes in spring 2006). However, they reported spending significantly less time than control teachers with mentors who were teachers (23 versus 59 minutes in fall 2005 and 26 versus 59 minutes in spring 2006). Figure IV.8 illustrates average minutes each week meeting with mentors who were teachers for treatment and control teachers over time.

Figure IV.7. Treatment-Control Differences in Total Minutes Spent in Mentoring per Week: One-Year Districts



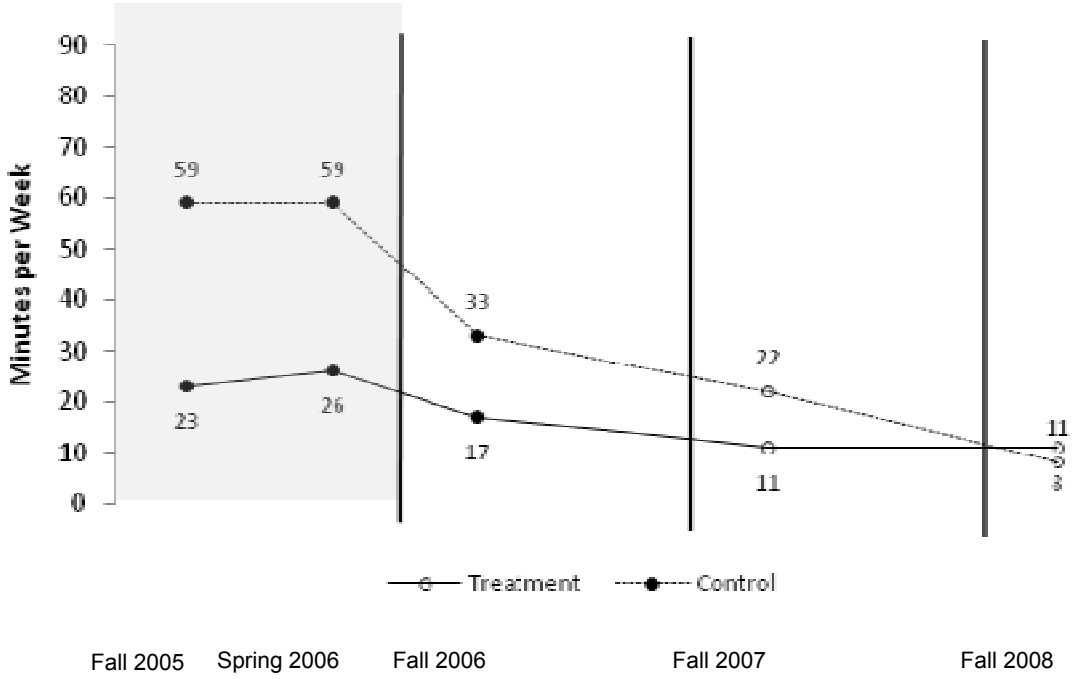
Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. N=503 teachers in fall 2005, 499 teachers in spring 2006, 472 teachers in fall 2006, 426 teachers in fall 2007, and 398 teachers in fall 2008.

Treatment-control differences are significantly different from zero at the 0.05 level except in fall 2007 and fall 2008.

In the fall after the comprehensive induction program ended, the time spent by both treatment and control teachers in mentor meetings dropped significantly (p-values both 0.000) to less than 40 minutes per week. However, the drop was larger for treatment teachers, who consequently spent significantly *less* total time than control teachers in mentor meetings (19 versus 39 minutes). Negative differences were found in both components of total mentor meeting time: usual scheduled meetings (10 versus 18 minutes) and informal meetings (9 versus 20 minutes). The significant difference in total meeting time was due to treatment teachers spending less time than control teachers with mentors who were teachers (17 versus 33 minutes) or mentors who were administrators (0 versus 2 minutes), since the differences in meeting time with full-time mentors (1 versus 3 minutes) and staff external to the district (1 versus 0 minutes) were insignificant. Two to three years after the comprehensive induction program was implemented (fall 2007 and fall 2008), differences between treatment and control teachers were insignificant in terms of total mentor meeting time (13 versus 23 minutes in fall 2007 and 13 versus 11 minutes in fall 2008) as well as time spent with full-time mentors and mentors who were teachers.

Figure IV.8. Treatment-Control Differences in Total Minutes Spent in Mentoring per Week with Mentors Who Are Teachers: One-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. N = 503 teachers in fall 2005, 499 teachers in spring 2006, 472 teachers in fall 2006, 426 teachers in fall 2007, and 398 teachers in fall 2008.

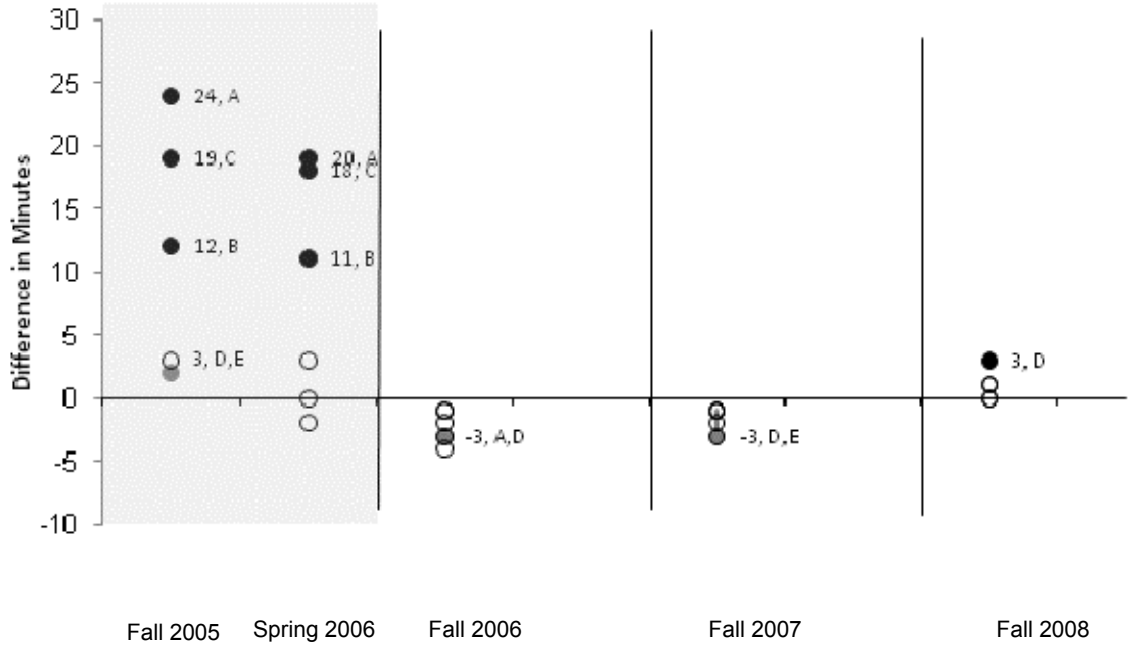
Treatment-control differences are significantly different from zero at the 0.05 level except in fall 2007 and fall 2008.

d. Mentor Activities

During the comprehensive induction program, treatment teachers reported spending 9 to 34 minutes in each of six types of mentoring activities during the most recent full week of teaching while control teachers reported spending 6 to 23 minutes. Figure IV.9 illustrates differences in levels of support for treatment and control teachers over time. A positive difference indicates that treatment teachers spent more time than control teachers in an activity, while a negative difference indicates that they spent less time; solid circles indicate that the difference is statistically significant.

Treatment teachers spent significantly more time than control teachers in certain of these activities during the comprehensive induction program: being observed by mentors (34 versus 10 minutes in fall 2005 and 27 versus 7 minutes in spring 2006), meeting one-on-one with mentors (34 versus 23 minutes in fall 2005 and 31 versus 21 minutes in spring 2006), meeting with mentors together with other first-year teachers (29 versus 9 minutes in fall 2005 and 24 versus 6 minutes in spring 2006), and having mentors model lessons (9 versus 6 minutes in fall 2005; the difference is insignificant in spring 2006). Treatment teachers did not spend significantly more time than control teachers meeting together with mentors and other teachers or co-teaching lessons with mentors. The total time spent in the six types of activities covered in the survey averaged 130 minutes per week for treatment teachers and 67 minutes per week for control teachers in fall 2005, a significant difference of 63 minutes per week. In spring 2006, the total time was 108 minutes for treatment teachers and 59 minutes for control teachers, a significant difference of 49 minutes.

Figure IV.9. Treatment-Control Differences in Time Spent in Six Mentoring Activities in the Last Full Week of Teaching: One-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Solid (open) circles indicate that treatment-control differences are (not) statistically significant at the 0.05 level. In fall 2005, E is significant but D is not. In fall 2006, A is significant but D is not. In fall 2007, E is significant but D is not (N = 503 teachers in fall 2005, 499 teachers in spring 2006, 472 teachers in fall 2006, 426 teachers in fall 2007, and 398 teachers in fall 2008).

Legend: A = Observing BT teaching, B = Meeting with BT one-on-one, C = Meeting with BT and other first-year teachers, D = Meeting with BT and other teachers, E = Modeling a lesson, F = Co-teaching a lesson., BT = Beginning Teacher.

In contrast, after the comprehensive induction program ended, treatment teachers reported spending 0 to 7 minutes per week in each of six types of mentoring activities during the most recent full week of teaching across fall 2006, fall 2007, and fall 2008, while control teachers reported spending 0 to 10 minutes. Treatment teachers spent significantly *less* total time than control teachers in all six activities covered in the fall 2006 and fall 2007 surveys (22 versus 36 minutes per week in fall 2006 and 12 versus 24 minutes per week in fall 2007). Differences in the time spent in each individual activity were insignificant, with two exceptions across the 12 activities and periods: treatment teachers spent less time being observed by mentors in fall 2006 (2 versus 6 minutes per week) and less time watching mentors model a lesson in fall 2007 (0.2 versus 3 minutes per week). There were no significant differences in the total time spent in all six activities or in the time spent in each activity in fall 2008.

e. Areas of Mentor Guidance

During the comprehensive induction program, a majority of treatment teachers reported receiving mentors' assistance during the last full week of teaching in all 10 topic areas covered in the survey and in both fall 2005 and spring 2006, while about 40 to 70 percent of control teachers reported receiving assistance. Figure IV.10 illustrates differences in levels of support for treatment and control teachers over time. The percentage of treatment teachers receiving each type of assistance ranged from 56 percent (sharing lesson plans, assignments, and other instructional activities) to 87 percent in fall 2005, and 53 percent (receiving help with teaching to meet state or district standards) to 78 percent in spring 2006. Among control teachers, the percentage reporting each type of assistance ranged from 44 percent (receiving guidance on how to assess students) to 66 percent in fall 2005, and 41 percent (discussing instructional goals and ways to achieve them) to 68 percent in spring 2006. The area in which both treatment and control teachers received the most guidance in each period was encouragement or moral support.

Examining differences in service receipt during the comprehensive induction program, we found treatment teachers were significantly more likely than control teachers to report receiving mentors' assistance during the last full week of teaching in all 10 topic areas covered in the survey, with one exception across the 20 areas and periods: treatment teachers were less likely than control teachers to have mentors share lesson plans, assignments, or other instructional activities in fall 2005. Significant differences ranged from 14 percentage points (receiving help with administrative and logistical issues) to 27 percentage points (receiving help identifying teaching challenges and solutions) in fall 2005 and 9 percentage points (sharing lesson plans, assignments, or other instructional activities) to 20 percentage points (discussing instructional goals and ways to achieve them) in spring 2006. Discussing instructional goals and ways to achieve them were among the two areas of guidance with the largest impacts during both the fall and spring of the comprehensive induction program; the following areas of guidance were among the five with the largest impacts in both the fall and spring: receiving suggestions to improve practice, having opportunities to raise issues and discuss concerns, and having the mentor act on a teacher's request.

During the fall after the comprehensive induction program ended, fewer than half of treatment or control teachers reported receiving each type of assistance during the last full week of teaching. Among treatment teachers, the percentage reporting each type of assistance ranged from 11 percent (receiving guidance on how to assess students) to 21 percent. Among control teachers, the percentage reporting each type of assistance ranged from 19 percent (receiving help teaching to state standards) to 33 percent. The area in which both treatment and control teachers received the most guidance in each period was encouragement or moral support. Estimating differences in service receipt in fall 2006, we found that treatment teachers were significantly *less* likely than control teachers to report receiving mentors' assistance in each of the 10 topic areas the survey covered. Differences ranged from 8 percentage points (receiving help teaching to meet state or district standards) to 14 percentage points (having opportunities to raise issues or discuss concerns).

In fall 2007 and fall 2008, no more than 20 percent of treatment or control teachers reported receiving each type of guidance. The percentage of treatment teachers receiving each type of guidance ranged from 7 percent (help teaching to meet state or district standards, having mentor act on the teacher's request) to 18 percent in fall 2007 and from 4 percent (help teaching to meet state or district standards) to 11 percent in fall 2008. Among control teachers, the percentage receiving each type of guidance ranged from 14 percent (receiving suggestions to improve practice) to 20 percent (encouragement and moral support, having an opportunity to raise issues or discuss concerns) in fall 2007 and from 5 percent (receiving guidance on how to assess students) to

11 percent in fall 2008. The area in which both treatment and control teachers received the most guidance in each period was encouragement or moral support. Turning to the service contrast, differences between treatment and control teachers in the receipt of guidance were insignificant in fall 2007 and fall 2008, with two exceptions across the 20 areas and periods: treatment teachers were less likely to receive guidance in teaching to meet state or district standards (7 versus 16 percent) or to have their mentors act on something they requested (7 versus 15 percent) in fall 2007.

f. Observations and Feedback

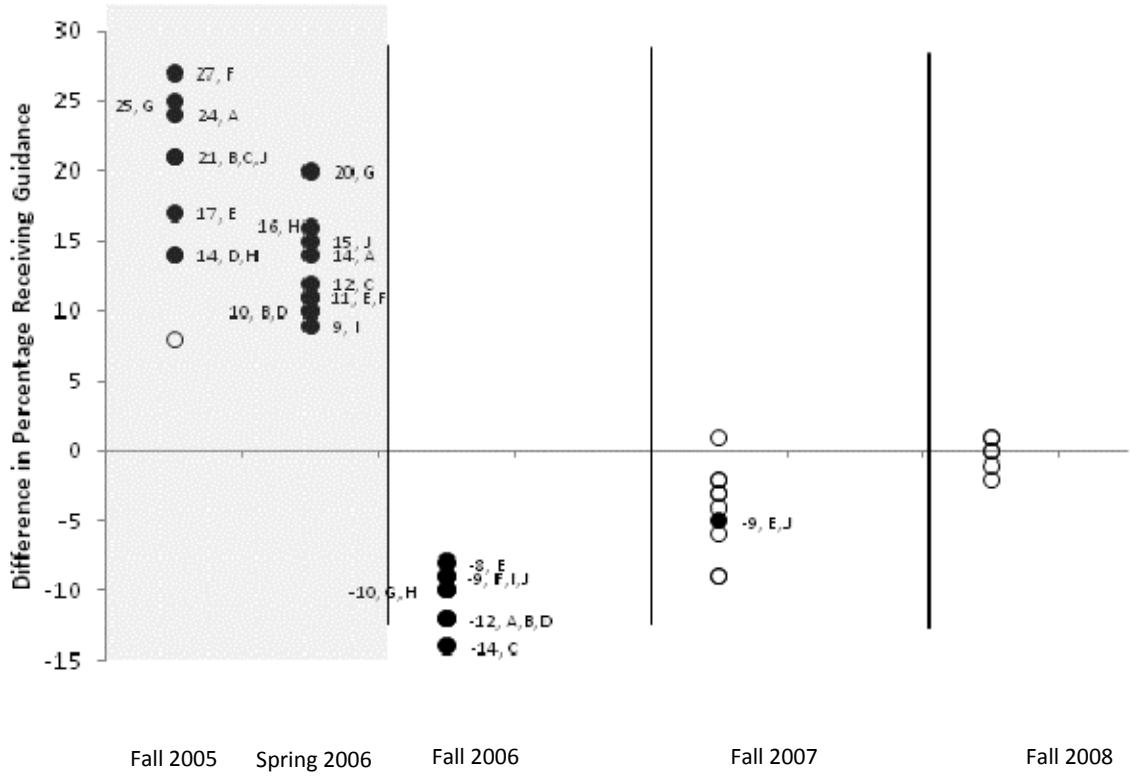
During the comprehensive induction program, both treatment and control teachers reported receiving two types of observations and three types of feedback at least once during the three months prior to the survey. Figure IV.11 illustrates differences in levels of support for treatment and control teachers over time.

Treatment teachers received certain types of observations and feedback significantly more frequently than control teachers during the comprehensive induction program.[38] Treatment teachers were significantly more frequently observed by mentors (4.0 versus 1.5 times in fall 2005 and 3.5 versus 1.5 times in spring 2006) and more frequently given informal feedback on teaching (3.2 versus 2.4 times in fall 2005 and 2.5 versus 1.9 times in spring 2006). Treatment and control teachers did not differ significantly in the frequency with which they were observed by their principals (2.3 versus 2.6 times in fall 2005 and 1.9 versus 2.1 times in spring 2006) or the frequency with which they received feedback on lesson plans (1.6 versus 1.7 times in fall 2005 and 1.3 versus 1.5 times in spring 2006) or as part of a formal evaluation (1.7 versus 1.4 times in fall 2005 and 1.6 versus 1.4 times in spring 2006) during the intervention.

After the comprehensive induction program ended, the frequency of the two types of observations during the three months prior to the survey ranged from 0 to 2 times and the frequency of the three types of feedback ranged from 1 to 2 times for both treatment and control teachers. Treatment and control teachers did not differ significantly in the frequency with which they engaged in these activities with one exception across the 15 activities and periods: treatment teachers were significantly *less* likely to be observed by a mentor in fall 2006 (0.3 versus 0.6 times).

[38] Reported levels of service receipt in the fall may be low relative to the spring as the three months prior to the survey may include part of the summer vacation for some teachers. The percent of teachers completing fall Induction Activities surveys prior to December, and thus reflecting on a period that begins before September 1st, is 32 percent in fall 2005, 83 percent in fall 2006, 75 percent in fall 2007, and 79 percent in fall 2008. The earliest surveys completed were on November 11, 2005, October 3, 2006, October 3, 2007, and September 8, 2008. However, the timing of fall survey completion does not affect the impact estimates as both treatment and control teachers completed their surveys at the same time on average. This issue pertains as well to participation in professional development activities and topic area sessions in one-year districts and to observations and feedback and participation in professional development activities and topic area sessions in two-year districts.

Figure IV.10. Treatment-Control Differences in Percent of Teachers Who Received 10 Types of Mentor Guidance in the Last Full Week of Teaching: One-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

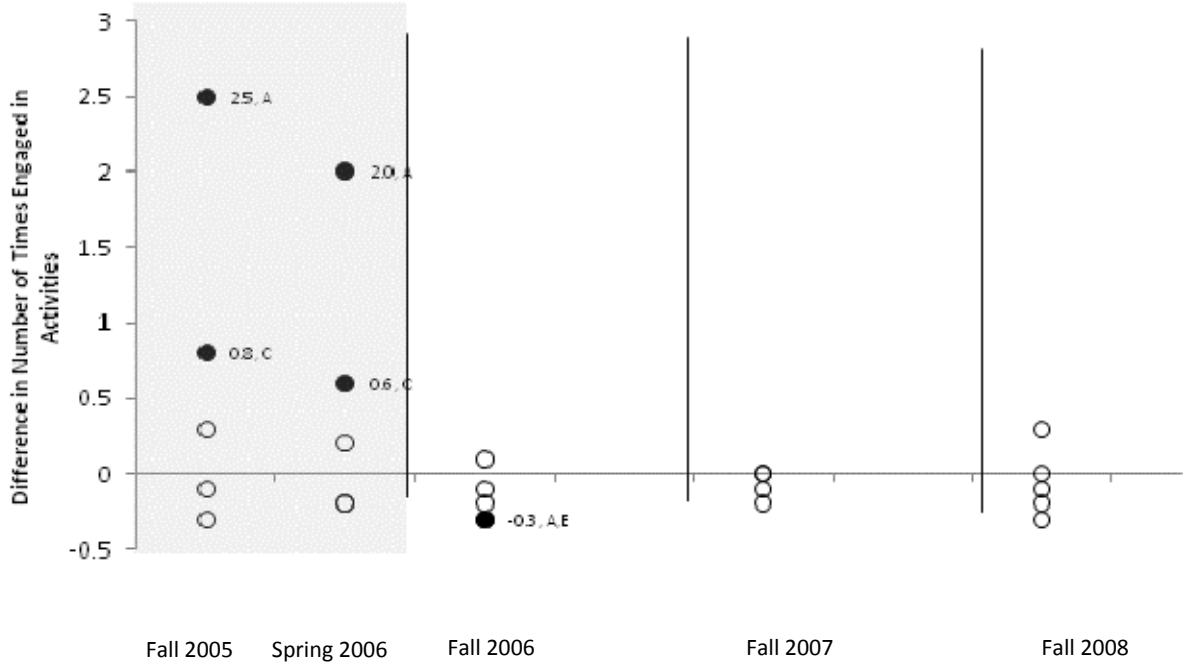
Note: Data pertain to teachers in one-year districts participating in the study. Solid (open) circles indicate that treatment-control differences are (not) statistically significant at the 0.05 level N = 503 teachers in fall 2005, 499 teachers in spring 2006, 472 teachers in fall 2006, 426 teachers in fall 2007, and 398 teachers in fall 2008.

Legend: A = Suggestions to improve practice, B = Encouragement or moral support, C = Opportunity to raise issues/discuss concerns, D = Help with administrative/logistical issues, E = Help with teaching to meet state or district standards, F = Help identifying teaching challenges and solutions, G = Discussed instructional goals and ways to achieve them, H = Guidance on how to assess students, I = Shared lesson plans, assignments, or other instructional activities, J = Acted on beginning teacher's request.

g. Professional Development Activities

During the comprehensive induction program, both treatment and control teachers reported participating in nine professional development activities during the three months prior to the survey. Among treatment teachers, participation rates across the 18 activities and periods ranged from 38 to 78 percent and did not reach 100 percent for any activity. Among control teachers, these participation rates ranged from 29 to 78 percent. Figure IV.12 illustrates differences in levels of support for treatment and control teachers over time.

Figure IV.11. Treatment-Control Differences in the Frequency of Selected Activities During Past Three Months: One-Year Districts



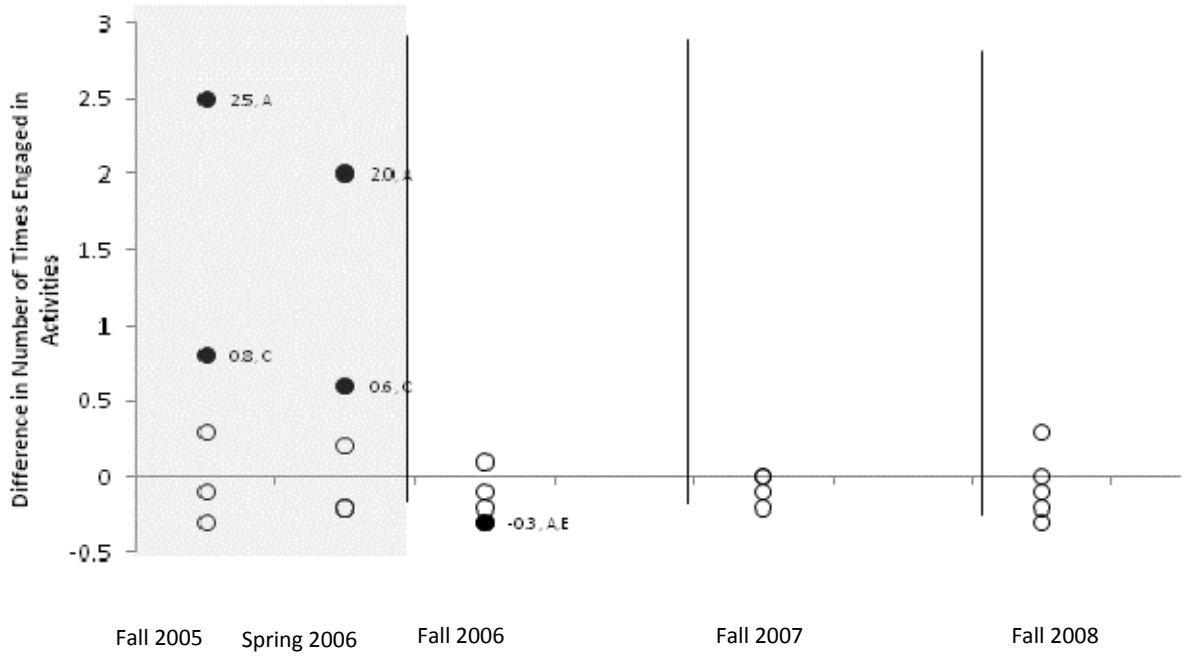
Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Solid (open) circles indicate that treatment-control differences are (not) statistically significant at the 0.05 level. In fall 2006, A is significant but E is not. N = 503 teachers in fall 2005, 499 teachers in spring 2006, 472 teachers in fall 2006, 426 teachers in fall 2007, and 398 teachers in fall 2008.

Legend: A = Teaching was observed by mentor, B = Teaching was observed by principal, C = Given feedback on your teaching, not as part of formal evaluation, D = Given feedback on your teaching as part of formal evaluation, E = Given feedback on your lesson plans.

Treatment teachers were more likely than control teachers to report participating in certain of these activities during the comprehensive induction program. Treatment teachers were significantly more likely than control teachers to report working with study groups of new teachers (66 versus 34 percent in fall 2005 and 71 versus 29 percent in spring 2006) and observing others teaching in their classrooms (61 versus 44 percent in fall 2005 and 68 versus 39 percent in spring 2006) at both time points during year 1, and they were significantly more likely than control teachers to report keeping written logs at one time point (38 versus 29 percent in spring 2006). However, treatment teachers were significantly less likely than control teachers to report meeting with a resource specialist to discuss the needs of particular students (66 versus 77 percent) in fall 2005. Treatment and control teachers did not differ significantly in their participation in the other five activities covered in the survey (keeping a portfolio and analysis of student work, working with a study group of new and experienced teachers, observing others teaching the beginning teacher's class, and meeting with a principal or with a literacy or mathematics coach or other curricular specialist).

Figure IV.12. Treatment-Control Differences in Percent of Teachers Who Completed Nine Types of PD Activities During Past Three Months: One-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Solid (open) circles indicate that treatment-control differences are (not) statistically significant at the 0.05 level N = 503 teachers in fall 2005, 499 teachers in spring 2006, 472 teachers in fall 2006, 426 teachers in fall 2007, and 398 teachers in fall 2008.

Legend: A = Kept written log, B = Kept portfolio and analysis of student work, C = Worked with study group of new teachers, D = Worked with study group of new and experienced teachers, E = Observed others teaching in their classrooms, F = Observed others teaching your class, G = Met with principal to discuss teaching, H = Met with literacy or mathematics coach or other curricular specialist, I = Met with a resource specialist to discuss needs of particular students.

After the comprehensive induction program ended, participation ranged widely from 13 to 78 percent among treatment teachers and from 12 to 80 percent among control teachers across the 27 activities and periods. In contrast to the differences in service receipt during the comprehensive induction program, there were no significant differences in participation between treatment and control teachers after the program ended in the nine activities covered in the survey with two exceptions across the 27 activities and periods: treatment teachers were significantly less likely to work with a study group of new teachers in fall 2006 (11 versus 21 percent) or to observe others teaching in the beginning teacher's classroom in fall 2008 (25 versus 35 percent).

h. Professional Development Topic Areas

During the comprehensive induction program, treatment teachers' reported attendance at professional development sessions in 12 topic areas during the three months prior to the survey ranged widely from 21 to 78 percent and did not reach 100 percent for any topic area. The range of reported attendance among control teachers was 21 to 74 percent. Figure IV.13 illustrates differences in levels of support for treatment and control teachers over time. For both treatment

and control teachers in both fall 2005 and spring 2006, the most attended topic area sessions were on instructional techniques while the least attended sessions were on understanding the composition of students in the teacher's class.

Figure IV.13. Treatment-Control Differences in Percent of Teachers Who Attended PD in 12 Topic Areas During the Past Three Months: One-Year Districts

Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Solid (open) circles indicate that treatment-control differences are (not) statistically significant at the 0.05 level N = 503 teachers in fall 2005, 499 teachers in spring 2006, 472 teachers in fall 2006, 426 teachers in fall 2007, and 398 teachers in fall 2008.

Legend: A = Parent and community relations, B = School policies on student disciplinary procedures, C = Instructional techniques/strategies, D = Understanding the composition of students in your class, E = Content area knowledge (language arts, mathematics, science), F = Lesson planning, G = Analyzing student work/assessment, H = Student motivation/engagement, I = Differentiated instruction, J = Using computers to support instruction, K = Classroom management techniques, L = Preparing students for standardized testing.

Treatment and control teachers did not differ significantly in their reported attendance at professional development sessions in 12 topic areas during the three months prior to the survey while the comprehensive induction program was being implemented, with five exceptions across the 24 areas and periods. Treatment teachers were significantly more likely to attend professional development sessions in lesson planning (33 versus 22 percent) in spring 2006 and analyzing student work and assessment (52 versus 43 percent) in spring 2006. Treatment teachers were significantly less likely to attend sessions in content area knowledge (61 versus 72 percent) in fall 2005, preparing students for standardized testing (30 versus 41 percent) in fall 2005, and school disciplinary policies (32 versus 45 percent) in spring 2006.

After the comprehensive induction program ended, participation rates remained variable among topic area sessions for both treatment and control teachers. Participation rates ranged from 12 to

77 percent among treatment teachers and from 15 to 82 percent among control teachers across the 36 areas and periods. For both treatment and control teachers in fall 2006, fall 2007, and fall 2008, the most attended topic area sessions were on instructional techniques while the least attended sessions were on parent and community relations.

Treatment and control teachers did not differ significantly in their reported attendance at professional development sessions after the comprehensive induction program ended, with two exceptions across the 36 areas and periods: treatment teachers were significantly less likely than control teachers to attend PD in differentiated instruction in fall 2007 (42 versus 55 percent) and analyzing student work and assessment in fall 2008 (42 versus 56 percent).

2. Two-Year Districts

a. Mentor Assignments

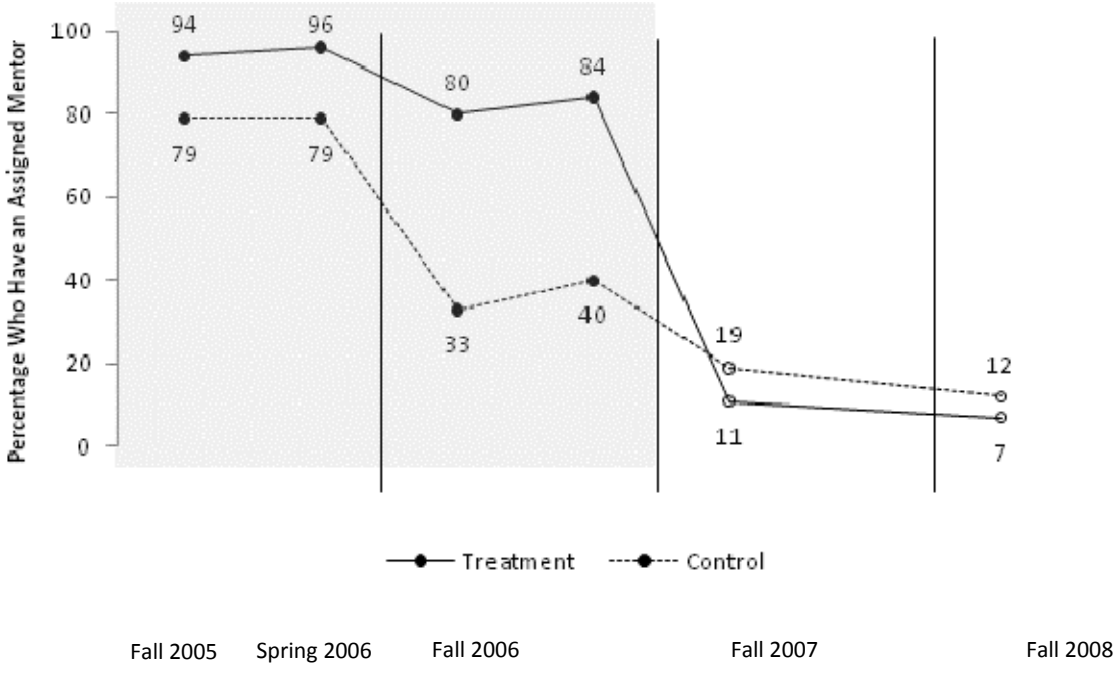
A majority of both treatment and control teachers reported having a mentor assigned to them during the comprehensive induction program, but treatment teachers were significantly more likely than control teachers to report this support. Figure IV.14 illustrates these levels of support for treatment and control teachers over time. The differences were 94 versus 79 percent in fall 2005, 96 versus 79 percent in spring 2006, 80 versus 34 percent in fall 2006, and 84 versus 40 percent in spring 2007. As in the one-year districts, while some treatment teachers do not report having a mentor assigned to them, all treatment teachers were assigned a mentor as part of the comprehensive induction program.

The percent of both treatment and control teachers with an assigned mentor dropped significantly from spring 2007 to fall 2007 (p-values both 0.000), after the comprehensive induction program ended. to less than 20 percent. In fall 2007, 11 percent of treatment teachers versus 19 percent of control teachers reported having an assigned mentor; these rates were 7 versus 12 percent in fall 2008. Differences between treatment and control teachers in the likelihood of having any mentor follow a similar pattern over time, but with a negative and significant difference in fall 2008.

b. Number and Types of Mentors

During the first year of the comprehensive induction program, treatment teachers were significantly more likely than control teachers to report having multiple mentors (38 versus 23 percent in fall 2005 and 38 versus 22 percent in spring 2006) and having two assigned mentors (31 versus 13 percent in fall 2005 and 31 versus 17 percent in spring 2006). During the second year of the comprehensive induction program and after the comprehensive induction program ended, however, treatment-control differences in these measures were insignificant.

Figure IV.14. Treatment-Control Differences in Percent Who Have an Assigned Mentor: Two-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. N = 395 teachers in fall 2005, 386 teachers in spring 2006, 360 teachers in fall 2006, 372 teachers in spring 2007, 326 teachers in fall 2007, and 321 teachers in fall 2008.

Treatment-control differences are significantly different from zero at the 0.05 level except in fall 2007 and fall 2008.

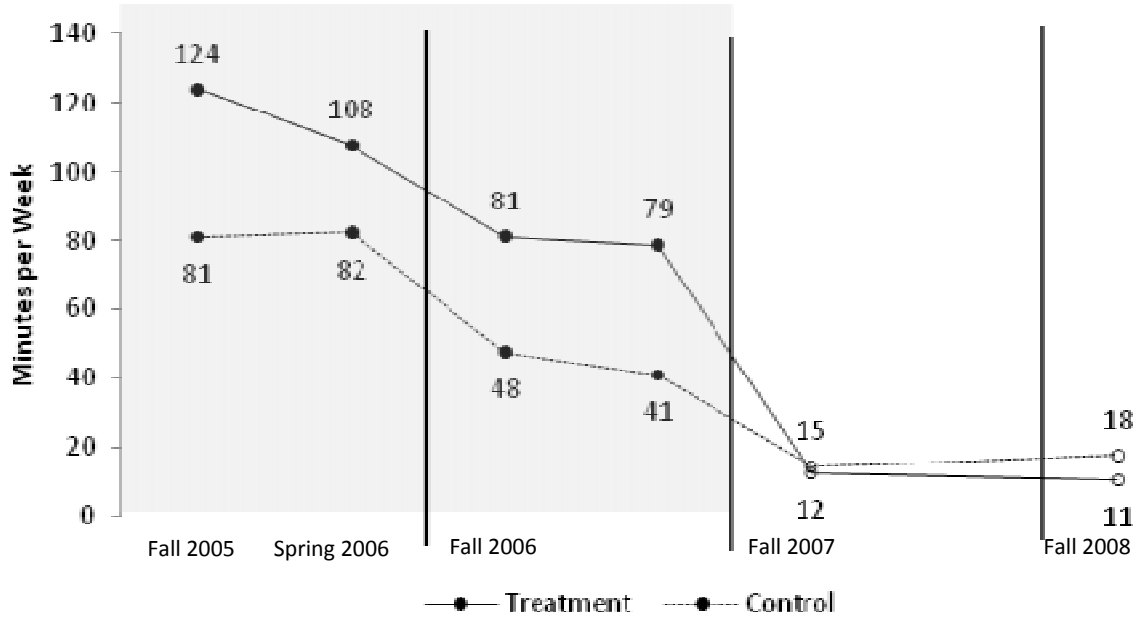
Differences between treatment and control teachers in their mentors' positions were significant during the comprehensive induction program. Treatment teachers were significantly more likely than control teachers to report having a full-time mentor (72 versus 16 percent in fall 2005, 75 versus 17 percent in spring 2006, 64 percent versus 7 percent in fall 2006, and 67 versus 15 percent in spring 2007) and significantly less likely to report having a mentor who was a teacher (38 versus 62 percent in fall 2005, 39 versus 65 percent in spring 2006, 12 versus 27 percent in fall 2006, and 16 versus 27 percent in spring 2007). After the comprehensive induction program ended, treatment teachers were significantly less likely to have a mentor who was a teacher in fall 2007 (8 versus 17 percent), but they did not differ significantly from control teachers in having a full-time mentor in either fall 2007 or fall 2008.

c. Meeting Time with Mentors

Treatment and control teachers both spent time per week in mentor meetings and activities during the comprehensive induction program, but treatment teachers spent significantly more time than control teachers in three out of four periods: from 79 to 124 minutes, compared to 41 to 82 minutes for control teachers. Figure IV.15 illustrates average total weekly mentor meeting time for treatment and control teachers over time. Combining usual scheduled time and informal time with all mentors during the most recent full week of teaching, we found that treatment teachers spent an average of 124 minutes in mentor meetings versus 81 minutes for control teachers in fall 2005, 108 versus 82 minutes in spring 2006, 82 versus 48 minutes in fall 2006, and 79 versus

41 minutes in spring 2007. Since total meeting time was not reported directly but had to be constructed from reports of the frequency and duration of usual scheduled meetings and the time spent in informal meetings, we could not determine precisely whether treatment teachers met with their study mentors for two hours per week as the ETS and NTC programs expected. The statistically significant differences in meeting time (43 minutes in fall 2005, 34 minutes in fall 2006, and 38 minutes in spring 2007) were attributable entirely to differences in the duration of the usual scheduled meetings (24 versus 12 minutes in fall 2005, 19 versus 7 minutes in fall 2006, and 20 versus 6 minutes in spring 2007). Differences in total meeting time in fall 2005 and fall 2006 are shown separately by district in Figures B.3–B.4 in Appendix B.

Figure IV.15. Treatment-Control Differences in Total Minutes Spent in Mentoring per Week: Two-Year Districts

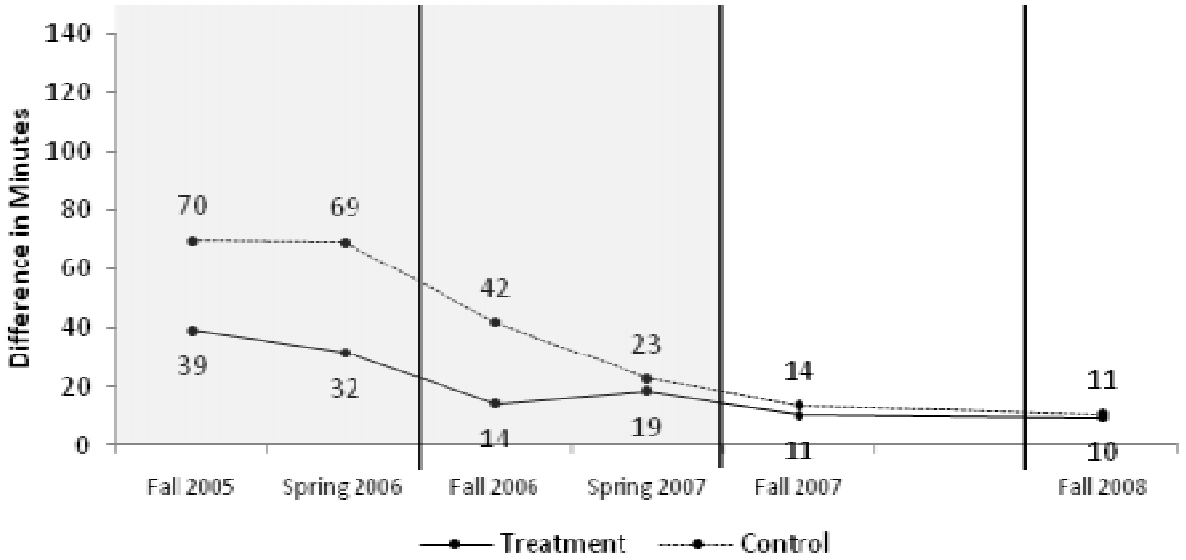


Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. N = 395 teachers in fall 2005, 386 teachers in spring 2006, 360 teachers in fall 2006, 372 teachers in spring 2007, 326 teachers in fall 2007, and 321 teachers in fall 2008.

Treatment-control differences are significantly different from zero at the 0.05 level except in spring 2006, fall 2007, and fall 2008 (N = 395 teachers in fall 2005, 386 teachers in spring 2006, 360 teachers in fall 2006, 372 teachers in spring 2007, 326 teachers in fall 2007, and 321 teachers in fall 2008).

Figure IV.16. Treatment-Control Differences in Total Minutes Spent in Mentoring per Week with Mentors Who Are Teachers: Two-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. N = 395 teachers in fall 2005, 386 teachers in spring 2006, 360 teachers in fall 2006, 372 teachers in spring 2007, 326 teachers in fall 2007, and 321 teachers in fall 2008.

Treatment-control differences are significantly different from zero at the 0.05 level except in spring 2007, fall 2007, and fall 2008.

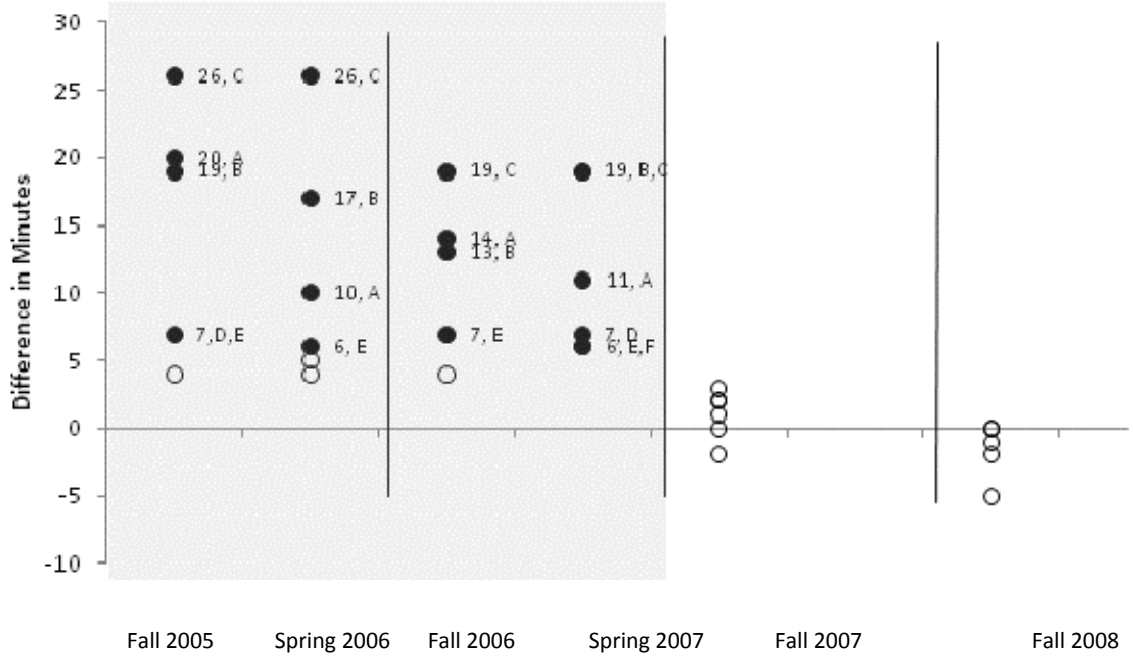
Differences in the time spent with full-time mentors and mentors who were teachers reflected differences in the types of mentors that treatment and control teachers reported. Treatment teachers reported spending significantly more time during the most recent week of teaching meeting with full-time mentors than did control teachers (75 versus 6 minutes in fall 2005, 71 versus 10 minutes in spring 2006, 59 versus 2 minutes in fall 2006, and 54 versus 6 minutes in fall 2007). However, they reported significantly less time than control teachers with mentors who were teachers (39 versus 70 minutes in fall 2005, 32 versus 70 minutes in spring 2006, 14 versus 42 minutes in fall 2006, and 19 versus 23 minutes in spring 2007). Figure IV.16 illustrates average minutes each week meeting with mentors who were teachers for treatment and control teachers over time.

After the comprehensive induction program ended, the time spent by both treatment and control teachers in mentor meetings dropped significantly (p-values both 0.000) to less than 15 minutes per week. There were no significant differences between treatment and control teachers in the time spent meeting with mentors during this period (12 versus 15 minutes in fall 2007 and 11 versus 18 minutes in fall 2008). No significant differences were found in either component of total mentor meeting time, usual scheduled meetings (7 versus 9 minutes in fall 2007 and 6 versus 9 minutes in fall 2008) or informal meetings (6 versus 6 minutes in fall 2007 and 5 versus 9 minutes in fall 2008). Differences between treatment and control teachers were also insignificant during this period in terms of time spent with full-time mentors and mentors who were teachers.

d. Mentor Activities

During the comprehensive induction program, treatment teachers reported spending 7 to 43 minutes in each of six types of mentoring activities during the most recent full week of teaching while control teachers reported spending 2 to 23 minutes. For both treatment and control teachers, time spent varied across activities and time periods. Figure IV.17 illustrates differences in levels of support for treatment and control teachers over time.

Figure IV.17. Treatment-Control Differences in Time Spent in Six Mentoring Activities in the Last Full Week of Teaching: Two-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Solid (open) circles indicate that treatment-control differences are (not) statistically significant at the 0.05 level. In fall 2005, E is significant but D is not. N = 395 teachers in fall 2005, 386 teachers in spring 2006, 360 teachers in fall 2006, 372 teachers in spring 2007, 326 teachers in fall 2007, and 321 teachers in fall 2008.

Legend: A = Observing BT teaching, B = Meeting with BT one-on-one, C = Meeting with BT and other first-year teachers, D = Meeting with BT and other teachers, E = Modeling a lesson, F = Co-teaching a lesson, BT = Beginning teacher.

Treatment teachers spent more time than control teachers in certain of these activities during certain time periods during the comprehensive induction program: being observed by mentors (38 versus 17 minutes in fall 2005, 26 versus 16 minutes in spring 2006, 22 versus 7 minutes in fall 2006, and 19 versus 8 minutes in spring 2007), meeting one-on-one with mentors (43 versus 23 minutes in fall 2005, 38 versus 21 minutes in spring 2006, 25 versus 12 minutes in fall 2006, and 29 versus 10 minutes in spring 2007), meeting together with mentors and other first-year teachers (38 versus 11 minutes in fall 2005, 35 versus 9 minutes in spring 2006, 25 versus 6 minutes in fall

2006, and 24 versus 5 minutes in spring 2007), and having mentors model lessons (16 versus 10 minutes in fall 2005, 14 versus 8 minutes in spring 2006, 12 versus 5 minutes in fall 2006, and 10 versus 4 minutes in spring 2007).[39] Treatment teachers did not spend significantly more time than control teachers meeting together with mentors and other teachers or co-teaching lessons with mentors, except in spring 2007 (16 versus 8 minutes and 8 versus 2 minutes, respectively). The total minutes per week spent by treatment and control teachers in the six types of activities covered in the survey was 170 versus 87 minutes in fall 2005, 146 versus 77 minutes in spring 2006, 106 versus 44 minutes in fall 2006, and 105 versus 36 minutes in spring 2007.

After the comprehensive induction program ended, both treatment and control teachers reported spending less than 10 minutes per week in each of the six types of mentoring activities during the most recent full week of teaching. Differences between treatment and control teachers in total time spent in the six activities and in all individual activities were not statistically significant.

e. Areas of Mentor Guidance

During the comprehensive induction program, at least half of all treatment teachers reported receiving mentors' assistance during the most recent full week of teaching in all 10 topic areas covered in the survey in each period, while roughly 20 to 75 percent of control teachers reported receiving assistance. Figure IV.18 illustrates differences in levels of support for treatment and control teachers over time. The percentage of treatment teachers receiving each type of assistance ranged from 66 to 92 percent in fall 2005, 70 to 92 percent in spring 2006, 50 to 72 percent in fall 2006, and 58 to 78 percent in spring 2007, but then dropped after the comprehensive induction program ended to 8 to 13 percent in fall 2007, and 3 to 6 percent in fall 2008. Among control teachers, the percentage reporting each type of assistance ranged from 48 to 73 percent in fall 2005, 44 to 70 percent in spring 2006, 21 to 30 percent in fall 2006, 19 to 38 percent in spring 2007, 8 to 17 percent in fall 2007, and 7 to 13 percent in fall 2008. The area in which both treatment and control teachers received most guidance in each period was encouragement or moral support, and the area in which they received least guidance was on how to assess students, with one exception: treatment teachers in fall 2006 received the least help in the area of teaching to meet state or district standards.

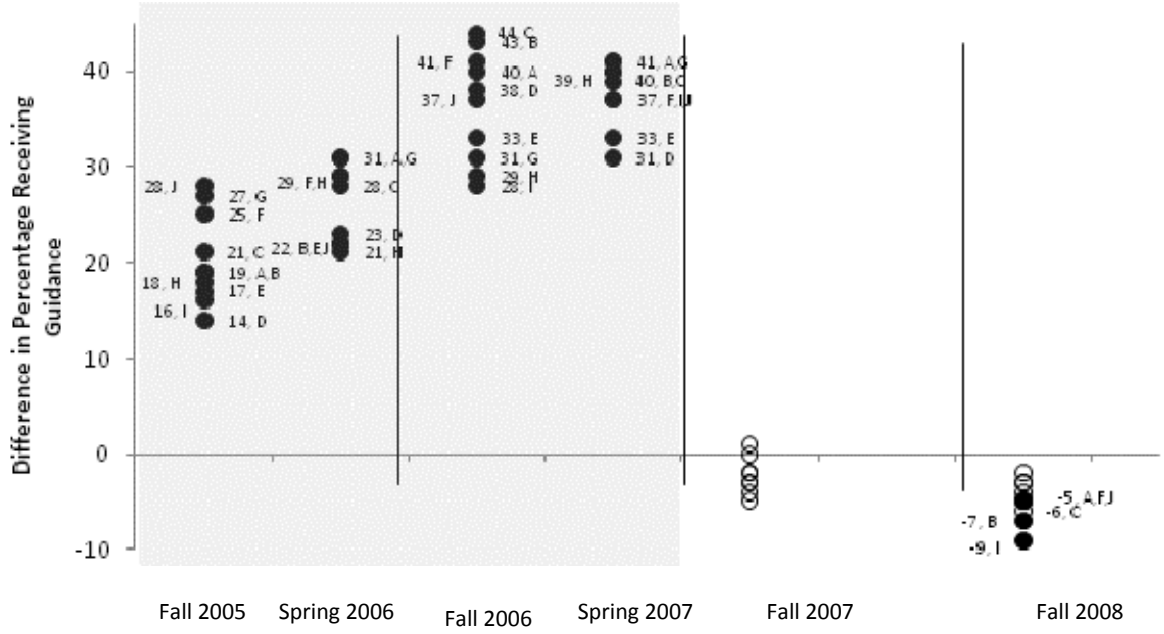
Examining differences in service receipt, during the comprehensive induction program, we found treatment teachers were significantly more likely than control teachers to report receipt of mentors' assistance during the most recent full week of teaching in all 10 topic areas covered in the survey. Differences ranged from 14 to 28 percentage points in fall 2005, 28 to 44 percentage points in spring 2006, 21 to 31 percentage points in fall 2006, and 31 to 41 percentage points in spring 2007. The following areas of guidance were among the top four in terms of impact size in three of the four periods: receiving suggestions to improve practice, having opportunities to raise issues or discuss concerns, receiving help identifying teaching challenges and solutions, and discussing instructional goals and ways to achieve them.

After the comprehensive induction program ended, fewer than 20 percent of treatment or control teachers reported receiving each type of assistance during the last full week of teaching.

[39] Implied differences in the time spent in some activities may not match the differences shown in Figure IV.17 due to rounding.

Among treatment teachers, the percentage reporting each type of assistance ranged from 4 to 13 percent, while among control teachers, the percentage reporting each type of assistance ranged from 8 to 17 percent. There were no significant differences between treatment and control in the receipt of mentors' assistance in any of the topic areas covered in the survey with one exception across the 20 areas and periods: treatment teachers were significantly less likely than control teachers to share lesson plans, assignments, or other instructional activities (5 versus 13 percent) in fall 2008.

Figure IV.18. Treatment-Control Differences in Percent of Teachers Who Received 10 Types of Mentor Guidance in the Last Full Week of Teaching: Two-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Solid (open) circles indicate that treatment-control differences are (not) statistically significant at the 0.05 level. In fall 2008, A is significant but F and J are not. N = 395 teachers in fall 2005, 386 teachers in spring 2006, 360 teachers in fall 2006, 372 teachers in spring 2007, 326 teachers in fall 2007, and 321 teachers in fall 2008.

Legend: A = Suggestions to improve practice, B = Encouragement or moral support, C = Opportunity to raise issues/discuss concerns, D = Help with administrative/logistical issues, E = Help with teaching to meet state or district standards, F = Help identifying teaching challenges and solutions, G = Discussed instructional goals and ways to achieve them, H = Guidance on how to assess students, I = Shared lesson plans, assignments, or other instructional activities, J = Acted on a beginning teacher's request.

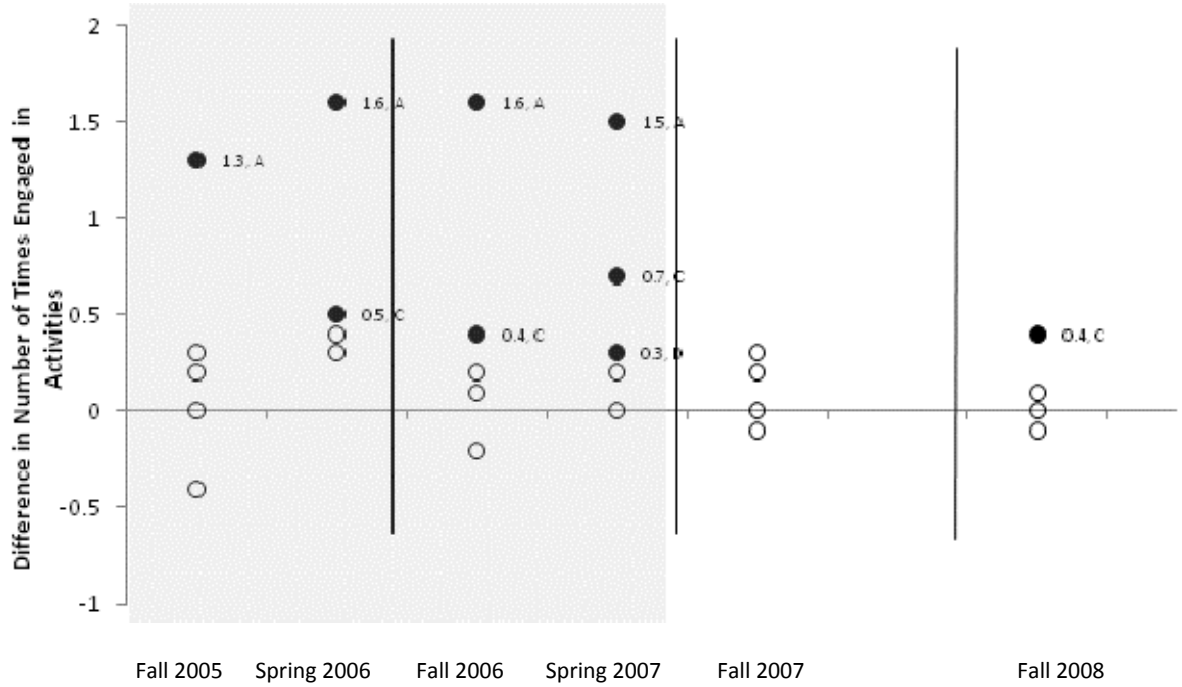
f. Observations and Feedback

During the comprehensive induction program, both treatment and control teachers reported receiving two types of observations and three types of feedback at least once during the three months prior to the survey. Treatment teachers received certain types of observations and feedback significantly more frequently than control teachers. Figure IV.19 illustrates differences in levels of support for treatment and control teachers over time.

Examining differences in service receipt during the comprehensive induction program, treatment teachers were significantly more likely than control teachers to be observed by mentors (3.4 versus 2.1 times in fall 2005, 3.2 versus 1.6 times in spring 2006, 2.3 versus 0.8 times in fall 2006, and 2.5 versus 1.0 times in spring 2007) and to receive informal feedback (2.5 versus 2.0 times

in spring 2006, 1.9 versus 1.5 times in fall 2006, and 2.2 versus 1.5 times in spring 2007) during the comprehensive induction program, with one exception across these 8 activities and periods: treatment teachers were not significantly more likely than control teachers to receive informal feedback during fall 2005. Treatment and control teachers did not differ significantly in the frequency with which they were observed by their principals or the frequency with which they received feedback on lesson plans or as part of a formal evaluation during the intervention with one exception across these 12 activities and periods: treatment teachers were more likely than control teachers to receive feedback as part of a formal evaluation in spring 2007 (1.6 versus 1.3 times).

Figure IV.19. Treatment-Control Differences in the Frequency of Selected Activities During Past Three Months: Two-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Solid (open) circles indicate that treatment-control differences are (not) statistically significant at the 0.05 level. N = 395 teachers in fall 2005, 386 teachers in spring 2006, 360 teachers in fall 2006, 372 teachers in spring 2007, 326 teachers in fall 2007, and 321 teachers in fall 2008.

Legend: A = Teaching was observed by mentor, B = Teaching was observed by principal, C = Given feedback on your teaching, not as part of formal evaluation, D = Given feedback on your teaching as part of formal evaluation, E = Given feedback on your lesson plans.

After the intervention, the frequency of the two types of observations during the three months prior to the survey ranged from 0 to 2 times among treatment teachers and 0 to 1 time among control teachers, while the frequency of the three types of feedback ranged from 1 to 2 times for both treatment and control teachers. Treatment and control teachers did not differ significantly in the frequency with which they engaged in these activities with one exception across the 10 activities and periods: treatment teachers were significantly more likely than control teachers to receive informal feedback in fall 2008 (1.5 versus 1.0 times).

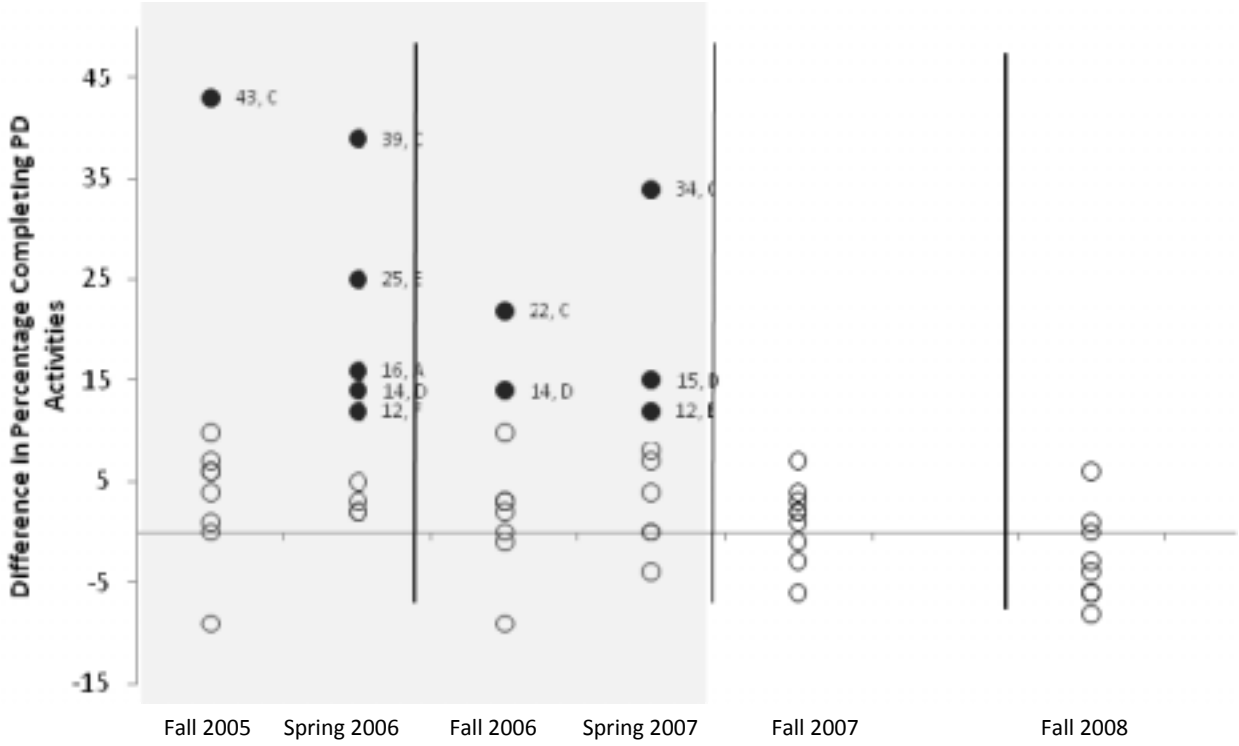
g. Professional Development Activities

During the comprehensive induction program, both treatment and control teachers reported participating in nine professional development activities during the three months prior to the survey. Figure IV.20 illustrates differences in levels of support for treatment and control teachers over time. Among treatment teachers, participation rates across the 36 activities and periods ranged from 33 to 86 percent and did not reach 100 percent for any activity. The most commonly reported activity among treatment teachers in each period was keeping a portfolio and analysis of student work, while the least commonly reported activity was keeping a written log. Among control teachers, these participation rates ranged from 14 to 84 percent, with the most commonly reported activity in each period being keeping a portfolio and analysis of student work and the least commonly reported activity in each period being working with a study group of new teachers.

Treatment teachers were more likely than control teachers to report participating in certain of these activities during the comprehensive induction program. Treatment teachers were significantly more likely to work with a study group of new teachers (67 versus 24 percent in fall 2005, 64 versus 25 percent in spring 2006, 42 versus 19 percent in fall 2006, and 48 versus 14 percent in spring 2007) or a study group of new and experienced teachers (48 versus 34 percent in spring 2006, 54 versus 40 percent in fall 2006, and 51 versus 36 percent in spring 2007), with one exception across these 8 activities and periods: treatment teachers were no more likely than control teachers to work with a group of new and experienced teachers in fall 2005. Treatment teachers were also significantly more likely than control teachers to observe others teaching in their classrooms in spring 2006 and spring 2007 (72 versus 47 percent and 47 versus 35 percent, respectively) or teaching the beginning teacher's class in spring 2006 (48 versus 36 percent). Treatment and control teachers did not differ significantly in their participation in the other five activities covered in the survey (keeping a written log; keeping a portfolio and analysis of student work; meeting with a principal; or meeting with a literacy or mathematics coach, other curricular specialist, or resource specialist) in any period with one exception across these 20 activities and periods: treatment teachers were more likely than control teachers to keep a written log in spring 2006 (42 versus 26 percent).

After the comprehensive induction program ended, participation ranged widely from 12 to 84 percent among treatment teachers and from 12 to 85 percent among control teachers across the 18 activities and periods. In contrast to the differences in service receipt during the comprehensive induction program, there were no significant differences between treatment and control teachers in any of the professional development activities covered in the survey during these periods.

Figure IV.20. Treatment-Control Differences in Percent of Teachers Who Completed Nine Types of PD Activities During the Past Three Months: Two-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Solid (open) circles indicate that treatment-control differences are (not) statistically significant at the 0.05 level. N = 395 teachers in fall 2005, 386 teachers in spring 2006, 360 teachers in fall 2006, 372 teachers in spring 2007, 326 teachers in fall 2007, and 321 teachers in fall 2008.

Legend: A = Kept written log, B = Kept portfolio and analysis of student work, C = Worked with study group of new teachers, D = Worked with study group of new and experienced teachers, E = Observed others teaching in their classrooms, F = Observed others teaching your class, G = Met with principal to discuss teaching, H = Met with literacy or mathematics coach or other curricular specialist, I = Met with a resource specialist to discuss needs of particular students.

h. Professional Development Topic Areas

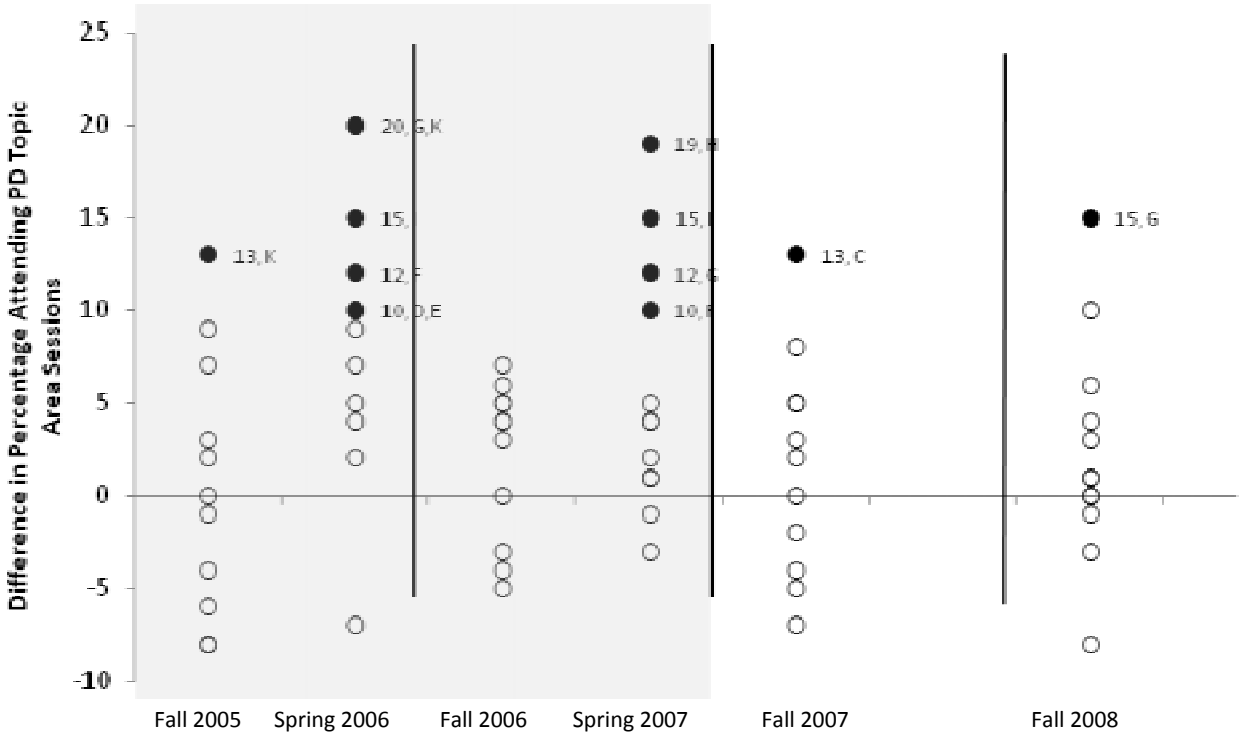
During the comprehensive induction program, the reported attendance of treatment teachers at professional development sessions in 12 topic areas during the three months prior to the survey ranged widely from 24 to 80 percent and did not reach 100 percent for any topic area. The range of reported attendance among control teachers was 18 to 79 percent. Figure IV.21 illustrates differences in levels of support for treatment and control teachers over time. For both treatment and control teachers in all four periods, the most attended topic sessions were on instructional techniques. The least attended sessions included understanding the composition of students in the teacher's class in three of the four periods for treatment teachers and in two of the four periods for control teachers.

Treatment teachers were significantly more likely than control teachers to attend professional development activities in certain topic areas during the three months prior to the survey while the

comprehensive induction program was being implemented. These areas included understanding the composition of students in your class in spring 2006 (32 versus 22 percent), content area knowledge in spring 2006 (70 versus 60 percent), lesson planning in spring 2006 and spring 2007 (43 versus 31 percent and 38 versus 28 percent, respectively), analyzing student work and assessment in spring 2006 and spring 2007 (60 versus 41 percent and 57 versus 45 percent, respectively), student motivation and engagement in spring 2007 (43 versus 24 percent), differentiated instruction in spring 2006 and spring 2007 (62 versus 47 percent and 58 versus 43 percent, respectively), and classroom management techniques in fall 2005 and spring 2006 (61 versus 48 percent and 53 versus 34 percent, respectively). Treatment teachers were not significantly more likely than control teachers to attend PD on parent and community relations, school policies on student disciplinary procedures, instructional techniques and strategies, using computers to support instruction, or preparing students for standardized testing during the intervention (20 areas and periods). The lack of significant differences between treatment and control teachers during the fall 2005 and fall 2006 surveys is not due to the recall period including part of the summer vacation for some teachers since treatment and control teachers completed their surveys at the same time on average.

After the comprehensive induction program ended, participation rates remained variable among topic area sessions for both treatment and control teachers. Participation rates ranged from 15 to 78 percent among treatment teachers and from 15 to 78 percent among control teachers across the 24 areas and periods. For both treatment and control teachers in fall 2007 and fall 2008, the most attended topic area sessions were on instructional techniques while the least attended sessions were on parent and community relations. Treatment teachers were not significantly more likely than control teachers to attend PD during this period, with two exceptions across the 24 areas and periods: treatment teachers were more likely to attend PD on instructional techniques and strategies in fall 2007 (78 versus 64 percent) and in analyzing student work and assessment in fall 2008 (55 versus 39 percent).

Figure IV.21. Treatment-Control Differences in Percent of Teachers Who Attended PD in 12 Topic Areas During the Past Three Months: Two-Year Districts



Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Solid (open) circles indicate that treatment-control differences are (not) statistically significant at the 0.05 level. N = 395 teachers in fall 2005, 386 teachers in spring 2006, 360 teachers in fall 2006, 372 teachers in spring 2007, 326 teachers in fall 2007, and 321 teachers in fall 2008.

Legend: A = Parent and community relations, B = School policies on student disciplinary procedures, C = Instructional techniques/strategies, D = Understanding the composition of students in your class, E = Content area knowledge (language arts, mathematics, science), F = Lesson planning, G = Analyzing student work/assessment, H = Student motivation/engagement, I = Differentiated instruction, J = Using computers to support instruction, K = Classroom management techniques, L = Preparing students for standardized testing.

V. IMPACT FINDINGS: CLASSROOM OUTCOMES

In this chapter, we present the estimated impacts of comprehensive induction on classroom outcomes. We hypothesize that treatment improves teachers' classroom practices and ultimately their students' achievement as reflected in standardized test scores. First, we examine the effect on teachers' classroom practices in the teaching of a literacy lesson in their first year of teaching, drawing on the results that were reported in the study's first interim report (Glazerman et al. 2008). Then, we test whether the students of teachers who received comprehensive induction performed better on standardized tests than did students of teachers who received the usual induction services (the control group). For both sets of outcomes, we present a summary of methods, findings, and sensitivity tests. For the classroom practices analysis, we focused on those teachers responsible for English language arts or literacy classes (698 teachers). For test score analyses, we focused on teachers in tested grades and subjects (about 200 teachers per year). Results pertaining to literacy instruction do not necessarily apply to teachers of other subjects. Similarly, results for teachers in tested grades do not necessarily apply to teachers of other grades or subjects. Readers may refer to Appendix A for a detailed description of analytic methods and to Appendix C for supplementary tables.

A. Classroom Practices

The conceptual framework presented in Chapter I suggests that for teacher induction to improve student achievement, it must first change the way teachers teach. To test for changes in teacher practices, we sent trained observers into treatment and control classrooms to administer the Diagnostic Classroom Observation (DCO), described in Chapter III and Appendix A. The DCO, formerly known as the Vermont Classroom Observation Tool, measures three domains of teaching: lesson implementation, lesson content, and classroom culture. It captures the degree to which the observed lesson reflects evidence of what are believed to be effective practices. We observed literacy lessons (or reading/language arts lessons) in more than 600 classrooms in spring 2006 (year 1 of the study). By design, we did not repeat the observations in later years. Teachers who were teaching special populations, were teaching subjects other than reading/English language arts, were no longer teaching, or had prior teaching experience were not observed (by design) and therefore not included in this component of the impact analysis. See Chapter III for a more detailed discussion of the sample exclusions.

We estimated impacts on classroom practices using the regression methods described in Appendix A. Because observations were conducted only in the first year of the study, we combined results from one-year and two-year districts. As discussed in Chapter III, observers scored teachers on a five-point scale in each of the three domains based on a set of 16 items believed to be indicators of effective practice. The three domains cover five, four, and seven of the indicators, respectively. The full set of 16 indicators is shown by domain in Appendix C (Table C.1), and covariates in the model are listed in Table A.1.

To summarize the information from the classroom observations across all 16 indicators, we produced three scores corresponding to the three domains captured by the observation protocol (into which the items had already been grouped). The benchmark estimates use the average score of the indicators within each domain and thus assume that the intervals between each category are equal. For example, the difference between "no evidence" and "limited evidence" is the same as the difference between "moderate evidence" and "consistent evidence." It also assigns equal weight to the indicators within each domain. In other words, we assume that a score of 3 on two indicators of

classroom culture—for example, "Classroom routines are clear and consistent" and "Behavior is respectful and appropriate"—is equivalent to a score of 4 on one of those indicators and a 2 on the other. Histograms for treatment and control teachers' performance in each of the three domains are included in Appendix C (Figures C.1–C.3). These histograms illustrate the pattern of variation (or distribution) of the classroom practices data. We also present impact estimates for literacy implementation scores separately by district in Figures C.4–C.5.

No Impact on Classroom Practices. There was no statistically significant impact of treatment on classroom practices for teachers in one-year and two-year districts combined (Table V.1). After controlling for important teacher and school characteristics, we observed no statistically significant differences between treatment and control teachers' performance on the implementation of a literacy lesson, content of a literacy lesson, or classroom culture. We express the impact on each domain of classroom practice as the difference in scores on the five-point scale. An impact of 0.5 points, for example, would suggest that the intervention moves the average teacher from being able to demonstrate "moderate" evidence of a particular practice in that domain half of the distance to being able to demonstrate "consistent" evidence of that practice. (The observed estimates of the impacts were smaller than the 0.5 points used in this example.)

Table V.1. Impacts on Classroom Practices (Average Score on a Five-Point Scale): One-Year Districts and Two-Year Districts Combined, 2005–2006 School Year

Outcome	Treatment	Control	Difference	Effect Size	P-value
Implementation of literacy lesson	2.7	2.6	0.0	0.02	0.766
Content of literacy lesson	2.4	2.4	0.0	-0.01	0.875
Classroom culture	3.1	3.0	0.0	0.04	0.629
Sample Size (Teachers)	**342**	**289**			

Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data are weighted and regression adjusted using ordinary least squares to account for differences in baseline characteristics and the study design. Scoring scale: (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence of effective teaching practice.

None of the differences is statistically significant at the 0.05 level.

Findings Were Robust. We reestimated the impacts using a variety of assumptions about item scoring and estimation and found that the results did not change substantially. The results were not sensitive to how we grouped the individual items into constructs, nor did they change when we collapsed the scale. We estimated the model separately for each classroom observation item after recoding each score from a five-point scale into a binary variable: (1) no, limited, or moderate evidence or (2) consistent or extensive evidence of good practice. The results support the same conclusions of no impact (see Table C.1 in Appendix C). The results were also not sensitive to the choice of summary score; when we substituted the observer summary scores for the computed average scores, we reached the same conclusions of no impact (see Table C.2). Finally, when we estimated the model separately for one-year and two-year districts (see Appendix C, Tables C.3 and C.4), we found that the impact estimates were not significantly different from zero.

B. Student Achievement

We compared the test scores for students of treatment teachers to those of control teachers, adjusted for pretest scores and background characteristics of students and teachers. Although district-administered test scores do not cover every domain of student achievement that induction might affect, they do capture the content that school districts or states deem important and worthy of assessing.

We present results for the teachers' third year, the 2007–2008 school year. Results for the first two years show no overall impacts in either subject, as documented in earlier reports (Glazerman et al. 2008; Isenberg et al. 2009). Although comprehensive teacher induction services ended after the 2005–2006 school year in one-year districts and after the 2006–2007 school year in two-year districts, there can be delayed impacts of induction programs because teachers may not be able to implement the advice they have been given immediately (Isenberg et al. 2010). The teachers' third year was two years following the end of the intervention in one-year districts, and one year following the end of the intervention in two-year districts.

For year 3, we found no evidence of a positive impact on test scores in either subject in one-year districts, but evidence of a positive and significant impact on both subjects in two-year districts. We checked the findings using different methods of aggregation, model specification, and model estimation, and we report on both the benchmark model and alternative specifications. The no-impact results for reading in one-year districts are robust to these sensitivity analyses, but impacts for math may be either statistically insignificant or negative and significant if we make alternative assumptions about the statistical model or data processing. For two-year districts, results for reading and math are not robust to all specifications and samples, as they may be either positive and significant or statistically insignificant, depending on the model.

Estimating impacts on student achievement required the use of test score data from 15 districts, which administered different tests under different conditions and followed different record-keeping practices. Although 10 one-year districts participated in the study, one of these districts was unable to match teachers in the study with student test scores. A second district declined to share its data in the third year of the study. All 7 two-year districts participated in all years of the study, but in the third year of the study, as a result of attrition, one district had no grades in which there was at least one treatment teacher and one control teacher who could be compared to each other.

We aggregated test scores across districts and grades by standardizing each test to a common metric called a z-score, which has a mean of zero and a standard deviation of one. The benchmark model accounts for the nesting of students within schools. As shown in Table A.1, the covariates in the benchmark model are (1) the normalized student pretest score (interacted with district-by-grade fixed effects so that each test has a different effect); (2) student characteristics; (3) teacher personal characteristics; (4) teacher professional characteristics; and (5) district-by-grade fixed effects. Appendix A describes in more detail the aggregation method, treatment of missing data, regression model, and estimation strategies.

1. One-Year Districts: No Impacts on Math or Reading

For one-year districts, the benchmark estimates of the impacts on math and reading scores were not significantly different from zero (see Table V.2).

Table V.2. Impacts on Test Scores: One-Year Districts, 2007–2008 School Year

Subject	Adjusted Mean Test Scores		Difference (Effect Size)	P-value	Sample Sizes		
	Treatment	Control			Students	Teachers	Districts
Reading	-0.24	-0.24	0.01	0.801	1,690	99	8
Math	-0.17	-0.07	-0.10	0.108	1,629	95	8

Source: Mathematica analysis using data from the 2006–2007 and 2007–2008 school years provided by participating school districts; Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study, Data are regression adjusted to account for pretest, student and teacher characteristics, district-by-grade fixed effects, and clustering of students within schools. Treatment and control group sample sizes are shown in Appendix Table C.5.

None of the differences is statistically significant at the 0.05 level.

We performed a sensitivity analysis by reestimating the impacts using different samples, sets of covariates, and estimation techniques:

- Disaggregating results by grade
- Removing teacher characteristics as control variables, so that only pretest, student background characteristics, and district-by-grade fixed effects remain (see Table A.1 in Appendix A for a list of control variables)
- Removing teacher and student characteristics as control variables so that only pretest and district-by-grade fixed effects remain (see Table A.1 in Appendix A for a list of control variables)
- Estimating impacts using a set of weights so that each school receives an equal weight in the analysis
- Estimating impacts using a set of weights so that each district receives an equal weight in the analysis
- Using specific information on teaching assignments gathered during telephone followup to surveys, to determine eligibility for analysis
- Including all students with a pretest and posttest, without imposing restrictions on the teachers and students included in the sample
- Estimating impacts without controlling for a pretest, using the same sample as the benchmark model
- Estimating impacts without controlling for a pretest, expanding the sample to include students with a posttest only
- Estimating impacts in a model that compares treatment and control teachers to each other within a district rather than within a district-grade combination
- Estimating impacts using the opposite-subject pretest as an instrumental variable to remove the influence of measurement error in the pretest

Grade-specific estimates are useful in that they can illustrate heterogeneity of impacts, and they do not require the assumption that increments of different types of learning be on the same scale. Keep in mind, however, that the study was not designed to detect significant impacts at the grade level. We also present estimates of the impacts separately by district in Figures C.6–C.9 in Appendix C. Details of the method that generated these results can be found in Appendix A.

Results for Reading in One-Year Districts Are Robust. When the reading results for one-year districts are disaggregated by grade, the impact estimate for grade 5 is significantly positive and the estimates are not significant for grades 2, 3, and 4. Results are shown in the top panel of Table V.3.

Table V.3. Impacts on Test Scores by Grade: One-Year Districts, 2007–2008 School Year

	Adjusted Mean Test Scores				Sample Sizes		
Subject/Grade	Treatment	Control	Difference (Effect Size)	P-value	Students	Teachers	Districts
Reading							
Grade 2	-0.04	0.03	-0.07	0.115	444	25	4
Grade 3	-0.14	-0.10	-0.04	0.650	421	27	5
Grade 4	-0.40	-0.41	0.01	0.901	479	32	7
Grade 5	-0.37	-0.55	0.18*	0.024	346	19	4
All Grades	-0.24	-0.24	0.01	0.801	1,690	99	8
Math							
Grade 2	0.31	0.45	-0.13	0.299	200	13	2
Grade 3	-0.10	0.25	-0.35*	0.002	422	27	5
Grade 4	-0.29	-0.21	-0.07	0.429	661	40	8
Grade 5	-0.30	-0.46	0.16	0.143	346	19	4
All Grades	-0.17	-0.07	-0.10	0.108	1,629	95	8

Source: Mathematica analysis using data from the 2006–2007 and 2007–2008 school years provided by participating school districts; Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study, Data are regression adjusted to account for pretest, student and teacher characteristics, district-by-grade fixed effects, and clustering of students within schools. Treatment and control group sample sizes are shown in Appendix Table C.6.

*Significantly different from zero at the 0.05 level.

A set of additional specification checks, shown in the top panel of Table V.4, confirmed that there was no statistically significant effect of treatment on reading in the third year of teaching. The top row of Table V.4 repeats the results of the benchmark analysis for reference. The second and third rows are estimated with fewer covariates than in the benchmark model. For the results shown in the second row, teacher background characteristics have been excluded so that we control for only pretest, student background characteristics, and district-grade fixed effects. In the third row, we exclude both teacher and student background characteristics so that we control only for pretest and district-grade fixed effects. In both cases, the impact estimates, consistent with the benchmark, are not significantly different from zero. See Appendix Table A.1 for a list of student and teacher covariates used in these models.

Table V.4. Impacts on Test Scores, Alternate Model Specifications: One-Year Districts, 2007–2008 School Year

Subject/Model	Adjusted Mean Test Scores		Difference (Effect Size)	P-value	Sample Sizes		
	Treatment	Control			Students	Teachers	Districts
Reading							
(1) Benchmark	-0.24	-0.24	0.01	0.801	1,690	99	8
(2) No teacher covariates	-0.23	-0.25	0.02	0.638	1,690	99	8
(3) No teacher or student covariates	-0.23	-0.25	0.02	0.669	1,690	99	8
(4) Schools weighted equally	-0.25	-0.24	-0.01	0.768	1,690	99	8
(5) Districts weighted equally	-0.23	-0.24	0.01	0.764	1,690	99	8
(6) Using specific information on teacher assignments	-0.25	-0.24	0.00	0.970	1,540	89	8
(7) Without imposing data restrictions	-0.26	-0.28	0.02	0.600	1,892	107	8
(8) No pretest, benchmark sample	-0.18	-0.29	0.10	0.108	1,690	99	8
(9) No pretest, expanded sample	-0.17	-0.27	0.10	0.085	2,856	151	8
(10) Compare teachers within districts, not district-grades	-0.27	-0.27	0.01	0.880	1,973	114	8
(11) Instrumental variables	-0.27	-0.24	-0.03	0.445	1,519	93	8
Math							
(1) Benchmark	-0.17	-0.07	-0.10	0.108	1,629	95	8
(2) No teacher covariates	-0.17	-0.06	-0.11*	0.044	1,629	95	8
(3) No teacher or student covariates	-0.18	-0.06	-0.12*	0.038	1,629	95	8
(4) Schools weighted equally	-0.20	-0.07	-0.13*	0.000	1,629	95	8
(5) Districts weighted equally	-0.18	-0.08	-0.10	0.092	1,629	95	8
(6) Using specific information on teacher assignments	-0.17	-0.03	-0.14*	0.037	1,487	87	8
(7) Without imposing data restrictions	-0.20	-0.08	-0.12*	0.042	1,700	97	8
(8) No pretest, benchmark sample	-0.10	-0.13	0.03	0.749	1,629	95	8
(9) No pretest, expanded sample	-0.09	-0.16	0.08	0.290	2,702	138	8
(10) Compare teachers within districts, not district-grades	-0.20	-0.12	-0.07	0.229	1,804	104	8
(11) Instrumental variables	-0.15	-0.04	-0.11	0.091	1,554	93	8

Source: Mathematica analysis using data from the 2006–2007 and 2007–2008 school years provided by participating school districts; Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are regression adjusted to account for clustering of students within schools. See Appendix Table A.1 for a list of student and teacher covariates used in these models. Treatment and control group sample sizes are shown in Appendix Table C.7.

*Significantly different from zero at the 0.05 level.

The fourth and fifth rows of Table V.4 show results from models that use the same sample as the benchmark model but apply a special set of weights to the students in the model. In the benchmark model, each student in the model receives an equal weight, implying that schools or districts with a greater number of students receive more weight in the overall results. To check how well the results generalize, for the fourth row, the student observations are weighted in such a way that each school in the model receives an equal weight. In the fifth row, the student observations are weighted in such a way that each district in the model receives an equal weight. Results from both of these specifications are consistent with the benchmark.

We confirmed that there was no impact on reading when we modified the rules for including or excluding teachers from the sample. We had to make several decisions about which teachers were correctly linked to students and exclude those believed to be incorrectly linked. As explained in Appendix A, we attempted to follow up by telephone with teachers whom we suspected should have been excluded from the analysis because, according to the data we received from the school district, they taught too many or too few students to be plausibly eligible for the analysis as a regular classroom teacher. We contacted teachers who met certain criteria and asked them how many students they taught in reading and math in the 2006-2007 and 2007-2008 school years. Based on the information we received, we constructed a set of general eligibility rules. Because treatment teachers were more likely to confirm their teaching assignments than control teachers, we applied the four restriction rules to all teachers in the benchmark analysis, even if they were an exception to these general rules. In this way, the application of the restrictions was not confounded with treatment status. As a check on our results, however, we estimated treatment effects based on a sample that used general rules for teachers whom we were unable to contact and particular information from teachers' responses if it contradicted the general rules.[1] The results are shown in the sixth row of Table V.4. As a further check, shown in the seventh row of Table V.4, we estimated the model based on a sample that did not impose any exclusion rules. The estimated impact of treatment in both of these models remained statistically insignificant.

Another set of models change the statistical model, thereby expanding the sample. We excluded from the benchmark analysis any student with missing pretest scores. Because of this, we may have excluded from the analysis mobile students and students who were in the lowest grade tested in the district, often third grade. The eighth row shows results from estimating impacts without controlling for a pretest but maintaining the same sample as the benchmark model. The ninth row shows results that do not control for pretest and expands the sample by including all students with a valid posttest score, including those who lack a pretest score. In both cases, the impact is not significantly different from zero.

The tenth row shows results from a model that uses district fixed effects (rather than district-grade fixed effects). Because comparisons between treatment and control teachers are made at the district level rather than the district-grade level, this model does not require that there be treatment/control overlap in each district-grade combination for teachers to be included. This is

[1] For reading for one-year districts, the additional information resulted in our excluding 10 teachers who were included when applying the general rules. For math for one-year districts, we excluded 8 teachers who had been included. For reading for two-year districts, the additional information did not change the sample. For math for two-year districts, we excluded fewer than three teachers.

because we are now estimating each teacher's effectiveness by implicitly comparing each teacher to all teachers in her or his state and then comparing the effectiveness of treatment teachers to control teachers for all grades within a district. With this estimation strategy, we can include a larger sample of teachers. This model also shows no significant treatment effect.

The eleventh row shows results from a regression model in which the math pretest is used as an instrumental variable to adjust for measurement error in the reading pretest. This also decreases the sample since students who lack a math pretest are excluded. In nonexperimental settings, if the students of teachers in the treatment and control groups are different in ways not easily observable to the researcher, this estimation strategy can correct bias in the estimates. Although we have conducted an experiment, these results are included to account for the possibility that principals may have assigned students to treatment and control teachers differently in the third year of the study than they did in the first year. For example, if principals believed that comprehensive teacher induction gave teachers a better ability to cope with disruptive students, they may have been more willing than usual to place potentially disruptive students in those teachers' classrooms in subsequent years. Using the instrumental variables model, however, did not change our findings.

Possible Negative Impact on Math in One-Year Districts. While the estimated math impact in the benchmark model was not statistically significant, the detailed analysis showed conditions under which the negative impact estimate was statistically significant. For the math results, the bottom panel of Table V.3 shows that grade-by-grade impacts are negative and statistically significant for grade 3 and not significantly different from zero for other grades.

Using all grades for math tests, the bottom panel of Table V.4 shows findings of no impact when we weight districts equally (line 5), exclude the pretest from the model (lines 8-9), use a model based on district fixed effects (line 10), or use an instrumental variables approach (line 11). Results are negative and significant when we include only student covariates (line 2), include only a single pretest measure plus district-grade fixed effects as covariates (line 3), weight schools equally (lines 4), use specific information from follow-up phone calls to teachers (line 6), or do not impose data restrictions on the sample (line 7).

2. Two-Year Districts: Positive Impact on Math and Reading for Benchmark Sample

For two-year districts, the benchmark estimates of the impacts on reading and math scores were positive and significant for the third year, one year following the end of the intervention in these districts (see Table V.5). The estimates suggest that assignment to two years of comprehensive teacher induction instead of a district's usual induction services increases student reading scores by 11 percent of a standard deviation and increases math scores by 20 percent of a standard deviation. These impacts are the equivalent of moving the average student from the 50th percentile up 4 percentile points in reading and 8 percentile points in math. As we show in the sensitivity analyses, however, if we reestimate the impacts without requiring test scores from the prior year, we do not find an impact on math or reading scores. This alternative approach nearly doubles the available sample of study teachers but the lack of data on students' prior achievement results in a less precise estimate. This means that we are less likely to detect a true impact if it exists, despite the larger sample size.

Impact Estimates for Reading in Two-Year Districts Not Robust to All Specifications and Samples. The sensitivity analysis for the reading test score impacts in two-year districts showed that the statistical significance of the result required pooling all grades and including covariates. Underlying the positive and significant overall impact on reading was a positive and significant effect in grade 3 and impact estimates for grades 2, 4, 5, and 6 that were not statistically significant, as shown in the top panel of Table V.6.

The top panel of Table V.7 shows mixed results when we apply alternate model specifications. The impact estimates remain positive and significant when we weight districts equally (line 5),

Table V.5. Impacts on Test Scores: Two-Year Districts, 2007–2008 School Year

	Adjusted Mean Test Scores				Sample Sizes		
Subject	Treatment	Control	Difference (Effect Size)	P-value	Students	Teachers	Districts
Reading	-0.14	-0.25	0.11*	0.018	1,347	74	6
Math	-0.06	-0.26	0.20*	0.000	1,198	68	6

Source: Mathematica analysis using data from the 2006–2007 and 2007–2008 school years provided by participating school districts; Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study, Data are regression adjusted to account for pretest, student and teacher characteristics, district-by-grade fixed effects, and clustering of students within schools. Treatment and control group sample sizes are shown in Appendix Table C.8.

*Significantly different from zero at the 0.05 level.

Table V.6. Impacts on Test Scores by Grade: Two-Year Districts, 2007–2008 School Year

	Adjusted Mean Test Scores		Difference (Effect		Sample Sizes		
Subject/Grade	Treatment	Control	Size)	P-value	Students	Teachers	Districts
Reading							
Grade 2	-0.26	-0.29	0.03	0.723	63	5	1
Grade 3	-0.21	-0.50	0.29*	0.000	261	18	2
Grade 4	-0.03	-0.07	0.04	0.595	741	39	5
Grade 5	-0.24	-0.44	0.19	0.098	246	12	2
Grade 6	-1.20	-1.02	-0.18	0.339	36	3	1
All Grades	-0.14	-0.25	0.11*	0.018	1,347	74	6
Math							
Grade 2	-0.23	-0.37	0.14	0.092	63	5	1
Grade 3	-0.14	-0.41	0.27*	0.000	279	19	2
Grade 4	0.12	-0.05	0.17*	0.033	614	35	5
Grade 5	-0.33	-0.46	0.13	0.090	206	11	2
Grade 6	-0.74	-1.25	0.51	0.078	36	3	1
All Grades	-0.06	-0.26	0.20*	0.000	1,198	68	6

Source: Mathematica analysis using data from the 2006–2007 and 2007–2008 school years provided by participating school districts; Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study, Data are regression adjusted to account for pretest, student and teacher characteristics, district-by-grade fixed effects, and clustering of students within schools. Treatment and control group sample sizes are shown in Appendix Table C.9.

*Significantly different from zero at the 0.05 level.

incorporate specific callback information (line 6), do not impose data restrictions on the sample (line 7), use a model with district fixed effects (line 10), and use an instrumental variables approach to remove the influence of measurement error (line 11).[41] On the other hand, we find results that are not statistically different from zero when we change the covariates used in the model, either by omitting teacher and/or student background characteristics (lines 2 and 3) or excluding the pretest (lines 8-9). The results are also statistically insignificant when we change the method used to estimate the standard errors of the model.[42] When we weight schools equally (line 4) the p-value rises above the 0.05 significance threshold to 0.052.

Impact Estimate for Math in Two-Year Districts Not Robust to All Specifications and Samples. The conclusion that treatment had a statistically significant positive impact on math did not change when we conducted sensitivity analysis except in some models where student covariates like pretest were ignored. Underlying the benchmark estimate in two-year districts are positive and significant impacts for grades 3 and 4. Results for grades 2, 5, and 6 are statistically insignificant, but the estimates for grades 2 and 6 were each derived using data from one district. See the bottom panel of Table V.6.

The bottom panel of Table V.7 shows that the initial finding of a positive and significant impact is robust to 7 of the 10 changes in covariates or specification that we present. The alternate specifications show a positive and significant effect of treatment except when we include only a single pretest measure plus district-grade fixed effects as covariates (line 3) or exclude the pretest from the model (lines 8-9).

3. Understanding the Year 3 Findings on Test Score Impacts in the Two-Year Districts

Positive and significant results emerge in two-year districts only in the third year of the study and depend on the definition of the analysis sample. In order to better understand this pattern of findings, we conducted two additional exploratory analyses: (1) we examined test score impacts for teachers who remained in the analysis sample in more than one study year; (2) we explored whether there are positive impacts on intermediary variables of interest for the benchmark sample of 74 teachers for reading and 68 teachers for math.

[41] For reading estimates for two-year districts, the sample did not change when we incorporated specific callback information (line 6), as none of the teachers contacted during the follow-up phase indicated a different teaching assignment from what we had assumed.

[42] As explained in Appendix A, there are multiple methods with which to estimate the standard errors of hierarchical models. The benchmark student achievement models use ordinary least squares with Huber-White robust standard errors. We use this method because we collect data from 15 school districts, and ordinary least squares does not require as many assumptions about the distribution of the error terms as a random effects model. Under the alternative methods discussed in Appendix A—generalized least squares estimates of a random effects model, maximum likelihood, and restricted maximum likelihood—the treatment effect of the reading estimate for two-year districts is not statistically significant.

Table V.7. Impacts on Test Scores, Alternate Model Specifications: Two-Year Districts, 2007–2008 School Year

Subject/Model	Adjusted Mean Test Scores		Difference (Effect Size)	P-value	Sample Sizes		
	Treatment	Control			Students	Teachers	Districts
Reading							
(1) Benchmark	**-0.14**	**-0.25**	**0.11***	**0.018**	**1,347**	**74**	**6**
(2) No teacher covariates	-0.17	-0.20	0.04	0.562	1,347	74	6
(3) No teacher or student covariates	-0.18	-0.19	0.01	0.853	1,347	74	6
(4) Schools weighted equally	-0.14	-0.24	0.09	0.052	1,347	74	6
(5) Districts weighted equally	-0.15	-0.27	0.12*	0.007	1,347	74	6
(6) Using specific information on teacher assignments	-0.14	-0.25	0.11*	0.018	1,347	74	6
(7) Without imposing data restrictions	-0.14	-0.25	0.11*	0.018	1,347	74	6
(8) No pretest, benchmark sample	-0.16	-0.21	0.05	0.496	1,347	74	6
(9) No pretest, expanded sample	-0.26	-0.18	-0.07	0.392	2,457	127	7
(10) Compare teachers within districts, not district-grades	-0.12	-0.27	0.16*	0.001	1,484	82	6
(11) Instrumental variables	-0.11	-0.27	0.16*	0.000	1,338	73	6
Math							
(1) Benchmark	**-0.06**	**-0.26**	**0.20***	**0.000**	**1,198**	**68**	**6**
(2) No teacher covariates	-0.08	-0.23	0.14*	0.019	1,198	68	6
(3) No teacher or student covariates	-0.09	-0.21	0.12	0.077	1,198	68	6
(4) Schools weighted equally	-0.07	-0.26	0.20*	0.000	1,198	68	6
(5) Districts weighted equally	-0.06	-0.26	0.19*	0.000	1,198	68	6
(6) Using specific information on teacher assignments	-0.03	-0.25	0.22*	0.000	1,161	66	6
(7) Without imposing data restrictions	-0.01	-0.23	0.23*	0.000	1,259	70	6
(8) No pretest, benchmark sample	-0.08	-0.23	0.15	0.054	1,198	68	6
(9) No pretest, expanded sample	-0.19	-0.16	-0.03	0.739	2,254	120	7
(10) Compare teachers within districts, not district-grades	-0.04	-0.17	0.13*	0.046	1,398	77	6
(11) Instrumental variables	-0.04	-0.28	0.24*	0.000	1,193	68	6

Source: Mathematica analysis using data from the 2006–2007 and 2007–2008 school years provided by participating school districts; Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are regression adjusted to account for clustering of students within schools. See Appendix Table A.1 for a list of student and teacher covariates used in these models. Treatment and control group sample sizes are shown in Appendix Table C.10.

*Significantly different from zero at the 0.05 level.

The impetus for the first analysis is to determine whether the progression to positive and statistically significant impacts in year 3 represents teacher improvement or a shift in the composition of the group of teachers in the study sample. Though the study follows a group of teachers over three years, the particular group of teachers included in the test score sample varies from year to year. This was due to teachers changing teaching assignment, leaving the district, leaving teaching, or teaching in a grade where their counterparts in the treatment or control group left the sample because each teacher needed a comparison teacher in the same grade and district. In this analysis, we restrict the data to a "common sample" of teachers who had students with valid test score data in either Year 1 and Year 3 or Year 2 and Year 3.The resulting sample sizes for these analyses are small, ranging from 37 to 52 teachers, so true changes in relative teacher performance may be difficult to detect.

Results, shown in Appendix C, Table C.11, include some evidence of treatment teachers improving relative to controls. Within the common samples, in reading there was a positive and significant improvement for treatment teachers relative to control teachers for the sample of teachers common to year 1 and year 3 (p-value = 0.00). In math, there was a positive but not statistically significant improvement (p-value = 0.14). Common sample results for year 2 and year 3, however, show no gain in reading and a small positive but statistically insignificant gain in math (p-value = 0.38).

The second exploratory analysis asks whether there were positive impacts on measures of induction activities and classroom practices for the samples of teachers who were eligible for the test score analysis, consistent with the conceptual framework of teacher induction (Figure I.1). This sample was referred to above as the "benchmark". Results are shown in Appendix C, Tables C.12 and C.13, and summarized as follows:

- Treatment teachers in this sample were more likely than controls to report having a mentor assigned to them in each of the four surveys during the first two years, while the treatment was in place. The differences, all statistically significant, were more than 20 percentage points in the first year and more than 50 points in the second year of implementation.

- There were no statistically significant differences for this sample in the overall amount of *time* spent with mentors, but treatment teachers were more likely by statistically significant margins of 23 percentage points or more to report receiving suggestions on improving instructional practices than their control group counterparts. Also, higher percentages of treatment than control teachers in this sample reported having received guidance with their particular subject area (math or reading), although the differences were not statistically significant for two of the three surveys for reading (p-values = 0.343, 0.000, 0.325 for fall 2005, spring 2006, and spring 2007, respectively) and were not significant for math during any of the surveys (p-values = 0.854, 0.073, and 0.126). When teachers reported the number of times they received feedback on their teaching outside of a formal evaluation, the treatment group reported greater numbers on average in at least one of the four surveys during the first two years for math and reading. The other differences were not statistically significant. Thus, while treatment teachers in the test score samples did not spend more time with mentors than controls, there is some suggestive evidence that they spent that time on the types of activities that could lead to impacts on student achievement.

- For this sample's performance on the classroom practice outcomes that we observed directly in spring 2006 (teachers' first year in the classroom), treatment teachers scored higher than control teachers on the measure of classroom culture, although the difference (effect size = 0.2) was not statistically significant (p=0.199). Differences for the other two teacher practice measures, one positive and one negative, were smaller in absolute value and had even higher p-values and thus were not statistically significant either. However, the sample of 56 teachers in two-year districts that had both valid test score data in year 3 *and* classroom observation data from year 1 (because they had been teaching English/language arts) was not large enough to be able to detect impacts (i.e. to attain statistical significance) unless the impacts had been very large.

Overall, these findings provide some evidence that is consistent with the theory that comprehensive induction improved student outcomes. However, the small sample sizes and lack of statistical significance mean that these findings are suggestive only and should be interpreted with caution.

This page is intentionally left blank.

VI. IMPACT FINDINGS: WORKFORCE OUTCOMES

The main goal of this study is to estimate the impact of comprehensive teacher induction on teacher and student outcomes. This chapter focuses on teacher outcomes, particularly teachers' attachment to the workforce. The goal is to determine whether the intervention makes teachers more likely to continue teaching in their original district or anywhere else. As a step along the way toward retention, we measure teachers' attitudes that relate to career decisions, including their satisfaction with teaching and their feelings of preparedness to deal with different aspects of their jobs. The ultimate policy goal is not 100 percent teacher retention, but retention of better-qualified and higher-quality teachers, so we also present evidence on how comprehensive teacher induction affects the mix of teachers who decide to stay in the district and in the profession.

A. Impact Findings: Teacher Attitudes

The impact of teacher induction on teacher attitudes is an important signal of whether the program is generating its intended effect—an intermediate step on the way to encouraging retention. The induction activities surveys indicated that comprehensive induction did not make teachers feel more satisfied with or more prepared to do their jobs. There were no statistically significant positive impacts of treatment on teacher satisfaction or teacher preparedness at any of the time points that we collected data: fall 2005, spring 2006, fall 2006, spring 2007, fall 2007, or fall 2008 for either one-year or two-year districts.

Using items from the induction activities surveys, we measured teachers' feelings of satisfaction in 19 areas and teachers' feelings of preparedness in 13 areas. Factor analysis suggested that teacher satisfaction and teacher preparedness could be grouped into three categories each: satisfaction with (1) school, (2) class, and (3) career; and preparedness to (1) instruct, (2) work with students, and (3) work with others (details are given in Appendix A). The constructed scales for each of these categories exhibited internal consistency ranging from 0.72 to 0.98, as tested by the Cronbach's alpha coefficient. Psychometric properties for each scale are given in Appendix A, Table A.5.

Benchmark estimates for teacher satisfaction and teacher preparedness are based on a hierarchical linear model. As shown in Appendix A, Table A.1, the model has district and grade fixed effects and no other covariates. The three satisfaction scales and three preparedness scales were entered into separate regression models with the same set of control variables. The results did not vary according to estimation method or to the set of control variables we used.

1. No Impact on Teacher Satisfaction

Overall, teachers from the treatment and control groups reported feelings of satisfaction that differed by 0.1 or less on a four-point scale in fall 2005, spring 2006, fall 2006, spring 2007, fall 2007, and fall 2008.[43] Out of the 15 differences examined among teachers in one-year districts (three measures at five points in time), and the 18 differences examined among teachers in two-year districts (three measures at six points in time), none were statistically significant. Figures VI.1 and VI.2 show treatment group means represented by a solid line and control group means represented

[43] Teacher satisfaction was not measured in one-year districts in spring 2007.

by a dotted line. See Tables D.1 and D.2 in Appendix D for detailed information on point estimates, differences, *p*-values, and sample sizes.

As a sensitivity test, we recoded the teacher satisfaction data from each time point into two categories and examined individual survey items separately, finding that the results were consistent with those based on collapsed scales. Out of the 95 differences examined among teachers in one-year districts (19 measures at five points in time), one difference was statistically significant. Treatment teachers were significantly less likely than control teachers to report satisfaction with salary and benefits in fall 2007.[44] Out of the 114 differences examined among teachers in two-year districts (19 measures at six points in time), three differences were statistically significant. The results showed that treatment teachers were significantly more likely than control teachers to report satisfaction with opportunities for professional development in fall 2006 and spring 2007 and were more likely than control teachers to report satisfaction with school facilities (buildings and grounds) in fall 2007. Detailed information from fall 2007 and fall 2008 is available for one-year districts in Table D.3 and for two-year districts in Table D.4 in Appendix D.[45]

2. No Impact on Teachers' Feeling of Preparedness

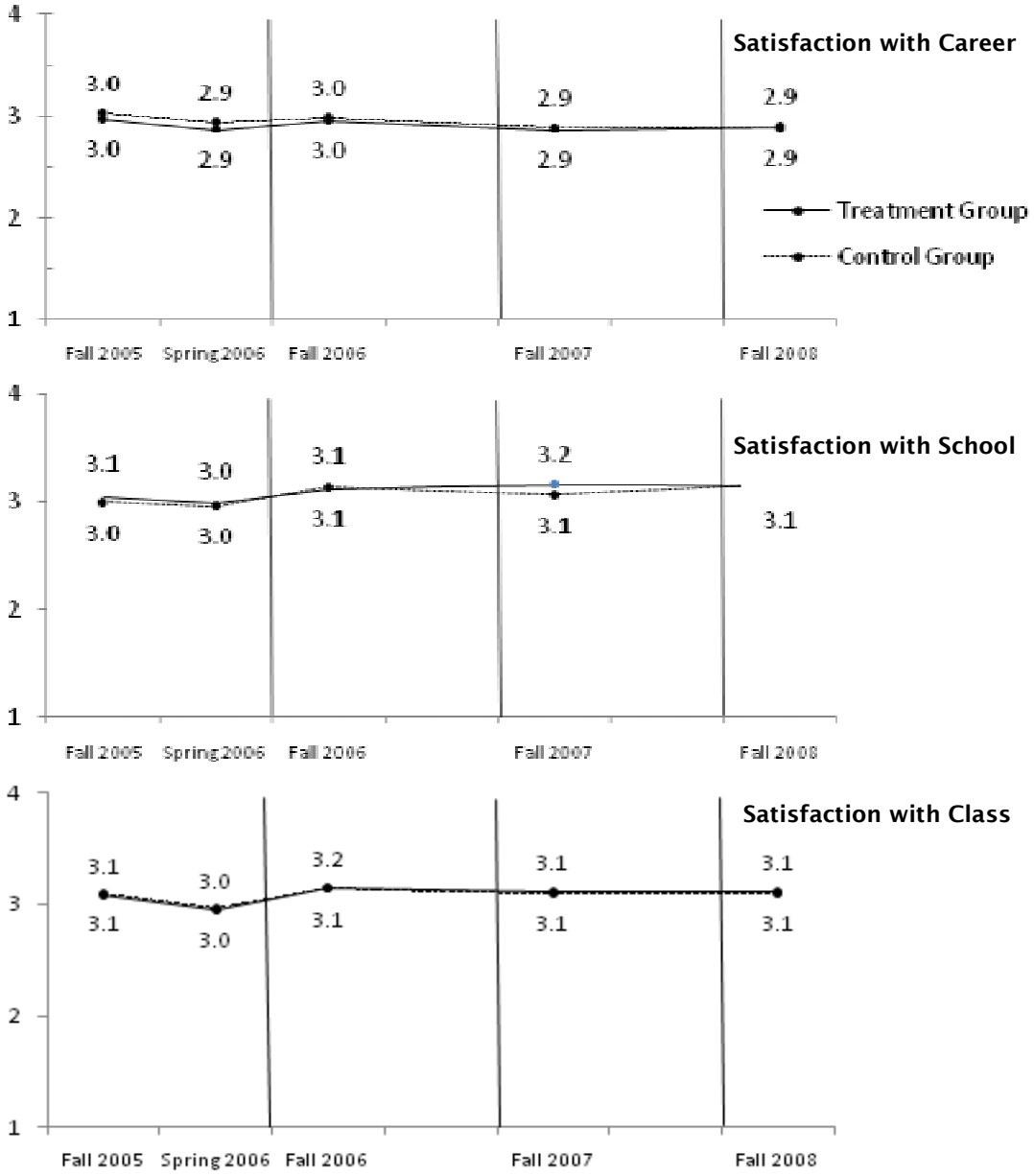
Overall, teachers from the treatment and control groups reported feelings of preparedness that differed by 0.1 or less on a four-point scale in fall 2005, spring 2006, spring 2007, and fall 2008 in both one-year and two-year districts.[46] Of the 9 differences examined among teachers in one-year districts (three measures at three points in time), none were statistically significant (Figure VI.3). Of the 12 differences examined among teachers in two-year districts (three measures at four points in time), one was statistically significant: treatment teachers were significantly less likely than control teachers to report being prepared to instruct in fall 2005 (Figure VI.4). See Appendix D for detailed information for one-year districts (Table D.5) and two-year districts (Table D.6).

[44] See Chapter II for a discussion of multiple comparisons and false discoveries.

[45] The item-specific impacts for fall 2005, spring 2006, fall 2006, and spring 2007 can be found in Isenberg et al. (2009).

[46] Teacher preparedness was not measured in one-year districts in fall 2006, spring 2007, or fall 2007. Teacher preparedness was not measured in two-year districts in fall 2006 or fall 2007.

Figure VI.1. Impacts on Teacher Satisfaction (Scores on a Four-Point Scale): One-Year Districts

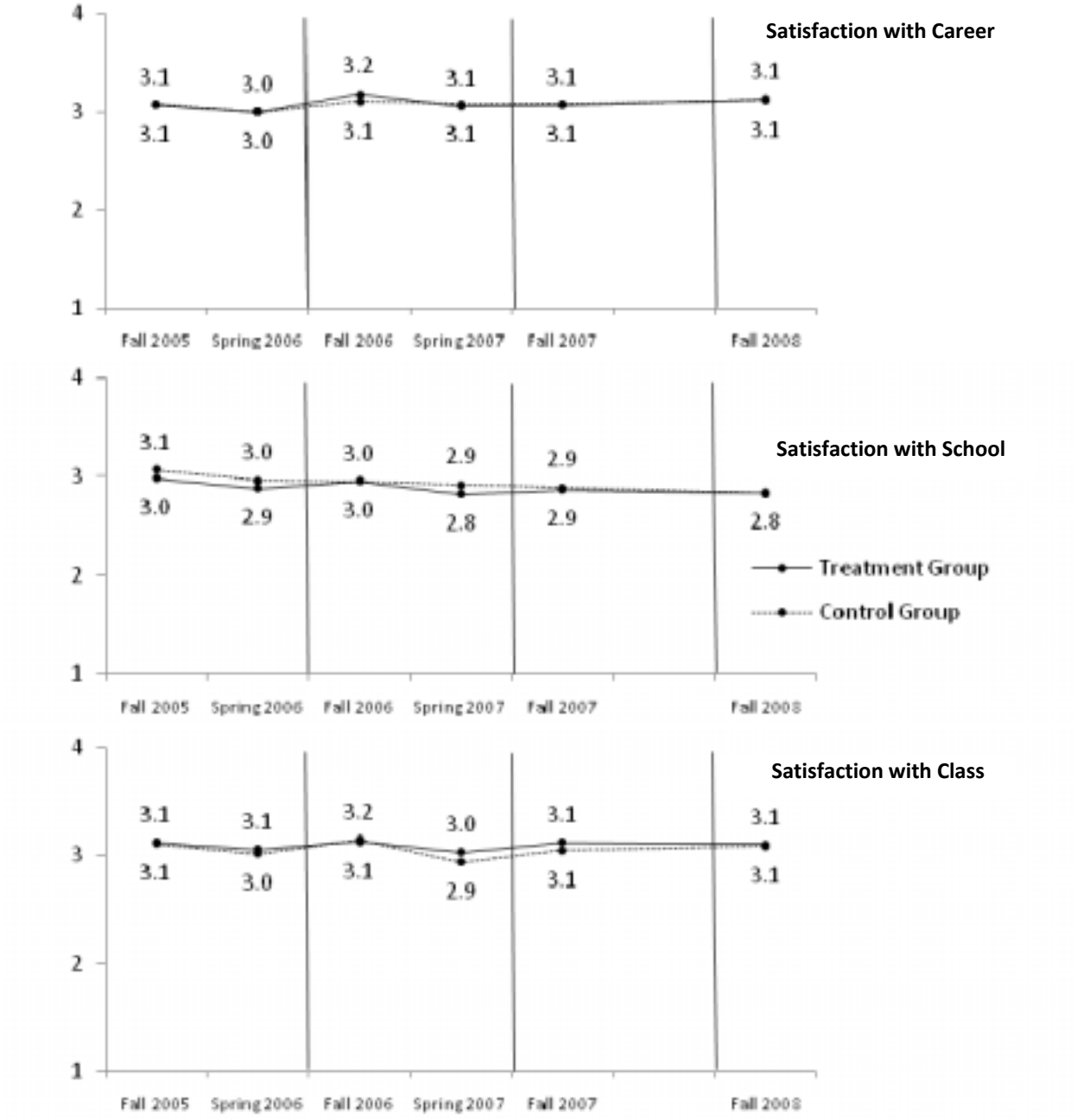


Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Satisfaction scale: (1) very dissatisfied, (2) somewhat dissatisfied, (3) somewhat satisfied, or (4) very satisfied. Sample sizes vary due to item nonresponse.

Treatment-control differences are not significantly different from zero at the 0.05 level (N = 498 teachers in fall 2005, 492 teachers in spring 2006, 472 teachers in fall 2006, 424 teachers in fall 2007, and 396 teachers in fall 2008).

Figure VI.2. Impacts on Teacher Satisfaction (Scores on a Four-Point Scale): Two-Year Districts

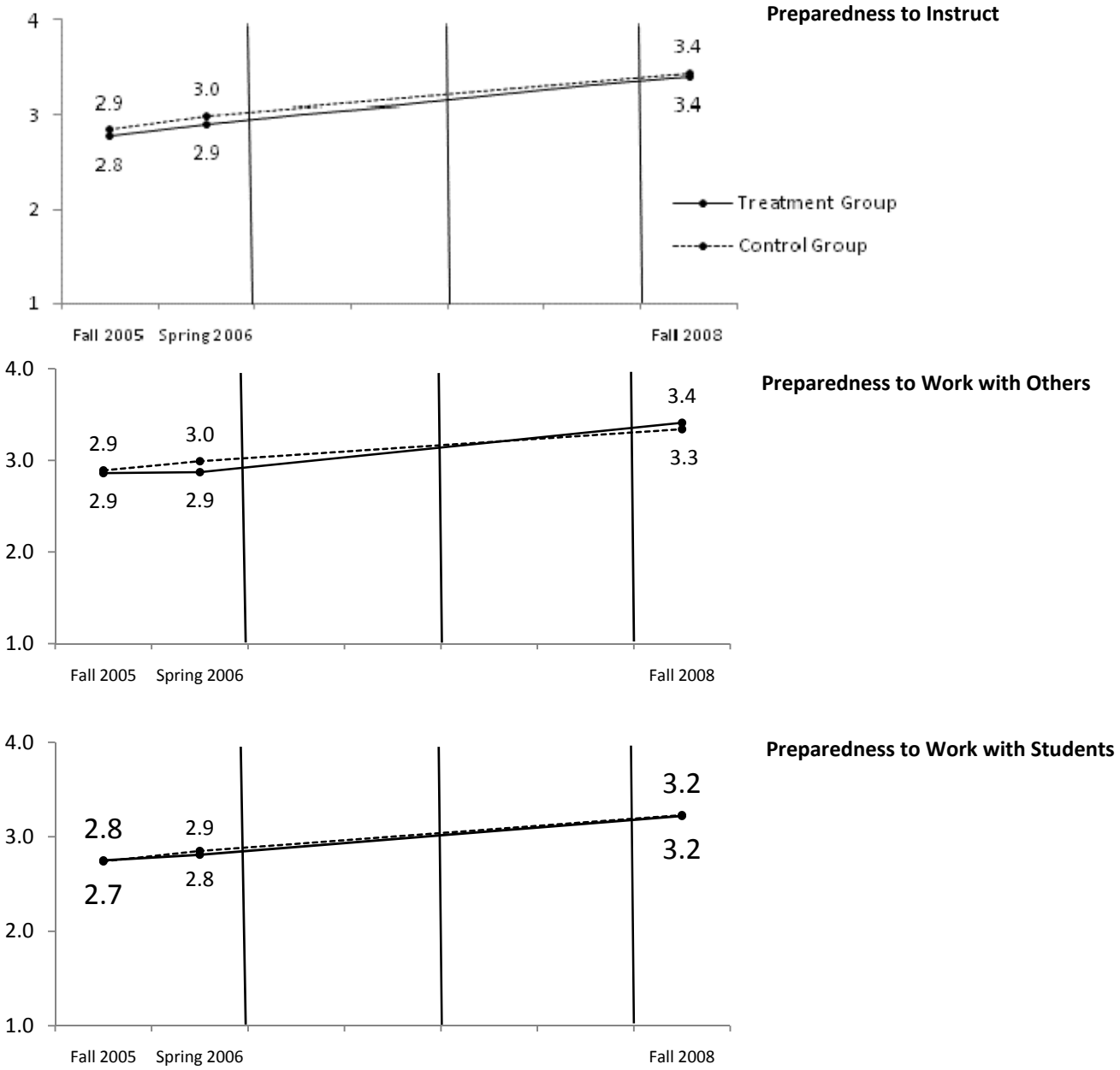


Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in all two-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Satisfaction scale: (1) very dissatisfied, (2) somewhat dissatisfied, (3) somewhat satisfied, or (4) very satisfied. Sample sizes vary due to item nonresponse.

Treatment-control differences are not significantly different from zero at the 0.05 level (N = 391 teachers in fall 2005, 384 teachers in spring 2006, 359 teachers in fall 2006, 370 teachers in spring 2007, 321 teachers in fall 2007, and 318 teachers in fall 2008).

Figure VI.3. Impacts on Teacher Preparedness (Scores on a Four-Point Scale): One-Year Districts

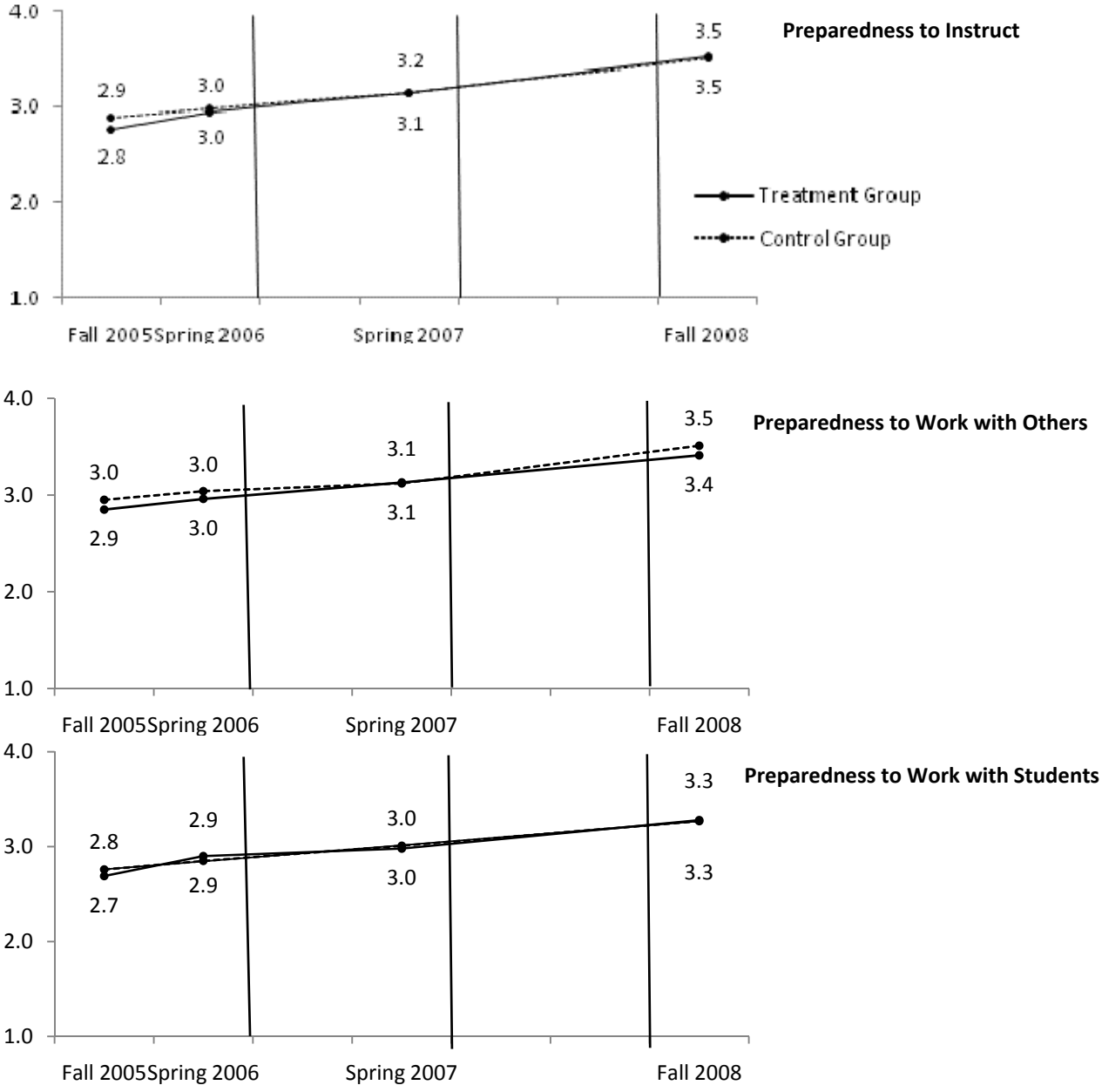


Source: Mathematica First, Second, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Preparedness scale: (1) not at all prepared, (2) somewhat prepared, (3) well prepared, or (4) very well prepared. Sample sizes vary due to item nonresponse.

Treatment-control differences are not significantly different from zero at the 0.05 level (N = 501 teachers in fall 2005, 493 teachers in spring 2006, and 386 teachers in fall 2008).

Figure VI.4. Impacts on Teacher Preparedness (Scores on a Four-Point Scale): Two-Year Districts



Source: Mathematica First, Second, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Preparedness scale: (1) not at all prepared, (2) somewhat prepared, (3) well prepared, or (4) very well prepared. Sample sizes vary due to item nonresponse.

Treatment-control differences are not significantly different from zero at the 0.05 level except for preparedness to instruct in fall 2005 (N = 394 teachers in fall 2005, 381 teachers in spring 2006, 371 teachers in spring 2007, and 308 teachers in fall 2008).

When we recoded the teacher preparedness indicators at each time point into two categories and examined individual survey items separately, the findings were consistent with the approach taken above. Of the 39 differences examined among teachers from one-year districts (13 measures at three points in time), none were statistically significant. Of the 52 differences examined among teachers from two-year districts (13 measures at four points in time), six differences were statistically significant. Treatment teachers were significantly less likely than control teachers to report being prepared to assess students, select/adapt curriculum and materials, plan effective lessons, and be an effective teacher in fall 2005; less likely than control teachers to report feeling prepared to work effectively with parents in spring 2006; and less likely than control teachers to report feeling prepared to work with other teachers to plan instruction in fall 2008.[47] See Appendix D for detailed information from fall 2008 for one-year districts (Table D.7) and two-year districts (Table D.8).[48]

B. Impact Findings: Teacher Retention

1. No Impact on Retention Rates

Neither exposure to one year nor exposure to two years of comprehensive induction had a significant impact on teacher retention over the first four years of the teachers' careers. We surveyed teachers annually to learn whether they had remained in their original school, remained in their original school district, or remained in the teaching profession. We use the terms *school stayer* and *district stayer* to refer to teachers who remained in their original school or district, respectively, after three years. (All school stayers are, by definition, district stayers as well.) We use the terms *mover* and *leaver*, respectively, to refer to teachers who, after three years, left the district but remained in teaching or were no longer teaching.

Figure VI.5 illustrates the lack of significant treatment-control differences. It shows a set of survival curves that plot the percentage of teachers retained in the district (Panel A) and in the profession (Panel B) in each year of the study for the one-year districts. By the end of the study period, 69 percent of teachers in one-year districts remained in their original district and 87 percent were still teaching. Differences between the treatment group, represented by a solid line, and the control group, represented by a dashed line, were statistically insignificant at each time point for both types of retention.

Figure VI.6 shows the same survival curves for two-year districts. The retention rates by the end of the study period were 63 percent in the district and 85 percent in teaching for two-year districts, but the treatment-control differences were not statistically significant.

When we conducted a similar analysis of retention in the same school, we found a similar pattern for both one- and two-year districts; those results are shown in Appendix D, Tables D.9 and D.10, which include more detail, including sample sizes, for all the final retention rates. Detailed data on retention rates from the first two years of the study period can be found in earlier reports (Glazerman et al. 2008; Isenberg et al. 2009).

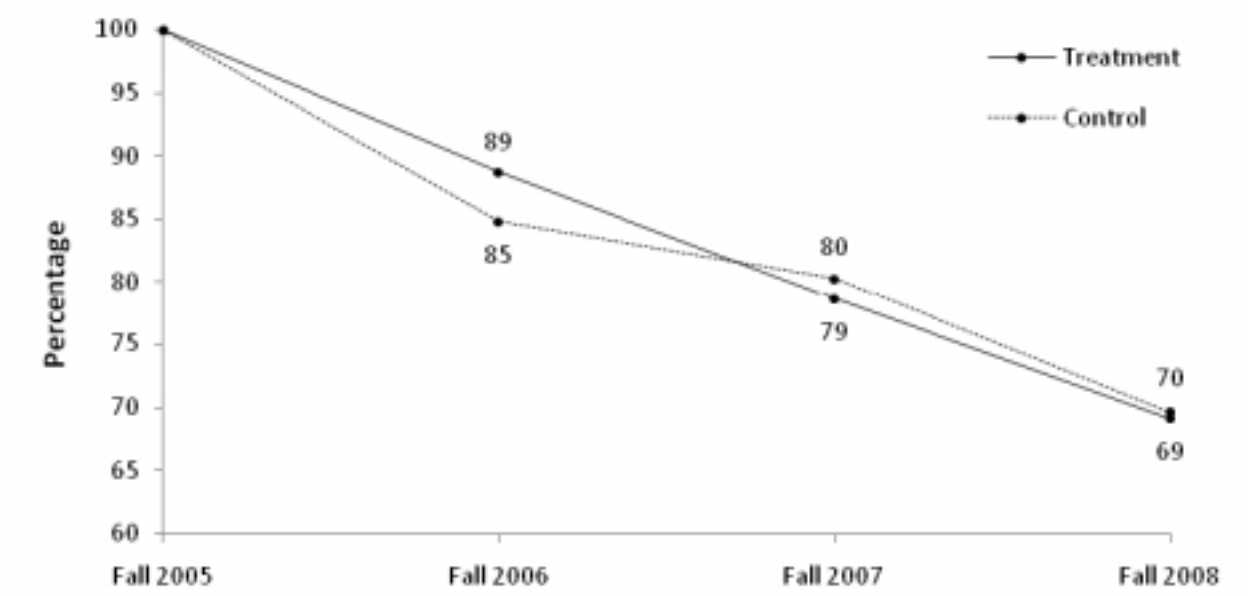
---

[47] See Chapter II for a discussion of multiple comparisons and false discoveries.

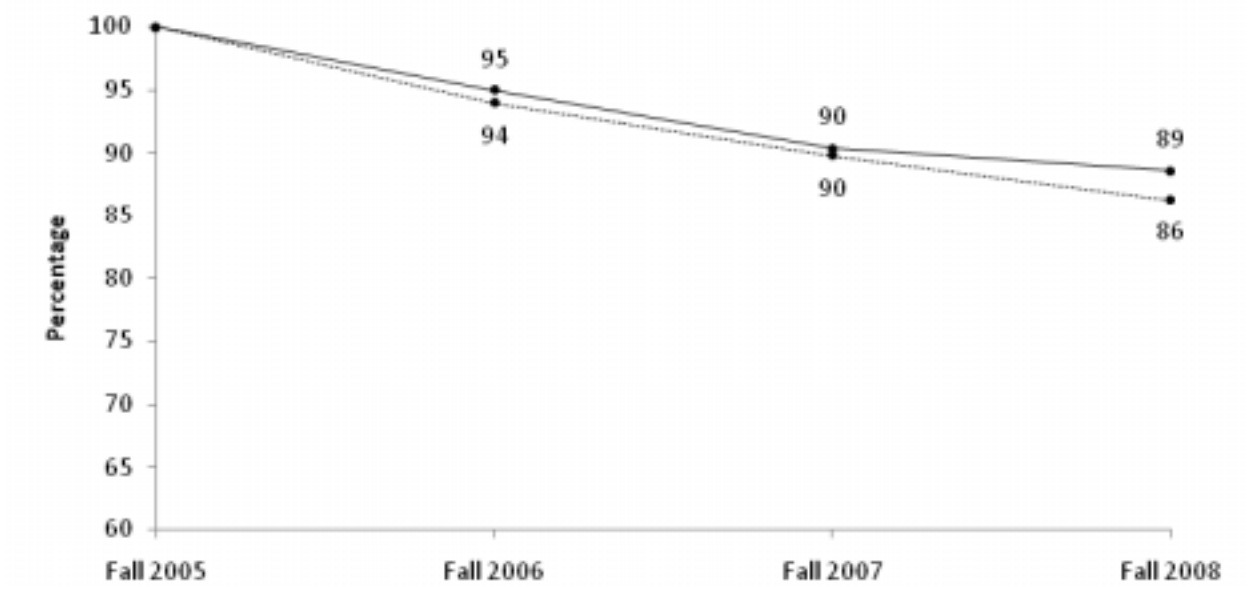
[48] The item-specific impacts for fall 2005 and spring 2006 can be found in Glazerman et al. (2008). The item-specific impacts for spring 2006 and spring 2007 can be found in Isenberg et al. (2009).

Figure VI.5. Survival Curves for One-Year Districts

Panel A. Percentage Remaining in the District



Panel B. Percentage Remaining in the Profession



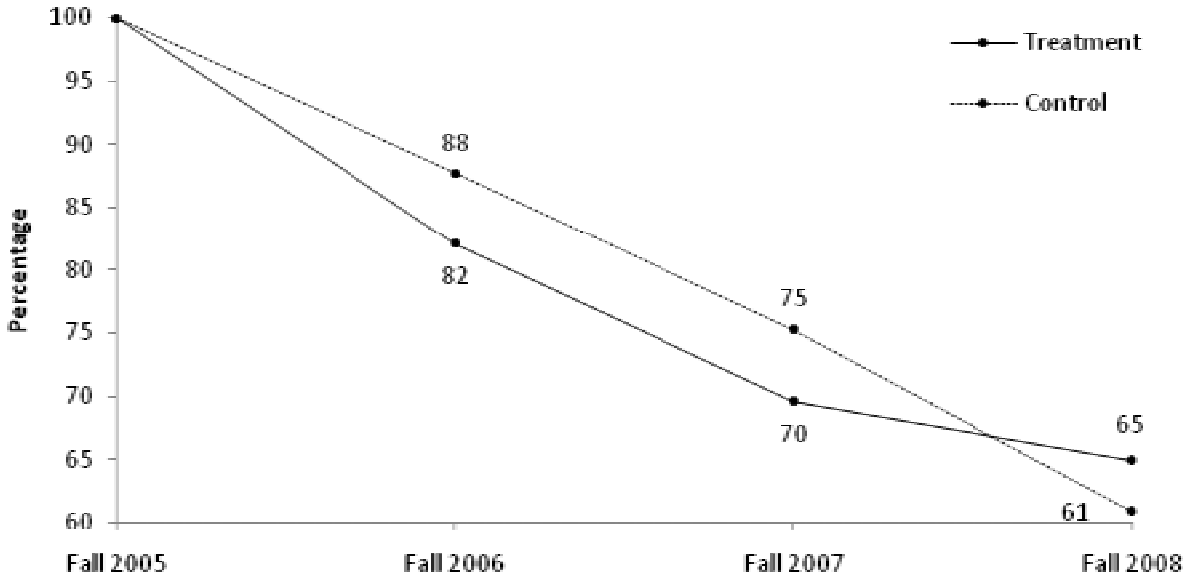
Source: Mathematica First, Second, and Third Teacher Mobility Surveys administered in fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study.

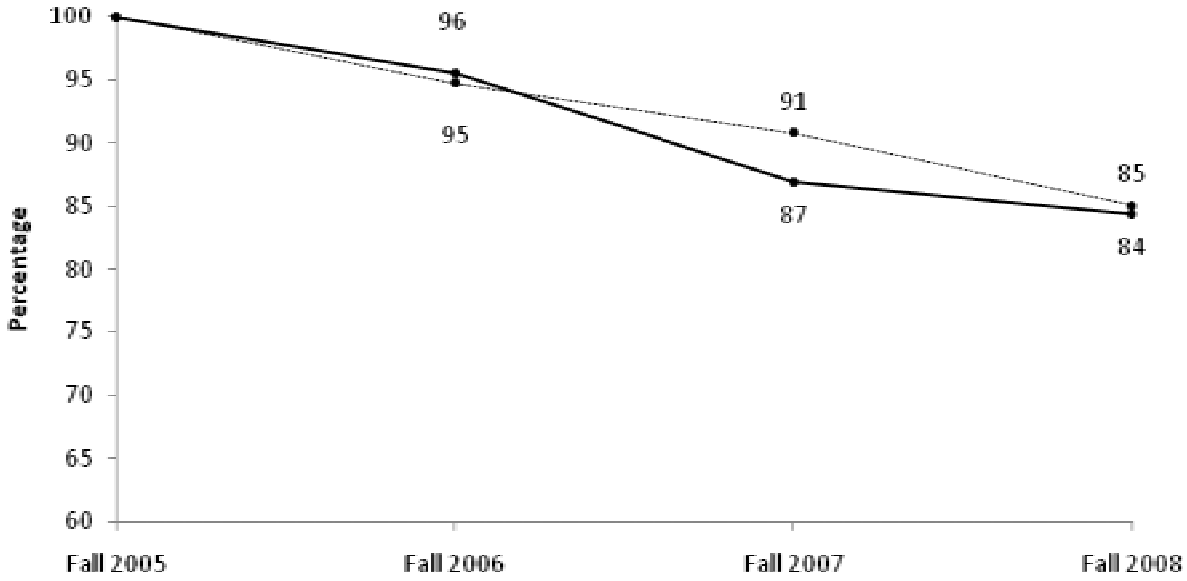
Treatment-control differences are not significantly different from zero at the 0.05 level (N = 561 teachers in fall 2005, 500 teachers in fall 2006, 476 teachers in fall 2007, and 417 in fall 2008 for Panel A and 464 teachers in fall 2008 for Panel B).

Figure VI.6. Survival Curves for Two-Year Districts

Panel A. Percentage Remaining in the District



Panel B. Percentage Remaining in the Profession



Source: Mathematica First, Second, and Third Teacher Mobility Surveys administered in fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study.

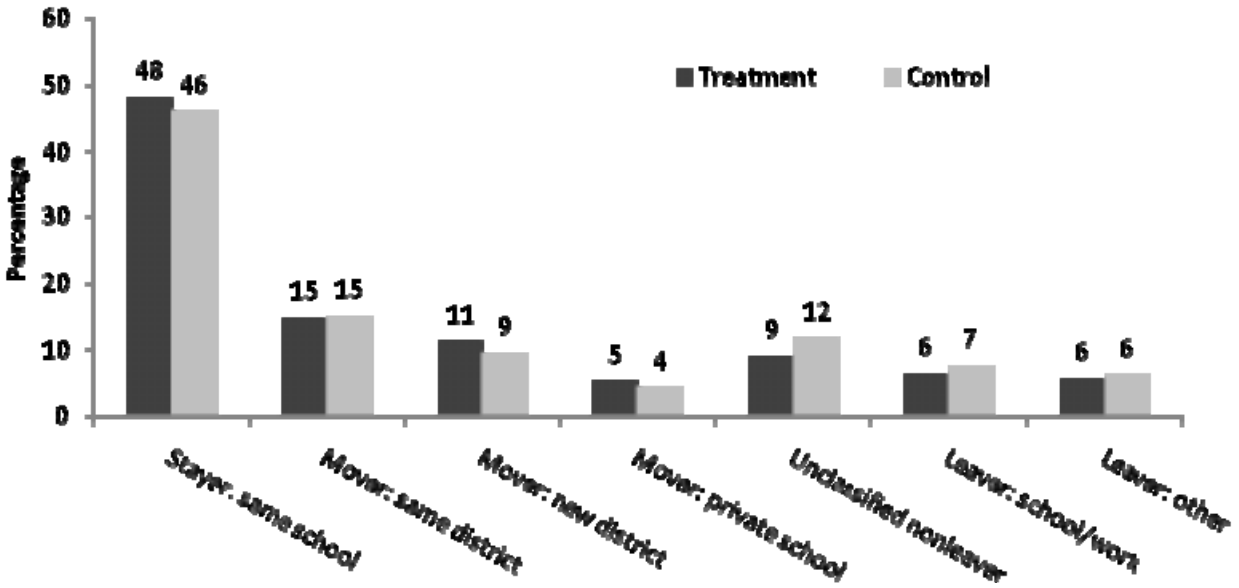
Treatment-control differences are not significantly different from zero at the 0.05 level (N = 448 teachers in fall 2005, 382 teachers in fall 2006, 364 teachers in fall 2007, and 345 teachers in fall 2008 for Panel A and 375 teachers in fall 2008 for Panel B).

2. No Impact on Mobility Patterns

When we examined in more detail where the movers and leavers went, we still did not find significant differences between treatment and control group mobility patterns. Figures VI.7 and VI.8 show the mobility outcomes in detail by treatment status for one-year and two-year districts, respectively. The hypothesis that the distributions were independent of treatment status could not be rejected (p-value = 0.929 for one-year districts and 0.447 for two-year districts), suggesting that the treatment-control differences were not statistically significant.

Statistically similar percentages of treatment and control teachers (about 47 percent) stayed in the same school in one-year districts, whereas 15 percent moved within the district, 10 percent moved to a new district (including charter schools), and 5 percent moved to private schools. Both groups included sample members we call *unclassified nonleavers* who were teaching but provided insufficient information for us to determine where they were (about 11 total). Similar percentages of treatment and control teachers left teaching to take another job or attend school or to do something else, such as work in the home. Results for two-year districts are shown in Figure VI.8.

Figure VI.7. Detailed Mobility Status, Teachers in One-Year Districts

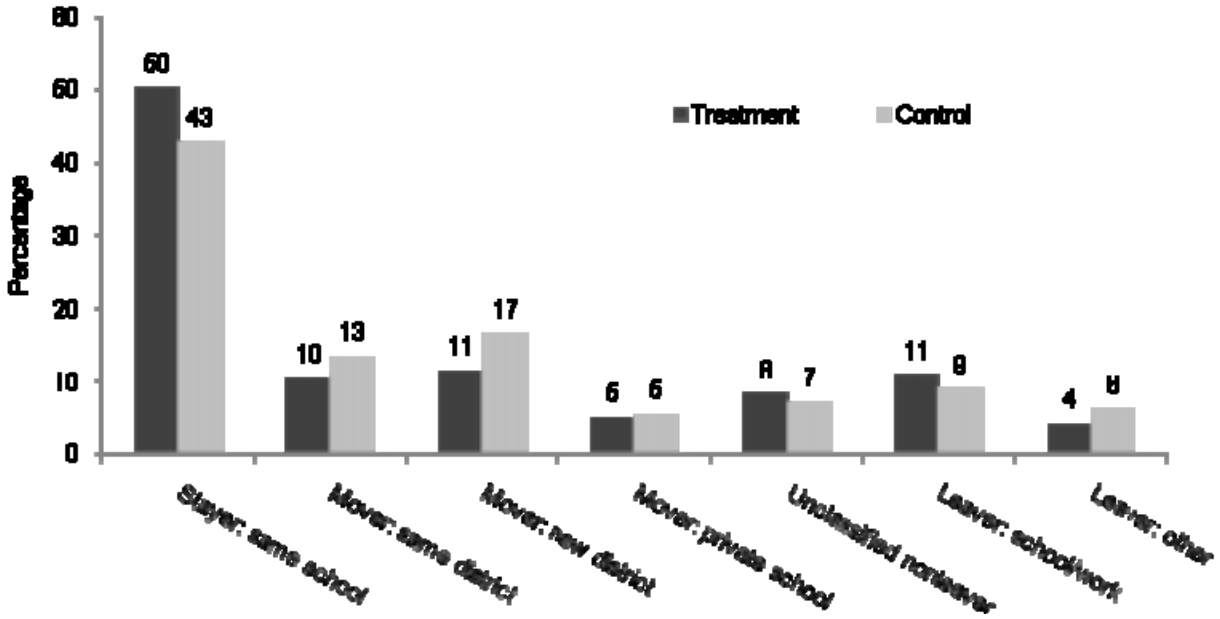


Source: Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Sample sizes are 239 treatment teachers and 230 control teachers.

Treatment-control differences are not significantly different from zero at the 0.05 level.

Figure VI.8. Detailed Mobility Status, Two-Year Districts



Source: Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Sample sizes are 212 treatment teachers and 171 control teachers.

Treatment-control differences are not significantly different from zero at the 0.05 level.

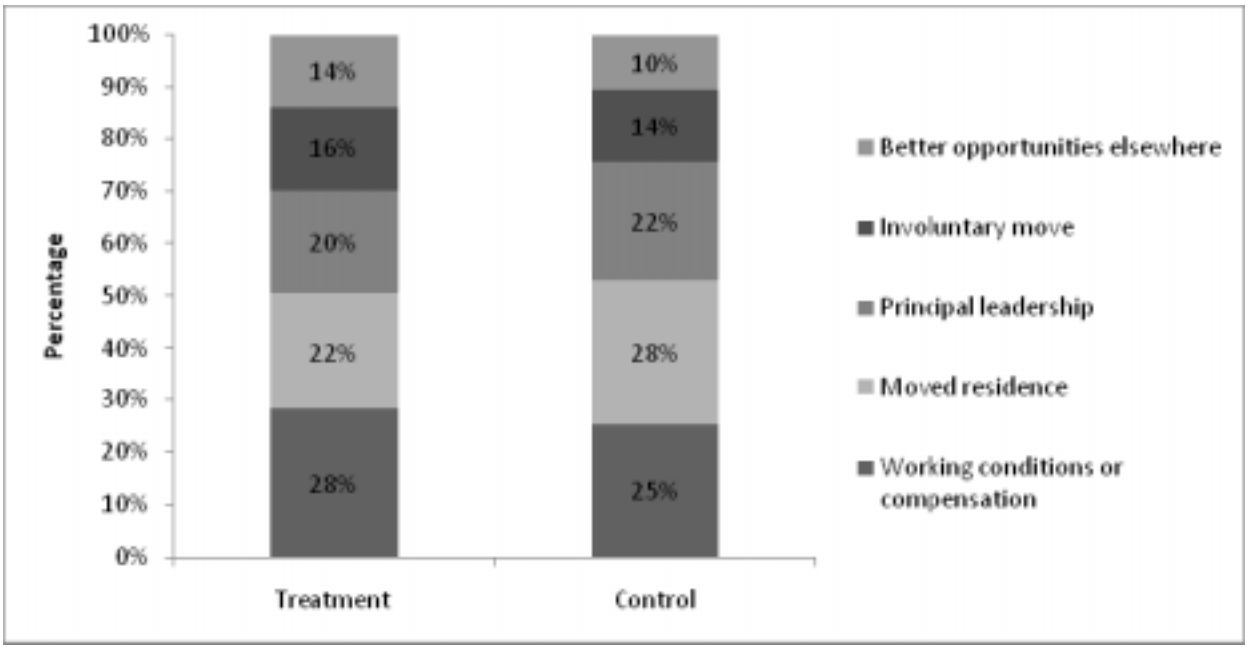
We also anticipated the possibility that leavers might return to teaching. Therefore, the mobility survey asked respondents when, if ever, they expected to return and, if so, with how much certainty. We did not find significant differences between the treatment and control leavers in their expectations about returning to teaching. The average probability that teachers in one-year districts reported for expecting to return to the profession was 60 percent for control leavers and 56 percent for treatment leavers, a difference that was not statistically significant (p-value = 0.755).[49] The corresponding numbers for two-year districts were 53 percent of control leavers and 64 percent of treatment leavers, a difference that was also not significant (p-value = .486).[50]

Next, we examined the reasons teachers gave for moving to a teaching position outside their original school or for leaving the profession and again found no significant differences. Figure VI.9 lists the reasons that teachers gave for moving, indicating which reasons were cited as the most important. Figure VI.10 presents the reasons cited by those who left teaching. In both cases, the

[49] The analysis is based on 21 control and 15 treatment leavers.

[50] The analysis is based on 11 control and 16 treatment leavers.

Figure VI.9. Self-Reported Reasons for Changing Schools

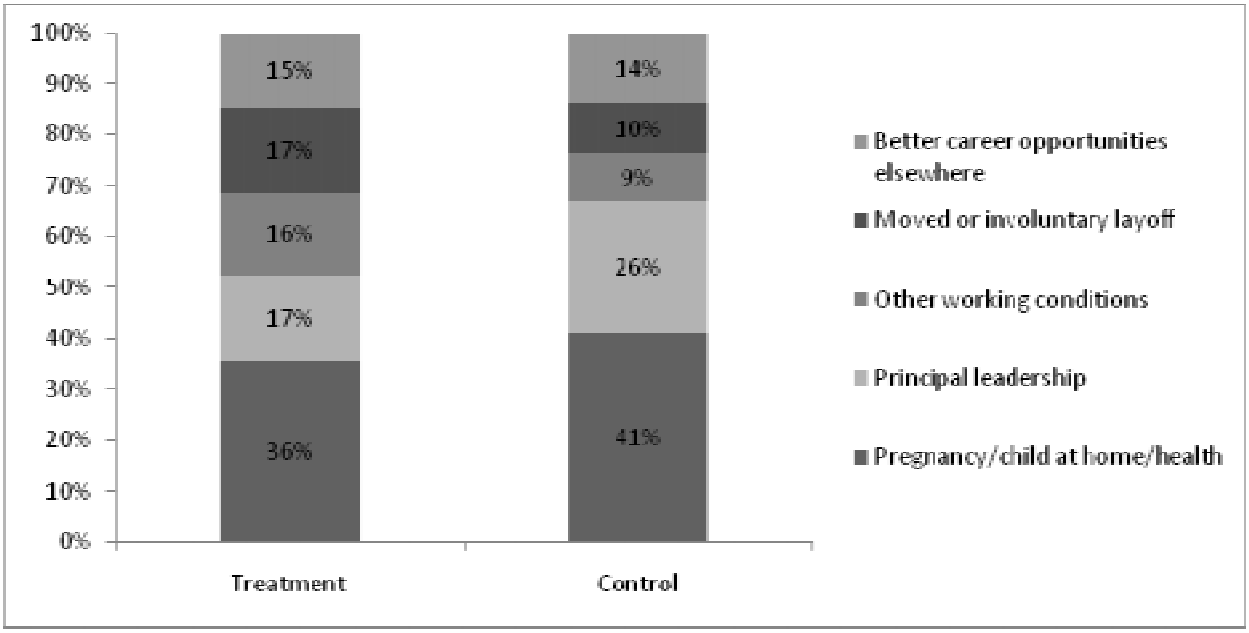


Source: Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

Note: Sample sizes are 113 treatment teachers and 114 control teachers.

Treatment-control differences are not significantly different from zero at the 0.05 level.

Figure VI.10. Self-Reported Reasons for Leaving Teaching



Source: Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

Note: Sample sizes are 48 treatment teachers and 49 control teachers.

Treatment-control differences are not significantly different from zero at the 0.05 level.

reasons were not significantly different between the treatment and control groups. Figures VI.9 and VI.10 show the data combined for one-year and two-year districts. The findings within district type (not shown) tell a similar story.

In order to go beyond self-reported reasons, we examined teacher mobility outcomes directly to identify the factors statistically associated with retention. Assignment to treatment (comprehensive teacher induction) did not explain teachers' decisions to leave the district or to leave teaching. A fuller discussion of these findings is presented in Chapter VII.

We conducted sensitivity analyses to confirm that the findings of no impacts on retention rates were robust. One concern with data that rely on longitudinal surveys is the possibility of systematic bias introduced by nonresponse. The response rate to the final teacher mobility survey was 85 percent overall, which exceeded the study's original goal of 80 percent, but differed by treatment status (90 percent for treatment teachers and 80 percent for control teachers). If nonrespondents are more likely than respondents to be leavers, then one might expect this differential nonresponse rate to translate into a negative bias in the estimated impact of treatment on retention. The size of this bias can be bounded by assuming that all nonrespondents are leavers or all are stayers. The findings, presented in detail in Appendix D, suggest that we would find a significant impact of treatment only under the most extreme assumptions, such as that 100 percent of nonrespondents to the survey were leavers or movers.

We also dealt with missing data by using information collected by data collectors in the field on the status of respondents. For example, we could code the mobility status of a survey respondent who refused to complete the survey based on where we found that individual (or failed to find the individual) when attempting to complete an in-person interview. When we used these survey field codes to augment the mobility variables from self-reports, we continued to find no impact on teacher mobility. Similarly, we reached the same conclusion when we attempted to augment the reported measures for item nonrespondents, people who gave incomplete responses to individual survey items, by using other information on respondents' records. For example, if a survey respondent failed to indicate whether he or she was currently teaching and did not specify the person's current school, but the respondent followed a survey skip instruction meant for current teachers or responded to an item about current teacher satisfaction, then the augmented variable included the teacher as a stayer.

Other sensitivity analyses focused on the statistical methods. We reestimated the impacts using a set of logistic regressions to control for teacher background characteristics and using a multinomial logit regression, which models mobility as a categorical outcome. In each case, the main findings were confirmed.

C. Composition of the Workforce

We investigated not only the impacts of treatment on retention, but also the impacts on the *composition* of the teaching force in the district. Although staff turnover can be disruptive and costly (Alliance for Excellent Education 2004; National Council on Teaching and America's Future 2003), some turnover is inevitable in teaching, as it is in most professions. A critical question is whether turnover raises quality by encouraging the weakest teachers to leave or lowers it by discouraging the strongest ones from staying (Goldhaber et al. 2007).

To test the impact of treatment on the mix of teachers, we used measures of teachers' professional qualifications, classroom practice ratings from their first year (only for those who taught

reading), and student test data from all three years (only for teachers in tested grades and subjects). The test score data were used to construct measures of treatment versus control teacher effectiveness at raising test scores, sometimes referred to as *value added*. Taken together, these measures can indicate whether there is a change in the mix of teachers that resulted from differential attrition.

The study's random assignment design allowed us to test the effects of treatment on the composition of the district's teaching force by comparing the characteristics of treatment stayers to control stayers. Because treatment assignment was random, the treatment and control teachers should be equivalent, on average, prior to the intervention. After three years, when some teachers had left the district (or left teaching), the average quality and qualifications of both groups of teachers may have changed. We examined the impacts on the teaching workforce in terms of both teachers' professional characteristics and teachers' observed performance in the classroom.

No Impact on Stayers' Background Characteristics. We found that the treatment did not significantly alter the mix of teachers in terms of their professional characteristics. The top panels of Tables VI.1 and VI.2 show average values of several teacher characteristics for stayers by treatment status for one- and two-year districts, respectively. For each teacher characteristic, we tested the hypothesis that comprehensive teacher induction had an impact on the percentage of stayers (people who stayed in the same district) with that characteristic because, from the perspective of a district administrator, the qualifications of those who remain are the most important. We examined qualifications such as SAT scores (or ACT equivalent), selectiveness of the teachers' undergraduate college, and highest degree obtained. We also examined the percentage of teachers who had a college major or minor in an education-related field, the amount of prior student teaching experience, certification status, and the route by which the teacher entered the profession. The tables show that there were no significant differences between the professional background characteristics of treatment stayers and control stayers.

No Impact or Negative Impact on Stayers' Prior Classroom Performance. We hypothesized that comprehensive teacher induction may have both productivity effects (helping teachers improve their practices) and composition effects (improving the mix of teachers by retaining strong teachers and encouraging weaker teachers to leave). In Chapter V, the experimental impacts on classroom practices that were measured in year 1 captured evidence on productivity effects because there had been no year-to-year attrition yet, and the impacts on test scores that were measured in year 3 captured evidence on combined productivity and composition effects. Here we explore pure composition effects by asking whether the performance of treatment teachers on these same outcomes was higher relative to control teachers for those who remained in the district until some later time point. The bottom panels of Tables VI.1 and VI.2 provide this type of evidence for one- and two-year districts, respectively.

The classroom practice measures from year 1 are contrasted for treatment and control teachers who remained in their districts through the end of the study period (the beginning of their fourth year). The test score outcomes from year 3 are contrasted for treatment and control teachers in that particular analysis who remained in their districts through the end of the study period as well.

Table VI.1. Characteristics of District Stayers After Three Years, by Treatment Status (Percentages Except Where Noted): One-Year Districts

Teacher Characteristic	Treatment	Control	Difference	P-value
Background Characteristic				
College entrance exam score (SAT combined score or equivalent)	1040	1013	27	0.325
Attended highly selective college	27.5	27.2	0.3	0.954
Major or minor in education	78.7	80.9	-2.1	0.665
Student teaching experience (weeks)	15.8	15.4	0.4	0.772
Highest degree Is master's or doctorate	22.4	28.2	-5.8	0.311
Entered the profession through traditional four-year program	67.6	58.9	8.7	0.171
Certified (regular or probationary)	94.7	94.7	0.0	0.999
Career changer	14.5	12.8	1.7	0.682
Sample Size (Teachers)	148	139		
Sample Size (Schools)	88	84		
Year 1 Classroom Observation Score (on 1 to 5 scale)				
Content of a literacy lesson	2.3	2.6	-0.3*	0.024
Implementation of a literacy lesson	2.6	2.8	-0.2	0.151
Classroom culture	3.0	3.1	-0.1	0.607
Sample Size (Teachers)	100	94		
Sample Size (Schools)	71	65		
Year 3 Test Scores (standard deviation units)				
Reading	-0.27	-0.28	0.02	0.764
Math	-0.26	-0.11	-0.15*	0.008
Sample Size (Teachers)	25	36		
Sample Size (Schools)	23	29		

Source: Mathematica analysis using data from the College Board and ACT, Inc.; data from the 2006-2007 and 2007-2008 school years provided by participating school districts; Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted to account for the study design. The analysis of college entrance exam scores relied on a smaller sample (84 treatment and 86 control teachers and 61 treatment and 62 control schools). The analysis of Year 3 Test Scores relied on a different sample for reading (26 treatment and 34 control teachers and 24 treatment and 27 control schools) and math (per table values).

*Significantly different from zero at the 0.05 level.

The findings suggest a mix of neutral and negative composition effects of comprehensive induction. Of the classroom practice measures, there was a significant negative impact of treatment on the average rating of teachers' lesson content in one-year districts (difference = -0.3, p-value = 0.024), but no significant differences between treatment and control stayers in the other two classroom practice measures or in the two-year districts.

Table VI.2. Characteristics of District Stayers After Three Years, by Treatment Status (Percentages Except Where Noted): Two-Year Districts

Teacher Characteristic	Treatment	Control	Difference	P-value
Background Characteristic				
College entrance exam scores (SAT combined score or equivalent)	905	935	-30	0.330
Attended highly selective college	23.7	21.4	2.3	0.703
Major or minor in education	67.8	66.6	1.2	0.874
Student teaching experience (weeks)	12.3	12.3	0.1	0.975
Highest degree Is master's or doctorate	16.7	10.2	6.5	0.196
Entered the profession through traditional four-year program	61.1	66.4	-5.4	0.443
Certified (regular or probationary)	95.8	92.9	2.9	0.366
Career changer	17.1	11.7	5.4	0.292
Sample Size (Teachers)	124	93		
Sample Size (Schools)	67	52		
Year 1 Classroom Observation Score (on 1 to 5 scale)				
Content of a literacy lesson	2.4	2.4	0.0	0.690
Implementation of a literacy lesson	2.7	2.6	0.1	0.583
Classroom culture	3.1	3.1	0.1	0.624
Sample Size (Teachers)	87	62		
Sample Size (Schools)	50	41		
Year 3 Test Scores (standard deviation units)				
Reading	-0.23	-0.27	0.05	0.302
Math	-0.12	-0.24	0.11	0.054
Sample Size (Teachers)	31	16		
Sample Size (Schools)	21	14		

Source: Mathematica analysis using data from the College Board and ACT, Inc.; data from the 2006-2007 and 2007-2008 school years provided by participating school districts; Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted to account for the study design. The analysis of college entrance exam scores relied on a smaller sample (56 treatment and 47 control teachers and 40 treatment and 35 control schools). The analysis of Year 3 Test Scores relied on a different sample for reading (33 treatment and 17 control teachers and 24 treatment and 15 control schools) and math (per table values).

None of the differences is statistically significant at the 0.05 level.

The estimates of impact on test score outcomes for stayers only were qualitatively different from the estimates derived for the whole sample of teachers who finished the third year. In Chapter V we showed that test score impact estimates were statistically insignificant in one-year districts in both subjects and positive and significant in two-year districts in both subjects. When we condition on teachers' returning for a fourth year, we find that the impact on math in one-year districts is negative and significant (difference = -0.15, p-value = 0.008) and the impact on reading in one-year districts is still insignificant (p-value = 0.764). Furthermore, the impacts on both math and reading scores, which had been significant for two-year districts, were no longer significant when we restricted the analysis to the district stayers (p-value = 0.302 for reading; p-value = 0.054 for math). Because the conclusions change, one might suspect a negative composition effect. However, statistical hypothesis testing of the differences between the two sets of findings suggests that these

apparent negative composition effects may in fact be due to chance variations in the teacher sample and not necessarily a true effect of the treatment.

Taken together, the findings on composition effects suggest that comprehensive teacher induction did not significantly improve teacher quality in the district. We reached the same conclusion when we used alternative methods for constructing hypothesis tests.[51] Sensitivity tests are presented in Appendix D.

[51] For example, rather than test the significance of the coefficient estimate for the treatment variable in a regression model predicting each characteristic, we also tested whether the interaction of treatment status and stayer status was a significant predictor, beyond stayer status and treatment status alone, of the variables listed in each row of Tables VI.1 and VI.2. These tests confirmed the results shown earlier.

This page is intentionally left blank.

VII. CORRELATIONAL ANALYSES

We have shown that the treatment and control groups in both one-year and two-year districts were equivalent on baseline characteristics (Chapter II) and then were exposed to different levels of beginning teacher support during their first two years (Chapter IV). We also showed, however, that treatment-control differences in induction services did not translate into a robust finding of positive impacts across both one-year and two-year districts as hypothesized in the conceptual framework in Figure I.1. By the end of Year 3, comprehensive induction showed no impacts on teacher attitudes or retention across one-year and two-year districts. The impact estimates for the benchmark test score sample showed mixed results: no impacts at the end of year 3 in one-year districts and positive impacts in two-year districts (Chapter V). For those continuing to the beginning of year 4, the benchmark impact estimates were statistically insignificant, or in the case of math scores in one-year districts, negative and significant (Chapter VI).

This chapter attempts to answer a new set of questions raised by these findings. First, we ask whether there is a relationship between these outcomes and the level or intensity of induction services more generally, even when there are no impacts associated with the treatment-control contrast ("comprehensive" versus "prevailing" induction services). Here we discuss how the full range of variation in induction activities was related to teacher attitudes, student test scores, and teacher retention, both within and between treatment and control groups.[52]

Second, we report on whether better outcomes are associated with matching between the mentor and mentee on two dimensions, race/ethnicity and grade. We conducted this analysis using the treatment group, which is the part of the sample for which we have detailed information on mentor background.

The results presented in this chapter should be interpreted with caution because the relationships we describe are correlational and not necessarily causal. Unlike with the randomized experiment that used variation in treatment status, the variation in induction services that we explore here can be caused by confounding factors that also explain teachers' workforce attachment and effectiveness in the classroom and is therefore nonexperimental. In particular, a nonexperimental estimate of the association of induction services with outcomes may be spurious, as it will confound the true (causal) impact of mentoring with the effect of the teacher's own ability or motivation. For example, a high level of services for a particular teacher may result from a principal's or mentor's decision to help struggling teachers who would likely have had poor outcomes anyway. Alternately, a high level might be obtained if an assertive, motivated teacher, who would have had positive outcomes regardless, takes the initiative and spends extra time with a mentor.

A. Nonexperimental Methods

Unlike the experimental analysis, which used a single variable—assignment to the treatment group—to measure the effect of induction on outcomes, the nonexperimental analysis uses four summary measures of induction support, all of which vary within each treatment group as well as between them. To construct these measures, we considered three primary dimensions on which

[52] For a correlational analysis of classroom practices, see Glazerman et al. (2008).

teacher induction programs vary: the breadth, intensity, and instructional focus of services (Ingersoll and Kralik 2004).[53] We constructed indices corresponding to each of these dimensions and selected the induction service components for each index based on suggestions in the literature (Portner 2005) and the emphasis that ETS and NTC placed on these components in their comprehensive induction programs (see Chapter IV). We added a fourth measure, a count of the number of years that beginning teachers had a mentor assigned to them within their first two years. The measures of internal consistency (Cronbach's alpha) for the three index variables range from 0.26 to 0.54. Lower internal consistency measures suggest that indices may measure multidimensional constructs, so the results presented in this chapter should be interpreted with caution. Details on the statistical properties of the variables are given in Table A.5 in Appendix A.

We constructed an **Induction Breadth Index** by counting how many of the following three activities the teacher reported in the three months prior to each survey, on average:

- Met with a literacy or math coach
- Worked with a study group (with new or both new and experienced teachers)
- Observed others teaching

We did this for three points in time: fall 2005, spring 2006, and fall 2006, and computed a weighted average of the values from the three reports that placed twice as much weight on the fall 2006 measure because it was the only measure for all teachers from their second year of teaching. This was equivalent to assuming that spring 2007 measures were identical to fall 2006 measures. For example, a teacher with no activities in fall 2005, two activities in spring 2006, and three activities in fall 2006 would have an Induction Breadth Index of 2.0, which equals the average over four time points of zero, two, and two times three. The breadth index ranges from 0.0 (never received any of the supports) to 3.0 (reported receiving all three supports at each of the three time points). The average value of the breadth index was 1.8 activities and the standard deviation was 0.7.

We constructed an **Instructional Focus Index** by examining another set of indicator variables at the same three time points averaged in the same way. These indicators measure whether the beginning teacher received:

- Suggestions from a mentor to improve his/her practices during the most recent full week of teaching
- A "moderate amount" or "a lot" of guidance in subject area content during the prior three months[54]
- Feedback on teaching, whether or not as part of a formal evaluation, during the prior three months

[53] Additional dimensions include the types of teachers served by a program (new to teaching or new to a school) and the process for selecting and training mentors.

[54] This variable was constructed using a survey question on math content if the outcome to be analyzed is math scores, literacy content if the outcome is reading scores, and math or reading if the outcome is teacher attitudes or mobility.

Because the question on subject area guidance (the second item above in the Instructional Focus Index) was not included in fall 2006 survey, the maximum score for the second year is 2.0 (not 3.0) and the index ranges from 0.0 to 2.5 (not 3.0). The index can be interpreted as measuring the strength of instructional support received by beginning teachers or more literally the number of instructionally focused supports reported in a given period. The average value was 1.6 supports and the standard deviation was 0.6.

For program duration and intensity, we constructed an **Induction Intensity Index** by averaging the number of hours per week[55] that beginning teachers reported spending in the following four activities in the fall 2005, spring 2006, and fall 2006:

- Mentoring sessions (both scheduled and informal)
- Being observed teaching by a mentor
- Professional development (for example, in-service workshops, study groups, seminars, and continuing education courses) learning instructional techniques and strategies
- Professional development learning content area knowledge, specifically language arts, math, and science

Since the questions on time spent on professional development activities were not included in the fall 2006 survey, the index only counts professional development in the first year and takes on values from 0 to the maximum reported time of 5.7 hours per week. The average value was 0.36 hours (21.5 minutes) per week, with a standard deviation of 0.36. To test the robustness of the main findings we also examined variations on this index where we used subsets of the activities, such as hours of informal time with a mentor (by itself) and total time with a mentor (by itself).

Finally, we used an indicator of whether the beginning teacher was **assigned a mentor** in the fall 2005, spring 2006, and fall 2006 (3 items) to create a variable that reflects the number of years (0, 1, or 2) the beginning teacher had an assigned mentor. As with the other indices, we averaged the first two time points to measure the first year and used the third time point to measure the second year. The average value of this measure was 1.2 years, standard deviation of 0.6. To test the robustness of the findings to different definitions of having a mentor we also examined a similar measure that used any mentor (assigned or not) and the number of mentors, since survey respondents were asked how many mentors they had at each time point.

The measures of induction support are closely related to each other and do not represent completely independent information. Table VII.1 shows the correlations among the four induction support variables. The correlations are all positive and range from 0.13 to 0.53, meaning that teachers who reported a greater breadth of induction supports also tended to report a greater intensity of support, instructional focus, and were more likely to report having an assigned mentor. The implication for our analysis is that it can be difficult to distinguish the unique effects of each dimension of teacher induction from their joint effect. Therefore we estimated the regressions in

[55] Time spent in mentoring sessions is measured during a typical week; time spent being observed by a mentor is measured during the most recent full week of teaching; time spent in the two types of professional development activities is measured during a three-month period. For the Induction Intensity Index, the professional development measures are converted to a weekly equivalent and added to the first two measures.

two ways. One way was to include all the induction measures in a single regression model. This has the advantage of allowing us to interpret the estimates as the unique contribution of each dimension, but the disadvantage of statistical imprecision, which makes it more likely that we falsely declare a true relationship to be insignificant. The other way we estimated the model was to conduct separate regressions for each dimension of induction support entered on its own. This approach provides us with more precise estimates, but ones that may conflate the dimensions. We report results from both approaches as well as the result of a hypothesis test of whether the four measures entered together are jointly statistically significant, which indicates whether or not there is an overall impact of induction supports on a particular outcome. Appendix A provides more details of the statistical model. We conducted a number of sensitivity analyses using alternate constructions of the indices and specifications of the regression model.

Table VII.1. Summary Measures of Induction Support

Statistic/Induction Measure	Breadth (Number of Activities)	Instructional Focus (Number of Supports)	Intensity (Hours per Week)	Assigned Mentor (Years)
Correlations				
Breadth	1.00			
Instructional Focus	0.314	1.00		
Intensity	0.163	0.435	1.00	
Assigned Mentor (Years)	0.127	0.530	0.435	1.00
Summary Statistics				
Minimum	0.0	0.0	0.00	0.0
Mean	1.8	1.6	0.36	1.2
Maximum	3.0	2.5	5.65	2.0
Standard Deviation	0.7	0.6	0.36	0.6
Observations (Teachers)	959	965	965	901

Source: Mathematica First, Second, and Third Induction Activities Surveys administered in fall 2005, spring 2006, and fall 2006 to all study teachers.

Note: Breadth is the average number of these activities over a three month period: (1) met with a literacy or math coach, (2) met with a study group, and (3) observed others teaching.

Instructional Focus is the average number of these three supports: (1) suggestions from a mentor to improve his/her teaching, (2) at least a moderate amount of guidance in subject area content, and (3) feedback on teaching.

Intensity is the average number of hours per week that beginning teachers reported spending: (1) in mentoring sessions, (2) being observed teaching by mentor, (3) in professional development learning instructional techniques and strategies, and (4) in professional development learning content area knowledge, specifically language arts, math, and science.

Race/ethnicity match and **grade match** were measured as dichotomous variables indicating whether the mentor and mentee matched on these characteristics. Because we have data on the characteristics of the mentors for the treatment group only, these measures were constructed only for the treatment group. The mentor survey and teacher background surveys were administered in fall 2005. We collapsed race/ethnicity into the following categories: white, African-American, and Hispanic. For these analyses, we replaced assignment to treatment status with a variable indicating with the mentor and mentee were of the same race/ethnicity and whether they taught the same grade in the first year of the study.

Although the induction activities measures are drawn from the first two years of the study, the outcome measures generally pertain to the third and fourth (final) year of the study. The student achievement results measure outcomes for students taught by study teachers in the 2007-2008 school year. Teacher attitudes reported here were measured in fall 2007, after the start of the teachers' third year, and fall 2008, after the start of the teachers' fourth year. Teacher retention was measured in fall 2008. It is important to recognize that all analyses pertain to teachers who persisted in the profession until at least the beginning of the second year of the study so we could observe two years of induction activities.[56]

The analyses use the same regression methods as the experimental analyses presented in Chapters V and VI, but instead of assignment to treatment status as the key explanatory variable, we used measures, described above, of induction support or matching between mentor and beginning teacher. For the first set of analyses, the key explanatory variables were the three indices of induction services and the number of years the teacher had an assigned mentor. Each regression model includes control variables that match the covariates used in the corresponding experimental analysis for the outcome under study (see Appendix A).

B. Nonexperimental Results

Teachers Receiving More Induction Support Also Reported Higher Satisfaction. As reported in Table VII.2, beginning teachers who received more induction support reported being more satisfied, on average, than those who received less. The four induction measures were collectively related to satisfaction with school (p-value = 0.02) and satisfaction with class (p-value = 0.01), but not satisfaction with the teaching career (p-value = 0.07).

Induction intensity and instructional focus stood out as the two aspects of support that were positively related to teacher attitudes. (Breadth of induction had a negative relationship with satisfaction, but was not significant when estimated on its own). Adding one instructionally focused support (among those in the instructional focus index) over a three-month period is associated with an increase in satisfaction of 0.13 points on a four point scale of satisfaction with school, 0.09 points for satisfaction with class, and 0.09 points for satisfaction with career. One way to describe the size of these relationships is to note that the estimate of 0.13 for satisfaction with school is equivalent to moving 13 percent of the way between "somewhat satisfied' and "very satisfied." An increase in one hour per week with a mentor in any of the activities in the induction intensity index is associated with increases in reported satisfaction of 0.04 for each of the three dimensions of satisfaction.

The relationship of induction services to teachers' reported feelings of preparedness exhibited a similar pattern but with fewer statistically significant relationships. We examined the same three dimensions of preparedness that we covered in the experimental analysis: preparedness to work with students, preparedness to instruct, and preparedness to work with other staff. The joint test

[56] We included teachers who did not complete all three induction activities surveys if they had completed the third induction survey in fall 2006 and at least one of the other induction surveys from fall 2005 and spring 2006. If necessary, we imputed missing values from the missing survey using data from the non-missing surveys.

Table VII.2. Association Between Induction Support and Teacher Satisfaction

	Satisfaction with School				Satisfaction with Class				Satisfaction with Teaching Career			
	Joint Model		Separate Regressions		Joint Model		Separate Regressions		Joint Model		Separate Regressions	
Induction Measure	Coef-ficient	P-value	Coef-ficient	P-value	Coef-ficient	P-value	Coef-ficient	P-value	Coef-ficient	P-value	Coef-ficient	P-value
Breadth	-0.01	0.870	0.03	0.366	-0.08*	0.021	-0.05	0.158	-0.02	0.532	0.00	0.893
Instructional Focus	0.11*	0.043	0.13*	0.002	0.09	0.096	0.09*	0.024	0.07	0.179	0.09*	0.029
Intensity	0.02	0.260	0.04*	0.012	0.03	0.104	0.04*	0.011	0.03	0.062	0.04*	0.009
Assigned Mentor (Years)	0.01	0.889	0.07	0.067	0.01	0.900	0.06	0.133	-0.02	0.600	0.04	0.318
Joint Test (all four measures different from zero)	*	0.021			*	0.010				.070		
Sample Size (Schools)		347				348				348		
Sample Size (Teachers)		689				693				695		

Source: Mathematica First, Second, and Third Induction Activities Surveys administered in fall 2005, spring 2006, and fall 2006 to all study teachers.

Note: Breadth is the average number of these activities over a three month period: (1) met with a literacy or math coach, (2) met with a study group, and (3) observed others teaching.

Instructional Focus is the average number of these three supports: (1) suggestions from a mentor to improve his/her teaching, (2) at least a moderate amount of guidance in subject area content, and (3) feedback on teaching.

Intensity is the average number of hours per week that beginning teachers reported spending: (1) in mentoring sessions, (2) being observed teaching by mentor, (3) in professional development learning instructional techniques and strategies, and (4) in professional development learning content area knowledge, specifically language arts, math, and science.

Regressions include pretest, district-by-grade fixed effects, and account for clustering of students within schools.

*Significantly different from zero at the 0.05 level.

suggested a significant relationship with preparedness to work with others (p-value = 0.01), but an insignificant relationship with the other two dimensions. The only one of the four induction services measures that was significantly associated with preparedness to work with others was Induction Intensity, with an association of 0.04 (p-value = 0.01).

None of the findings for teacher attitudes changed when we varied the method for combining information from the four induction activities surveys conducted in the first two years. That is, we applied different weights to the fall of the first year, spring of the first year, fall of the second year, and spring of the second year (which was only measured for teachers in two-year districts). One set of weights counted the first three time points equally. Another counted only the first two time points. Another used all four time points, but used only two-year districts, since the fourth survey was not administered in one-year districts.

Nor did the findings change when we examined teacher attitudes measured in the fourth year as opposed to the same outcomes measured in the third year. The fourth year outcomes are more recent and hence pick up longer term trends, but they also must be interpreted differently because they only pertain to the teachers who remained in the profession through the beginning of the fourth year. One might expect that this sample of stayers excludes those whose induction experiences reduced their satisfaction and consequently influenced their decision to leave.

Increased Induction Support Was Not Associated with Higher Test Scores. We did not find consistent evidence that induction support was related to student achievement in math or reading, as shown in Table VII.3. For math, the joint test using all four induction measures was insignificant. There were three insignificant relationships (breadth, instructional focus, and intensity) and one positive and significant relationship (assigned mentor) when we estimated the effect of the four induction indices on math scores simultaneously. When we entered the indices one at a time, none of the induction measures has a significant association with math scores.

For reading scores, the test of a joint effect of induction measures was not significant and the regression coefficient estimates for all of the individual index measures were statistically insignificant. Taken together, these findings do not present a pattern of significant impacts and therefore do not support the hypothesis that teachers who received more induction services produced higher student test scores.

The relationships between induction services and test scores were also statistically insignificant when we used different modeling assumptions. When we measured induction services received in a variety of different ways and used a variety of model specifications, we did not find evidence of a significant association.

Increased Induction Support Was Not Associated With Higher Retention. We found that none of the four measures of beginning teacher support was related to retention in the district or in the profession (Table VII.4). We conducted many of the same specification checks described above and reached the same conclusion regardless of how we specified the model, defined the variables, or delineated the sample.

Table VII.3. Association Between Induction Support and Test Scores

	Math				Reading			
	Joint Model		Separate Regressions		Joint Model		Separate Regressions	
Induction Measure	Coef-ficient	P-value	Coef-ficient	P-value	Coef-ficient	P-value	Coef-ficient	P-value
Breadth	0.02	0.650	0.01	0.902	-0.04	0.180	-0.04	0.172
Instructional Focus	-0.07	0.126	-0.02	0.622	-0.01	0.834	-0.01	0.787
Intensity	0.00	0.644	0.00	0.546	0.00	0.713	0.00	0.633
Assigned Mentor (Years)	0.09*	0.046	0.06	0.130	0.03	0.448	0.02	0.560
Joint Test (all four measures different from zero)		0.270				0.635		
Sample Size (Teachers)	141				149			
Sample Size (Students)	2,405				2,607			

Source: Mathematica analysis using data from the 2005-2006 and 2006-2007 school years provided by participating school districts; Mathematica First, Second, and Third Induction Activities Surveys administered in fall 2005, spring 2006, and fall 2006 to all study teachers.

Note: Breadth is the average number of these activities over a three month period: (1) met with a literacy or math coach, (2) met with a study group, and (3) observed others teaching.

Instructional Focus is the average number of these three supports over a three month period: (1) suggestions from a mentor to improve his/her teaching, (2) at least a moderate amount of guidance in subject area content, and (3) feedback on teaching.

Intensity is the average number of hours per week that beginning teachers reported spending: (1) in mentoring sessions, (2) being observed teaching by mentor, (3) in professional development learning instructional techniques and strategies, and (4) in professional development learning content area knowledge, specifically language arts, math, and science.

Regressions include pretest, district-by-grade fixed effects, and account for clustering of students within schools.

*Significantly different from zero at the 0.05 level.

Table VII.4. Association Between Induction Support and Teacher Mobility

	Remains in District				Remains in Teaching			
	Joint Model		Separate Regressions		Joint Model		Separate Regressions	
Induction Measure	Effect	P-value	Effect	P-value	Effect	P-value	Effect	P-value
Breadth	0.03	0.343	0.03	0.333	0.00	0.855	0.00	0.964
Instructional Focus	-0.03	0.480	0.01	0.868	0.00	0.644	0.00	0.598
Intensity	0.01	0.312	0.01	0.294	0.00	0.414	0.00	0.207
Assigned Mentor (Years)	0.02	0.675	0.02	0.612	0.01	0.194	0.01	0.119
Joint Test (all four measures different from zero)		0.623				0.488		
Sample Size (Schools)			338				359	
Sample Size (Teachers)			678				750	

Source: **Mathematica Teacher Background Survey administered in fall 2005, Mathematica Third Teacher Mobility Survey administered in fall 2008, and Mathematica First, Second, and Third Induction Activities Surveys administered in fall 2005, spring 2006, and fall 2006 to all study teachers.**

Note: **Breadth is the average number of these activities over a three month period: (1) met with a literacy or math coach, (2) met with a study group, and (3) observed others teaching.**

Instructional Focus is the average number of these three supports over a 3 month period: (1) suggestions from a mentor to improve his/her teaching, (2) at least a moderate amount of guidance in subject area content, and (3) feedback on teaching.

Intensity is the average number of hours per week that beginning teachers reported spending: (1) in mentoring sessions, (2) being observed teaching by mentor, (3) in professional development learning instructional techniques and strategies, and (4) in professional development learning content area knowledge, specifically language arts, math, and science.

Regressions use a logit model to account for baseline characteristics and robust standard errors to account for clustering of teachers within schools. Marginal effects are reported instead of coefficients.

***Significantly different from zero at the 0.05 level.**

Match With Mentor on Race or Grade Was Associated with Negative Retention Outcomes. Beginning teachers who had the same race/ethnicity as their mentor or taught the same grade as had their mentor had lower rates of retention in the district and in the profession than those who did not have such a match, contradicting the hypothesis that better matching would produce better outcomes. The regression-adjusted differences were 17 percent for a match on race/ethnicity and 16 percent for grade level (see Table VII.5). When we examined the other two outcomes, teacher attitudes and student achievement, we found no evidence of a statistically significant relationship with a mentor match.

Table VII.5. Association Between Mentor Match and Attitudes, Test Scores, and Retention

	Outcome						
	Attitudes, Satisfaction with School, Class, and Career			Test Scores		Retention	
	School	Class	Career	Math	Reading	In District	In Teaching
Race/Ethnicity Match							
Coefficient	-0.04	0.12	0.06	0.03	-0.06	-0.17*	-0.06*
P-value	0.722	0.303	0.589	0.768	0.428	0.037	0.006
Sample Size (Schools)	159	160	160	--	--	166	171
Sample Size (Teachers)	341	344	345	73	76	363	396
Sample Size (Students)	--	--	--	1,296	1,415	--	--
Grade Match							
Coefficient	-0.07	-0.01	-0.05	-0.01	-0.08	-0.16*	0.01
P-value	0.311	0.885	0.487	0.841	0.107	0.003	0.533
Sample Size (Schools)	182	183	183	--	--	185	193
Sample Size (Teachers)	376	380	381	79	83	392	431
Sample Size (Students)	--	--	--	1,392	1,549	----	

Source: Mathematica analysis using data from the 2005-2006 and 2006-2007 school years provided by participating school districts; Mathematica First, Second, and Third Induction Activities Surveys administered in fall 2005, spring 2006, and fall 2006 to all study teachers.

Note: Data are regression-adjusted to account for pretest, district-by-grade fixed effects, and clustering of students within schools.

*Significantly different from zero at the 0.05 level.

REFERENCES

Alliance for Excellent Education. *Tapping the Potential: Retaining and Developing High-Quality New Teachers*. New York: Carnegie Corporation, 2004.

Alt, Martha Naomi, and Robin R. Henke. "To Teach or Not to Teach: Teaching Experience and Preparation Among 1992-93 Bachelor's Degree Recipients 10 Years After College." NCES 2007-163. Washington, DC: U.S. Department of Education, National Center for Education Statistics. July 2007.

American Association of State Colleges and Universities. "Teacher Induction Programs: Trends and Opportunities," vol. 3, no. 10, October, 2006.

Angrist, Joshua, Guido W. Imbens, and Donald B. Rubin. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, vol. 91, no. 434, June 1996, pp. 444–455.

Barnes, Gary, Edward Crowe, and Benjamin Schaefer. "The Cost of Teacher Turnover in Five School Districts: A Pilot Study." Washington, DC: National Commission on Teaching and America's Future, 2007.

Barron's Profiles of American Colleges. Hauppauge, NY: Barron's Educational Series, Inc., 2003.

Benjamini, Yoav, and Yosef Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society,* vol. 57, 1995, pp. 289–300.

Berry, Barnett, Peggy Hopkins-Thompson, and Mandy Hoke. "Assessing and Supporting New Teachers: Lessons from the Southeast." Chapel Hill, NC: Southeast Center for Teaching Quality, December 2002.

Borman, Geoffrey, and N. Maritza Dowling. "Teacher Attrition and Retention: A Meta-Analytic and Narrative Review of the Research." *Review of Educational Research,* vol. 78, 2008, pp. 367-409.

Boyd, Donald, Pam Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. "Who Leaves? Teacher Attrition and Student Achievement." NBER Working Paper 14022. Available at [http://www.nber.org/papers/w14022]. Accessed May 2008.

California Commission on Teacher Credentialing. "Preliminary Report on Teacher Retention in California." Sacramento, CA: California Commission on Teacher Credentialing, September 17, 2002.

California Commission on Teacher Credentialing. *California's Standards for the Teaching Profession*. Sacramento, CA: California Commission on Teacher Credentialing, 1997.

Danielson, Charlotte. *Enhancing Professional Practice: A Framework for Teaching*. Alexandria, VA: Association for Supervision and Curriculum Development, 1996.

Darling-Hammond, Linda. "Teacher Quality and Student Achievement: A Review of State Policy Evidence." *Education Policy Analysis Archives*, vol. 8, no. 1, 2000.

Dorans, Neil J., C. Felicia Lyu, Mary Pommerich, and Walter M Houston. "Concordance Between ACT Assessment and Recentered SAT I Sum Scores." *College and University*, vol. 73, 1997, pp. 24-31.

Ferguson, Ronald F., and Helen F. Ladd. "How and Why Money Matters: An Analysis of Alabama Schools." *Holding Schools Accountable*, edited by H.F. Ladd. Washington, DC: Brookings Institution, 1996, pp. 265-298.

Fuller, Edward J. "Beginning Teacher Retention Rates for TxBESS and Non-TexBESS Teachers." Austin, TX: State Board for Educator Certification, 2003.

Fuller, Edward J. "The Cost of Teacher Turnover." Austin, TX: State Board for Educator Certification, November 2000.

Glazerman, Steven, Sarah Dolfin, Martha Bleeker, Amy Johnson, Eric Isenberg, Julieta Lugo-Gil, Mary Grider, and Edward Britton. "Impacts of Comprehensive Teacher Induction: Results from the First Year of a Randomized Controlled Study." NCEE 2009-4034. Washington, DC: U.S. Department of Education, National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, 2008.

Glazerman, Steven M., Sarah Senesky, Neil Seftor, and Amy Johnson. "Design of an Impact Evaluation of Teacher Induction Programs." Princeton, NJ: Mathematica Policy Research, September 2005.

Goldhaber, Dan, Betheny Gross, and Daniel Player. "Are Public Schools Really Losing Their 'Best'? Assessing the Career Transitions of Teachers and Their Implications for the Quality of the Teacher Workforce." Washington, DC: Urban Institute, October 2007.

Greenwald, R., Larry V. Hedges, and Richard D. Laine. "The Effect of School Resources on Student Achievement. *Review of Educational Research*, vol. 66, no. 3, Autumn 1996, pp. 361-396.

Henke, Robin R., Xianglei Chen, and Sonya Geis. "Progress Through the Teacher Pipeline: 1992–93 College Graduate and Elementary/Secondary School Teaching as of 1997." Washington, DC: U.S. Department of Education, National Center for Education Statistics, 2000.

Howard, Tyrone C. "Who Receives the Short End of the Shortage? Implications of the U.S. Teacher Shortage on Urban Schools." *Journal of Curriculum and Supervision,* vol. 18, no. 2, 2003, pp. 142-160.

Huber, Peter J. "The Behavior of Maximum Likelihood Estimation Under Nonstandard Conditions." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1*, edited by L. M. LeCam and J. Neyman. Berkeley. CA: University of California Press, 1967.

Ingersoll, Richard M. "Is There Really a Teacher Shortage?" Seattle, WA: Center for the Study of Teaching and Policy, 2003.

Ingersoll, Richard M. "The Teacher Shortage: A Case of Wrong Diagnosis and Wrong Prescription." *NASSP Bulletin,* vol. 86, no. 631, 2002, pp. 16-31.

Ingersoll, Richard M. "Teacher Turnover and Teacher Shortages: An Organizational Analysis." *American Educational Research Journal,* vol. 38, no. 3, 2001, pp. 499-534.

Ingersoll, Richard M., and Jeffrey M. Kralik. "The Impact of Mentoring on Teacher Retention: What the Research Says." Denver, CO: Education Commission of the States, 2004, pp. 1-23.

Ingersoll, Richard M., and Thomas M. Smith. "Do Teacher Induction and Mentoring Matter?" *NASSP Bulletin*, vol. 88, no. 638, 2004, pp. 28-40.

Ingersoll, Richard, and Thomas M. Smith. "The Wrong Solution to the Teacher Shortage." *Educational Leadership*, vol. 60, 2003, pp. 30-33.

Interstate New Teacher Assessment and Support Consortium. *Model Standards for Beginning Teacher Licensing, Assessment and Development: A Resource for State Dialogue.* Washington, DC: Council of Chief State School Officers, 1992.

Isenberg, Eric, Steven Glazerman, Martha Bleeker, Amy Johnson, Julieta Lugo-Gil, Mary Grider, Sarah Dolfin, and Edward Britton. Impacts of Comprehensive Teacher Induction: Results from the Second Year of a Randomized Controlled Study. NCEE 2009-4072. Washington, DC: U.S. Department of Education, National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, 2009.

Isenberg, Eric, Steven Glazerman, Amy Johnson, Sarah Dolfin, and Martha Bleeker. "Linking Induction to Student Achievement." In *Past, Present, and Future Research on Teacher Induction: An Anthology for Researchers, Policy Makers, and Practitioners,* edited by J. Wang, S.J. Odell, and R.T. Clift. Lanham, MD: Rowman and Littlefield Education, 2010.

Kapadia, Kavita, Vanessa Coca, and John Q. Easton. "Keeping New Teachers: A First Look at the Influences of Induction in the Chicago Public Schools." Chicago, IL: Consortium on Chicago School Research at the University of Chicago, January 2007.

Kauffman, David, Susan Moore Johnson, Susan M. Kardos, Edward Liu, and Heather G. Peske. "'Lost at Sea': New Teachers' Experiences with Curriculum and Assessment." *Teachers College Record,* vol. 104, no. 2, 2002, pp. 273-300.

Kelly, Linda Molner. "Why Induction Matters." *Journal of Teacher Education*, vol. 55, no. 5, winter 2004, pp. 438-448.

King, M. Bruce, and Fred M. Newmann. "Will Teacher Learning Advance School Goals?" *Phi Delta Kappan,* vol. 81, no. 8, 2000, pp. 576-580.

Levin, Jessica, and Meredith Quinn. "Missed Opportunities: How We Keep High-Quality Teachers Out of Urban Schools." New York, NY: New Teacher Project, 2003.

Lopez, Alejandra, Andrea Lash, Monika Schaffner, Patrick Shields, and Mary Wagner. "Review of Research on the Impact of Beginning Teacher Induction on Teacher Quality and Retention." Menlo Park, CA: SRI International, 2004.

Milanowski, Anthony, and Allen Odden. "A New Approach to the Cost of Teacher Turnover." Seattle, WA: Daniel J. Evans School of Public Affairs, April 6, 2007.

Murphy, Patrick, Michael DeArmond, and Kacey Guin. "A National Crisis or Localized Problem? Getting Perspective on the Scope and Scale of the Teacher Shortage." *Education Analysis and Policy Archives,* vol. 11, no. 23, 2003.

National Commission on Teaching and America's Future. "The High Cost of Teacher Turnover." Washington, DC: National Commission on Teaching and America's Future, 2007.

National Commission on Teaching and America's Future. "No Dream Denied: A Pledge to America's Children." Washington, DC: National Commission on Teaching and America's Future, 2003.

New Teacher Center. *Continuum of Teacher Development*. Santa Cruz, CA: New Teacher Center, 2002.

Ng, Jennifer C. "Teacher Shortages in Urban Schools: The Role of Traditional and Alternative Certification in Filling the Voids." *Education and Urban Society*, vol. 35, no. 4, 2003, pp. 380-398.

Odell, Sandra J., and Douglas P. Ferraro. "Teacher Mentoring and Teacher Retention." *Journal of Teacher Education,* vol. 43, no. 3, 1992, pp. 200-204.

Portner, Hal (ed.). *Teacher Mentoring and Induction: the State of the Art and Beyond.* Thousand Oaks, CA: SAGE Publications, 2005.

Rockoff, Jonah. "Does Mentoring Reduce Turnover and Improve Skills of New Employees? Evidence from Teachers in New York City." NBER Working Paper 13868. Available at [http://www.nber.org/papers/w13868]. 2008.

Saginor, Nicole. *Diagnostic Classroom Observation: Moving Beyond Best Practice.* Thousand Oaks, CA: Corwin Press, 2008.

Scafidi, Benjamin, David L. Sjoquist, and Todd R. Stinebrickner. "Do Teachers Really Leave for Higher Paying Jobs in Alternative Occupations?" *Advances in Economic Analysis and Policy,* vol. 6, no. 1, 2006.

Serpell, Zewelanji, and L.A. Bozeman. "Beginning Teacher Induction: A Report on Beginning Teacher Effectiveness and Retention." Washington, DC: American Association of Colleges for Teacher Education, 2000.

Shen, Jianping. "Teacher Retention and Attrition in Public Schools: Evidence From SASS91." *Journal of Educational Research*, vol. 91, November/December 1997, pp. 81-88.

Smith, Thomas M., and Richard M. Ingersoll. "What Are the Effects of Induction and Mentoring on Beginning Teacher Turnover?" *American Educational Research Journal*, vol. 41, no. 3, 2004, pp. 681-714.

Stansbury, Kendyll, and Joy Zimmerman. "Lifelines to the Classroom: Designing Support for Beginning Teachers." San Francisco, CA: WestEd, 2000.

Totterdell, Michael, Lynda Woodroffe, Sara Bubb, and Karen Hanrahan. "The Impact of NQT Induction Programmes on the Enhancement of Teacher Expertise, Professional Development, Job Satisfaction or Retention Rates: A Systematic Review of Research Literature on Induction." London: University of London Institute of Education, EPPI-Centre, 2004.

Tushnet, Naida C., Danielle Briggs, Jeanne Elliot, Camille Esch, Don Havilland, Daniel C. Humphrey, Nada Rayyes, Lori M. Riehl, and Vicki Young. "Final Report of the Independent Evaluation of the Beginning Teacher Assessment Program." San Francisco, CA: WestEd, 2002.

Villar, Anthony, and Michael Strong. "Is Mentoring Worth the Money? A Benefit-Cost Analysis and Five-Year Rate of Return of a Comprehensive Mentoring Program for Beginning Teachers." *ERS Spectrum*, vol. 25, no. 3, 2007, pp. 1-17.

Wechsler, Marjorie E., Kyra Caspary, Daniel C. Humphrey, and Kavita Kapadia Matsko. "Examining the Effects of New Teacher Induction." Menlo Park, CA: SRI International, April 2010.

West, Martin R., and Matthew M. Chingos. "Teacher Effectiveness, Mobility, and Attrition in Florida." In *Performance Incentives: Their Growing Impact on American K-12 Education*, edited by Matthew Springer. Washington, DC: Brookings Institution Press, 2009, pp. 251-271.

White, H. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica,* vol. 48, 1980, pp. 817-830.

Youngs, Peter A. "How Elementary Principals' Beliefs and Actions Influence New Teachers' Experiences." *Educational Administration Quarterly*, vol. 43, no. 1, 2007, pp. 101-137.

Youngs, Peter A. "State and District Policy Related to Mentoring and New Teacher Induction in Connecticut." New York, NY: National Commission on Teaching and America's Future, 2002.

This page is intentionally left blank.

APPENDIX A

SUPPLEMENTAL INFORMATION FOR CHAPTERS II AND III

This appendix provides technical details of the impact estimation method and discusses the test score data in greater depth.

A. Impact Estimation

Basic Model. To estimate the effects of comprehensive teacher induction on outcomes, we implemented a two-level regression model. The first level corresponds to teachers (for the classroom practices, teacher attitudes and retention analyses) and the second level to schools. Treatment effects are estimated in the level two model, in which the sample size is dictated by the number of schools, not teachers. The basic form of the model for the teacher attitudes and retention analyses is presented in Equations (A.1) and (A.2), which express teacher-level analyses (A.1) and school-level analyses (A.2):

$$Y_{ij} = c_j + \beta' X_{ij} + e_{ij} \quad \text{(A.1)}$$

$$c_j = \mu + \delta T_j + \gamma' Z_j + u_j \quad \text{(A.2)}$$

where Y_{ij} is the outcome of interest for teacher *i* in school *j*; c_j is a school-specific intercept; X_u is a vector that includes baseline teacher characteristics; e_{ij} is an independently and identically distributed teacher-level random error term that captures the effects of unobserved factors that influence the outcome; T_j is an indicator that equals 1 if school *j* was randomly assigned to the treatment group (receiving services from one of the two comprehensive induction programs) and equals 0 otherwise; Z_j includes school characteristics; u_j is a random component representing unobserved factors that vary by school (the random "school effect"); and β, μ, δ, and γ are parameters or vectors of parameters to be estimated. We also must estimate the variance of the school effects u_j.

By substituting Equation (A.2) into Equation (A.1), we can express the unified model as Equation (A.3):

$$Y_{ij} = \mu + \delta T_j + \beta' X_{ij} + \gamma' Z_j + [u_j + e_{ij}] \quad \text{(A.3)}$$

In Equation (A.3), in place of the generic outcome Y_{ij}, we substitute classroom practices, teacher satisfaction, or teacher retention data. Teacher mobility outcomes are binary or categorical. In one model specification, we use an indicator for whether the teacher returned for a fourth year of teaching. In another, we use a variable with separate categories for remaining in, moving within, or leaving the teaching profession. In the case of categorical outcome variables, we use bivariate or multinomial logistic regression to estimate the parameters of Equation (A.3).

The student achievement analysis is mathematically similar. Equations (A.4) and (A.5) express the basic student achievement model, with the unified model expressed by Equation (A.6):

$$Y_{hij,t} = c_j + \lambda' Y_{hij,t-1} + \beta' X_{ij} + \varphi' W_{hij} + e_{hij} \quad \text{(A.4)}$$

$$c_j = \mu + \delta T_j + \gamma' Z_j + u_j \quad \text{(A.5)}$$

$$Y_{hij,t} = \mu + \delta T_j + \lambda' Y_{hij,t-1} + \beta' X_{ij} + \varphi' W_{hij} + \gamma' Z_j + [u_j + e_{hij}] \quad \text{(A.6)}$$

Equation (A.6) differs from Equation (A.3) in two main ways. First, the level one units are students, represented by the *h* subscript, rather than teachers. Second, Equation (A.6) includes a lagged measure of the dependent variable $Y_{hij,t-1}$,which makes it a growth model of student achievement, sometimes referred to as a value added model. In other words, $Y_{hij,t}$ can be thought of as the posttest, which depends on $Y_{hij,t-1}$, the pretest. Other variables in the model include teacher background characteristics X_{ij}, student background characteristics W_{hij}, an indicator T_j for random assignment to the treatment group, and a set of district-by-grade fixed effects which make up the vector of school level variables Z_j. We allow the coefficient on the pretest λ to vary by district-grade (that is, to vary by test). We substitute data for both math and reading test scores for the outcomes $Y_{hij,t}$ and $Y_{hij,t-1}$.

In Equations (A.3) and (A.6), the coefficient δ for the treatment group indicator represents the impact of the receipt of comprehensive induction services and is the main parameter of interest. The standard error of this impact estimate accounts for the design effects attributable to the clustering of teachers and students within schools, which occurs because teachers or students within schools tend to have similar outcomes. The school at the time of random assignment is always used when clustering teachers and students, even if teachers have changed schools.

Equations (A.3) and (A.6) can be thought of as mixed models or as hierarchical models. They are "mixed" because they contain fixed effects (represented by $\mu\delta\beta\gamma\lambda\varphi$) as well as random effects (represented by *e* and *u*). They are hierarchical because they embed a school-level model (indexed by *j*) within an individual-level model (indexed by *h* or *i*). Several techniques are available for estimating such a model, including ordinary least squares (OLS) with robust standard errors (see Huber 1967; White 1980); generalized least squares (GLS) estimates of a random effects model; maximum likelihood; and restricted maximum likelihood. We estimated the standard errors of the model by using each of these methods, but the findings did not change, except for one student achievement result noted in Chapter V. We report findings based on GLS estimates of a random effects model for classroom practices, teacher attitudes, and teacher retention. For student achievement models, in which we collect data from each school district, we use ordinary least squares with robust standard errors, which does not require as many assumptions about the distribution of the error terms as a random effects model.

A teacher background questionnaire, discussed in Chapter III, provides a long list of potential explanatory variables for inclusion in the model (the *X* vector), including demographic and

household characteristics, information on teachers' education and professional background, and teaching assignment. In addition, we include school-level variables (the *Z* vector) from the Common Core of Data (CCD) of the National Center for Education Statistics.[57] For the student achievement analyses, districts provided student pretest scores (Y_{t-1}) and student demographic characteristics that could be included in the model (the *W* vector).

We used a separate set of covariates for each type of outcome we analyzed. Table A.1 presents the lists by analysis type. The analysis of classroom practices (Table V.1) used teacher personal characteristics, teacher professional characteristics, school characteristics, and district and grade fixed effects. The student achievement benchmark analyses (Tables V.2 and V.5) had normalized student pretest score, student background characteristics, teacher personal characteristics, teacher professional characteristics, and district-by-grade fixed effects. The analysis of teacher attitudes (Tables D.1 to D.8) had district and grade fixed effects and no other covariates. Finally, the benchmark teacher retention analysis (Tables D.11 and D.12) included teacher personal characteristics, teacher professional characteristics, teacher neighborhood characteristics, school characteristics, and district and grade fixed effects.

We generate results for each district by regressing the outcome variable on the benchmark covariates, an indicator variable for random assignment of the school to the treatment group, and a set of interaction terms between the treatment indicator and each district. The district-specific impact is the coefficient on its interaction term. To derive these results, we pool one-year and two-year districts. See Figures B.1–B.4, C.4–C.9, and D.2–D.5. An analogous method is used to derive grade-by-grade results for student achievement shown in Tables V.3 and V.6.

Instrumental Variable Estimation to Correct for Measurement Error. One feature of achievement growth models is that pretest is used as an explanatory variable. However, pretest score is an imperfect measure of prior ability, so measurement error in the pretest induces a bias in the pretest coefficient estimate and anything correlated with pretest. By design, treatment status, whose coefficient δ is the main parameter of interest, is statistically independent of pretest, but even chance correlations can influence the impact estimate. In addition, the possibility that school principals of treatment teachers could have assigned different types of students to their teachers than principals of control teachers raises the possibility of a relationship between treatment and pretest.

Therefore, as a specification check of the benchmark student achievement results, we estimated a regression model using the method of instrumental variables to adjust for the measurement error problem. An instrumental variable is one that explains variation in the variable measured with error but is not correlated with the final outcome. We used the opposite subject test score as an instrument in this exercise (math for reading and reading for math). We therefore estimate this system of equations, using M for math test score and R for the reading test score:

$$M_{hij,t-1} = \alpha + \theta R_{hij,t-1} + \tau_1' X_{ij} + \tau_2' W_{hij} + \tau_3 T_j + \tau_4' Z_j + v_{hij} \quad \text{(A.7)}$$

$$M_{hij,t} = \mu + \lambda \hat{M}_{hij,t-1} + \beta' X_{ij} + \varphi' W_{hij} + \delta T_j + \gamma' Z_j + e_{hij} \quad \text{(A.8)}$$

[57] CCD data are reported with a lag; therefore, the school-level information describes schools in 2004–2005, one year before the study began.

Table A.1. Covariates Included in Impact Estimation Models by Analysis Type

Analysis	Tables	Covariates Included in the Impact Estimation Model
Classroom practices	V.1, C.1, C.2, C.3, C.4	Teacher personal characteristics: • Age • Gender Teacher professional characteristics: • Route into teaching • Certification status • Highest degree • Months of teaching experience • Grade level School characteristics: • Percentage of students eligible to receive a free or reduced-price lunch • Percentage of students who are white District fixed effects Grade fixed effects
Student achievement (benchmark model)	V.2, V.3, V.4 (lines 1, 4-7, 11), V.5, V.6, V.7 (lines 1, 4-7, 11), C.11	Student characteristics: • Gender • Race/ethnicity • Special education status • English-language learner status • Free/reduced-price lunch status • Over age for grade Teacher personal characteristics: • Age • Age squared • Gender • Race/ethnicity • Teacher race/ethnicity matches that of a majority of students Teacher professional characteristics: • Route into teaching • Highest degree • Holds a degree in an education-related field • First-year teacher • Hired after the school year began • Attended a competitive college • Held a nonteaching job for five or more years Normalized pretest score District-by-grade fixed effects
Student achievement (alternative model 1)	V.4 (line 2), V.7 (line 2)	Student characteristics: • Gender • Race/ethnicity • Special education status • English-language learner status • Free/reduced-price lunch status • Over age for grade Normalized pretest score District-by-grade fixed effects
Student achievement (alternative model 2)	V.4 (line 3), V.7 (line 3)	Normalized pretest score District-by-grade fixed effects

Table A.1 ***(continued)***

Analysis	Tables	Covariates Included in the Impact Estimation Model
Student achievement (alternative model 3)	V.4 (lines 8-9), V.7 (lines 8-9)	Student characteristics: • Gender • Race/ethnicity • Special education status • English-language learner status • Free/reduced-price lunch status • Over age for grade Teacher personal characteristics: • Age • Age squared • Gender • Race/ethnicity • Teacher race/ethnicity matches that of a majority of students Teacher professional characteristics: • Route into teaching • Highest degree • Holds a degree in an education-related field • First-year teacher • Hired after the school year began • Attended a competitive college • Held a nonteaching job for five or more years District-by-grade fixed effects
Student achievement (alternative model 4)	V.4 (line 10), V.7 (line 10)	Student characteristics: • Gender • Race/ethnicity • Special education status • English-language learner status • Free/reduced-price lunch status • Over age for grade Teacher personal characteristics: • Age • Age squared • Gender • Race/ethnicity • Teacher race/ethnicity matches that of a majority of students Teacher professional characteristics: • Route into teaching • Highest degree • Holds a degree in an education-related field • First-year teacher • Hired after the school year began • Attended a competitive college • Held a nonteaching job for five or more years Normalized pretest score District fixed effects
Teacher attitudes	D.1, D.2, D.3, D.4, D.5, D.6, D.7, D.8	District fixed effects Grade fixed effects

Table A.1 ***(continued)***

Analysis	Tables	Covariates Included in the Impact Estimation Model
Teacher mobility	D.9, D.10, D.11, D.12	Teacher personal characteristics: • Age • Gender • Race/ethnicity • Teacher race/ethnicity matches that of a majority of students • Marital status • Teacher has children Teacher professional characteristics: • Months of relevant teaching experience • Certification status • Holds a degree in an education-related field • Hired after the school year began • Attended a competitive college • Held a nonteaching job for five or more years • Taught a single grade level Teacher neighborhood characteristics: • Commuting distance • Teacher is a homeowner • Teacher lives in the school district • Teacher attended an elementary school in which the socioeconomic status of students was similar to the school taught in School characteristics: • Percentage of students eligible to receive a free or reduced-price lunch • Percentage of students who are white District fixed effects Grade fixed effects
Teacher mobility (alternative model 1)	D.11, D.12	District fixed effects
Teacher mobility (alternative model 2)	D.11, D.12	Teacher personal characteristics: • Age • Gender • Race/ethnicity • Teacher race/ethnicity matches that of a majority of students • Teacher has children District fixed effects Grade fixed effects

The first stage equation (A.7) regresses $M_{hij,t-1}$ on all of the other independent variables from the outcome equation (A.8) plus an instrumental variable, the opposite-subject pretest, $R_{hij,t-1}$ (that is, we use the reading pretest as an instrument for the math pretest and vice versa). In the second stage (outcome) equation (A.8), $M_{hij,t-1}$ is replaced by its predicted value, which is generated from equation (A.7) by setting the error term v_{hij} to zero. Instrumental variable results are reported on line 11 in the top and bottom panels of Tables V.7 and V.8.

Difference-in-Differences Analysis of the Change in the Treatment Effect for Student Achievement in Multiple Years. To measure the improvement of treatment teachers relative to control teachers from one study year to another, we employ a difference-in-differences estimator. In particular, we compare the difference in student outcomes between treatment and control teachers in Year 3 to the corresponding differences in Year 2 and separately compare treatment/control differences in Year 3 and Year 1.[58] Both comparisons follow the same method. We pool data on all students taught by the common sample of teachers in the data in both years and estimate the following model:

$$Y_{hij,t} = \mu + \pi C_{hij} + \delta_1 T_j + \delta_2 \left(T_j * C_{hij}\right) + \lambda_1 ' Y_{hij,t-1} + \lambda_2 ' \left(Y_{hij,t-1} * C_{hij}\right) + \beta_1 ' X_{ij} + \beta_2 ' \left(X_{ij} * C_{hij}\right)$$
$$+ \varphi_1 ' W_{hij} + \varphi_2 ' \left(W_{hij} * C_{hij}\right) + \gamma_1 ' Z_j + \gamma_2 ' \left(Z_j * C_{hij}\right) + [u_j + e_{hij}] \qquad \text{(A.9)}$$

In this model, the student posttest is regressed on an indicator variable for cohort C_{ij} (the Year 1, Year 2, or Year 3 cohort of students), student pretest, teacher background characteristics, student background characteristics, district-by-grade fixed effects, assignment to the treatment group T_j, and the interaction of all variables in the model with cohort.

Students in the Year 1 or Year 2 cohort are assigned weights in order to equalize the weight for a teacher in the earlier cohort and the Year 3 cohort. For example, if a teacher has 20 students in cohort two and 10 students in cohort three, each student in cohort two will receive a weight of 0.5 so that the total weight for that teacher in cohort two is 10 (since 20*0.5 = 10). Conversely, for a teacher with 10 students in cohort two and 20 students in cohort three, each student in cohort two receives a weight of 2.

The key parameter of interest is δ_2, which estimates the effect of the interaction of treatment status and cohort. This parameter estimates the difference of the treatment/control contrast in the teacher effect on student test scores between Year 1 and Year 3 (or between Year 2 and Year 3). We use robust standard errors to account for correlation in outcomes for students clustered within schools.

Non-Experimental Analysis. Chapter VII presents findings from non-experimental analyses that are very similar in structure to the experimental analyses. Those analyses are based on Equations (A.3) and (A.6), except that we replace the treatment status indicator with a vector of variables that are indices describing the level or intensity of teacher induction services reported by the teacher. Alternately, we replace the treatment status indicator with a single index. The result, presented in Equation (A.9), is an extension of the retention analysis. The student achievement model (not shown) is analogous.

$$Y_{ij} = \mu + \theta ' Q_{ij} + \beta ' X_{ij} + \gamma ' Z_j + [u_j + e_{ij}] \qquad \text{(A.10)}$$

where we replace T, the indicator variable for assignment to the treatment group in Equation (A.3), with Q_{ij}, representing a vector of indices (or a single index) describing the breadth, intensity, or

[58] Due to small sample size, we do not present estimates of a model in which the sample is composed of teachers common to all three cohorts.

nature of induction services. Each coefficient in the θ vector captures the relationship between an induction index and the outcome *Y*. The same vector of *X* and *Z* variables used in the experimental section is used in the corresponding nonexperimental analysis unless specified otherwise. The psychometric properties of the indices are presented in Table A.5.

B. Analysis Weights

Most analyses in the report use weights that accounted for two aspects of the study design. One is nonresponse to the surveys, and the other is the unequal probability across districts of a teacher being in the treatment group. This appendix explains the nature of these problems and how weights were used to address them.

The response rates for this study's surveys exceeded the targets set in the study design, but we did observe statistically significant differences between treatment and control groups. A concern with differential response rates is that, if nonresponse is not random with respect to outcomes, the degree to which nonresponse affects the average outcomes will differ by treatment status, and the impact estimates—which are differences in mean outcomes for respondents only—will be biased. If, for example, nonrespondents have worse outcomes than respondents, then we would expect the lower response rates for the control group to translate into an upwardly biased estimate of the counterfactual outcome and therefore a downwardly biased estimate of the impact.

To mitigate the potential bias, we constructed nonresponse adjustment weights, calculated separately for each data collection instrument as follows. First, we used a logistic regression model to estimate the relationship between the likelihood of responding to the survey and the baseline variables, such as the teacher's age, level of education, and preparation route. We estimated separate prediction models for the treatment and control groups. Then we computed the weight as the inverse of the predicted probability of responding. This procedure is equivalent to letting the respondents in each treatment group who look most like nonrespondents carry a greater weight so that they can stand in for their missing counterparts. We used these weights in all impact estimations with teacher outcomes, although the weights did not substantially change the findings.

We made one adjustment to the weights to deal with potential confounding of district characteristics with treatment status. As with most multisite studies, the probability of assignment to treatment was not identical across districts. Therefore, we tailored the random assignment procedure slightly to each district based on (1) the number of schools that the district contributed to the study and (2) the cluster size (number of eligible teachers per school), resulting in some variation in the ratio of treatment to control teachers. Thus, when we report averages based on data pooled across districts, we must use weights to account for differential treatment-control ratios; otherwise, the treatment-control comparisons for the full study would confound treatment differences with site differences. For example, if we had assigned 60 percent of the teachers to the treatment group in an extremely low-income district and 50 percent of the teachers to the treatment group in all other districts, the low-income students would be overrepresented in the overall treatment group, even though random assignment produced equivalent groups within each district. To correct for such overrepresentation, we divided the weights described above by the number of observations in each treatment group within each site and multiplied by the average number of observations in the two treatment groups in the district. The result is Equation (A.10):

$$WEIGHT_{ikm} \propto (1/\hat{p}_i) * \frac{1}{n_{km}} \frac{(n_{kT} + n_{kC})}{2} \qquad \text{(A.11)}$$

where i indexes teachers, k indexes districts, and m indexes experimental group (treatment or control). The term p_i represents the predicted probability of teacher i being a respondent.

We developed enhanced weights for use with follow-up surveys to take advantage of the detailed list of background variables available from the background (baseline) survey. The enhanced weights made no difference in the estimates; therefore, we did not use them in the benchmark analyses presented in this report.

C. Outcome Variables

1. Classroom Practices

Classroom observers were trained to use the Diagnostic Classroom Observation (DCO) protocol to assess instruction practices. The DCO, originally called the Vermont Classroom Observation Tool (VCOT), is a proprietary tool for classroom observations developed by the Vermont Institutes (see Saginor and Hyjek 2005; Saginor 2008). Researchers who first worked with Science and Math Program Improvement (SAMPI), a research group at Western Michigan University, developed the tool over several years. SAMPI had developed an instrument to measure the quality of standards-based, investigative science and mathematics instruction based on research conducted by Horizon Research, Inc.

In developing the DCO, the Vermont Institutes staff used the SAMPI Observation Tool as a starting point and carefully reviewed Charlotte Danielson's Framework for Teaching (1996), on which the widely used Praxis III observational assessment (and ETS induction program) is based (Dwyer 1994). In parallel with the Praxis III content, the DCO developers included examples of evidence for each indicator, added systematic and ongoing formative and summative assessment of student learning as a major indicator, and simplified and shortened the tool. The tool underwent further refinement through its use in the field by a group of trained teacher-leaders who observed classrooms. In 2004, several of those involved in the original design of the tool adapted it for use in the observation of literacy lessons. The standards and practices included in the National Council of Teachers of English (NCTE) Standards and the National Reading Panel (NICHHD 2000) also helped inform development of the literacy version of the DCO.

The DCO describes teaching practices in four areas:

1. Planning and Organization of a Lesson
2. Implementation of a Lesson
3. Content of a Lesson
4. Classroom Culture

In this study, we attempted to measure all but the first construct, lesson planning and organization. The procedure for assessing lesson planning and organization is more suited for individual teacher feedback than for research and requires measurement of activities before the start of a lesson and a separate teacher interview of varying length and content.

Implementing the DCO. Staff from the Vermont Institutes trained the classroom observers. Much of the training relied on videotaped classes, but it also included practice observations conducted in pairs in "live" school settings. During the practice observations, observers scored independently and then debriefed to reach consensus on any individual items for which the

discrepancy exceeded a single point. In addition to practice observations, observers participated in training for a total of nine days over the course of three training sessions.

After observing and scoring a videotaped class, observers were deemed "certified" to conduct the observations based on a comparison of their 16-item scores to the observations of a "gold standard" panel. The gold standard panel consisted of the tool's developer and two trained observers who demonstrated a clear understanding of the items measured in the tool and showed high rates of agreement in scoring. Trainees had two opportunities to come within 0.75 points of the gold standard average score for the three constructs (implementation, content, and culture) during a test observation. Trainees who did not meet the standard were not allowed to conduct observations. To address the possibility that observers' scoring would start to drift in one direction or another after conducting observations, we asked the tool developer to observe a classroom with each observer in the field at least once to verify scoring after each observer had completed several observations. As mentioned in Chapter III, observers were always blind to teachers' treatment status and therefore did not know if they were observing someone who had received the comprehensive induction support.

Interpreting DCO Scores. To summarize the information from the classroom observations across all 16 indicators, we produced three scores corresponding to the three domains captured by the observation protocol (into which the items had already been grouped). We performed a factor analysis of the 16 classroom observation items to explore the degree to which the theoretical groupings were empirically justified. In finding the groupings justified, we maintained the three-construct scoring method (implementation of literacy lesson, content of literacy lesson, and classroom culture) described above. Factor loadings for the 16 class observation items are shown in Table A.2. Psychometric details for each construct are presented in Table A.5.

Table A.2. DCO Classroom Practices Constructs: Factor Loadings

Variable	Factor Loading 1	Factor Loading 2
Literacy Implementation		
Best practices	0.808	0.364
Institutional choices	0.719	0.509
Student choices	0.805	0.241
Pace	0.595	0.581
Literacy Content		
Understanding content and close reading	0.756	0.321
Assessment	0.473	0.275
Skill development	0.784	0.332
Connections between reading and writing	0.771	0.138
Literacy Classroom Culture		
Maximizes learning opportunities	0.315	0.868
Routines clear and consistent	0.256	0.817
Respectful behavior, safe atmosphere	0.278	0.867
Literacy valued	0.644	0.439
Teacher works collaboratively with students	0.536	0.652
Students work collaboratively with students	0.458	0.654
Equal access to teacher and resources	0.285	0.776

Source: Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in all districts participating in the study. The extraction method was principal components analysis, and the rotation method was varimax with Kaiser normalization.

N = 631 teachers.

The estimated impacts on classroom practices described in Chapter V can be better understood by relating the DCO scores to student test scores. We conducted correlational analyses to explore whether there is a relationship between student achievement gains and DCO scores. After fielding the DCO for this study and comparing the results with student achievement gains, we found that the association between each of the three classroom practices indices and student test scores was positive and statistically significant (regression coefficients = 0.065 to 0.085, p-values = 0.001 to 0.024.) Because the test scores have been standardized to have a mean of zero and a standard deviation of one, the magnitude of each estimate can be interpreted as an effect size. That is, a one-unit change in a DCO score, for example from "limited evidence" to "moderate evidence," is associated with an increase in reading test scores of between 6.5 percent and 8.5 percent of a standard deviation.

2. Teacher Attitude Measures

Using items from the induction activities surveys, we measured teachers' feelings of satisfaction in 19 areas (such as satisfaction with their workload) and preparedness in 13 areas (such as preparedness to work with students with special challenges). The surveys asked teachers to respond along a four-point scale (ranging from "very dissatisfied" to "very satisfied" and from "not at all prepared" to "very well prepared"). For both satisfaction and preparedness, we conducted a factor analysis on fall 2005 data to explore how items could be sensibly grouped together. The factor analyses suggested that teacher satisfaction consisted of satisfaction with (1) school, (2) class, and (3) career, and teacher preparedness consisted of preparedness to (1) instruct, (2) work with students, and (3) work with others. We used these domains to summarize the data. Factor loadings for the teacher satisfaction items are shown in Table A.3 and for teacher preparedness items in Table A.4. Psychometric properties for each scale are given in Table A.5.

Table A.3. Teacher Satisfaction Constructs: Factor Loadings

Variable	Factor Loading 1	2	3
Satisfaction with School			
Support from administration for beginning teachers	0.757	0.330	0.043
Availability of resources and materials/equipment for your classroom	0.576	0.264	0.153
Input into school policies and practices	0.665	0.296	0.202
Opportunities for professional development	0.473	0.250	0.338
Principals' leadership and vision	0.765	0.281	0.015
Professional caliber of colleagues	0.709	0.046	0.251
Supportive atmosphere among faculty/collaboration with colleagues	0.728	0.075	0.191
School facilities such as the building or grounds	0.557	0.215	0.141
School policies	0.631	0.449	0.183
Satisfaction with Class			
Autonomy or control over own classroom	0.397	0.551	0.038
Student motivation to learn	0.194	0.736	0.194
Student discipline and behavior	0.167	0.795	0.177
Parental involvement in the school	0.210	0.498	0.336
Grade assignment	0.239	0.558	-0.021
Students assigned	0.156	0.734	0.143
Satisfaction with Teaching Career			
Salary and benefits	0.035	0.008	0.851
Professional prestige	0.425	0.271	0.623
Intellectual challenge	0.414	0.346	0.460
Workload	0.313	0.386	0.475

Source: Mathematica First Induction Activities Survey administered in fall 2005 to all study teachers.

Note: Data pertain to teachers in all districts participating in the study. Emphasis on standardized test scores was not included in factor analyses or subscales. The extraction method was principal components analysis, and the rotation method was varimax with Kaiser normalization.

N = 889 teachers.

3. Test Score Data

Aggregation of Test Scores Across Grades, Subjects, and Districts. Districts and even grades within some districts varied with respect to types of tests administered. Aggregating test scores across different tests posed a challenge for the analysis because it is important that achievement be measured in a common metric in order to combine the results across districts and grades. In anticipation of this problem, we designed the random assignment of schools to yield an approximately even mix of teachers in the treatment and control groups by grade level within district. Therefore, treatment-control comparisons within any grade level and district became "apples-to-apples" comparisons. Instead of trying to compare different increments of learning across grades and districts, this approach only requires combining treatment-control differences (impact estimates) from all district-grade combinations to a single number in order to summarize the findings and draw on as large a sample as possible.

Table A.4. Teacher Preparedness Constructs: Factor Loadings

Variable	Factor Loading 1	2	3
Prepared to Instruct			
Managing classroom activities, transitions, and routines	0.677	0.397	0.045
Using a variety of instructional methods	0.747	0.182	0.225
Assessing your students	0.621	0.211	0.399
Selecting and adapting curriculum and instructional materials	0.690	0.154	0.345
Planning effective lessons	0.644	0.148	0.497
Being an effective teacher	0.693	0.340	0.298
Addressing the needs of a diversity of learners	0.621	0.337	0.292
Prepared to Work with Students			
Handling a range of classroom behavior or discipline situations	0.573	0.599	0.001
Motivating students	0.448	0.604	0.133
Working effectively with parents	0.077	0.725	0.447
Working with students who have special behavioral, emotional, developmental, or physical challenges	0.264	0.691	0.226
Prepared to Work with Other School Staff			
Working with other teachers to plan instruction	0.268	0.166	0.809
Working with the principal or other instructional leaders	0.282	0.287	0.779

Source: Mathematica First Induction Activities Survey administered in fall 2005 to all study teachers.

Note: Data pertain to teachers in all districts participating in the study. The following items were not included in factor analyses or subscales: teaching reading/language arts, teaching mathematics, and working with English-language learners. The extraction method was principal components analysis, and the rotation method was varimax with Kaiser normalization.

N = 895 teachers.

To facilitate aggregation by grade and district, we converted all test scores to a common metric called a z-score, which is obtained by subtracting the mean from each value and dividing by the standard deviation of a "universe" of test-takers, approximated by a state or national norm group. The resulting score can be interpreted as the distance from the average score as a fraction of a standard deviation difference from the average for the reference group; therefore, a z-score of -0.5, for example, means that the score was one-half of a standard deviation below the state or national mean. We used the mean and standard deviation of the norm sample for each test as published by state agencies or test developers.

As an example, consider the hypothetical case where we compare the gains for a fourth-grade teacher named Ms. Smith in Seattle with those of a fifth-grade teacher named Mr. Cone in Cleveland.[59] Assume that Ms. Smith's students scored at the average level for Seattle third graders in the pretest year and 10 percent of a standard deviation above the fourth-grade average at the end of the posttest year on a Washington State math assessment. Also assume that students in Mr. Cone's class in Cleveland who performed at one-half of a standard deviation above the mean at the end of grade four on Ohio's state math assessment subsequently scored 0.6 of a standard deviation at the end of grade five. These would be considered equivalent, as both sets of students moved up one-

[59] Seattle and Cleveland are listed as hypothetical examples. They are not in the study.

tenth of a standard deviation relative to their local reference groups on their own state's assessment ($0.1 - 0.0 = 0.6 - 0.5$).

It is also possible to aggregate by subject matter. We kept two broad subject areas distinct—math and reading (which includes English/language arts)—and we present the findings separately for those two subjects. We dropped reading z-scores from two districts because the tests were scored within the district. We excluded other subjects from the main impact analysis, such as foreign languages, social studies, or science, which are not available in enough districts to yield meaningful findings. Psychometric properties of the test score measures are given in Table A.5.

Missing Data. Not every student that a teacher was responsible for during the year had a valid, usable test score for the analysis. For example, students might have been exempt from testing, be missing a test score because of repeated absence, not have been enrolled during the test period, or have repeated the grade in which they were enrolled in 2007–2008. In districts for which these data were available, we did not use pretest scores for students whose grade in the 2006–2007 school year matched their grade in the 2007–2008 school year. These students were presumed to be grade repeaters, meaning that the pretest score was given on a different scale than that of the main group of students.

In addition, we excluded students from a subject if their test scores were out of range for the test for that district, grade, and subject, as this was evidence that the test score provided was taken from an alternative assessment. Of the students with a valid reading posttest in district-grade combinations included in the benchmark model, 88 percent of treatment students had a pretest score compared to 85 percent of control students, a difference of almost 4 percentage points (after accounting for rounding). For students with a valid math posttest in district-grade combinations included in the benchmark model, 86 percent of treatment students had a pretest score compared to 84 percent of control students, a difference of 2 percentage points. We assumed that the data were missing at random. Finally, we dropped teachers whose average student gain scores were greater than 1.5 standard deviations above or below the mean for the reference group (state or norm sample). This resulted in the loss of one classroom, whose students had gains in math scores that would have placed most of them in the 94th percentile or above for the state, even though the same students' gains in reading were slightly below average.

Restrictions. Based on the data provided by school districts, we excluded some students from the model if it appeared implausible based on objective criteria that the teacher linked to them was their full-time teacher for one or both subjects. Our first step was to check the second Teacher Mobility Survey, conducted in fall 2007, to ensure that all teachers in the test score data claimed to teach reading, math, or special education/resource in response to a survey question about teaching assignment. Furthermore, we checked this survey to ensure that there were no mismatches between the grade of the majority of students linked to the teacher and the grade(s) the teacher claimed to have taught on this survey. In this way, we excluded two teachers, both of whom were linked to more than 100 students because they taught physical education/health.

Even after excluding these teachers, the test score samples included some teachers who were linked to an implausibly high or low number of students to be a regular classroom teacher. Two districts provided the subjects taught by teachers with student test score data. To help determine teaching assignments in other districts, we followed up with teachers whose responses on the second Teacher Mobility Survey were outliers, especially teachers who were linked to 30 or more students or to 12 or fewer students. We asked these teachers to verify their teaching assignments by giving the number of students for whom they were principally responsible for math and reading outcomes in

the 2006–2007 school year and the 2007–2008 school year. We compared their responses to the data we had been given by the district and, in some cases, clarified discrepancies by querying the school district. Of those identified, we successfully contacted 24 of 46 treatment teachers (52 percent) and 18 of 43 of control teachers (42 percent).

Based in part on their responses, we used these criteria to restrict the sample for the test score analysis:

- If a teacher was linked to 30 or more students and indicated on the Teacher Background Survey that she was not responsible for reading or math outcomes, students were excluded from whatever subject the teacher said she did not teach. This affected 5 teachers and 223 students in reading and 3 teachers and 144 students in math.
- We excluded all students from a teacher who was linked to 40 or more students and responded on the Teacher Background Survey that she was responsible for both reading and math outcomes on the background survey. This affected 4 teachers, 329 students in reading, and 330 students in math. Because treatment teachers were more likely to confirm their teaching assignments than control teachers, we applied the four restriction rules to all teachers, even if we learned during the followup interview that they were an exception to these general rules. In this way, the application of the restrictions was not confounded with treatment status. As part of the sensitivity analysis, however, we also estimated the model (1) based on a sample that used general rules for teachers whom we were unable to contact and particular information from teachers' responses if it contradicted the general rules; and (2) based on a sample that did not impose the two rules.

The test score analysis pertains only to teachers in tested grades and subjects. Because students were included in the benchmark model only if they had a valid pretest from the prior year, we excluded the youngest grade at which testing began in a district. For example, in districts that test students, as mandated by the No Child Left Behind Act, in grades three through eight, and operate elementary schools that enroll students in kindergarten through grade five (the most common case), we were able to estimate impacts on achievement for grades four and five only. As part of the sensitivity analyses, we excluded the pretest covariate from the analysis and thus were able to consider more grades and include more students in the analysis.

Table A.5. Psychometric Properties of Measures

Outcome	Number of Items	Mean	Median	SD	Minimum Value	Maximum Value	Sample Size	Cronbach's Alpha
Teacher Satisfaction								
Satisfaction with career								
Fall 2005	4	3.01	3.00	0.60	1.00	4.00	889	0.77
Spring 2006	4	2.91	3.00	0.63	1.00	4.00	876	0.78
Fall 2006	4	2.96	3.00	0.63	1.00	4.00	831	0.73
Spring 2007	4	2.87	3.00	0.66	1.00	4.00	370	0.78
Fall 2007	4	2.88	3.00	0.63	1.00	4.00	752	0.72
Fall 2008	4	2.87	3.00	0.71	1.00	4.00	715	0.90
Satisfaction with class								
Fall 2005	6	3.05	3.17	0.61	1.00	4.00	889	0.84
Spring 2006	6	2.99	3.00	0.64	1.00	4.00	876	0.85
Fall 2006	6	3.14	3.17	0.58	1.17	4.00	832	0.78
Spring 2007	6	3.09	3.17	0.59	1.00	4.00	370	0.82
Fall 2007	6	3.11	3.17	0.59	1.00	4.00	749	0.77
Fall 2008	6	3.15	3.17	0.59	1.00	4.00	715	0.93
Satisfaction with school								
Fall 2005	9	3.10	3.11	0.63	1.00	4.00	889	0.90
Spring 2006	9	2.99	3.00	0.66	1.00	4.00	876	0.91
Fall 2006	9	3.14	3.22	0.59	1.11	4.00	832	0.88
Spring 2007	9	2.99	3.00	0.64	1.00	4.00	370	0.89
Fall 2007	9	3.10	3.11	0.59	1.11	4.00	745	0.87
Fall 2008	9	3.11	3.22	0.62	1.11	4.00	716	0.96
Teacher Preparedness								
Preparedness to instruct								
Fall 2005	7	2.80	2.86	0.56	1.00	4.00	895	0.90
Spring 2006	7	2.95	3.00	0.56	1.00	4.00	876	0.92
Spring 2007	7	3.14	3.00	0.54	1.00	4.00	371	0.90
Fall 2008	7	3.46	3.57	0.48	1.71	4.00	694	0.98
Preparedness to work with others								
Fall 2005	2	2.88	3.00	0.74	1.00	4.00	895	0.82
Spring 2006	2	2.95	3.00	0.71	1.00	4.00	874	0.82
Spring 2007	2	3.12	3.00	0.68	1.00	4.00	371	0.73
Fall 2008	2	3.41	3.50	0.62	1.00	4.00	694	0.94
Preparedness to work with students								
Fall 2005	4	2.73	2.75	0.59	1.00	4.00	895	0.78
Spring 2006	4	2.84	2.75	0.61	1.00	4.00	876	0.83
Spring 2007	4	2.99	3.00	0.57	1.00	4.00	371	0.75
Fall 2008	4	3.25	3.25	0.55	1.25	4.00	694	0.95
Classroom Practices								
Implementation of literacy lesson	5	2.66	2.60	0.84	1.00	4.80	631	0.89
Content of literacy lesson	4	2.37	2.25	0.79	1.00	5.00	631	0.80
Classroom culture	7	3.06	3.14	0.87	1.00	5.00	631	0.93
Student Achievement								
Reading posttest scores, 2008	1	-0.21	-0.15	0.93	-4.40	3.40	3,037	n/a
Math posttest scores, 2008	1	-0.13	-0.12	1.00	-4.44	3.47	2,827	n/a
Induction Support								
Full sample of teachers								
Years BT had an assigned mentor	3	1.23	1.00	0.62	0.00	2.01	901	n/a
Induction Breadth Index	9	1.83	2.00	0.70	0.00	3.00	959	n/a
Fall 2005	3							0.26
Spring 2006	3							0.38
Fall 2006	3							0.33

Table A.5 *(continued)*

Outcome	Number of Items	Mean	Median	SD	Minimum Value	Maximum Value	Sample Size	Cronbach's Alpha
Instructional Focus Index	8	1.55	1.50	0.57	0.00	2.50	965	n/a
Fall 2005	3							0.52
Spring 2006	3							0.54
Fall 2006	2							0.29
Induction Intensity Index	10	1.44	1.08	1.43	0.00	22.61	965	n/a
Fall 2005	4							0.28
Spring 2006	4							0.28
Fall 2006	2							0.42
Sample of teachers in student math test scores analyses								
Years BT had an assigned mentor	3	1.19	1.00	0.61	0.00	2.01	171	n/a
Induction Breadth Index	9	1.92	2.00	0.65	0.00	3.00	172	n/a
Fall 2005	3							0.26
Spring 2006	3							0.41
Fall 2006	3							0.35
Instructional Focus Index	8	1.50	1.50	0.54	0.00	2.50	172	n/a
Fall 2005	3							0.55
Spring 2006	3							0.33
Fall 2006	2							0.25
Induction Intensity Index	10	1.57	1.17	2.03	0.00	22.61	172	n/a
Fall 2005	4							0.22
Spring 2006	4							0.30
Fall 2006	2							0.46
Sample of teachers in student reading test scores analyses								
Years BT had an assigned mentor	3	1.21	1.00	0.62	0.00	2.01	175	n/a
Induction Breadth Index	9	1.92	2.00	0.67	0.00	3.00	177	n/a
Fall 2005	3							0.34
Spring 2006	3							0.44
Fall 2006	3							0.33
Instructional Focus Index	8	1.57	1.50	0.56	0.00	2.50	178	n/a
Fall 2005	3							0.57
Spring 2006	3							0.40
Fall 2006	2							0.27
Induction Intensity Index	10	1.63	1.23	2.01	0.00	22.61	178	n/a
Fall 2005	4							0.21
Spring 2006	4							0.32
Fall 2006	2							0.46

Source: Mathematica analysis using data from the 2007–2008 school year provided by participating school districts; Mathematica First, Second, and Third Induction Activities Surveys administered in fall 2005, spring 2006, and fall 2006 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Cronbach's alpha was calculated separately for variables within each time point for the Induction Breadth Index, Instructional Focus Index, and Induction Intensity Index. Due to missing data, some values for induction support have been imputed, which can cause maximum values for a particular variable to exceed the value associated with full-time induction support.

BT = beginning teacher; n/a = not applicable.

D. Supplementary Table for Chapter III

Chapter III presented information on the survey instruments and other methods by which we collected data. Table A.6 provides sample size information for Table III.2.

Table A.6. Sample Sizes for Table III.2, Response Rates to Teacher Surveys by Subgroup and Treatment Status

	Number of Teachers	
	Treatment	Control
District Type (Years of Implementation)		
One year	275	286
Two year	231	217
Grade Level		
K or pre-K	80	72
1	73	71
2	84	78
3	81	57
4	60	60
5	46	52
Other/multiple	82	113
School Type (Percent in Free Lunch Program)		
Unknown	30	29
0–49.9%	37	29
50–74.9%	98	128
75–100%	341	317

Source: Mathematica teacher induction survey management system; Mathematica Teacher Background Survey (fall 2005), Induction Activities/Teacher Mobility Surveys (fall 2006 and 2007) administered to all study teachers; Induction Activities Survey (spring 2007) administered to teachers in two-year districts.

Note: The Induction Activities Survey and Mobility Survey were administered together in fall 2006 and fall 2007.

n/a = not applicable.

APPENDIX B

SUPPLEMENTAL INFORMATION FOR CHAPTER IV

This appendix presents estimates of the receipt of induction activities as reported by teachers. First, tables show treatment and control group means and service contrast estimates using data from each induction activities survey: fall 2005, spring 2006, fall 2006, spring 2007 (for teachers in two-year districts), fall 2007, and fall 2008. Estimates for one-year and two-year districts are presented separately. Tables B.1-B.15 pertain to one-year districts and Tables B.16-B.30 pertain to two-year districts. The figures of service contrast estimates shown in Chapter IV are based on the estimates shown in these tables.

Second, figures show treatment-control differences in total minutes spent in mentoring by district. Estimates are shown for fall 2005 and fall 2006 and for one-year and two-year districts separately. Figures B.1 and B.2 pertain to one-year districts and Figures B.3 and B.4 pertain to two-year districts. Within each figure, the districts are ordered according to the size of the difference.

Table B.1. Teacher Reports on Professional Support and Duties (Percentages), Fall 2005 and Fall 2006: One-Year Districts

	Fall 2005				Fall 2006			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
BT has mentor	93.1	77.5	15.6*	0.000	24.5	37.7	-13.2*	0.003
BT has assigned mentor	89.8	69.9	20.0*	0.000	19.7	29.2	-9.5*	0.017
Sample Size (Teachers)	258	245	503		241	231	472	

Source: Mathematica First and Third Induction Activities Surveys administered in fall 2005 and fall 2006 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher.

Table B.2. Impacts on Teacher-Reported Mentor Profiles (Percentages), Fall 2005 and Fall 2006: One-Year Districts

	Fall 2005				Fall 2006			
Mentoring Characteristic	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Number of Mentors								
Multiple Mentors	25.4	14.6	10.8*	0.006	5.9	9.7	-3.8	0.106
Number of Mentors								
None	6.9	22.5	-15.6*	0.000	75.5	62.3	13.2*	0.003
One	67.7	62.9	4.8	0.333	18.6	28.0	-9.4*	0.021
Two	20.9	8.4	12.5*	0.000	5.9	9.7	-3.8	0.106
Number of Mentors Assigned								
None	10.1	30.1	-20.0*	0.000	80.3	70.8	9.5*	0.017
One	71.0	62.6	8.4	0.093	18.3	23.5	-5.2	0.186
Two	18.9	7.3	11.6*	0.001	1.5	5.8	-4.3*	0.010
Mentor Positions								
Positions of All Mentors								
Full-time mentor	73.7	7.5	66.3*	0.000	1.5	3.7	-2.2	0.201
Teacher	24.5	63.8	-39.3*	0.000	20.8	30.7	-9.9*	0.014
School or district administrator or staff external to district	10.5	9.1	1.4	0.575	2.9	4.2	-1.3	0.379
No mentor	6.9	22.5	-15.6*	0.000	75.5	62.3	13.2*	0.003
Sample Size (Teachers)	258	245	503		241	231	472	

Source: Mathematica First and Third Induction Activities Surveys administered in fall 2005 and fall 2006 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Table B.3. Impacts on Teacher-Reported Mentor Services Received in the Most Recent Full Week of Teaching, Fall 2005 and Fall 2006: One-Year Districts

	Fall 2005					Fall 2006				
Mentor Service	Treatment	Control	Difference	Effect Size[a]	P-value	Treatment	Control	Difference	Effect Size[b]	P-value
"Usual" Meetings with Mentors										
Frequency (number of meetings)	1.3	1.2	0.1	0.03	0.730	0.3	0.7	-0.3*	-0.25	0.015
Average duration (minutes)	23.2	9.9	13.3*	0.74	0.000	2.5	4.6	-2.1*	-0.23	0.014
Total time[b] (minutes)	56.4	33.3	23.1*	0.36	0.000	9.9	18.4	-8.4*	-0.20	0.043
Informal Meetings with Mentors										
Total time (minutes)	30.4	33.4	-3.0	-0.08	0.372	9.2	20.1	-10.9*	-0.33	0.001
Total Usual and Informal Time with Mentors (Minutes)	**86.8**	**66.7**	**20.0***	**0.24**	**0.007**	**19.1**	**38.5**	**-19.4***	**-0.30**	**0.002**
Meeting Time with Mentors in the Following Positions (Minutes)										
Full-time mentor	60.3	4.2	56.2*	0.99	0.000	0.6	2.6	-2.0	-0.17	0.109
Teacher	23.0	59.2	-36.2*	-0.46	0.000	16.6	32.6	-15.9*	-0.26	0.009
Administrator	4.1	2.0	2.1	0.13	0.145	0.3	2.3	-2.0*	-0.23	0.028
Staff external to district	1.4	1.4	0.0	0.00	0.976	1.1	0.0	1.1	0.13	0.164
Mentor Time in the Following Activities (Minutes)										
Observing BT teaching	33.5	10.0	23.5*	0.75	0.000	2.3	5.7	-3.3*	-0.22	0.021
Meeting with BT one-on-one	34.4	22.7	11.7*	0.38	0.000	6.1	10.1	-4.0	-0.19	0.056
Meeting with BT and other first-year teachers	28.5	9.2	19.4*	0.54	0.000	2.3	3.6	-1.2	-0.09	0.285
Meeting with BT and other teachers	18.8	15.4	3.3	0.09	0.320	6.8	10.1	-3.3	-0.14	0.138
Modeling a lesson	9.0	5.6	3.3*	0.18	0.032	2.1	4.0	-1.8	-0.12	0.208
Co-teaching a lesson	5.8	4.2	1.6	0.09	0.314	1.9	2.6	-0.7	-0.04	0.665
All six activities (all mentors)	130.0	67.1	62.9*	0.58	0.000	21.5	35.8	-14.3*	-0.19	0.049
All six activities (study mentor only)	110.6	0.0	110.6*	1.19	0.000	n/a	n/a	n/a	n/a	n/a
Types of Assistance a Mentor Provided (Percentage)										
Suggestions to improve practice	77.4	53.1	24.4*	n/a	0.000	14.9	26.9	-12.1*	n/a	0.001
Encouragement or moral support	86.8	65.5	21.3*	n/a	0.000	20.7	32.8	-12.1*	n/a	0.004
Opportunity to raise issues/discuss concerns	85.9	64.7	21.3*	n/a	0.000	17.7	31.6	-13.9*	n/a	0.000
Help with administrative/logistical issues	67.2	52.9	14.3*	n/a	0.001	12.4	24.6	-12.2*	n/a	0.001
Help teaching to meet state or district standards	61.1	44.1	17.0*	n/a	0.000	10.9	19.3	-8.4*	n/a	0.010
Help identifying teaching challenges and solutions	82.2	54.8	27.4*	n/a	0.000	15.9	25.0	-9.1*	n/a	0.013
Discussed instructional goals and ways to achieve them	72.6	48.1	24.5*	n/a	0.000	14.0	24.4	-10.4*	n/a	0.004
Guidance on how to assess students	58.1	43.7	14.4*	n/a	0.000	10.9	21.2	-10.4*	n/a	0.002
Shared lesson plans, assignments, or other instructional activities	55.9	48.4	7.5	n/a	0.110	13.4	22.5	-9.1*	n/a	0.014
Acted on BT's request[c]	71.9	50.7	21.1*	n/a	0.000	12.0	20.5	-8.6*	n/a	0.015
Sample Size (Teachers)	258	245	503			241	231	472		

Source: Mathematica First and Third Induction Activities Surveys administered in fall 2005 and fall 2006 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures but are not indicated for dichotomous variables that are reported as percentages.

[b]The product of the mean frequency and mean average duration does not necessarily equal the mean of total time.

[c]Total sample size is 396 in fall 2005, 441 in fall 2006. The question did not apply to teachers who did not make a request to their mentors.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher; n/a = not applicable.

Table B.4. Impacts on Teacher-Reported Professional Development Activities During Past Three Months, Fall 2005 and Fall 2006: One-Year Districts

	Fall 2005					Fall 2006				
Aspect of Professional Development	Treatment	Control	Difference	Effect Size[a]	P-value	Treatment	Control	Difference	Effect Size[a]	P-value
Activities Completed (Percentages)										
Kept a written log	39.9	32.5	7.5	n/a	0.072	27.0	28.5	-1.5	n/a	0.718
Kept a portfolio and analysis of student work	71.6	77.5	-5.9	n/a	0.121	75.2	74.7	0.5	n/a	0.897
Worked with a study group of new teachers	65.5	34.4	31.0*	n/a	0.000	10.5	20.9	-10.4*	n/a	0.003
Worked with a study group of new and experienced teachers	47.8	42.1	5.7	n/a	0.182	37.8	39.8	-1.9	n/a	0.669
Observed others teaching in their classrooms	61.3	44.2	17.1*	n/a	0.000	28.0	26.3	1.7	n/a	0.685
Observed others teaching your class	51.1	50.6	0.5	n/a	0.913	26.9	32.1	-5.2	n/a	0.239
Met with principal to discuss teaching	68.8	70.4	-1.6	n/a	0.693	45.0	51.0	-6.0	n/a	0.232
Met with literacy or mathematics coach or other curricular specialist	77.5	77.1	0.4	n/a	0.900	77.8	75.8	1.9	n/a	0.668
Met with a resource specialist to discuss needs of particular students	65.5	77.2	-11.7*	n/a	0.005	70.8	77.8	-7.0	n/a	0.067
Frequency of Selected Activities (Number of Times During Past Three Months)										
Teaching was observed by mentor	4.0	1.5	2.5*	0.98	0.000	0.3	0.6	-0.3*	-0.21	0.024
Teaching was observed by principal	2.3	2.6	-0.3	-0.13	0.218	1.9	1.8	0.1	0.03	0.758
Given feedback on your teaching, not as part of formal evaluation	3.2	2.4	0.8*	0.37	0.000	1.4	1.6	-0.2	-0.11	0.259
Given feedback on your teaching, as part of formal evaluation	1.7	1.4	0.3	0.17	0.077	0.7	0.7	-0.1	-0.04	0.659
Given feedback on your lesson plans	1.6	1.7	-0.1	-0.04	0.683	1.0	1.4	-0.3	-0.17	0.079
Sample Size (Teachers)	258	245	503			241	231	472		

Source: Mathematica First and Third Induction Activities Surveys administered in fall 2005 and fall 2006 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures, but are not indicated for dichotomous variables that are reported as percentages.

*Significantly different from zero at the 0.05 level.

n/a = not applicable.

Table B.5. Impacts on Teacher-Reported Areas of Professional Development During the Past Three Months, Fall 2005 and Fall 2006: One-Year Districts

	Attended Professional Development Activities (Percentages)							
	Fall 2005				Fall 2006			
Professional Development Topic	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Parent and community relations	37.3	28.9	8.3	0.052	17.1	17.2	0.0	0.997
School policies on student disciplinary procedures	46.1	54.4	-8.3	0.052	47.6	47.9	-0.3	0.951
Instructional techniques/strategies	77.7	82.0	-4.3	0.297	71.0	68.9	2.1	0.664
Understanding the composition of students in your class	24.9	26.0	-1.1	0.773	21.1	23.5	-2.5	0.546
Content area knowledge (language arts, mathematics, science)	61.1	72.1	-10.9*	0.008	67.5	65.2	2.3	0.617
Lesson planning	30.2	32.1	-1.9	0.641	22.1	24.3	-2.1	0.591
Analyzing student work/assessment	44.7	50.1	-5.4	0.239	41.9	44.1	-2.2	0.635
Student motivation/engagement	36.2	35.5	0.7	0.876	24.5	24.5	-0.1	0.991
Differentiated instruction	52.5	49.0	3.6	0.466	42.0	45.9	-3.9	0.392
Using computers to support instruction	26.7	34.7	-7.9	0.062	38.7	38.6	0.1	0.984
Classroom management techniques	52.7	54.5	-1.8	0.711	23.7	30.2	-6.5	0.105
Preparing students for standardized testing	30.2	40.9	-10.8*	0.018	29.2	34.9	-5.8	0.177
Sample Size (Teachers)	258	245	503		241	231	472	

Source: Mathematica First and Third Induction Activities Surveys administered in fall 2005 and fall 2006 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Table B.6. Teacher Reports on Professional Support and Duties (Percentages), Spring 2006: One-Year Districts

	Spring 2006			
	Treatment	Control	Difference	P-value
BT has a mentor	90.3	80.3	10.0*	0.001
BT has an assigned mentor	89.7	72.2	17.5*	0.000
Sample Size (Teachers)	258	241	499	

Source: Mathematica Second Induction Activities Survey administered in spring 2006 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher.

Table B.7. Impacts on Teacher-Reported Mentor Profiles (Percentages), Spring 2006: One-Year Districts

	Spring 2006			
Mentoring Characteristic	Treatment	Control	Difference	P-value
Number of Mentors				
Multiple Mentors	21.7	11.8	10.0*	0.009
Number of Mentors				
None	9.7	19.7	-10.0*	0.001
One	68.6	68.6	0.0	0.998
Two	19.0	9.6	9.4*	0.008
Number of Mentors Assigned				
No mentor assigned	10.3	27.8	-17.5*	0.000
One mentor assigned	71.3	65.4	5.9	0.209
Two mentors assigned	18.4	6.8	11.6*	0.001
Mentor Positions				
Positions of All Mentors				
Full-time mentor	72.1	10.4	61.8*	0.000
Teacher	24.5	66.1	-41.6*	0.000
School or district administrator or staff external to district	10.9	6.3	4.6	0.072
No mentor	9.7	19.7	-10.0*	0.001
Sample Size (Teachers)	258	241	499	

Source: Mathematica Second Induction Activities Survey administered in spring 2006 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Table B.8. Impacts on Teacher-Reported Mentor Services Received in the Most Recent Full Week of Teaching, Spring 2006: One-Year Districts

	Spring 2006				
Mentor Service	Treatment	Control	Difference	Effect Size[a]	P-value
"Usual" Meetings with Mentors					
Frequency (number of meetings)	1.2	1.2	0.1	0.03	0.750
Average duration (minutes)	23.4	11.0	12.4*	0.68	0.000
Total time[b] (minutes)	55.9	33.8	22.1*	0.35	0.000
Informal Meetings with Mentors Total time (minutes)	28.8	33.7	-4.9	-0.12	0.197
Total Usual and Informal Time with Mentors (Minutes)	**84.7**	**67.5**	**17.2***	**0.20**	**0.039**
Meeting Time with Mentors in the Following Positions (Minutes)					
Full-time mentor	52.4	6.0	46.4*	0.88	0.000
Teacher	25.9	59.0	-33.1*	-0.41	0.000
Administrator	3.9	2.5	1.4	0.08	0.427
Staff external to district	2.3	0.2	2.1*	0.16	0.046
Mentor Time in the Following Activities (Minutes)					
Observing BT teaching	26.8	7.4	19.4*	0.65	0.000
Meeting with BT one-on-one	31.4	20.8	10.6*	0.33	0.000
Meeting with BT and other first-year teachers	23.8	6.0	17.8*	0.51	0.000
Meeting with BT and other teachers	12.2	13.7	-1.5	-0.05	0.548
Modeling a lesson	8.4	5.4	3.0	0.16	0.077
Co-teaching a lesson	5.1	5.6	-0.4	-0.02	0.820
All six activities (all mentors)	107.8	58.8	48.9*	0.49	0.000
All six activities (study mentor only)	95.0	0.0	95.0*	1.15	0.000
Types of Assistance a Mentor Provided (Percentage)					
Suggestions to improve practice	66.2	52.0	14.2*	n/a	0.001
Encouragement or moral support	77.7	67.8	9.9*	n/a	0.010
Opportunity to raise issues/discuss concerns	76.8	65.2	11.6*	n/a	0.003
Help with administrative/logistical issues	60.4	50.7	9.7*	n/a	0.022
Help with teaching to meet state or district standards	52.8	41.6	11.3*	n/a	0.007
Help identifying teaching challenges and solutions	63.6	52.4	11.2*	n/a	0.007
Discussed instructional goals and ways to achieve them	61.3	40.9	20.4*	n/a	0.000
Guidance on how to assess students	53.3	37.0	16.3*	n/a	0.000
Shared lesson plans, assignments, or other instructional activities	55.7	46.9	8.8*	n/a	0.049
Acted on BT's request[c]	61.2	46.3	14.8*	n/a	0.002
Sample Size (Teachers)	258	241	499		

Source: Mathematica Second Induction Activities Survey administered in spring 2006 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures but are not indicated for dichotomous variables that are reported as percentages.

[b]The product of the mean frequency and mean average duration does not necessarily equal the mean of total time.

[c]Total sample size is 390. The question did not apply to teachers who did not make a request to their mentors.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher; n/a = not applicable.

Table B.9. Impacts on Teacher-Reported Professional Development Activities During Past Three Months, Spring 2006: One-Year Districts

	Spring 2006				
Aspect of Professional Development	Treatment	Control	Difference	Effect Size[a]	P-value
Activities Completed (Percentages)					
Kept a written log	38.1	29.4	8.7*	n/a	0.036
Kept a portfolio and analysis of student work	77.3	72.7	4.6	n/a	0.250
Worked with a study group of new teachers	71.1	29.1	42.0*	n/a	0.000
Worked with a study group of new and experienced teachers	45.1	40.2	4.9	n/a	0.286
Observed others teaching in their classrooms	67.5	38.7	28.8*	n/a	0.000
Observed others teaching your class	44.7	39.3	5.4	n/a	0.264
Met with principal to discuss teaching	63.7	68.8	-5.1	n/a	0.288
Met with a literacy or mathematics coach or other curricular specialist	69.9	68.4	1.5	n/a	0.737
Met with a resource specialist to discuss needs of particular students	57.6	65.3	-7.7	n/a	0.085
Frequency of Selected Activities (Number of Times During Past Three Months)					
Teaching was observed by mentor	3.5	1.5	2.0*	0.83	0.000
Teaching was observed by principal	1.9	2.1	-0.2	-0.09	0.377
Given feedback on your teaching, not as part of formal evaluation	2.5	1.9	0.6*	0.30	0.001
Given feedback on your teaching, as part of formal evaluation	1.6	1.4	0.2	0.13	0.153
Given feedback on your lesson plans	1.3	1.5	-0.2	-0.13	0.187
Sample Size (Teachers)	258	241	499		

Source: Mathematica Second Induction Activities Survey administered in spring 2006 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures but are not indicated for dichotomous variables that are reported as percentages.

*Significantly different from zero at the 0.05 level.

n/a = not applicable.

Table B.10. Impacts on Teacher-Reported Areas of Professional Development During the Past Three Months, Spring 2006: One-Year Districts

	Attended Professional Development Activities (Percentages)			
	Spring 2006			
Professional Development Topic	Treatment	Control	Difference	P-value
Parent and community relations	24.7	22.9	1.8	0.635
School policies on student disciplinary procedures	32.1	44.9	-12.9*	0.006
Instructional techniques/strategies	70.1	73.5	-3.4	0.380
Understanding the composition of students in your class	20.6	21.3	-0.6	0.861
Content area knowledge (language arts, mathematics, science)	59.2	67.7	-8.5	0.051
Lesson planning	33.0	21.6	11.4*	0.005
Analyzing student work/assessment	52.4	42.7	9.7*	0.041
Student motivation/engagement	30.1	29.0	1.2	0.783
Differentiated instruction	49.1	44.3	4.8	0.283
Using computers to support instruction	24.3	32.0	-7.7	0.082
Classroom management techniques	33.2	39.9	-6.7	0.162
Preparing students for standardized testing	48.2	52.7	-4.5	0.218
Sample Size (Teachers)	258	241	499	

Source: Mathematica Second Induction Activities Survey administered in spring 2006 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Table B.11. Teacher Reports on Professional Support and Duties (Percentages), Fall 2007 and Fall 2008: One-Year Districts

	Fall 2007				Fall 2008			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
BT has mentor	20.0	25.3	-5.3	0.194	12.7	14.1	-1.4	0.699
BT has assigned mentor	11.3	13.6	-2.3	0.478	7.8	8.3	-0.6	0.826
Sample Size (Teachers)	219	207	426		206	192	398	

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher.

Table B.12. Impacts on Teacher-Reported Mentor Profiles (Percentages), Fall 2007 and Fall 2008: One-Year Districts

	Fall 2007				Fall 2008			
Mentoring Characteristic	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Number of Mentors								
None	80.0	74.7	5.3	0.194	87.3	85.9	1.4	0.699
One	16.3	19.5	-3.2	0.386	7.8	12.3	-4.4	0.145
Two or more	3.7	5.8	-2.1	0.353	5.3	2.2	3.1	0.118
Number of Mentors Assigned								
None	88.7	86.4	2.3	0.478	92.2	91.6	0.6	0.818
One or more	11.3	13.6	-2.3	0.478	7.8	8.4	-0.6	0.818
Mentor Positions								
Teacher	15.1	19.9	-4.9	0.184	10.8	9.8	0.9	0.767
Other Position	3.6	4.8	-1.2	0.572	2.5	5.9	-3.4	0.111
No mentor	80.0	74.7	5.3	0.194	87.3	85.9	1.4	0.699
Sample Size (Teachers)	219	207	426		206	192	398	

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

None of the differences is statistically significant at the 0.05 level.

Table B.13. Impacts on Teacher-Reported Mentor Services Received in the Most Recent Full Week of Teaching, Fall 2007 and Fall 2008: One-Year Districts

	Fall 2007					Fall 2008				
Mentor Service	Treatment	Control	Difference	Effect Size[a]	P-value	Treatment	Control	Difference	Effect Size[a]	P-value
"Usual" Meetings with Mentors										
Frequency (number of meetings)	0.3	0.4	-0.1	-0.12	0.267	0.2	0.1	0.1	0.10	0.354
Average duration (minutes)	1.4	2.8	-1.4*	-0.18	0.047	1.4	2.0	-0.6	-0.09	0.397
Total time[b] (minutes)	6.7	12.7	-6.0	-0.14	0.164	5.7	4.3	1.4	0.06	0.566
Informal Meetings with Mentors										
Total time (minutes)	6.7	10.5	-3.8	-0.13	0.198	7.0	7.0	0.0	0.00	0.992
Total Usual and Informal Time with Mentors (Minutes)	**13.3**	**23.1**	**-9.8**	**-0.15**	**0.138**	**12.7**	**11.3**	**1.5**	**0.03**	**0.770**
Meeting Time with Mentors in the Following Positions (Minutes)										
Full-time mentor	0.5	1.6	-1.2	-0.11	0.337	0.6	0.9	-0.3	-0.05	0.675
Teacher	11.4	22.1	-10.7	-0.16	0.128	10.6	7.9	2.6	0.07	0.560
Administrator	0.3	0.0	0.3	0.05	0.379	0.6	2.2	-1.7	-0.18	0.092
Staff external to district	0.0	0.0	0.0	0.10	0.316	0.0	0.2	-0.2	-0.11	0.323
Mentor Time in the Following Activities (Minutes)										
Observing BT teaching	1.6	2.6	-1.0	-0.10	0.350	1.2	0.4	0.8	0.10	0.227
Meeting with BT one-on-one	4.9	6.9	-2.0	-0.11	0.236	3.6	3.0	0.6	0.04	0.695
Meeting with BT and other first-year teachers	0.7	2.1	-1.4	-0.17	0.108	1.2	0.1	1.1	0.16	0.099
Meeting with BT and other teachers	4.4	7.7	-3.3	-0.17	0.094	3.7	1.6	2.0	0.15	0.134
Modeling a lesson	0.2	3.0	-2.8*	-0.25	0.021	0.7	0.3	0.4	0.08	0.566
Co-teaching a lesson	0.2	1.5	-1.3	-0.20	0.056	0.3	0.2	0.2	0.04	0.603
All six activities (all mentors)	12.1	23.8	-11.7*	-0.22	0.000	10.7	5.0	5.7	0.16	0.097
All six activities (study mentor only)	0.0	0.0	0.0			0.6	0.0	0.6	0.07	0.332
Types of Assistance a Mentor Provided (Percentage)										
Suggestions to improve practice	14.9	14.3	0.6	n/a	0.857	7.8	7.7	0.1	n/a	0.972
Encouragement or moral support	17.9	20.3	-2.4	n/a	0.534	11.0	10.8	0.2	n/a	0.963
Opportunity to raise issues/discuss concerns	17.1	20.2	-3.0	n/a	0.413	10.4	11.1	-0.7	n/a	0.830
Help with administrative/logistical issues	14.7	16.5	-1.8	n/a	0.590	7.5	9.1	-1.6	n/a	0.564
Help teaching to meet state or district standards	7.0	15.5	-8.5*	n/a	0.007	3.8	5.6	-1.8	n/a	0.414
Help identifying teaching challenges and solutions	12.9	16.3	-3.4	n/a	0.318	7.7	7.2	0.5	n/a	0.861
Discussed instructional goals and ways to achieve them	9.1	15.0	-5.9	n/a	0.062	5.8	5.7	0.2	n/a	0.945
Guidance on how to assess students	12.0	15.5	-3.5	n/a	0.299	5.1	5.2	-0.1	n/a	0.965
Shared lesson plans, assignments, or other instructional activities	11.8	16.8	-5.0	n/a	0.140	8.6	7.3	1.3	n/a	0.665
Acted on BT's request[c]	6.5	15.1	-8.6*	n/a	0.007	6.4	5.5	0.9	n/a	0.716
Sample Size (Teachers)	219	207	426			206	192	398		

Table B.13 ***(continued)***

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures but are not indicated for dichotomous variables that are reported as percentages.

[b]The product of the mean frequency and mean average duration does not necessarily equal the mean of total time.

[c]Total sample size is 393 in fall 2007 and 390 in fall 2008. The question did not apply to teachers who did not make a request to their mentors.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher; n/a = not applicable.

Table B.14. Impacts on Teacher-Reported Professional Development Activities During Past Three Months, Fall 2007 and Fall 2008: One-Year Districts

	Fall 2007					Fall 2008				
Aspect of Professional Development	Treatment	Control	Difference	Effect Size[a]	P-value	Treatment	Control	Difference	Effect Size[a]	P-value
Activities Completed (Percentages)										
Kept a written log	23.3	24.1	-0.8	n/a	0.833	21.3	20.3	0.9	n/a	0.818
Kept a portfolio and analysis of student work	77.7	73.9	3.7	n/a	0.356	73.2	79.7	-6.4	n/a	0.094
Worked with a study group of new teachers	13.1	12.2	0.9	n/a	0.772	13.7	16.5	-2.9	n/a	0.466
Worked with a study group of new and experienced teachers	41.8	41.5	0.4	n/a	0.935	38.0	36.3	1.7	n/a	0.735
Observed others teaching in their classrooms	29.2	31.7	-2.5	n/a	0.551	30.7	32.6	-1.9	n/a	0.686
Observed others teaching your class	27.1	28.8	-1.7	n/a	0.713	25.0	35.0	-10.0*	n/a	0.036
Met with principal to discuss teaching	44.1	49.3	-5.2	n/a	0.288	53.1	52.6	0.4	n/a	0.935
Met with literacy or mathematics coach or other curricular specialist	74.5	75.5	-0.9	n/a	0.832	70.2	73.3	-3.1	n/a	0.537
Met with a resource specialist to discuss needs of particular students	75.2	73.9	1.4	n/a	0.745	70.5	73.6	-3.0	n/a	0.482
Frequency of Selected Activities (Number of Times During Past Three Months)										
Teaching was observed by mentor	0.3	0.4	0.0	-0.04	0.697	0.2	0.3	-0.2	-0.18	0.096
Teaching was observed by principal	1.9	1.9	0.0	0.00	0.981	1.8	1.7	0.1	0.05	0.661
Given feedback on your teaching, not as part of formal evaluation	1.5	1.6	-0.1	-0.07	0.550	1.3	1.5	-0.2	-0.09	0.364
Given feedback on your teaching, as part of formal evaluation	0.7	0.8	0.0	-0.04	0.718	0.6	0.6	0.0	-0.03	0.814
Given feedback on your lesson plans	1.0	1.3	-0.2	-0.12	0.251	0.9	1.4	-0.4	-0.24	0.027
Sample Size (Teachers)	219	207	426			206	192	398		

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures but are not indicated for dichotomous variables that are reported as percentages.

*Significantly different from zero at the 0.05 level.

n/a = not applicable.

Table B.15. Impacts on Teacher-Reported Areas of Professional Development During the Past Three Months, Fall 2007 and Fall 2008: One-Year Districts

	Attended Professional Development Activities (Percentages)							
	Fall 2007				Fall 2008			
Professional Development Topic	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Parent and community relations	14.3	15.2	-0.9	0.808	12.3	15.1	-2.8	0.431
School policies on student disciplinary procedures	45.5	47.8	-2.3	0.643	48.8	52.0	-3.3	0.530
Instructional techniques/strategies	76.7	81.6	-4.9	0.191	74.8	80.4	-5.6	0.170
Understanding the composition of students in your class	22.4	26.0	-3.7	0.394	19.3	20.9	-1.5	0.676
Content area knowledge (language arts, mathematics, science)	63.5	72.2	-8.7	0.060	65.7	68.2	-2.5	0.607
Lesson planning	21.8	24.2	-2.5	0.565	20.4	28.0	-7.6	0.069
Analyzing student work/assessment	46.8	54.3	-7.5	0.116	42.1	55.1	-13.1*	0.009
Student motivation/engagement	21.1	28.2	-7.1	0.122	15.4	21.1	-5.6	0.138
Differentiated instruction	41.5	51.6	-10.2*	0.029	43.3	49.6	-6.3	0.211
Using computers to support instruction	44.0	38.9	5.1	0.290	39.5	42.1	-2.5	0.631
Classroom management techniques	20.2	21.1	-0.8	0.838	20.5	24.9	-4.4	0.290
Preparing students for standardized testing	24.8	26.2	-1.4	0.721	23.8	30.8	-7.0	0.147
Sample Size (Teachers)	219	207	426		206	192	398	

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Table B.16. Teacher Reports on Professional Support and Duties (Percentages), Fall 2005 and Fall 2006: Two-Year Districts

	Fall 2005				Fall 2006			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
BT has mentor	97.5	85.7	11.8*	0.001	80.4	41.0	39.4*	0.000
BT has assigned mentor	93.9	78.7	15.2*	0.000	80.0	33.5	46.6*	0.000
Sample Size (Teachers)	213	182	395		191	169	360	

Source: Mathematica First and Third Induction Activities Surveys administered in fall 2005 and fall 2006 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher.

Table B.17. Impacts on Teacher-Reported Mentor Profiles (Percentages), Fall 2005 and Fall 2006: Two-Year Districts

	Fall 2005				Fall 2006			
Mentoring Characteristic	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Number of Mentors								
Multiple Mentors	38.2	22.8	15.4*	0.002	10.6	13.2	-2.6	0.528
Number of Mentors								
None	2.5	14.3	-11.8*	0.001	19.5	59.0	-39.4*	0.000
One	59.3	63.0	-3.6	0.537	69.9	27.9	42.0*	0.000
Two	32.1	17.7	14.4*	0.001	10.6	13.2	-2.6	0.528
Number of Mentors Assigned								
No mentor assigned	6.1	21.3	-15.2*	0.000	20.0	66.5	-46.6*	0.000
One mentor assigned	62.8	65.7	-2.9	0.630	72.8	26.5	46.2*	0.000
Two mentors assigned	31.1	13.1	18.1*	0.000	7.3	6.9	0.3	0.905
Mentor Positions								
Positions of All Mentors	71.5	15.8	55.7*	0.000	63.6	6.5	57.1*	0.000
Full-time mentor								
Teacher	38.2	61.9	-23.7*	0.000	11.9	26.8	-14.8*	0.002
School or district administrator or staff external to district	13.2	14.7	-1.4	0.709	10.0	8.9	1.1	0.723
No mentor	2.5	14.3	-11.8*	0.001	19.5	59.0	-39.4*	0.000
Sample Size (Teachers)	213	182	395		191	169	360	

Source: Mathematica First and Third Induction Activities Surveys administered in fall 2005 and fall 2006 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Table B.18. Impacts on Teacher-Reported Mentor Services Received in Most Recent Full Week of Teaching, Fall 2005 and Fall 2006: Two-Year Districts

	Fall 2005					Fall 2006				
Mentor Service	Treatment	Control	Difference	Effect Size[a]	P-value	Treatment	Control	Difference	Effect Size[a]	P-value
"Usual" Meetings with Mentors										
Frequency (number of meetings)	1.7	1.4	0.4*	0.21	0.049	1.3	0.8	0.5*	0.29	0.011
Average duration (minutes)	24.4	11.5	12.9*	0.71	0.000	18.8	7.0	11.8*	0.68	0.000
Total time[b] (minutes)	78.5	43.3	35.2*	0.40	0.001	54.8	29.5	25.3*	0.28	0.032
Informal Meetings with Mentors										
Total time (minutes)	45.5	37.7	7.8	0.17	0.127	27.0	18.2	8.8	0.23	0.051
Total Usual and Informal Time with Mentors (Minutes)	**124.0**	**80.9**	**43.0***	**0.38**	**0.002**	**81.8**	**47.7**	**34.1***	**0.29**	**0.024**
Meeting Time with Mentors in the Following Positions (Minutes)										
Full-time mentor	74.8	6.4	68.4*	0.85	0.000	59.3	1.9	57.4*	0.90	0.000
Teacher	39.3	69.9	-30.6*	-0.34	0.003	14.2	41.9	-27.7*	-0.28	0.043
Administrator	6.5	2.4	4.1	0.21	0.093	6.2	3.2	3.0	0.14	0.173
Staff external to district	5.2	1.9	3.3	0.09	0.384	2.6	0.4	2.2	0.10	0.241
Mentor Time in the Following Activities (Minutes)										
Observing BT teaching	37.5	17.4	20.1*	0.55	0.000	21.8	7.4	14.3*	0.53	0.000
Meeting with BT one-on-one	42.5	23.2	19.2*	0.57	0.000	25.1	11.7	13.4*	0.42	0.000
Meeting with BT and other first-year teachers	37.7	11.4	26.3*	0.64	0.000	24.8	5.8	19.0*	0.52	0.000
Meeting with BT and other teachers	23.3	15.8	7.5	0.23	0.055	15.1	11.4	3.7	0.10	0.330
Modeling a lesson	16.3	9.7	6.6*	0.23	0.016	11.9	4.7	7.1*	0.30	0.003
Co-teaching a lesson	12.8	9.2	3.6	0.12	0.215	7.3	3.0	4.2	0.22	0.080
All six activities (all mentors)	169.9	86.8	83.2*	0.60	0.000	105.8	44.1	61.8*	0.48	0.000
All six activities (study mentor only)	118.7	0.0	118.7*	1.17	0.000	92.8	0.0	92.8*	0.97	0.000
Types of Assistance Mentor Provided (Percentages)										
Suggestions to improve practice	81.1	62.4	18.8*	n/a	0.000	62.4	22.9	39.5*	n/a	0.000
Encouragement or moral support	91.8	73.0	18.8*	n/a	0.000	72.3	29.5	42.8*	n/a	0.000
Opportunity to raise issues/discuss concerns	89.6	69.0	20.6*	n/a	0.000	71.9	28.1	43.8*	n/a	0.000
Help with administrative/logistical issues	73.6	59.7	13.9*	n/a	0.004	62.5	24.1	38.4*	n/a	0.000
Help teaching to meet state or district standards	67.8	50.8	16.9*	n/a	0.002	55.2	22.1	33.0*	n/a	0.000
Help identifying teaching challenges and solutions	81.9	57.5	24.5*	n/a	0.000	63.9	23.3	40.5*	n/a	0.000
Discussed instructional goals and ways to achieve them	75.4	48.4	27.0*	n/a	0.000	56.9	25.7	31.1*	n/a	0.000
Guidance on how to assess students	65.7	48.1	17.5*	n/a	0.001	49.6	21.0	28.6*	n/a	0.000
Shared lesson plans, assignments, or other instructional activities	69.9	53.7	16.3*	n/a	0.004	53.5	25.1	28.4*	n/a	0.000
Acted on BT's request[c]	77.9	50.0	27.9*	n/a	0.000	59.7	23.0	36.7*	n/a	0.000
Sample Size (Teachers)	213	182	395			191	169	360		

Table B.18 ***(continued)***

Source: Mathematica First and Third Induction Activities Surveys administered in fall 2005 and fall 2006 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures but are not indicated for dichotomous variables that are reported as percentages.

[b]The product of the mean frequency and mean average duration does not necessarily equal the mean of total time.

[c]Total sample size is 315 in fall 2005 and 313 in fall 2006. The question did not apply to teachers who did not make a request to their mentors.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher; n/a = not applicable.

Table B.19. Impacts on Teacher-Reported Professional Development Activities During Past Three Months, Fall 2005 and Fall 2006: Two-Year Districts

	Fall 2005					Fall 2006				
Aspect of Professional Development	Treatment	Control	Difference	Effect Size[a]	P-value	Treatment	Control	Difference	Effect Size[a]	P-value
Activities Completed (Percentages)										
Kept written log	40.3	33.5	6.7	n/a	0.221	33.5	31.6	1.9	n/a	0.699
Kept portfolio and analysis of student work	82.4	78.6	3.8	n/a	0.362	86.3	83.8	2.5	n/a	0.561
Worked with study group of new teachers	67.0	24.2	42.8*	n/a	0.000	41.6	19.2	22.4*	n/a	0.000
Worked with study group of new and experienced teachers	48.1	41.8	6.3	n/a	0.237	54.3	40.2	14.1*	n/a	0.008
Observed others teaching in their classrooms	58.2	48.6	9.6	n/a	0.084	48.7	38.3	10.3	n/a	0.090
Observed others teaching your class	46.9	47.0	0.0	n/a	0.995	38.5	38.5	0.1	n/a	0.991
Met with principal to discuss teaching	74.5	73.5	1.0	n/a	0.817	55.9	53.5	2.5	n/a	0.665
Met with literacy or mathematics coach or other curricular specialist	67.8	76.6	-8.9	n/a	0.087	67.8	68.4	-0.6	n/a	0.901
Met with a resource specialist to discuss needs of particular students	67.6	61.2	6.4	n/a	0.173	60.2	68.9	-8.7	n/a	0.072
Frequency of Selected Activities (Number of Times During Past Three Months)										
Teaching was observed by mentor	3.4	2.1	1.3*	0.56	0.000	2.3	0.8	1.6*	0.73	0.000
Teaching was observed by principal	2.0	2.4	-0.4	-0.22	0.062	1.8	1.7	0.1	0.05	0.674
Given feedback on your teaching, not as part of formal evaluation	2.8	2.5	0.3	0.12	0.266	1.9	1.5	0.4*	0.24	0.031
Given feedback on your teaching as part of formal evaluation	1.6	1.5	0.2	0.14	0.185	0.9	0.7	0.2	0.17	0.079
Given feedback on your lesson plans	2.0	2.0	0.0	-0.02	0.886	1.5	1.7	-0.2	-0.09	0.459
Sample Size (Teachers)	213	182	395			191	169	360		

Source: Mathematica First and Third Induction Activities Surveys administered in fall 2005 and fall 2006 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures but are not indicated for dichotomous variables that are reported as percentages.

*Significantly different from zero at the 0.05 level.

n/a = not applicable.

Table B.20. Impacts on Teacher-Reported Areas of Professional Development During Past Three Months, Fall 2005 and Fall 2006: Two-Year Districts

	Attended Professional Development Activities (Percentages)							
	Fall 2005				Fall 2006			
Professional Development Topic	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Parent and community relations	33.2	30.5	2.6	0.580	24.9	17.5	7.3	0.138
School policies on student disciplinary procedures	43.6	51.3	-7.7	0.151	38.4	43.2	-4.8	0.378
Instructional techniques/strategies	75.3	79.3	-4.0	0.337	65.6	69.0	-3.4	0.467
Understanding the composition of students in your class	30.3	23.1	7.2	0.142	23.8	18.8	5.0	0.268
Content area knowledge (language arts, mathematics, science)	63.5	71.8	-8.3	0.064	59.7	55.7	4.0	0.411
Lesson planning	36.8	37.0	-0.2	0.976	32.8	27.9	4.9	0.306
Analyzing student work/assessment	44.7	42.8	1.9	0.716	42.2	38.5	3.7	0.488
Student motivation/engagement	47.5	38.8	8.8	0.116	28.4	24.7	3.7	0.433
Differentiated instruction	55.9	46.8	9.1	0.121	41.6	41.2	0.4	0.939
Using computers to support instruction	35.0	36.3	-1.3	0.798	37.3	34.0	3.3	0.510
Classroom management techniques	60.8	47.8	13.0*	0.012	28.1	22.2	5.9	0.155
Preparing students for standardized testing	30.3	35.7	-5.5	0.261	28.0	31.6	-3.7	0.476
Sample Size (Teachers)	213	182	395		191	169	360	

Source: Mathematica First and Third Induction Activities Surveys administered in fall 2005 and fall 2006 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Table B.21. Teacher Reports on Professional Support and Duties (Percentages), Spring 2006 and Spring 2007: Two-Year Districts

	Spring 2006				Spring 2007			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
BT has a mentor	98.4	85.5	12.9*	0.000	87.4	47.1	40.3*	0.000
BT has an assigned mentor	95.9	78.9	17.0*	0.000	83.8	39.6	44.3*	0.000
Sample Size (Teachers)	210	176	386		203	169	372	

Source: Mathematica Second Induction Activities Survey administered in spring 2006 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher.

Table B.22. Impacts on Teacher-Reported Mentor Profiles (Percentages), Spring 2006 and Spring 2007: Two-Year Districts

	Spring 2006				Spring 2007			
Mentoring Characteristic	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Number of Mentors								
Multiple Mentors	37.8	22.4	15.4*	0.005	17.6	12.6	5.1	0.197
Number of Mentors								
None	1.6	14.5	-12.9*	0.000	12.6	52.9	-40.3*	0.000
One	61.0	63.2	-2.2	0.722	69.7	34.6	35.2*	0.000
Two	31.5	18.4	13.1*	0.016	13.4	12.6	0.8	0.829
Number of Mentors Assigned								
No mentor assigned	4.1	21.1	-17.0*	0.000	16.2	60.4	-44.3*	0.000
One mentor assigned	64.6	61.9	2.7	0.668	73.6	31.7	41.9*	0.000
Two mentors assigned	31.3	17.0	14.3*	0.003	10.3	7.9	2.4	0.470
Mentor Positions								
Positions of All Mentors								
Full-time mentor	74.5	16.6	57.9*	0.000	67.4	15.0	52.3*	0.000
Teacher	38.8	65.4	-26.6*	0.000	15.7	26.9	-11.2*	0.025
School or district administrator or staff external to district	14.1	12.5	1.6	0.671	10.9	8.5	2.4	0.444
No mentor	1.6	14.5	-12.9*	0.000	12.6	52.9	-40.3*	0.000
Sample Size (Teachers)	210	176	386		203	169	372	

Source: Mathematica Second Induction Activities Survey administered in spring 2006 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Table B.23. Impacts on Teacher-Reported Mentor Services Received in the Most Recent Full Week of Teaching, Spring 2006 and Spring 2007: Two-Year Districts

	Spring 2006					Spring 2007				
Mentor Service	Treatment	Control	Difference	Effect Size[a]	P-value	Treatment	Control	Difference	Effect Size[a]	P-value
"Usual" Meetings with Mentors										
Frequency (number of meetings)	1.5	1.2	0.3	0.15	0.172	1.1	0.7	0.5*	0.33	0.003
Average duration (minutes)	23.4	11.2	12.1*	0.66	0.000	19.5	6.1	13.4*	0.81	0.000
Total time[b] (minutes)	62.3	43.2	19.1	0.21	0.101	50.3	21.5	28.8*	0.43	0.000
Informal Meetings with Mentors										
Total time (minutes)	45.3	39.1	6.2	0.14	0.188	28.4	19.5	8.9*	0.25	0.028
Total Usual and Informal Time with Mentors (Minutes)	**107.6**	**82.4**	**25.3**	**0.21**	**0.087**	**78.7**	**41.0**	**37.7***	**0.42**	**0.001**
Meeting Time with Mentors in the Following Positions (Minutes)										
Full-time mentor	70.8	9.6	61.2*	0.80	0.000	54.3	6.1	48.2*	0.77	0.000
Teacher	31.6	69.3	-37.7*	-0.35	0.006	18.5	22.8	-4.3	-0.08	0.524
Administrator	3.9	2.5	1.4	0.10	0.417	6.2	3.6	2.5	0.12	0.239
Staff external to district	0.8	1.9	-1.1	-0.11	0.281	2.0	1.7	0.3	0.03	0.806
Mentor Time in the Following Activities (Minutes)										
Observing BT teaching	25.6	15.6	10.0*	0.30	0.003	19.1	8.1	11.1*	0.48	0.000
Meeting with BT one-on-one	38.2	21.2	17.0*	0.53	0.000	29.2	10.1	19.1*	0.66	0.000
Meeting with BT and other first-year teachers	34.8	9.1	25.8*	0.65	0.000	23.7	4.5	19.2*	0.56	0.000
Meeting with BT and other teachers	22.4	17.0	5.4	0.16	0.137	15.5	8.1	7.3*	0.24	0.024
Modeling a lesson	14.1	8.0	6.1*	0.23	0.027	10.0	3.6	6.4*	0.33	0.005
Co-teaching a lesson	10.8	6.5	4.3	0.16	0.082	7.8	1.5	6.3*	0.40	0.000
All six activities (all mentors)	146.3	76.9	69.3*	0.50	0.000	105.3	36.0	69.4*	0.62	0.000
All six activities (study mentor only)	108.7	0.0	108.7*	0.99	0.000	82.3	0.0	82.3*	0.87	0.000
Types of Assistance a Mentor Provided (Percentage)										
Suggestions to improve practice	83.2	52.5	30.7*	n/a	0.000	68.0	27.4	40.6*	n/a	0.000
Encouragement or moral support	92.4	70.4	22.0*	n/a	0.000	77.9	37.6	40.4*	n/a	0.000
Opportunity to raise issues/discuss concerns	90.0	62.3	27.7*	n/a	0.000	76.1	36.5	39.7*	n/a	0.000
Help with administrative/logistical issues	76.6	53.2	23.4*	n/a	0.000	59.6	29.0	30.6*	n/a	0.000
Help with teaching to meet state or district standards	69.6	47.7	21.9*	n/a	0.000	58.5	25.3	33.3*	n/a	0.000
Help identifying teaching challenges and solutions	80.7	51.8	28.9*	n/a	0.000	66.0	29.5	36.5*	n/a	0.000
Discussed instructional goals and ways to achieve them	79.1	48.1	31.0*	n/a	0.000	65.5	24.4	41.0*	n/a	0.000
Guidance on how to assess students	72.3	43.5	28.8*	n/a	0.000	58.3	19.1	39.2*	n/a	0.000
Shared lesson plans, assignments, or other instructional activities	71.0	50.5	20.5*	n/a	0.000	59.7	22.3	37.4*	n/a	0.000
Acted on BT's request[c]	75.9	54.2	21.7*	n/a	0.000	60.4	23.5	37.0*	n/a	0.000
Sample Size (Teachers)	210	176	386			203	169	372		

Table B.23 ***(continued)***

Source: Mathematica Second Induction Activities Survey administered in spring 2006 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures but are not indicated for dichotomous variables that are reported as percentages.

[b]The product of the mean frequency and mean average duration does not necessarily equal the mean of total time.

[c]Total sample size is 306 in spring 2006 and 325 in spring 2007. The question did not apply to teachers who did not make a request to their mentors.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher; n/a = not applicable.

Table B.24. Impacts on Teacher-Reported Professional Development Activities During Past Three Months, Spring 2006 and Spring 2007: Two-Year Districts

	Spring 2006					Spring 2007				
Aspect of Professional Development	Treatment	Control	Difference	Effect Size[a]	P-value	Treatment	Control	Difference	Effect Size[a]	P-value
Activities Completed (Percentages)										
Kept a written log	41.6	26.1	15.5*	n/a	0.003	33.1	24.7	8.4	n/a	0.081
Kept a portfolio and analysis of student work	78.6	76.3	2.3	n/a	0.580	83.1	76.4	6.7	n/a	0.131
Worked with a study group of new teachers	64.1	24.9	39.2*	n/a	0.000	48.2	14.1	34.1*	n/a	0.000
Worked with a study group of new and experienced teachers	48.3	33.9	14.4*	n/a	0.003	51.1	35.9	15.2*	n/a	0.004
Observed others teaching in their classrooms	71.5	46.7	24.9*	n/a	0.000	47.0	35.2	11.9*	n/a	0.020
Observed others teaching your class	48.0	36.0	12.0*	n/a	0.029	36.1	36.4	-0.3	n/a	0.953
Met with principal to discuss teaching	73.2	70.1	3.1	n/a	0.541	68.9	65.4	3.5	n/a	0.460
Met with a literacy or mathematics coach or other curricular specialist	66.2	63.9	2.3	n/a	0.665	63.1	63.1	0.0	n/a	0.999
Met with a resource specialist to discuss needs of particular students	64.0	59.1	4.9	n/a	0.347	62.2	65.8	-3.6	n/a	0.474
Frequency of Selected Activities (Number of Times During Past Three Months)										
Teaching was observed by mentor	3.2	1.6	1.6*	0.69	0.000	2.5	1.0	1.5*	0.66	0.000
Teaching was observed by principal	2.3	1.9	0.4	0.19	0.121	2.0	1.8	0.2	0.10	0.354
Given feedback on your teaching, not as part of formal evaluation	2.5	2.0	0.5*	0.24	0.031	2.2	1.5	0.7*	0.37	0.001
Given feedback on your teaching, as part of formal evaluation	1.8	1.5	0.3	0.18	0.093	1.6	1.3	0.3*	0.21	0.046
Given feedback on your lesson plans	1.9	1.6	0.3	0.15	0.175	1.5	1.5	0.0	0.01	0.964
Sample Size (Teachers)	210	176	386			203	169	372		

Source: Mathematica Second Induction Activities Survey administered in spring 2006 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures but are not indicated for dichotomous variables that are reported as percentages.

*Significantly different from zero at the 0.05 level.

n/a = not applicable.

Table B.25. Impacts on Teacher-Reported Areas of Professional Development During the Past Three Months, Spring 2006 and Spring 2007: Two-Year Districts

	Attended Professional Development Activities (Percentages)							
	Spring 2006				Spring 2007			
Professional Development Topic	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Parent and community relations	28.2	24.3	3.9	0.423	27.7	23.6	4.0	0.438
School policies on student disciplinary procedures	39.0	34.4	4.5	0.320	36.0	36.7	-0.7	0.893
Instructional techniques/strategies	80.4	73.2	7.2	0.154	74.4	72.2	2.1	0.662
Understanding the composition of students in your class	31.6	21.5	10.1*	0.033	24.9	23.7	1.2	0.811
Content area knowledge (language arts, mathematics, science)	69.9	60.3	9.6*	0.040	62.2	57.5	4.7	0.355
Lesson planning	42.9	31.2	11.7*	0.019	37.5	27.8	9.6*	0.038
Analyzing student work/assessment	60.4	40.6	19.8*	0.000	56.5	45.0	11.5*	0.034
Student motivation/engagement	42.7	33.5	9.1	0.071	42.7	23.6	19.0*	0.000
Differentiated instruction	62.0	47.0	15.0*	0.010	58.4	43.3	15.1*	0.006
Using computers to support instruction	36.0	34.3	1.7	0.727	37.9	40.9	-3.0	0.601
Classroom management techniques	53.3	33.8	19.5*	0.000	26.1	21.9	4.2	0.347
Preparing students for standardized testing	43.8	50.5	-6.8	0.152	49.1	48.0	1.1	0.838
Sample Size (Teachers)	210	176	386		203	169	372	

Source: Mathematica Second Induction Activities Survey administered in spring 2006 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Table B.26. Teacher Reports on Professional Support and Duties (Percentages), Fall 2007 and Fall 2008: Two-Year Districts

	Fall 2007				Fall 2008			
	Treatment	**Control**	**Difference**	**P-value**	**Treatment**	**Control**	**Difference**	**P-value**
BT has mentor	14.5	23.3	-8.8	0.066	9.7	17.0	-7.3*	0.043
BT has assigned mentor	11.3	18.8	-7.5	0.077	6.9	11.8	-4.8	0.146
Sample Size (Teachers)	179	147	326		178	143	321	

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher.

Table B.27. Impacts on Teacher-Reported Mentor Profiles (Percentages), Fall 2007 and Fall 2008: Two-Year Districts

	Fall 2007				Fall 2008			
Mentoring Characteristic	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Number of Mentors								
None	85.5	76.7	8.8	0.066	90.3	83.0	7.3*	0.043
One	11.3	19.0	-7.7	0.091	9.1	13.7	-4.5	0.162
Two or more	3.3	5.0	-1.6	0.495	2.4	6.1	-3.7	0.126
Number of Mentors Assigned								
None	88.6	81.1	7.6	0.076	93.0	87.9	5.1	0.135
One or more	11.4	18.9	-7.6	0.076	7.0	12.1	-5.1	0.135
Mentor Positions								
Teacher	7.6	17.1	-9.5*	0.023	6.4	10.6	-4.2	0.179
Other position	4.2	6.6	-2.4	0.368	2.4	6.2	-3.8	0.122
No mentor	85.5	76.7	8.8	0.066	90.3	83.0	7.3*	0.043
Sample Size (Teachers)	179	147	326		178	143	321	

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

n/a = not applicable.

Table B.28. Impacts on Teacher-Reported Mentor Services Received in Most Recent Full Week of Teaching, Fall 2007 and Fall 2008: Two-Year Districts

	Fall 2007					Fall 2008				
Mentor Service	Treatment	Control	Difference	Effect Size[a]	P-value	Treatment	Control	Difference	Effect Size[a]	P-value
"Usual" Meetings with Mentors										
Frequency (number of meetings)	0.2	0.3	-0.1	-0.11	0.337	0.1	0.3	-0.2	-0.19	0.182
Average duration (minutes)	1.7	2.7	-1.0	-0.12	0.352	1.2	2.1	-0.9	-0.13	0.334
Total time[b] (minutes)	6.7	8.7	-2.0	-0.06	0.630	5.7	8.5	-2.8	-0.09	0.540
Informal Meetings with Mentors										
Total time (minutes)	5.7	5.9	-0.2	-0.01	0.939	4.9	9.4	-4.5	-0.22	0.114
Total Usual and Informal Time with Mentors (Minutes)	**12.4**	**14.5**	**-2.1**	**-0.04**	**0.731**	**10.5**	**17.8**	**-7.3**	**-0.15**	**0.263**
Meeting Time with Mentors in the Following Positions (Minutes)										
Full-time mentor	0.3	1.6	-1.3	-0.12	0.307	0.2	2.5	-2.3	-0.21	0.171
Teacher	10.6	13.7	-3.0	-0.06	0.642	9.6	10.8	-1.2	-0.03	0.845
Administrator	0.2	0.4	-0.2	-0.07	0.401	0.0	2.7	-2.7	-0.26	0.081
Staff external to district	0.0	0.0	0.0			0.5	0.8	-0.3	-0.04	0.775
Mentor Time in the Following Activities (Minutes)										
Observing BT teaching	3.7	1.4	2.3	0.17	0.183	3.0	3.0	-0.1	-0.01	0.969
Meeting with BT one-on-one	7.0	4.2	2.8	0.16	0.217	2.6	4.4	-1.8	-0.14	0.332
Meeting with BT and other first-year teachers	3.8	1.4	2.4	0.18	0.160	1.2	3.0	-1.9	-0.17	0.282
Meeting with BT and other teachers	2.9	5.2	-2.3	-0.11	0.394	1.4	6.2	-4.8	-0.30	0.058
Modeling a lesson	2.0	1.6	0.3	0.03	0.813	0.7	1.4	-0.7	-0.09	0.562
Co-teaching a lesson	1.3	0.4	0.9	0.12	0.306	0.0	0.2	-0.1	-0.13	0.326
All six activities (all mentors)	20.6	14.2	6.4	0.10	0.422	9.1	17.4	-8.3	-0.17	0.266
All six activities (study mentor only)	5.1	0.0	5.1	0.23	0.201	0.2	0.0	0.2	0.08	0.338
Types of Assistance Mentor Provided (Percentage)										
Suggestions to improve practice	7.9	12.8	-4.9	n/a	0.186	5.6	10.1	-4.5	n/a	0.150
Encouragement or moral support	13.1	17.3	-4.2	n/a	0.334	8.2	13.8	-5.6	n/a	0.096
Opportunity to raise issues/discuss concerns	12.2	16.4	-4.2	n/a	0.301	7.2	12.6	-5.4	n/a	0.102
Help with administrative/logistical issues	10.3	13.1	-2.8	n/a	0.412	4.8	9.3	-4.5	n/a	0.127
Help teaching to meet state or district standards	7.8	10.8	-2.9	n/a	0.410	5.4	8.8	-3.4	n/a	0.243
Help identifying teaching challenges and solutions	10.3	12.4	-2.1	n/a	0.558	4.1	8.9	-4.9	n/a	0.069
Discussed instructional goals and ways to achieve them	10.1	10.2	-0.1	n/a	0.966	4.3	8.8	-4.5	n/a	0.106
Guidance on how to assess students	8.0	10.8	-2.9	n/a	0.383	5.4	8.1	-2.7	n/a	0.324
Shared lesson plans, assignments, or other instructional activities	9.1	11.6	-2.4	n/a	0.500	4.9	13.4	-8.5*	n/a	0.005
Acted on BT's request[c]	8.6	8.1	0.5	n/a	0.888	5.7	10.0	-4.3	n/a	0.132
Sample Size (Teachers)	179	147	326			178	143	321		

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Table B.28 ***(continued)***

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures but are not indicated for dichotomous variables that are reported as percentages.

[b]The product of the mean frequency and mean average duration does not necessarily equal the mean of total time.

[c]Total sample size is 305 in fall 2007 and 318 in fall 2008. The question did not apply to teachers who did not make a request to their mentors.

*Significantly different from zero at the 0.05 level.

BT = beginning teacher; n/a = not applicable.

Table B.29. Impacts on Teacher-Reported Professional Development Activities During Past Three Months, Fall 2007 and Fall 2008: Two-Year Districts

	Fall 2007					Fall 2008				
Aspect of Professional Development	Treatment	Control	Difference	Effect Size[a]	P-value	Treatment	Control	Difference	Effect Size[a]	P-value
Activities Completed (Percentages)										
Kept written log	26.6	24.8	1.8	n/a	0.724	26.9	20.6	6.3	n/a	0.169
Kept portfolio and analysis of student work	84.4	81.1	3.3	n/a	0.475	80.5	84.5	-4.0	n/a	0.324
Worked with study group of new teachers	13.8	11.6	2.2	n/a	0.539	12.2	17.9	-5.7	n/a	0.221
Worked with study group of new and experienced teachers	35.9	34.9	1.0	n/a	0.863	38.9	38.4	0.5	n/a	0.932
Observed others teaching in their classrooms	29.7	32.8	-3.2	n/a	0.569	29.5	37.8	-8.4	n/a	0.136
Observed others teaching your class	28.0	34.2	-6.1	n/a	0.323	23.3	28.8	-5.5	n/a	0.261
Met with principal to discuss teaching	59.7	52.5	7.2	n/a	0.198	50.4	50.7	-0.3	n/a	0.953
Met with literacy or mathematics coach or other curricular specialist	62.6	63.7	-1.2	n/a	0.845	67.3	70.5	-3.2	n/a	0.500
Met with a resource specialist to discuss needs of particular students	66.9	63.2	3.8	n/a	0.476	64.2	69.0	-4.8	n/a	0.367
Frequency of Selected Activities (Number of Times During Past Three Months)										
Teaching was observed by mentor	0.3	0.4	-0.1	-0.07	0.497	0.2	0.2	-0.1	-0.08	0.528
Teaching was observed by principal	1.7	1.4	0.3	0.18	0.068	1.5	1.4	0.1	0.05	0.657
Given feedback on your teaching, not as part of formal evaluation	1.3	1.1	0.2	0.12	0.303	1.5	1.0	0.5*	0.30	0.007
Given feedback on your teaching as part of formal evaluation	0.7	0.7	0.0	-0.03	0.819	0.7	0.6	0.1	0.10	0.369
Given feedback on your lesson plans	1.3	1.4	0.0	-0.02	0.886	1.5	1.5	0.0	-0.01	0.905
Sample Size (Teachers)	179	147	326			178	143	321		

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Effect sizes are reported for continuous measures but are not indicated for dichotomous variables that are reported as percentages.

*Significantly different from zero at the 0.05 level.

n/a = not applicable.

Table B.30. Impacts on Teacher-Reported Areas of Professional Development During Past Three Months, Fall 2007 and Fall 2008: Two-Year Districts

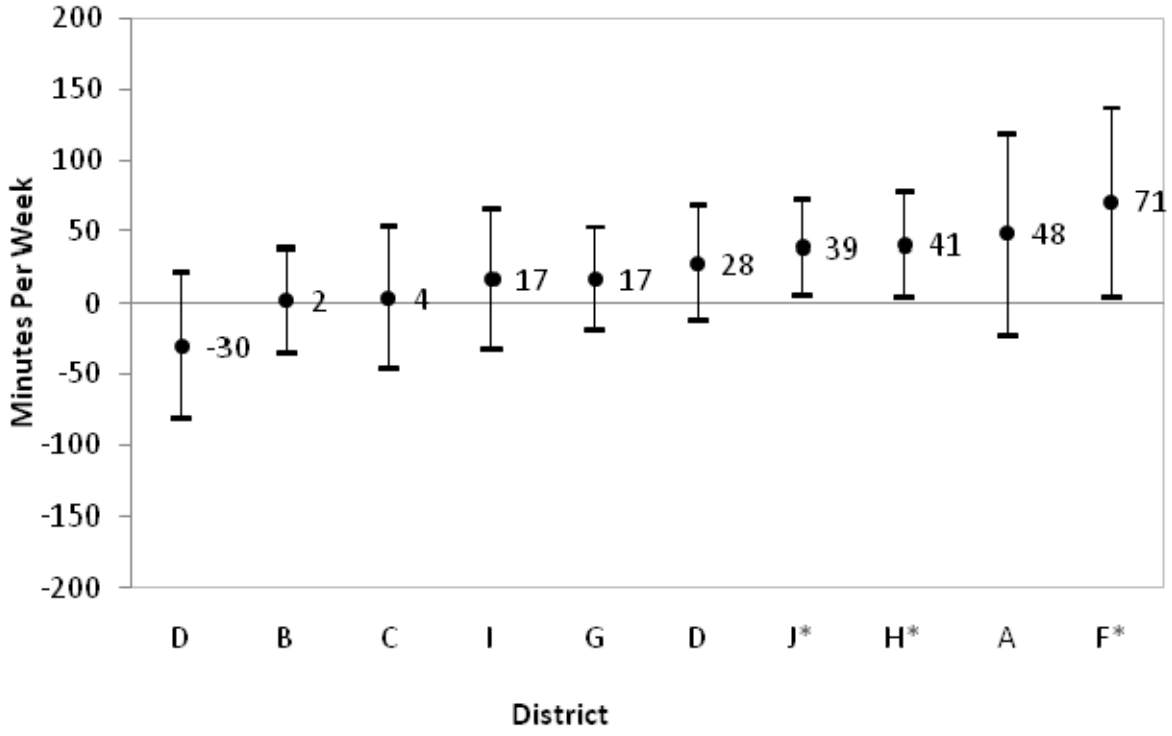
	Attended Professional Development Activities (Percentages)							
	Fall 2007				Fall 2008			
Professional Development Topic	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Parent and community relations	15.4	19.5	-4.1	0.340	19.0	15.2	3.7	0.388
School policies on student disciplinary procedures	43.7	41.4	2.3	0.714	44.1	52.0	-7.8	0.178
Instructional techniques/strategies	77.5	64.3	13.3*	0.008	78.2	77.7	0.5	0.920
Understanding the composition of students in your class	21.4	21.0	0.4	0.933	24.7	24.8	-0.1	0.982
Content area knowledge (language arts, mathematics, science)	66.9	61.8	5.2	0.319	67.7	57.9	9.8	0.060
Lesson planning	30.3	31.8	-1.5	0.785	27.6	29.0	-1.3	0.795
Analyzing student work/assessment	46.0	38.0	7.9	0.131	54.7	39.4	15.3*	0.011
Student motivation/engagement	24.0	21.4	2.6	0.568	29.3	23.3	6.0	0.203
Differentiated instruction	53.0	50.2	2.8	0.640	46.1	46.3	-0.1	0.984
Using computers to support instruction	38.1	42.6	-4.5	0.440	39.7	42.3	-2.7	0.642
Classroom management techniques	26.3	21.6	4.6	0.346	23.2	22.5	0.7	0.889
Preparing students for standardized testing	20.9	28.3	-7.4	0.082	33.1	30.5	2.6	0.628
Sample Size (Teachers)	179	147	326		178	143	321	

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Figure B.1. Treatment-Control Differences in Total Minutes Spent in Mentoring per Week, Fall 2005: One-Year Districts

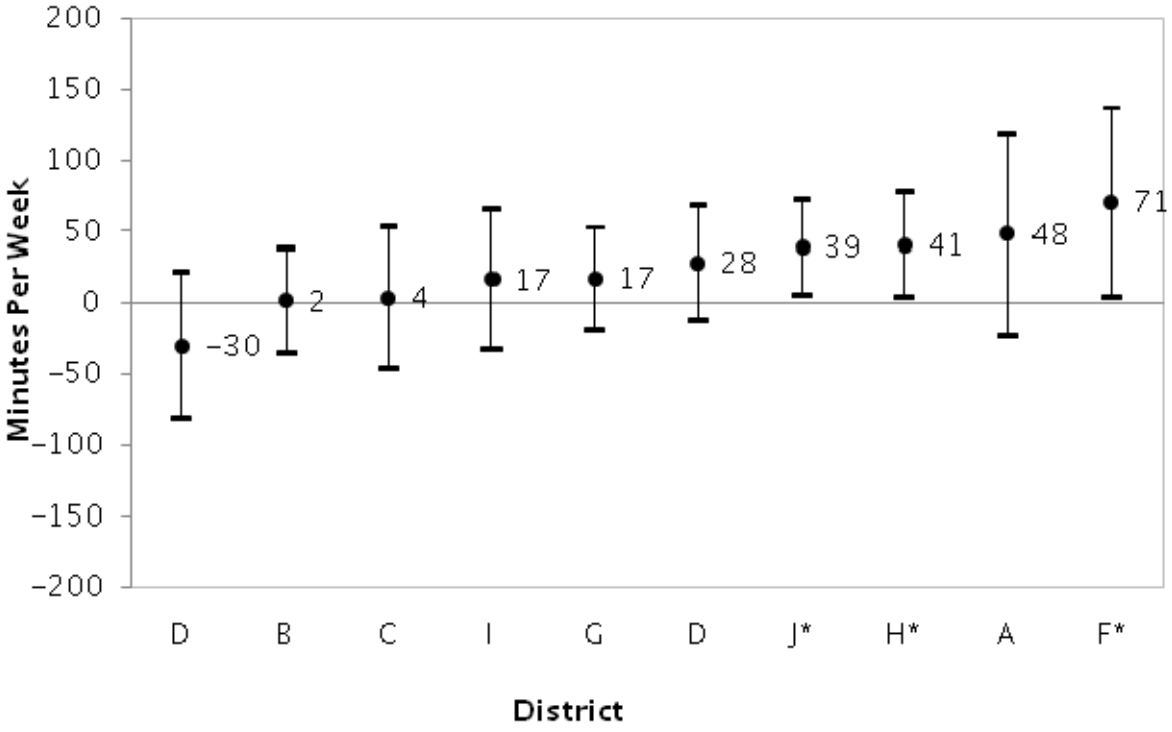


Source: Mathematica First Induction Activities Survey administered in fall 2005 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes A through J are arbitrary. Districts are ordered according to the size of the treatment-control difference. N=503 teachers.

*Significantly different from zero at the 0.05 level. (No adjustment is applied for multiple comparisons.)

Figure B.2. Treatment-Control Differences in Total Minutes Spent in Mentoring per Week, Fall 2006: One-Year Districts

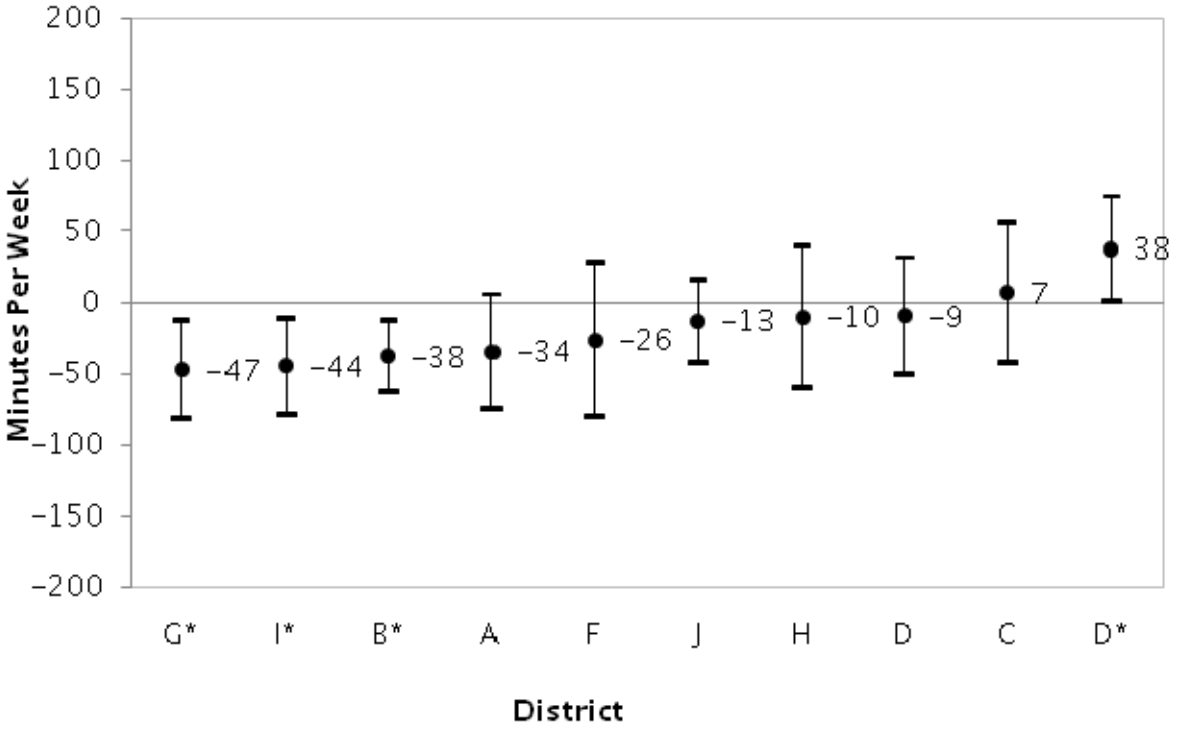


Source: Mathematica Third Induction Activities Survey administered in fall 2006 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes A through J are arbitrary. Districts are ordered according to the size of the treatment-control difference. N=472 teachers.

*Significantly different from zero at the 0.05 level. (No adjustment is applied for multiple comparisons.)

Figure B.3. Treatment-Control Differences in Total Minutes Spent in Mentoring per Week, Fall 2005: Two-Year Districts

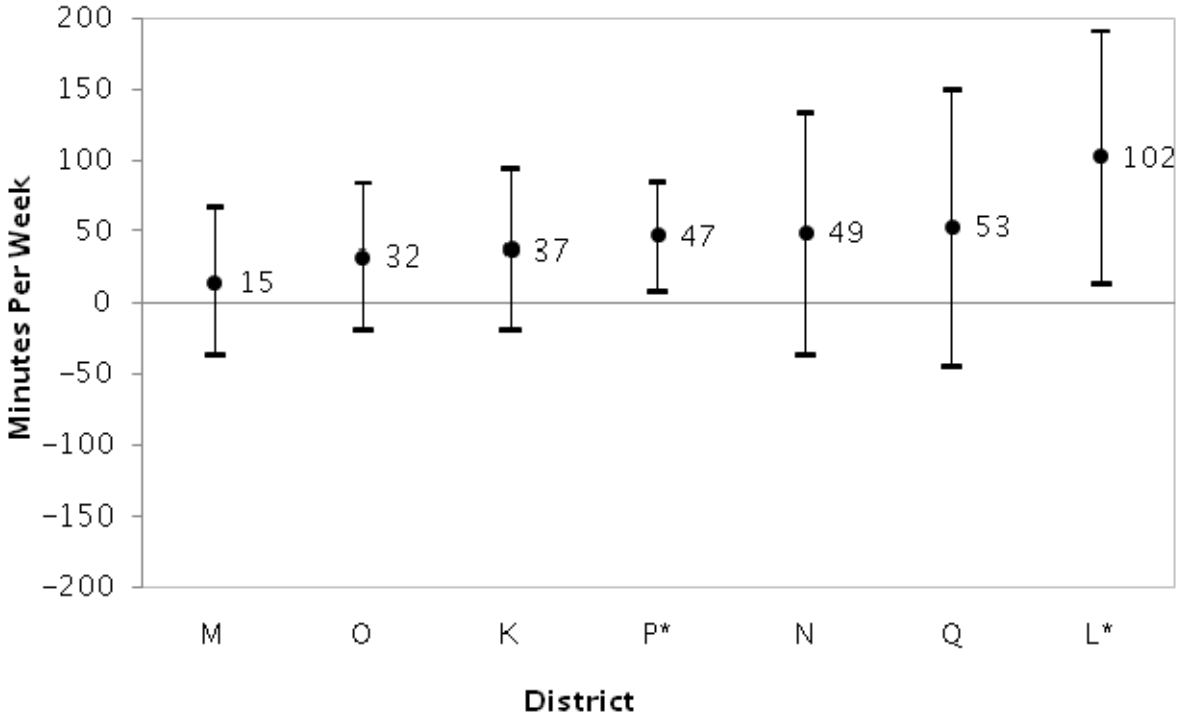


Source: Mathematica First Induction Activities Survey administered in fall 2005 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes K through Q are arbitrary. Districts are ordered according to the size of the treatment-control difference. N=395 teachers.

*Significantly different from zero at the 0.05 level. (No adjustment is applied for multiple comparisons.)

Figure B.4. Treatment-Control Differences in Total Minutes Spent in Mentoring per Week, Fall 2006: Two-Year Districts



Source: Mathematica Third Induction Activities Survey administered in fall 2006 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes K through Q are arbitrary. Districts are ordered according to the size of the treatment-control difference. N=360 teachers.

*Significantly different from zero at the 0.05 level. (No adjustment is applied for multiple comparisons.)

This page is intentionally left blank.

APPENDIX C

SENSITIVITY ANALYSES AND SUPPLEMENTAL INFORMATION FOR CHAPTER V

Chapter V presented estimates of the impacts of comprehensive induction on classroom practices and student test scores. This appendix supplements that analysis with additional results on the sensitivity of the benchmark findings to alternative assumptions and with additional tables providing detail on the sample sizes used to produce the estimates.

A. Supplementary Information and Sensitivity Analysis for Classroom Practices

We re-estimated the impacts on classroom practices by using a variety of assumptions about item scoring and estimation and found that the results did not change substantially.

The results were not sensitive to how we grouped the individual items into constructs. We performed a factor analysis of the 16 classroom observation items to explore the degree to which the theoretical groupings were empirically justified. In finding the groupings justified, we maintained the three-construct scoring method (implementation of literacy lesson, content of literacy lesson, and classroom culture) described in Chapter V. Although the factor analyses were consistent with the theoretical groupings, they did suggest that the implementation and content items could be grouped together, forming one construct rather than two. (Factor loadings for the 16 class observation items are shown in Table A.5 in Appendix A.) When we substituted a single construct that included all implementation and content items in place of two constructs, there were no statistically significant impacts.

The findings did not change when we collapsed the scale or divided the sample into two subgroups. As part of our sensitivity analyses, we estimated the model separately for each classroom observation item after recoding each score from a five-point scale into a binary variable: (1) no, limited, or moderate evidence or (2) consistent or extensive evidence of good practice. This dichotomous coding scheme allowed us to compare the percentages of treatment and control teachers who demonstrated "consistent" or "extensive" evidence of good practice in the classroom. The results, however, support the same conclusions of no impact. Impact estimates for each of the 16 class observation items are shown in Table C.1.

The results were not sensitive to the choice of summary score. In addition to scoring individual items under each domain, classroom observers reported a summary score for each of the three domains. They based the summary score on a five-point scale that could differ from our constructed domain scores in two ways. First, they reported the score as an integer such that they had to round off to the nearest whole number and thus could have recorded numbers that differ from the average score. Second and more significant, observers could exercise their discretion in assigning an overall domain score. Thus, if indicator scores were 3, 3, 3, 4, and 4 for the five indicators, respectively, then an observer, in reporting the overall domain score, could round up to 4 instead of down to 3 based on a judgment that the last two domains are more important than the first three. Observers could also justify an overall score of 4 if the item scores of 3 were actually rounded down from, say, 3.4 and the item scores of 4 had been rounded down from 4.4. (The average of 3.4, 3.4, 3.4, 4.4, and 4.4 is 3.8, which rounds up to 4.)

The two types of summary scores were not identical. Given that each has advantages and disadvantages, we had to choose one arbitrarily to include for the benchmark estimates presented in Chapter V. When we substituted the observer summary scores for the computed average scores, we reached the same conclusions: no statistically significant impact of treatment (see Table C.2).

We also estimated the model separately for one-year and two-year districts (see Tables C.3 and C.4) and found that the impact estimates were not significantly different from zero.

Histograms for treatment and control teachers' performance in each of the three domains are shown in Figures C.1–C.3. These histograms illustrate the pattern of variation (or distribution) of the classroom practices data.

Impact estimates for literacy implementation scores are presented separately by district in Figures C.4–C.5.

Table C.1. Impacts on Classroom Practices (Percentages with Consistent or Extensive Evidence of Practice): One-Year Districts and Two-Year Districts Combined, 2005–2006 School Year

Classroom Observation Item	Treatment	Control	Difference	P-value
Implementation of Literacy Lesson				
Best practices	23.4	27.2	-3.8	0.306
Instructional choices	28.8	30.7	-1.8	0.614
Student choices	18.2	18.4	-0.2	0.952
Pace	24.2	26.3	-2.1	0.559
Student-student interaction	16.8	15.5	1.3	0.682
Content of Literacy Lesson				
Understanding content and close reading	23.5	25.4	-1.9	0.593
Assessment	7.2	7.4	-0.2	0.935
Skill development	17.9	17.8	0.1	0.983
Connections between reading and writing	15.9	17.0	-1.1	0.737
Classroom Culture				
Maximizes learning opportunities	44.4	46.4	-2.0	0.619
Routines clear and consistent	46.1	49.4	-3.3	0.434
Behavior respectable, atmosphere safe	45.3	44.0	1.2	0.756
Literacy valued	28.1	31.1	-3.0	0.429
Teacher works collaboratively with students	39.5	37.2	2.2	0.594
Students work collaboratively with other students	25.0	23.8	1.2	0.735
Equal access to teacher and resources	41.3	46.0	-4.6	0.291
Sample Size (Teachers)	**342**	**289**		

Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression adjusted to account for differences in baseline characteristics and the study design.

None of the differences is statistically significant at the 0.05 level.

Table C.2. Impacts on Classroom Practices (Observer Summary Scores): One-Year Districts and Two-Year Districts Combined, 2005–2006 School Year

Outcome	Treatment	Control	Difference	Effect Size	P-value
Implementation of literacy lesson	2.7	2.7	0.0	-0.01	0.942
Content of literacy lesson	2.5	2.5	0.0	-0.01	0.859
Classroom culture	3.1	3.0	0.0	0.02	0.804
Sample Size (Teachers)	**342**	**289**			

Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression adjusted to account for differences in baseline characteristics and the study design. Scoring scale: (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence.

None of the differences is statistically significant at the 0.05 level.

Table C.3. Impacts on Classroom Practices (Average Score on a Five-Point Scale): One-Year Districts, 2005–2006 School Year

Outcome	Treatment	Control	Difference	Effect Size	P-value
Implementation of literacy lesson	2.7	2.7	0.0	-0.05	0.646
Content of literacy lesson	2.4	2.4	0.0	-0.03	0.774
Classroom culture	3.1	3.1	0.0	0.04	0.720
Sample Size (Teachers)	**178**	**164**			

Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted to account for differences in baseline characteristics and the study design. Scoring scale: (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence of effective teaching practice.

None of the differences is statistically significant at the 0.05 level.

Table C.4. Impacts on Classroom Practices (Average Score on a 5-Point Scale): Two-Year Districts, 2005–2006 School Year

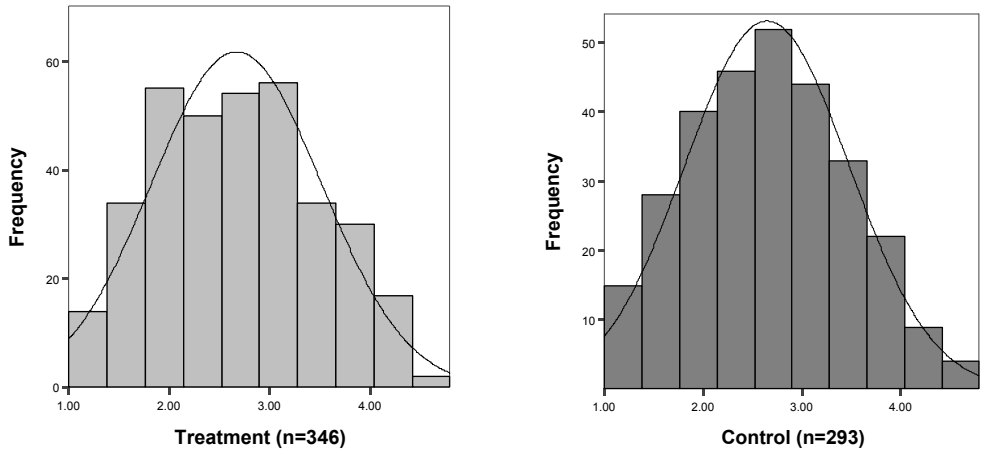
Outcome	Treatment	Control	Difference	Effect Size	P-value
Implementation of literacy lesson	2.7	2.6	0.1	0.08	0.467
Content of literacy lesson	2.3	2.3	0.0	0.01	0.935
Classroom culture	3.0	3.0	0.0	0.04	0.774
Sample Size (Teachers)	**164**	**125**			

Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted to account for differences in baseline characteristics and the study design. Scoring scale: (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence of effective teaching practice.

None of the differences is statistically significant at the 0.05 level.

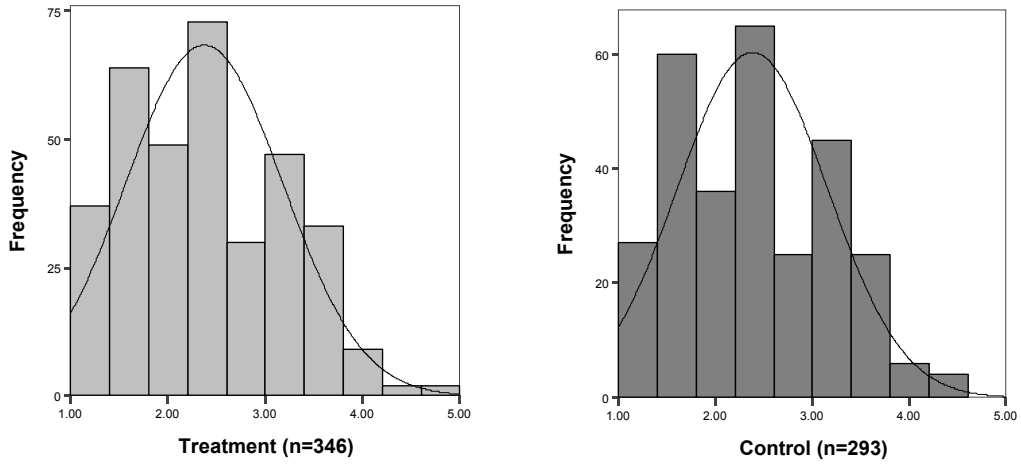
Figure C.1. Distribution of Literacy Implementation Scores: One-Year Districts and Two-Year Districts Combined, 2005–2006 School Year



Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression adjusted to account for differences in baseline characteristics and the study design. Scoring scale: (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence of effective teaching practice.

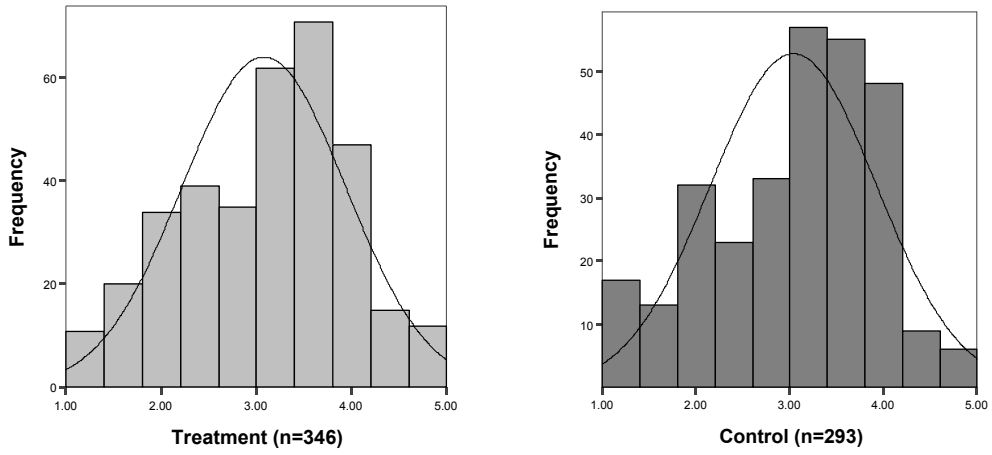
Figure C.2. Distribution of Literacy Content Scores: One-Year Districts and Two-Year Districts Combined, 2005–2006 School Year



Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression adjusted to account for differences in baseline characteristics and the study design. Scoring scale: (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence of effective teaching practice.

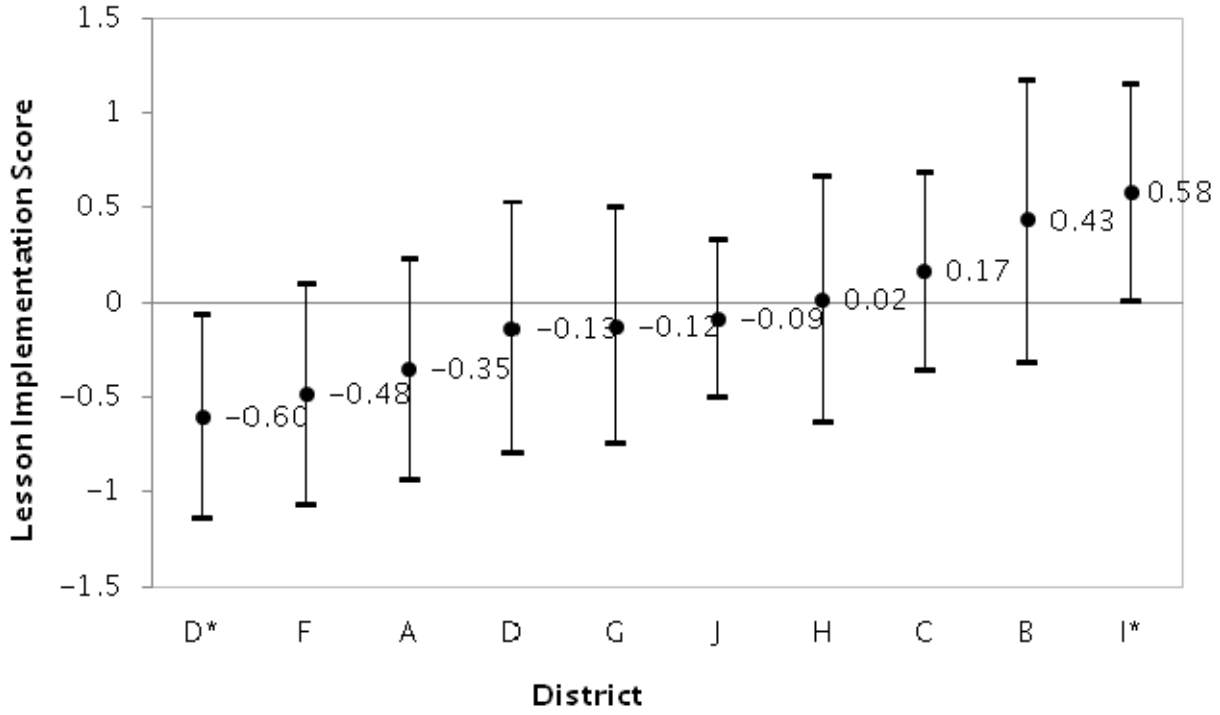
Figure C.3. Distribution of Literacy Culture Scores: One-Year Districts and Two-Year Districts Combined, 2005–2006 School Year



Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression adjusted to account for differences in baseline characteristics and the study design. Scoring scale: (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence of effective teaching practice.

Figure C.4. Impacts on Literacy Implementation Scores by District: One-Year Districts, 2005–2006 School Year

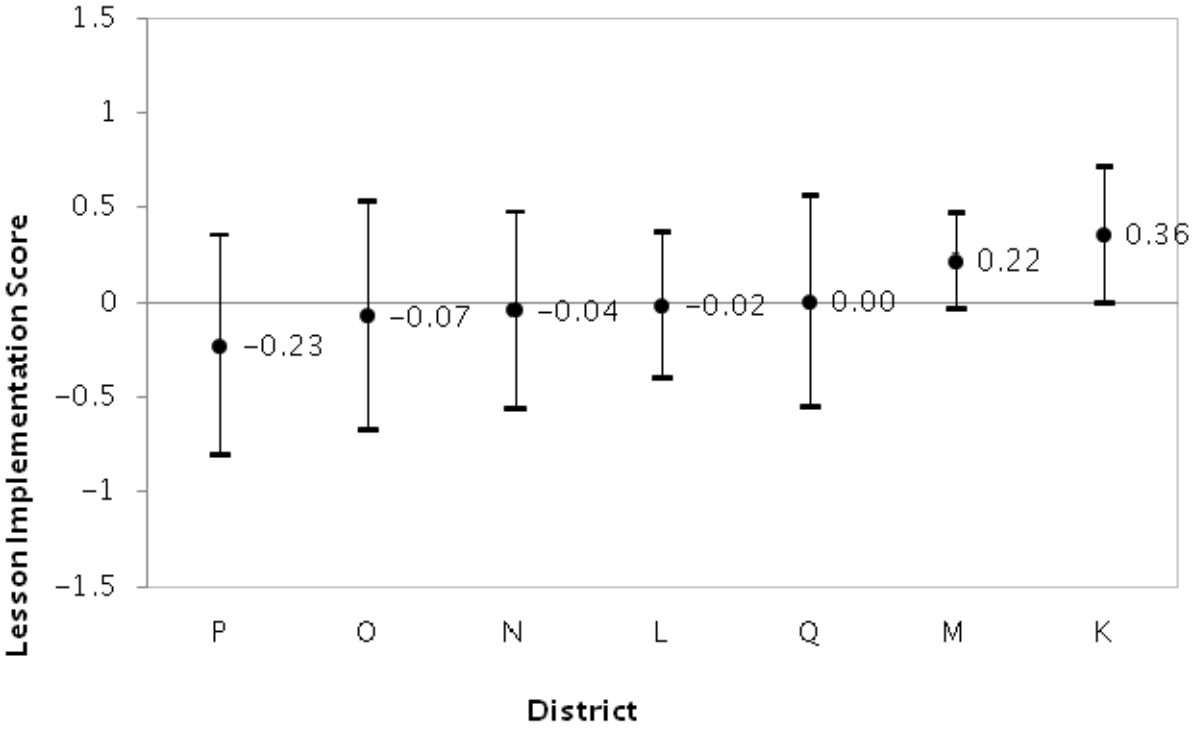


Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes A through J are arbitrary. Districts are ordered according to the size of the impact. N=342 teachers.

*Significantly different from zero at the 0.05 level. (No adjustment is applied for multiple comparisons.)

Figure C.5. Impacts on Literacy Implementation Scores by District: Two-Year Districts, 2005–2006 School Year



Source: Mathematica Teacher Background Survey administered in fall 2005 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes K through Q are arbitrary. Districts are ordered according to the size of the impact. N=289 teachers.

None of the differences is statistically significant at the 0.05 level.

B. Supplementary Information for Student Achievement

Tables C.5–C.10 show treatment and control sample sizes for the results shown in Tables V.2–V.7. Figures C.6–C.9 illustrate district-by-district effects for reading and math in one-year districts (C.6 and C.7) and two-year districts (C.8 and C.9).

Table C.5. Treatment and Control Sample Sizes for Impacts on Test Scores: One-Year Districts, 2007-2008 School Year

	Sample Sizes: Treatment Group				Sample Sizes: Control Group			
Subject:	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Reading	776	44	37	8	914	55	42	8
Math	766	44	35	8	863	51	41	8

Source: Mathematica analysis using data from the 2006-2007 and 2007-2008 school years provided by participating school districts.

Table C.6. Treatment and Control Sample Sizes for Impacts on Test Scores by Grade: One-Year Districts, 2007-2008 School Year

	Sample Sizes: Treatment Group		Sample Sizes: Control Group	
Subject: Grade	Students	Districts	Students	Districts
Reading				
Grade 2	182	4	262	4
Grade 3	200	5	221	5
Grade 4	242	7	237	7
Grade 5	152	4	194	4
All Grades	776	8	914	8
Math				
Grade 2	94	2	106	2
Grade 3	199	5	223	5
Grade 4	321	8	340	8
Grade 5	152	4	194	4
All Grades	766	8	863	8

Source: Mathematica analysis using data from the 2006-2007 and 2007-2008 school years provided by participating school districts.

Table C.7. Treatment and Control Sample Sizes for Impacts on Test Scores, Alternate Model Specifications: One-Year Districts, 2007-2008 School Year

	Sample Sizes: Treatment Group				Sample Sizes: Control Group			
Subject/Model	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Reading								
1. Benchmark	776	44	37	8	914	55	42	8
2. No teacher covariates	776	44	37	8	914	55	42	8
3. No teacher or student covariates	776	44	37	8	914	55	42	8
4. Schools weighted equally	776	44	37	8	914	55	42	8
5. Districts weighted equally	776	44	37	8	914	55	42	8
6. Using specific information on teacher assignments	721	41	35	8	819	48	36	8
7. Without imposing data restrictions	874	48	39	8	1,018	59	43	8
8. No pretest, benchmark sample	776	44	37	8	914	55	42	8
9. No pretest, expanded sample	1,383	70	50	8	1,473	81	57	8
10. Compare teachers within districts, not district-grades	921	52	41	8	1,052	62	49	8
11. Instrumental variables	708	42	36	8	811	51	41	8
Math								
1. Benchmark	766	44	35	8	863	51	41	8
2. No teacher covariates	766	44	35	8	863	51	41	8
3. No teacher or student covariates	766	44	35	8	863	51	41	8
4. Schools weighted equally	766	44	35	8	863	51	41	8
5. Districts weighted equally	766	44	35	8	863	51	41	8
6. Using specific information on teacher assignments	746	43	34	8	741	44	35	8
7. Without imposing data restrictions	809	45	36	8	891	52	41	8
8. No pretest, benchmark sample	766	44	35	8	863	51	41	8
9. No pretest, expanded sample	1,297	65	46	8	1,405	73	54	8
10. Compare teachers within districts, not district-grades	803	46	37	8	1,001	58	48	8
11. Instrumental variables	730	43	35	8	824	50	40	8

Source: Mathematica analysis using data from the 2006-2007 and 2007-2008 school years provided by participating school districts.

Table C.8. Treatment and Control Sample Sizes for Impacts on Test Scores: Two-Year Districts, 2007-2008 School Year

	Sample Sizes: Treatment Group				Sample Sizes: Control Group			
Subject	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Reading	807	43	30	6	540	31	26	6
Math	699	39	27	6	499	29	25	6

Source: Mathematica analysis using data from the 2006-2007 and 2007-2008 school years provided by participating school districts.

Table C.9. Treatment and Control Sample Sizes for Impacts on Test Scores by Grade: Two-Year Districts, 2007-2008 School Year

	Sample Sizes: Treatment Group		Sample Sizes: Control Group	
Subject/Grade	Students	Districts	Students	Districts
Reading				
Grade 2	51	1	12	1
Grade 3	173	2	88	2
Grade 4	379	5	362	5
Grade 5	192	2	54	2
Grade 6	12	1	24	1
All Grades	807	6	540	6
Math				
Grade 2	51	1	12	1
Grade 3	208	2	71	2
Grade 4	274	5	340	5
Grade 5	154	2	52	2
Grade 6	12	1	24	1
All Grades	699	6	499	6

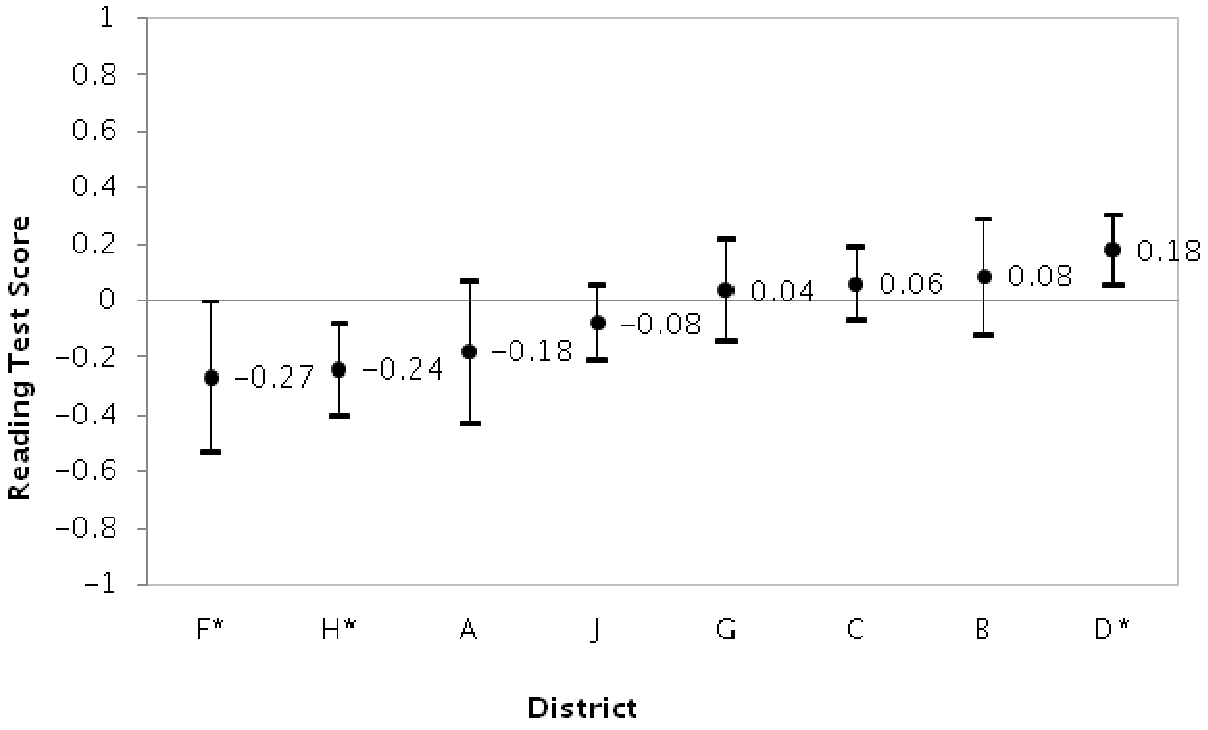
Source: Mathematica analysis using data from the 2006-2007 and 2007-2008 school years provided by participating school districts.

Table C.10. Treatment and Control Sample Sizes for Impacts on Test Scores, Alternate Model Specifications: Two-Year Districts, 2007-2008 School Year

	Sample Sizes: Treatment Group				Sample Sizes: Control Group			
Subject/Model	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Reading								
1. Benchmark	807	43	30	6	540	31	26	6
2. No teacher covariates	807	43	30	6	540	31	26	6
3. No teacher or student covariates	807	43	30	6	540	31	26	6
4. Schools weighted equally	807	43	30	6	540	31	26	6
5. Districts weighted equally	807	43	30	6	540	31	26	6
6. Using specific information on teacher assignments	807	43	30	6	540	31	26	6
7. Without imposing data restrictions	807	43	30	6	540	31	26	6
8. No pretest, benchmark sample	807	43	30	6	540	31	26	6
9. No pretest, expanded sample	1,464	73	40	7	993	54	38	7
10. Compare teachers within districts, not district-grades	908	49	34	6	576	33	27	6
11. Instrumental variables	805	42	29	6	533	31	26	6
Math								
1. Benchmark	699	39	27	6	499	29	25	6
2. No teacher covariates	699	39	27	6	499	29	25	6
3. No teacher or student covariates	699	39	27	6	499	29	25	6
4. Schools weighted equally	699	39	27	6	499	29	25	6
5. Districts weighted equally	699	39	27	6	499	29	25	6
6. Using specific information on teacher assignments	662	37	26	6	499	29	25	6
7. Without imposing data restrictions	730	40	28	6	529	30	26	6
8. No pretest, benchmark sample	699	39	27	6	499	29	25	6
9. No pretest, expanded sample	1,290	68	39	7	964	52	38	7
10. Compare teachers within districts, not district-grades	800	45	31	6	598	32	27	6
11. Instrumental variables	697	39	27	6	496	29	25	6

Source: Mathematica analysis using data from the 2006-2007 and 2007-2008 school years provided by participating school districts.

Figure C.6. Impacts on Reading Test Scores by District: One-Year Districts, 2007–2008 School Year

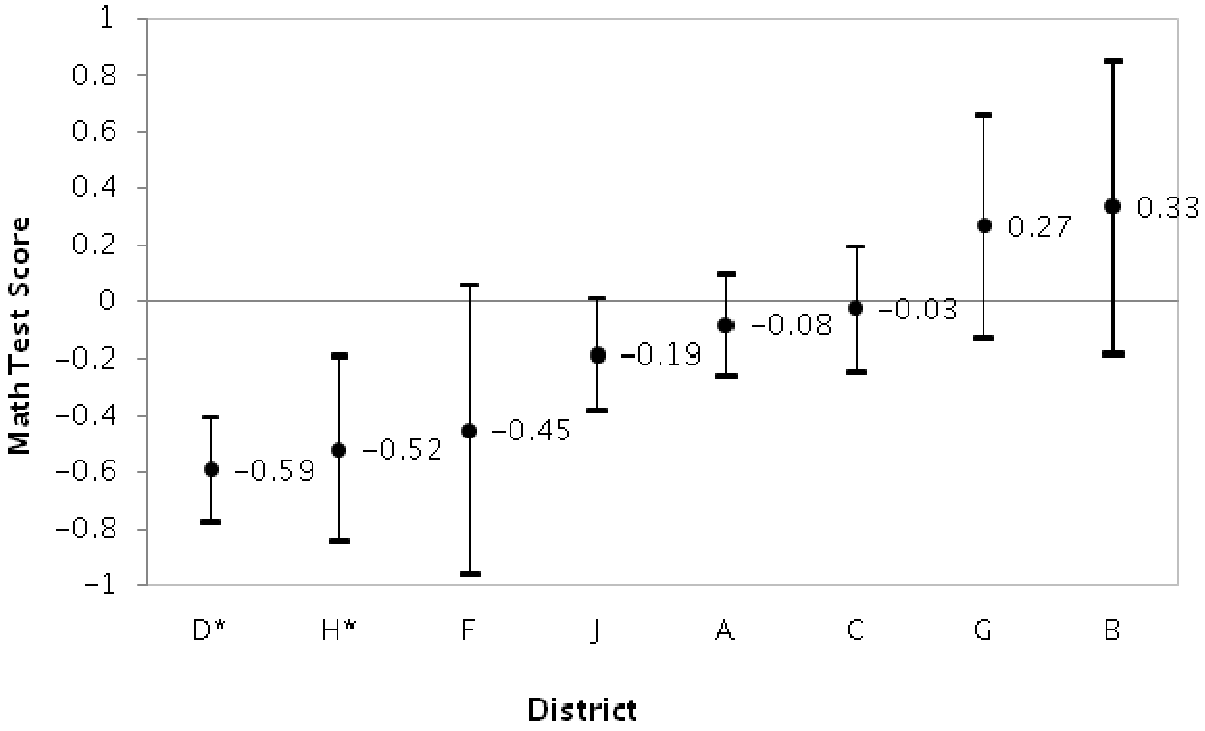


Source: Mathematica analysis using data from the 2006–2007 and 2007–2008 school years provided by participating school districts.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes A through J are arbitrary. Districts are ordered according to the size of the impact. Impacts are expressed as a fraction of a standard deviation in scores, where the standard deviation is based on all study students in the same grade and district. N=99 teachers and 1,690 students.

***Significantly different from zero at the 0.05 level. (No adjustment is applied for multiple comparisons.)**

Figure C.7. Impacts on Math Test Scores by District: One-Year Districts, 2007–2008 School Year

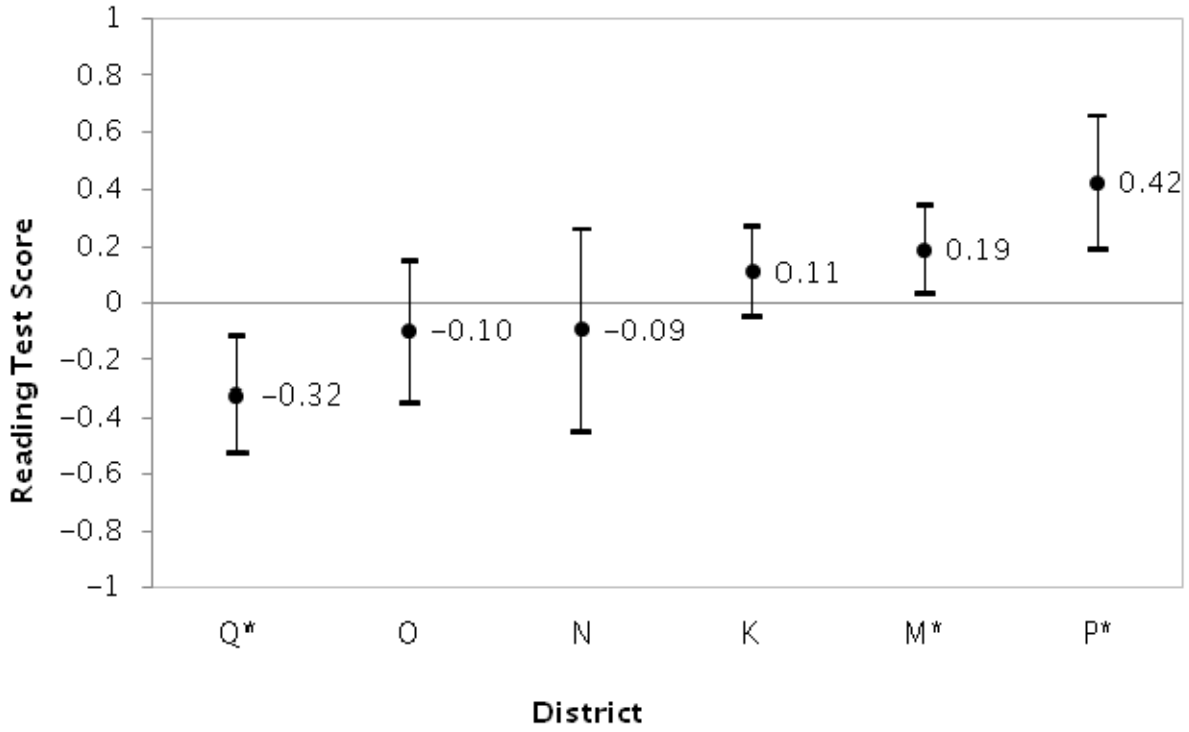


Source: Mathematica analysis using data from the 2006–2007 and 2007–2008 school years provided by participating school districts.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes A through J are arbitrary. Districts are ordered according to the size of the impact. Impacts are expressed as a fraction of a standard deviation in scores, where the standard deviation is based on all study students in the same grade and district. N=95 teachers and 1,629 students.

***Significantly different from zero at the 0.05 level. (No adjustment is applied for multiple comparisons.)**

Figure C.8. Impacts on Reading Test Scores by District: Two-Year Districts, 2007–2008 School Year

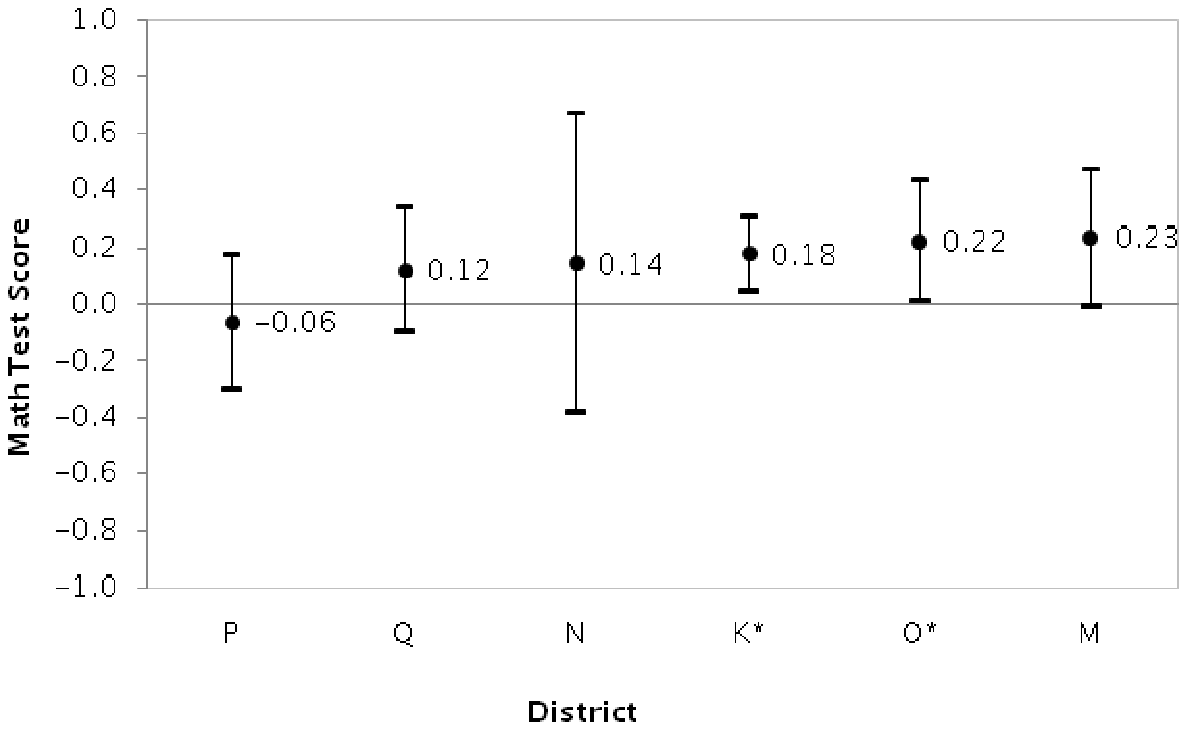


Source: Mathematica analysis using data from the 2006–2007 and 2007–2008 school years provided by participating school districts.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes K through Q are arbitrary. Districts are ordered according to the size of the impact. Impacts are expressed as a fraction of a standard deviation in scores, where the standard deviation is based on all study students in the same grade and district. N=74 teachers and 1,347 students.

***Significantly different from zero at the 0.05 level. (No adjustment is applied for multiple comparisons.)**

Figure C.9. Impacts on Math Test Scores by District: Two-Year Districts, 2007–2008 School Year



Source: Mathematica analysis using data from the 2006–2007 and 2007–2008 school years provided by participating school districts.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes K through Q are arbitrary. Districts are ordered according to the size of the impact. Impacts are expressed as a fraction of a standard deviation in scores, where the standard deviation is based on all study students in the same grade and district. N=68 teachers and 1,198 students.

***Significantly different from zero at the 0.05 level. (No adjustment is applied for multiple comparisons.)**

Table C.11. Changes in Impacts on Test Scores Over Time, Common Sample of Teachers: Two-Year Districts

Subject	Adjusted Mean Test Scores		Difference (Effect Size)	P-value	Sample Sizes		
	Treatment	Control			Students	Teachers	Districts
Reading							
Year 1	-0.14	0.02	-0.15	0.18	637	37	5
Year 3	-0.08	-0.32	0.24*	0.00	676	37	5
Year 3–Year 1	0.05	-0.34	0.39*	0.00			
Year 2	-0.13	-0.18	0.04	0.59	947	52	6
Year 3	-0.16	-0.20	0.04	0.38	947	52	6
Year 3–Year 2	-0.02	-0.02	0.00	0.99			
Math							
Year 1	-0.04	0.09	-0.13	0.32	567	37	5
Year 3	-0.04	-0.13	0.10	0.22	684	37	5
Year 3–Year 1	0.00	-0.22	0.22	0.14			
Year 2	-0.21	-0.22	0.01	0.88	902	52	6
Year 3	-0.09	-0.19	0.10	0.18	961	52	6
Year 3–Year 2	0.11	0.02	0.09	0.38			

Source: Mathematica analysis using data from the 2004-2005, 2005-1006, 2006–2007 and 2007–2008 school years provided by participating school districts; Mathematica Teacher Background Survey administered in fall 2005 to all study teachers.

Notes: Data pertain to teachers in two-year districts participating in the study. Data are regression adjusted to account for pretest, student and teacher characteristics, district-by-grade fixed effects, and clustering of students within schools. The common sample is the subsample of teachers who had students with valid test score data in either Year 1 and Year 3 or Year 2 and Year 3.

*Significantly different from zero at the 0.05 level.

Table C.12. Impacts on Intermediate Outcomes, Reading Test Score Sample

Outcome	Treatment	Control	Difference	P-value		N
Any mentor assigned (percentage)						
Year 1, fall	93.96	73.85	20.1	0.030	*	66
Year 1, spring	98.37	75.20	23.2	0.007	*	67
Year 2, fall	84.52	31.14	53.4	0.000	*	63
Year 2, spring	92.20	39.57	52.6	0.000	*	64
Total time spent meeting with mentors (minutes per week)						
Year 1, fall	63.69	73.07	-9.4	0.783		66
Year 1, spring	62.00	74.11	-12.1	0.851		67
Year 2, fall	29.32	42.65	-13.3	0.852		63
Year 2, spring	68.56	40.98	27.6	0.556		65
Suggestions to improve practice (percentage)						
Year 1, fall	80.48	57.21	23.3	0.030	*	66
Year 1, spring	88.99	52.25	36.7	0.001	*	67
Year 2, fall	55.32	25.14	30.2	0.013	*	57
Year 2, spring	54.94	27.35	27.6	0.016	*	65
Guidance in teaching reading (percentage)						
Year 1, fall	56.58	43.32	13.3	0.343		66
Year 1, spring	80.09	34.19	45.9	0.001	*	66
Year 2, fall (question not asked)						
Year 2, spring	31.68	19.89	11.8	0.325		65
Given feedback on teaching, not as part of formal evaluation (number of times)						
Year 1, fall	2.67	2.42	0.2	0.648		65
Year 1, spring	3.03	1.96	1.1	0.025	*	66
Year 2, fall	1.35	1.52	-0.2	0.727		62
Year 2, spring	2.71	1.53	1.2	0.002	*	65
Classroom practices in year 1						
Literacy lesson content	2.34	2.38	-0.0	0.831		56
Literacy lesson implementation	2.74	2.65	0.1	0.705		56
Classroom culture	3.27	3.04	0.2	0.199		56

Source: Mathematica analysis using data from the 2007–2008 school year provided by participating school districts; Mathematica Teacher Background Survey administered in fall 2005 and Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, and spring 2007.

Notes: Data pertain to teachers in two-year districts participating in the study who were also in the benchmark reading test score analysis. Classroom practice data are weighted and regression adjusted using ordinary least squares to account for differences in baseline characteristics and the study design.

*Significantly different from zero at the 0.05 level.

Table C.13. Impacts on Intermediate Outcomes, Math Test Score Sample

Outcome	Treatment	Control	Difference	P-value		N
Any mentor assigned (percentage)						
Year 1, fall	92.11	73.85	18.3	0.053		59
Year 1, spring	96.24	75.20	21.0	0.016	*	62
Year 2, fall	86.55	31.14	55.4	0.000	*	56
Year 2, spring	89.50	39.57	49.9	0.000	*	56
Total time spent meeting with mentors (minutes per week)						
Year 1, fall	34.92	73.07	-38.2	0.345		59
Year 1, spring	71.29	74.11	-2.8	0.961		62
Year 2, fall	26.62	42.65	-16.0	0.839		56
Year 2, spring	72.04	40.98	31.1	0.506		57
Suggestions to improve practice (percentage)						
Year 1, fall	78.82	57.21	21.6	0.050	*	59
Year 1, spring	86.44	52.25	34.2	0.002	*	62
Year 2, fall	57.31	25.14	32.2	0.015	*	49
Year 2, spring	49.52	27.35	22.2	0.065		57
Guidance in teaching math (percentage)						
Year 1, fall	37.21	34.74	2.5	0.854		59
Year 1, spring	59.36	31.76	27.6	0.073		62
Year 2, fall (question not asked)						
Year 2, spring	35.88	16.92	19.0	0.126		57
Given feedback on teaching, not as part of formal evaluation (number of times)						
Year 1, fall	2.96	2.42	0.5	0.250		58
Year 1, spring	2.63	1.96	0.7	0.201		62
Year 2, fall	1.25	1.52	-0.3	0.606		55
Year 2, spring	2.40	1.53	0.9	0.012	*	57

Source: Mathematica analysis using data from the 2007–2008 school year provided by participating school districts; Mathematica Teacher Background Survey administered in fall 2005; Mathematica Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, and spring 2007.

Notes: Data pertain to teachers in two-year districts participating in the study who were also in the benchmark math test score analysis.

*Significantly different from zero at the 0.05 level.

APPENDIX D

SENSITIVITY ANALYSES AND SUPPLEMENTAL INFORMATION FOR CHAPTER VI

Chapter VI presented estimates of the impacts of comprehensive induction on workforce outcomes (teacher attitudes and teacher mobility). This appendix presents supplemental results related to those findings.

A. Supplementary Information and Sensitivity Analysis for Teacher Satisfaction

Table D.1 presents results for teacher satisfaction for one-year districts in fall 2005, spring 2006, fall 2006, fall 2007 and fall 2008.[60] Table D.2 presents results for teacher satisfaction for two-year districts in fall 2005, spring 2006, fall 2006, spring 2007, fall 2007, and fall 2008. These results are summarized in Chapter VI, Figures VI.1 and VI.2.

One concern with the analysis of the teacher attitudes data is that the summary scores may mask impacts for individual items that make up the three summary scores within each domain. Another concern is that self-reported attitude measures rely on scales that may not have equal intervals. For example, the difference between the first and second categories may be larger than those between the third and fourth. We recoded teacher satisfaction into two categories: (1) "very dissatisfied" or "somewhat dissatisfied" or (2) "somewhat satisfied" or "very satisfied." We then examined item-specific impacts on the outcome defined by these dichotomous variables. Of the 38 differences examined among teachers in one-year districts in fall 2007 and fall 2008 shown in Table D.3, one was statistically significant. Treatment teachers were significantly less likely than control teachers to report satisfaction with salary and benefits in fall 2007. Of the 38 differences examined among teachers in two-year districts in fall 2007 and fall 2008 shown in Table D.4, one was statistically significant. Treatment teachers were significantly more likely than control teachers to report satisfaction with school facilities (buildings and grounds) in fall 2007.[61]

[60] Teacher satisfaction was not measured in one-year districts in spring 2007.

[61] The item-specific impacts for fall 2005, spring 2006, fall 2006, and spring 2007 can be found in Isenberg et al. (2009).

Table D.1. Impacts on Teacher Satisfaction (Scores on a Four-Point Scale): One-Year Districts

	Treatment	Control	Difference	P-value	Sample Size (Teachers)
Fall 2005					
Feel Satisfied with:					
School	3.1	3.1	0.0	0.751	498
Class	3.0	3.0	0.1	0.339	498
Teaching career	3.0	3.0	-0.1	0.290	498
Spring 2006					
Feel Satisfied with:					
School	3.0	3.0	0.0	0.927	492
Class	3.0	3.0	0.0	0.720	492
Teaching career	2.9	2.9	-0.1	0.201	492
Fall 2006					
Feel Satisfied with:					
School	3.2	3.1	0.0	0.843	472
Class	3.1	3.1	0.0	0.812	472
Teaching career	3.0	3.0	0.0	0.615	472
Fall 2007					
Feel Satisfied with:					
School	3.1	3.1	0.0	0.944	424
Class	3.2	3.1	0.1	0.155	425
Teaching career	2.9	2.9	0.0	0.701	426
Fall 2008					
Feel Satisfied with:					
School	3.1	3.1	0.0	0.786	397
Class	3.1	3.2	0.0	0.609	397
Teaching career	2.9	2.9	0.0	0.910	396

Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Satisfaction scale: (1) very dissatisfied, (2) somewhat dissatisfied, (3) somewhat satisfied, or (4) very satisfied. Sample sizes vary due to item nonresponse.

None of the differences is statistically significant at the 0.05 level.

Table D.2. Impacts on Teacher Satisfaction (Scores on a Four-Point Scale): Two-Year Districts

	Treatment	Control	Difference	P-value	Sample Size (Teachers)
Fall 2005					
Feel Satisfied with:					
School	3.1	3.1	0.0	0.908	391
Class	3.1	3.1	0.0	0.895	391
Teaching career	3.0	3.1	-0.1	0.127	391
Spring 2006					
Feel Satisfied with:					
School	3.1	3.0	0.0	0.596	384
Class	3.0	3.0	0.0	0.977	384
Teaching career	2.9	3.0	-0.1	0.286	384
Fall 2006					
Feel Satisfied with:					
School	3.1	3.2	0.0	0.793	360
Class	3.2	3.1	0.1	0.280	360
Teaching career	3.0	3.0	0.0	0.999	359
Spring 2007					
Feel Satisfied with:					
School	3.0	2.9	0.1	0.207	370
Class	3.1	3.1	0.0	0.797	370
Teaching career	2.8	2.9	-0.1	0.214	370
Fall 2007					
Feel Satisfied with:					
School	3.1	3.0	0.1	0.339	321
Class	3.1	3.1	0.0	0.912	324
Teaching career	2.9	2.9	0.0	0.856	326
Fall 2008					
Feel Satisfied with:					
School	3.1	3.1	0.0	0.874	319
Class	3.1	3.1	0.0	0.897	318
Teaching career	2.8	2.8	0.0	0.852	319

Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Satisfaction scale: (1) very dissatisfied, (2) somewhat dissatisfied, (3) somewhat satisfied, or (4) very satisfied. Sample sizes vary due to item nonresponse.

None of the differences is statistically significant at the 0.05 level.

Table D.3. Impacts on Teacher Satisfaction (Percentage "Somewhat Satisfied" or "Very Satisfied"): One-Year Districts, Fall 2007 and Fall 2008

	Fall 2007					Fall 2008				
Area of Satisfaction	Treatment	Control	Difference	Effect Size	P-value	Treatment	Control	Difference	Effect Size	P-value
Satisfaction with School										
Administration support for beginning teachers	69.3	73.4	-4.1	-0.09	0.261	64.6	66.5	-1.9	-0.04	0.600
Availability of resources and materials/equipment for your classroom	67.9	66.7	1.3	0.03	0.750	59.7	60.4	-0.7	-0.01	0.868
Input into school policies and practices	67.1	68.8	-1.7	-0.04	0.642	60.3	63.9	-3.6	-0.08	0.355
Opportunities for professional development	75.5	74.3	1.3	0.03	0.720	68.9	73.5	-4.6	-0.10	0.186
Principal's leadership and vision	68.6	65.4	3.2	0.07	0.396	63.9	59.1	4.8	0.10	0.210
Professional caliber of colleagues	70.2	71.7	-1.5	-0.03	0.678	65.9	67.8	-2.0	-0.04	0.581
Supportive atmosphere among faculty/collaboration with colleagues	73.0	71.3	1.7	0.04	0.609	70.5	67.4	3.1	0.07	0.370
School facilities such as the building or grounds	66.8	62.0	4.7	0.10	0.275	60.5	63.9	-3.4	-0.07	0.375
School policies	71.6	74.3	-2.7	-0.06	0.415	67.8	66.1	1.7	0.04	0.642
Satisfaction with Students										
Autonomy or control over own classroom	77.9	74.7	3.3	0.08	0.279	72.3	72.6	-0.3	-0.01	0.943
Student motivation to learn	65.6	59.5	6.1	0.13	0.115	61.0	64.8	-3.7	-0.08	0.330
Student discipline and behavior	64.9	58.7	6.3	0.13	0.116	56.0	60.0	-4.0	-0.08	0.352
Parental involvement in the school	46.8	39.7	7.1	0.14	0.142	45.7	44.3	1.3	0.03	0.766
Grade assignment	84.4	82.3	2.2	0.06	0.308	78.2	78.7	-0.5	-0.01	0.833
Students assigned	78.1	79.3	-1.2	-0.03	0.686	71.3	75.7	-4.4	-0.10	0.157
Satisfaction with Teaching Career										
Salary and benefits	54.2	64.6	-10.4*	-0.21	0.009	55.3	57.0	-1.7	-0.03	0.676
Professional prestige	68.9	65.4	3.5	0.08	0.334	64.4	62.2	2.3	0.05	0.559
Intellectual challenge	78.5	78.5	0.0	0.00	0.998	72.4	75.7	-3.2	-0.07	0.331
Workload	46.8	46.8	0.0	0.00	0.997	45.4	42.6	2.8	0.06	0.520
Sample Size (Teachers)	**219**	**207**	**426**			**206**	**192**	**398**		

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Table D.4. Impacts on Teacher Satisfaction (Percentage "Somewhat Satisfied" or "Very Satisfied"): Two-Year Districts, Fall 2007 and Fall 2008

	Fall 2007					Fall 2008				
Area of Satisfaction	Treatment	Control	Difference	Effect Size	P-value	Treatment	Control	Difference	Effect Size	P-value
Satisfaction with School										
Administration support for beginning teachers	70.7	73.2	-2.5	-0.05	0.563	67.5	70.0	-2.5	-0.05	0.512
Availability of resources and materials/equipment for your classroom	69.9	63.4	6.5	0.14	0.165	64.5	61.8	2.8	0.06	0.509
Input into school policies and practices	68.3	59.8	8.5	0.18	0.102	58.5	62.9	-4.4	-0.09	0.371
Opportunities for professional development	77.4	71.3	6.1	0.14	0.115	69.8	68.2	1.6	0.03	0.685
Principal's leadership and vision	70.8	67.7	3.1	0.07	0.467	65.3	65.3	0.0	0.00	0.995
Professional caliber of colleagues	71.1	75.6	-4.6	-0.10	0.256	65.2	67.7	-2.5	-0.05	0.596
Supportive atmosphere among faculty/collaboration with colleagues	74.5	73.2	1.4	0.03	0.748	64.5	67.7	-3.1	-0.07	0.539
School facilities such as the building or grounds	69.5	55.5	14.0*	0.29	0.004	62.5	58.8	3.7	0.08	0.474
School policies	69.8	74.4	-4.5	-0.10	0.295	69.0	67.7	1.3	0.03	0.753
Satisfaction with Students										
Autonomy or control over own classroom	77.3	76.8	0.4	0.01	0.901	69.8	75.3	-5.5	-0.12	0.190
Student motivation to learn	61.3	65.2	-3.9	-0.08	0.393	61.5	63.5	-2.1	-0.04	0.668
Student discipline and behavior	60.3	60.4	-0.1	0.00	0.989	58.2	58.2	-0.1	0.00	0.990
Parental involvement in the school	45.5	45.7	-0.3	-0.01	0.957	47.8	47.7	0.2	0.00	0.975
Grade assignment	79.4	81.1	-1.7	-0.04	0.614	75.7	76.5	-0.8	-0.02	0.805
Students assigned	73.2	77.4	-4.2	-0.10	0.274	67.6	71.8	-4.2	-0.09	0.317
Satisfaction with Teaching Career										
Salary and benefits	61.1	56.7	4.4	0.09	0.349	43.8	46.5	-2.6	-0.05	0.629
Professional prestige	72.4	65.2	7.2	0.15	0.094	64.7	59.4	5.3	0.11	0.278
Intellectual challenge	78.8	78.7	0.1	0.00	0.977	71.3	72.3	-1.1	-0.02	0.772
Workload	50.0	47.6	2.5	0.05	0.630	44.4	48.8	-4.4	-0.09	0.372
Sample Size (Teachers)	**179**	**147**	**326**			**178**	**143**	**321**		

Source: Mathematica Fifth and Sixth Induction Activities Surveys administered in fall 2007 and fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

B. Supplementary Information and Sensitivity Analysis for Teacher Preparedness

Table D.5 presents results for teacher preparedness for one-year districts in fall 2005, spring 2006, and fall 2008.[62] Table D.6 presents results for teacher preparedness for two-year districts in fall 2005, spring 2006, spring 2007, and fall 2008.[63] These results are also shown in Chapter VI, Figures VI.3 and VI.4.

As described earlier, one concern with the analysis of the teacher attitudes data is that the summary scores may mask impacts for individual items that make up the three summary scores within each domain. Another concern is that self-reported attitude measures rely on scales that may not have equal intervals. We recoded teacher preparedness into two categories: (1) "not at all prepared" or "somewhat prepared" or (2) "well prepared" or "very well prepared." We then examined item-specific impacts on the outcomes defined by the dichotomous variables. Of the 13 differences among teachers examined in fall 2008, none were statistically significant in one-year districts (Table D.7) and one was significant in two-year districts (Table D.8). Treatment teachers in two-year districts were significantly less likely than control teachers to report feeling prepared to work with other teachers to plan instruction.[64,65]

[62] Teacher preparedness was not measured in one-year districts in fall 2006, spring 2007, or fall 2007.

[63] Teacher preparedness was not measured in two-year districts in fall 2007.

[64] See Chapter II for a discussion of multiple comparisons and false discoveries.

[65] The item-specific impacts for fall 2005 and spring 2006 can be found in Glazerman et al. (2008). The item-specific impacts for spring 2006 and spring 2007 can be found in Isenberg et al. (2009).

Table D.5. Impacts on Teacher Preparedness (Scores on a Four-Point Scale): One-Year Districts

	Treatment	Control	Difference	P-value	Sample Size (Teachers)
Fall 2005					
Feel Prepared to:					
Instruct	2.8	2.8	-0.1	0.179	501
Work with others	2.9	2.9	0.0	0.688	501
Work with students	2.8	2.7	0.0	0.786	501
Spring 2006					
Feel Prepared to:					
Instruct	2.9	3.0	-0.1	0.088	493
Work with others	2.9	3.0	-0.1	0.067	493
Work with students	2.8	2.8	0.0	0.511	493
Fall 2008					
Feel Prepared to:					
Instruct	3.4	3.4	0.0	0.619	386
Work with others	3.4	3.3	0.1	0.281	386
Work with students	3.2	3.2	0.0	0.923	386

Source: Mathematica First, Second, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, and fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Satisfaction scale: (1) very dissatisfied, (2) somewhat dissatisfied, (3) somewhat satisfied, or (4) very satisfied. Sample sizes vary due to item nonresponse.

None of the differences is statistically significant at the 0.05 level.

Table D.6. Impacts on Teacher Preparedness (Scores on a Four-Point Scale): Two-Year Districts

	Treatment	Control	Difference	P-value	Sample Size (Teachers)
Fall 2005					
Feel Prepared to:					
Instruct	2.8	2.9	-0.1*	0.030	394
Work with others	2.8	3.0	-0.1	0.178	394
Work with students	2.7	2.8	-0.1	0.219	394
Spring 2006					
Feel Prepared to:					
Instruct	3.0	3.0	0.0	0.703	383
Work with others	3.0	3.0	-0.1	0.338	381
Work with students	2.9	2.8	0.1	0.472	383
Spring 2007					
Feel Prepared to:					
Instruct	3.2	3.1	0.0	0.869	371
Work with others	3.1	3.1	0.0	0.933	371
Work with students	3.0	3.0	0.0	0.614	371
Fall 2008					
Feel Prepared to:					
Instruct	3.5	3.5	0.0	0.891	308
Work with others	3.4	3.5	-0.1	0.255	308
Work with students	3.3	3.3	0.0	0.824	308

Source: Mathematica First, Second, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, and fall 2008 to all study teachers and Fourth Induction Activities Survey administered in spring 2007 to study teachers in two-year districts.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Satisfaction scale: (1) very dissatisfied, (2) somewhat dissatisfied, (3) somewhat satisfied, or (4) very satisfied. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

Table D.7. Impacts on Teacher Preparedness (Percentage "Somewhat Prepared" or "Very Prepared"): One-Year Districts, Fall 2008

	Fall 2008				
Area of Preparedness	Treatment	Control	Difference	Effect Size	P-value
Prepared to Instruct					
Managing classroom activities, transitions, and routines	76.1	78.3	-2.2	-0.05	0.457
Using variety of instructional methods	73.5	74.3	-0.8	-0.02	0.783
Assessing your students	74.1	74.8	-0.6	-0.01	0.838
Selecting and adapting instructional materials	72.6	72.6	0.0	0.00	0.992
Planning effective lessons	75.5	78.3	-2.7	-0.07	0.299
Being an effective teacher	74.9	78.3	-3.3	-0.08	0.246
Addressing needs of a diversity of learners	67.8	73.9	-6.2	-0.14	0.070
Prepared to Work with Students					
Handling range of classroom behavior or discipline situations	74.0	73.9	0.1	0.00	0.983
Motivating students	73.5	73.0	0.4	0.01	0.888
Working effectively with parents	68.9	71.3	-2.4	-0.05	0.465
Working with students with special challenges	48.8	54.8	-6.0	-0.12	0.194
Prepared to Work with Other School Staff					
Working with other teachers to plan instruction	72.3	74.8	-2.5	-0.06	0.448
Working with the principal or other instructional leaders	72.7	69.6	3.1	0.07	0.340
Sample Size (Teachers)	**206**	**192**	**398**		

Source: Mathematica Sixth Induction Activities Survey administered in fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

None of the differences is statistically significant at the 0.05 level.

Table D.8. Impacts on Teacher Preparedness (Percentage "Somewhat Prepared" or "Very Prepared"): Two-Year Districts, Fall 2008

	Fall 2008				
Area of Preparedness	Treatment	Control	Difference	Effect Size	P-value
Prepared to Instruct					
Managing classroom activities, transitions, and routines	75.8	80.1	-4.4	-0.11	0.260
Using variety of instructional methods	74.3	74.8	-0.6	-0.01	0.852
Assessing your students	75.4	78.9	-3.6	-0.09	0.283
Selecting and adapting instructional materials	72.0	74.3	-2.3	-0.05	0.588
Planning effective lessons	76.3	79.5	-3.3	-0.08	0.322
Being an effective teacher	74.6	80.1	-5.6	-0.13	0.112
Addressing needs of a diversity of learners	72.0	74.3	-2.3	-0.05	0.541
Prepared to Work with Students					
Handling range of classroom behavior or discipline situations	73.0	71.3	1.7	0.04	0.690
Motivating students	72.8	74.8	-2.1	-0.05	0.584
Working effectively with parents	68.9	76.0	-7.1	-0.16	0.055
Working with students with special challenges	48.9	52.6	-3.7	-0.07	0.434
Prepared to Work with Other School Staff					
Working with other teachers to plan instruction	68.7	78.9	-10.3*	-0.23	0.015
Working with the principal or other instructional leaders	69.6	73.7	-4.1	-0.09	0.354
Sample Size (Teachers)	**178**	**143**	**321**		

Source: Mathematica Sixth Induction Activities Survey administered in fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted to account for differences in districts, teacher grade assignments, study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

*Significantly different from zero at the 0.05 level.

C. Sensitivity and Supplemental Analysis for Teacher Retention

1. Detail on Retention Measures

Chapter VI presented figures that summarized the retention rates for treatment and control teachers (Figures VI.7 and VI.8). Here we include tables with the numbers underlying those results for the final year of followup (Tables D.9 and D.10). The tables present additional detail on retention rates, along with sample sizes, to complement the findings shown in Figures VI.7 and VI.8. They include the regression-adjusted percentages of teachers who were retained in the same school for one-year districts (53 percent) and two-year districts (51 percent). The corresponding numbers for years 1 and 2 of the study are presented in earlier reports (see Glazerman et al. 2008; Isenberg et al. 2009).

Table D.9. Impacts on Teacher Retention Rates After Three Years (Percentages): One-Year Districts

Outcome	All Teachers	Treatment	Control	Difference	P-value
Retained in the same school	53.3	53.9	52.7	1.2	0.804
Retained in the same district	69.3	69.1	69.6	-0.5	0.912
Retained in the teaching profession	87.4	88.6	86.3	2.3	0.442
Sample Size (Teachers)	464	237	227		
Sample Size (Schools)	224	114	110		

Source: Mathematica Teacher Background Survey administered in fall 2005 and Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are regression adjusted using a logit model with robust standard errors to account for baseline characteristics and clustering of teachers within schools.

None of the differences is statistically significant at the 0.05 level.

Table D.10. Impacts on Teacher Retention Rates After Three Years (Percentages): Two-Year Districts

Outcome	All Teachers	Treatment	Control	Difference	P-value
Retained in the same school	50.9	54.2	47.1	7.1	0.159
Retained in the same district	63.0	64.9	60.9	4.0	0.388
Retained in the teaching profession	84.7	84.4	85.1	-0.7	0.850
Sample Size (Teachers)	375	208	167		
Sample Size (Schools)	152	82	70		

Source: Mathematica Teacher Background Survey administered in fall 2005 and Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are regression adjusted using a logit model with robust standard errors to account for baseline characteristics and clustering of teachers within schools.

None of the differences is statistically significant at the 0.05 level.

Cumulative attrition of control group members from the teaching profession was 15 percent after three years, on a pace to reach 25 percent after five years if we were to make a linear extrapolation of five percentage points per year. However, policy discussions about the problem of teacher turnover frequently cite attrition rates that are twice as high. For example, Stansbury and Zimmerman (2000) noted that "a third of beginning teachers quit within their first three years." The National Council on Teaching and America's Future (2003) noted that "within three years, 33 percent will leave and after five years…half of all new teachers will have exited the profession." The Alliance for Excellent Education (2004) reported that "after just three years it is estimated that almost a third of the new entrants to teaching have left the field, and after five years almost half are gone." The American Association of State Colleges and Universities (2006) report, "nearly half of new teachers leave the classroom within the first five years." The source for these statements is the 1991 Schools and Staffing Survey (SASS), conducted by the National Center for Education Statistics, yet the most authoritative analysis of those data indicated that 24 percent leave after two years and 46 percent leave after five years (Ingersoll 2002), rates that are still higher than those obtained in the current study.

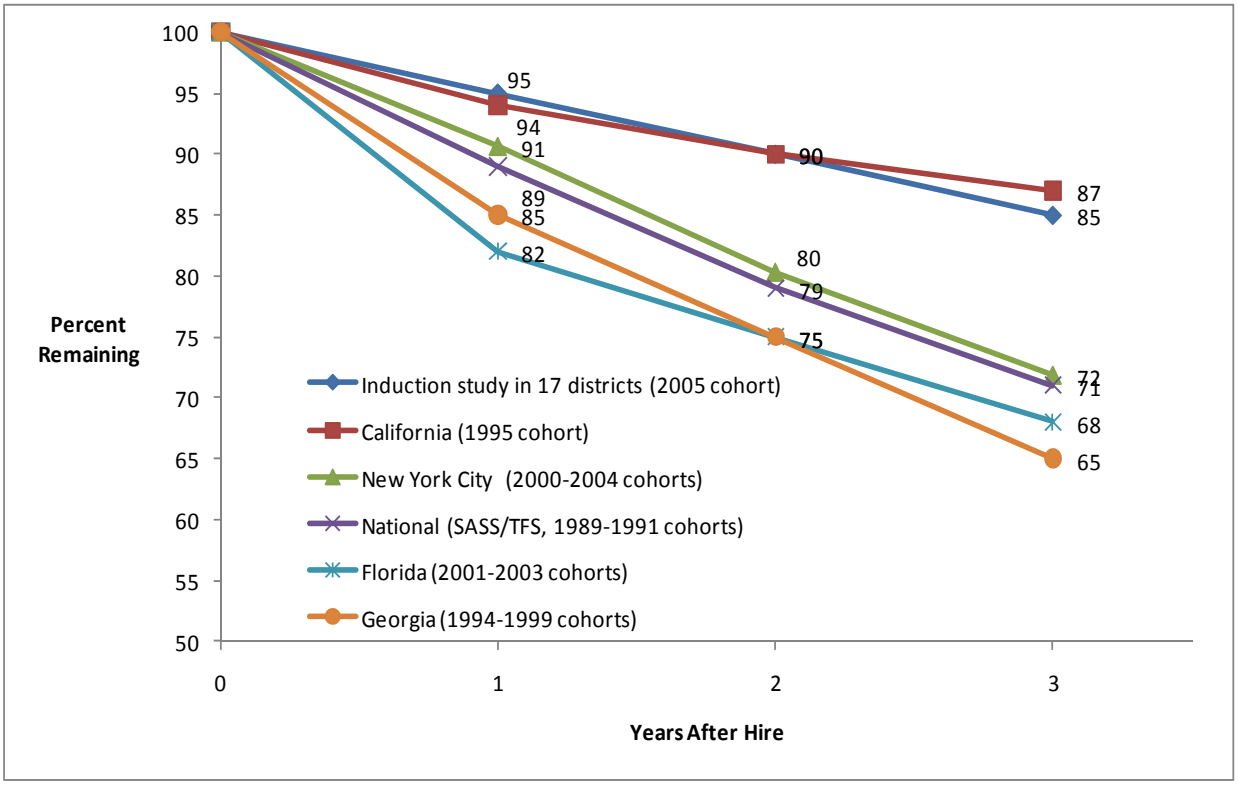
Given the concern about lower estimates of prevailing teacher retention rates, we compared the retention analysis in this study to similar analyses published by other researchers using a variety of data sources. The comparisons are meant to help clarify any discrepancies and place our study in perspective. Figure D.1 presents the survival curves for the current study alongside those that other researchers have calculated in different settings and time points around the country.

The data sources include one set of estimates based on SASS, the national survey, which followed teachers of all experience levels from 1991 to 1992, and several others from administrative datasets in three states (California, Florida, and Georgia) plus one district (New York City). The administrative data included various cohorts of teachers who began their careers in the late 1990s and early 2000s.

In contrast, this study followed the 1,009 teachers from the start of their careers in 2005 in traditional public schools in urban districts meeting our size and poverty criteria. We followed this single cohort of teachers into (what would be) the fourth year of their teaching career using surveys with response rates of at least 85 percent. We did not follow any teachers who were hired after the first week of school, because of concerns that the hiring of these teachers could be affected by treatment status. Nor did we follow temporary teachers, such as long-term substitutes. The retention rates for the study were approximately 95, 90, and 85 percent for the first three years, respectively.

The study of teachers in Georgia in the late 1990s by Scafidi et al. (2008) presented survival curves in which retention rates were about 85, 75, and 65 percent after each of the teachers' first three years, lower than comparable numbers for the teacher induction study. The difference, however, could be due to time period, geographic area, or method of measurement. In the Georgia study, figures were reported for women only and pertain to a period almost ten years prior to the current induction study. This study treated a teacher as a leaver if she left the state, even if she took a teaching position in a different state, or if she remained in state but was teaching in a private school.

Figure D.1. Teacher Survival Curves, Published Estimates



Source: Mathematica First, Second, and Third Teacher Mobility Surveys administered in fall 2006, fall 2007, and fall 2008 to all study teachers; California Teaching Commission 2002; Boyd et al. 2008; Ingersoll 2002; West and Chingos 2009; Scafidi et al. 2008.

A study of beginning teachers in Florida (West and Chingos 2009) reported retention rates of 82, 75, and 68 percent, all within 3 points of the corresponding Georgia estimates. Again, the measure of retention is from state administrative data and only includes retention in public schools in the state. Teachers who moved to private or charter schools are treated as if they had left the profession. The Florida data pertain to teachers who began in 2001 through 2003.[66]

A study of beginning teachers in New York City (Boyd et al. 2008) reported retention rates of 91, 80, and 72 percent. They focused on the three cohorts that started teaching between 2000 and 2004. Like the Florida study, the New York City study counts a teacher as being retained if she continues teaching public school in the state and as a leaver if she works in a private school, out of state school, or leaves teaching entirely.

A study by California's Commission on Teacher Credentialing reported rates of retention in the state's public education system that were 94, 90, and 87, closer to those reported in the current study. The California study, which used a state credentialing database and a state employment database, followed a cohort that began teaching 10 years before the induction study and measured

[66] The teachers are only observed through 2005, so the two-year retention estimate is based on the 2001 and 2002 cohorts only and the three-year retention estimate is based on the 2001 cohort only.

retention in public education in the state, although California is a larger state than Georgia or Florida.

Finally, Ingersoll (2002), the source frequently cited, is based on the SASS)conducted in 1990-1991), and the Teacher Followup Survey (TFS) which re-interviewed SASS leavers and a sample of SASS stayers one year later. Because it used survey methods instead of administrative data, the SASS/TFS analysis was able to count teachers who moved between sectors (public/private) and between states as stayers, but it covers an earlier time period from the induction study and includes suburban, rural, charter, and private schools.[67]

These comparisons show that estimates of teacher retention can vary over time and across different settings and methodologies. Only survey-based estimates are able to account for moves out of state and moves into private and charter schools, but estimates based on survey data can vary depending on when the data were collected, and whether public school teachers or a broader set of teachers form the population of interest.

2. Sensitivity of Impact Findings

We conducted several sensitivity tests to examine the robustness of the findings regarding the impacts of comprehensive induction on teacher retention. The results of these tests, shown in Tables D.11 and D.12, suggest that under a wide range of assumptions, we confirmed the finding that there was no significant impact of the treatment.

The conclusions did not change when we used an enhanced weight that incorporated information from the teacher background survey or when no weights were used.[68] Nor did they change when information was incorporated from data sources other than the mobility survey. For example, we coded the mobility status of nonrespondents who appeared in the student test score databases provided by the districts, reclassifying such teachers as district stayers. Similarly, we recoded the mobility status of nonrespondents who were flagged as unlocatable by the data collectors who called and visited the schools, reclassifying such teachers as district leavers. The variables edited in this way used more of the sample but led to the same conclusion of no significant impact of treatment.

[67] The SASS/TFS sample included only one year of followup, but the sample had teachers of all experience levels. Thus, the author was able to calculate cumulative retention rates using the probabilities of turnover for second- and third-year teachers in the same time period. It therefore combines information from cohorts that began teaching in 1988, 1989, and 1990.

[68] Unlike the enhanced weights, the benchmark weights rely only on school characteristics from the Common Core of Data compiled by the U.S. Department of Education. The enhanced weights used information on teacher's gender, age, race/ethnicity, home ownership, residence in the district, ACT/SAT score, preparation (whether completed a traditional four-year teacher training program), prior career, prior experience teaching, whether the teacher was hired after the school year began, whether the teacher attended a selective college/university, whether the teacher majored in an education-related field, and the amount of student-teaching experience.

Table D.11. Mobility Impacts After Three Years Under Alternative Assumptions: One-Year Districts

Outcome and Assumption	Treatment Group Mean	Control Group Mean	Difference (Estimated Impact)
Retention in the District			
Respondents			
Benchmark weights (benchmark estimates)	69.0	69.5	-0.5
No weights	68.9	70.2	-1.3
Enhanced weights	69.3	69.1	0.2
Alternative variance estimation (random effects model)	69.5	68.7	0.8
Alternative set of control variables 1	68.5	69.7	-1.3
Alternative set of control variables 2	69.7	68.9	0.8
Linear probability model	68.9	70.0	-1.2
Multinomial logit model	69.2	69.9	-0.1
Respondents and Nonrespondents			
Assume 100% of treatment nonrespondents are movers, 0% of controls	55.4	66.8	-11.4*
Assume 0% of nonrespondents are movers	66.6	66.8	-0.2
Assume 25% of nonrespondents are movers	69.0	71.4	-2.4
Assume 50% of nonrespondents are movers	65.9	66.8	-0.9
Assume 100% of nonrespondents are movers	55.3	52.7	2.6
Assume 0% of treatment nonrespondents are movers, 100% of controls	66.6	52.8	13.8*
Respondents and Selected Nonrespondents			
Recode selected nonrespondents from other data sources	70.6	71.7	-1.0
Recode selected nonrespondents and assume 100% of other nonrespondents are movers	60.5	57.2	3.2
Retention in the Teaching Profession			
Respondents			
Benchmark weights (benchmark estimates)	88.5	86.3	2.3
No weights	88.4	86.4	2.1
Enhanced weights	88.6	85.9	2.7
Alternative variance estimation (random effects model)	88.3	85.9	2.4
Alternative set of control variables 1	88.0	86.4	1.6
Alternative set of control variables 2	88.1	86.5	1.6
Linear probability model	88.1	86.8	1.3
Multinomial logit model	87.3	84.4	2.9
Respondents and Nonrespondents			
Assume 100% of treatment nonrespondents are leavers, 0% of controls	77.7	88.4	-10.7*
Assume 0% nonrespondents are leavers	89.2	88.2	1.0
Assume 25% of nonrespondents are leavers	86.1	85.4	0.7
Assume 50% of nonrespondents are leavers	**83.4**	**81.3**	**2.1**
Assume 100% of nonrespondents are leavers	77.7	74.2	3.4
Assume 0% of treatment nonrespondents are leavers, 100% of controls	89.2	74.2	14.9*
Respondents and Selected Nonrespondents			
Recode selected nonrespondents from other data sources	**88.8**	**87.0**	**1.8**
Recode selected nonrespondents and assume 100% of other nonrespondents are leavers	**82.9**	**78.7**	**4.1**
Sample Size (Teachers)			
Respondents	**215**	**201**	**416**
Respondents and Selected Nonrespondents	**228**	**215**	**443**
Respondents and Nonrespondents	**267**	**265**	**532**

Source: Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

*Significantly different from zero at the 0.05 level.

Table D.12. Mobility Impacts After Three Years Under Alternative Assumptions: Two-Year Districts

Outcome and Assumption	Treatment Group Mean	Control Group Mean	Difference (Estimated Impact)
Retention in the District			
Respondents			
Benchmark weights (benchmark estimates)	64.8	60.8	4.0
No weights	64.9	60.7	4.2
Enhanced weights	64.8	60.8	4.0
Alternative variance estimation (random effects model)	66.0	60.7	5.3
Alternative set of control variables 1	65.9	61.3	4.6
Alternative set of control variables 2	64.5	61.3	3.2
Linear probability model	65.3	61.1	4.2
Multinomial logit model	64.9	60.6	4.3
Respondents and Nonrespondents			
Assume 100% of treatment nonrespondents are movers, 0% of controls	55.5	63.3	-7.8
Assume 0% of nonrespondents are movers	62.1	62.9	-0.8
Assume 25% of nonrespondents are movers	67.4	63.4	4.0
Assume 50% of nonrespondents are movers	65.3	58.3	6.9
Assume 100% of nonrespondents are movers	55.9	46.5	9.4*
Assume 0% of treatment nonrespondents are movers, 100% of controls	62.5	46.0	16.5*
Respondents and Selected Nonrespondents			
Recode selected nonrespondents from other data sources	66.4	64.6	1.8
Recode selected nonrespondents and assume 100% of other nonrespondents are movers	58.5	55.3	3.2
Retention in the Teaching Profession			
Respondents			
Benchmark weights (benchmark estimates)	84.4	85.1	-0.7
No weights	84.4	84.9	-0.4
Enhanced weights	84.4	84.9	-0.5
Alternative variance estimation (random effects model)	85.1	84.7	0.3
Alternative set of control variables 1	85.0	85.1	-0.1
Alternative set of control variables 2	84.4	85.2	-0.8
Linear probability model	84.8	85.0	-0.2
Multinomial logit model	83.4	84.0	-0.6
Respondents and Nonrespondents			
Assume 100% of treatment nonrespondents are leavers, 0% of controls	79.2	87.4	-8.2*
Assume 0% of nonrespondents are leavers	85.9	87.1	-1.2
Assume 25% of nonrespondents are leavers	85.4	83.1	2.3
Assume 50% of nonrespondents are leavers	83.3	78.4	4.9
Assume 100% of nonrespondents are leavers	79.9	70.7	9.2*
Assume 0% of treatment nonrespondents are leavers, 100% of controls	86.5	70.1	16.4*
Respondents and Selected Nonrespondents			
Recode selected nonrespondents from other data sources	85.3	85.9	-0.6
Recode selected nonrespondents and assume 100% of other nonrespondents are leavers	82.2	79.6	2.6
Sample Size (Teachers)			
Respondents	190	154	344
Respondents and Selected Nonrespondents	196	172	368
Respondents and Nonrespondents	222	199	421

Source: Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

*Significantly different from zero at the 0.05 level.

When we re-estimated the treatment effects using a linear probability model and a multinomial logit model, we reached the same conclusions as when we used the benchmark model, which consisted of separate logistic regressions to predict the probability of being a stayer or a nonleaver (see Tables D.11 and D.12).

Tables D.11 and D.12 also present the results of imposing alternative assumptions regarding nonrespondents to the mobility survey. For that exercise, we recalculated the impact estimates under a range of different assumptions about the retention rates for nonrespondents. Under a wide range of assumptions, the treatment-control differences were not statistically significant. Under the most extreme assumptions, however, there could be significant treatment effects on retention ranging from -11 to +16 percentage points. Since there is no evidence to support these extreme assumptions, we report them here for context only.

Breaking down the overall impact estimates on retention into separate impacts by district shows that the results are not driven by one or two outlier districts. Figures D.2 and D.3 show impacts on retention of teachers in their original district, while Figures D.4 and D.5 show impacts on retention in the profession. Although the study was not designed to detect district-specific estimates, these results illustrate the possible heterogeneity across districts. However, the lack of obvious discontinuities in the distributions is consistent with the hypothesis that the variation in district-specific results represents a single treatment effect along with sampling variation.

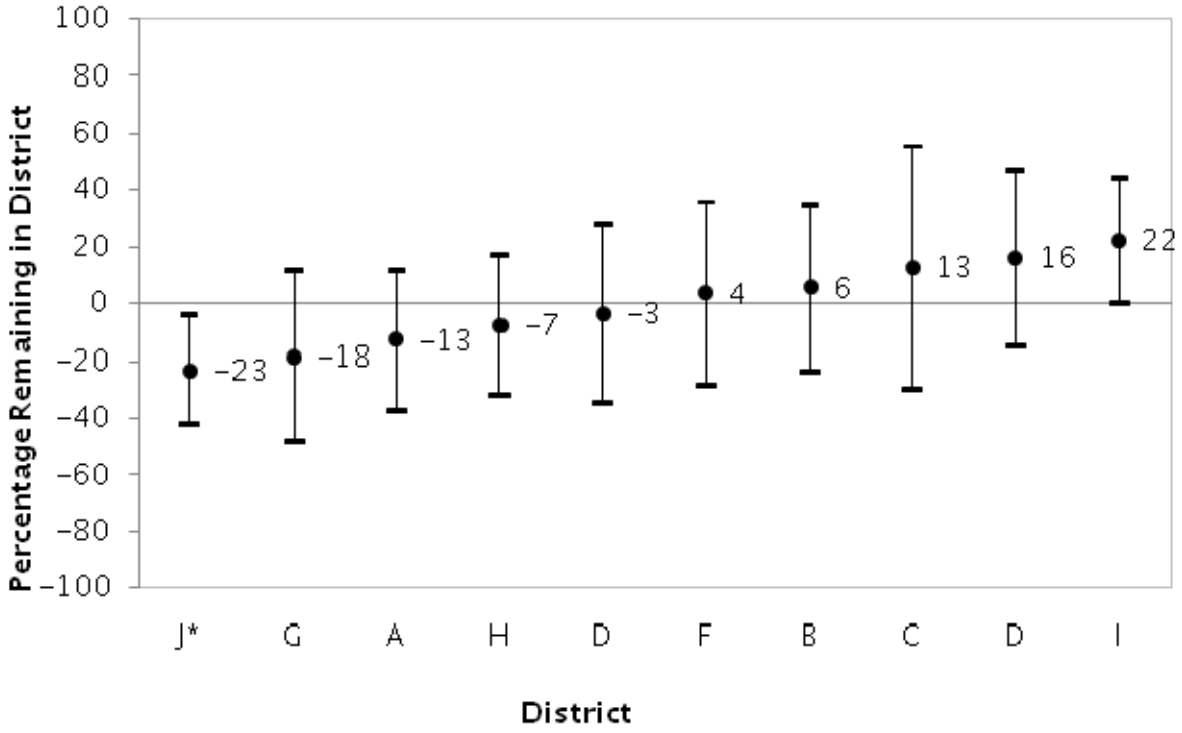
D. Supplementary Analysis of Composition of the Workforce

Another set of data from Chapter VI that we expand on here is the comparison of treatment and control stayers. The rationale for comparing stayers to stayers only without regard to the movers or leavers, for whom we also have data, can be summarized as follows. The composition effect that we are attempting to estimate is the difference between the average teacher quality under the treatment and the average teacher quality that what would have been realized in the absence of treatment. We recognize that the average teacher quality in both cases depends on the quality of teachers who stay and the quality of teachers who replace those who leave.

The true experimental impact is the difference between (a) the average outcome/characteristics of treatment stayers plus teachers who replace treatment leavers and (b) the average outcome of control stayers plus teachers who replace control leavers. Replacement teachers presumably come from a common pool of candidates who replace the leavers. We assume that the treatment has no impact on the quality of replacement teachers or the school principal's hiring process, in other words, that replacement teachers in treatment and control schools are equally well qualified. This way of formulating the experimental impact can be expressed in Equation (D.1):

$$\begin{aligned}\Delta &= [\lambda_t Y_t + (1-\lambda_t)Y_r] - [\lambda_c Y_c + (1-\lambda_c)Y_r] \\ &= \lambda_t Y_t - \lambda_c Y_c + (\lambda_c - \lambda_t)Y_r\end{aligned} \qquad \text{(D.1)}$$

Figure D.2. Impacts on Teacher Retention in the District After Three Years by District: One-Year Districts



Source: Mathematica Teacher Background Survey administered in fall 2005 and Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes A through J are arbitrary. Districts are ordered according to the size of the impact. N=464 teachers.

*Significantly different from zero at the 0.05 level. (No adjustment is applied for multiple comparisons.)

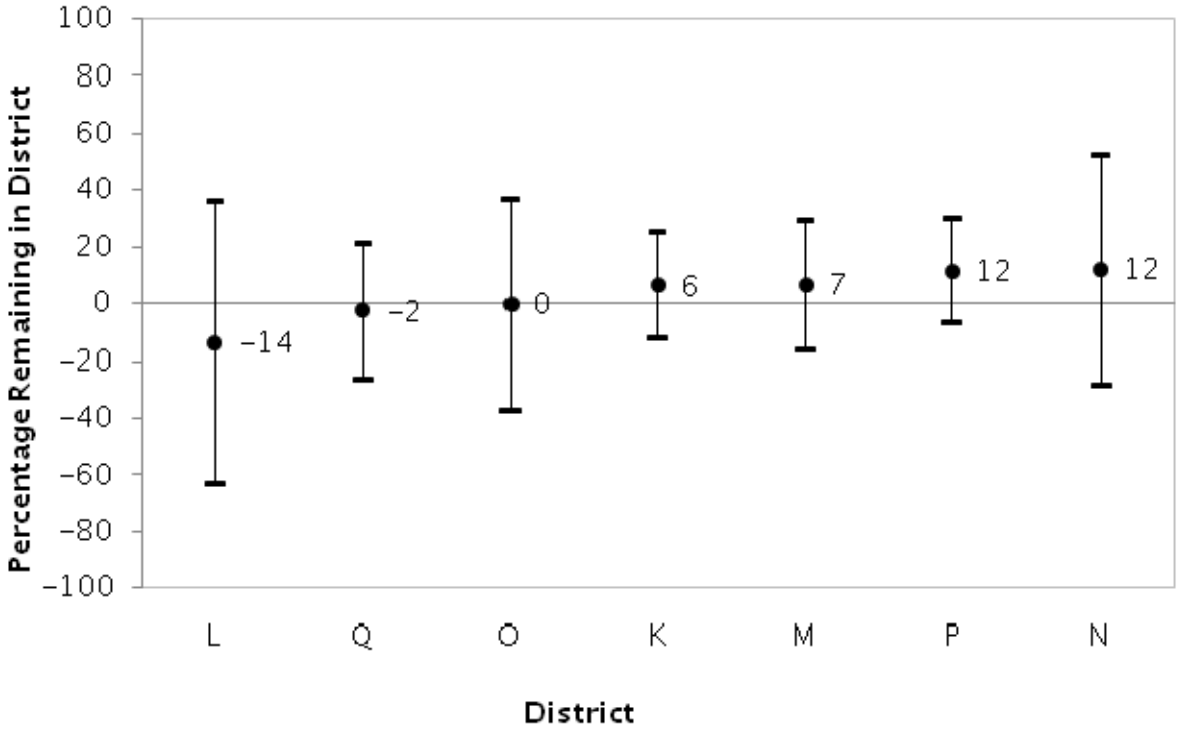
In this equation $Y_{t,}$ $Y_{c,}$ $Y_{r,}$ and $Y_{r,}$ represent the mean outcomes for treatment stayers, control stayers, and replacement teachers, respectively; Δ represents the impact of interest; and λ_t and λ_c represent the retention rates for the treatment and control group respectively.

We do not have data on replacement teachers, but we do not need to measure their outcomes explicitly. If we assume further that replacement teachers are similar to control teachers, because they all came from the same stream of teaching candidates, in other words that $Y_c = Y_r$ then the impact reduces to the difference between treatment stayers' mean and control stayers' mean (see Equation D.2), which is what we reported in Chapter VI.

$$\Delta = Y_t - Y_c \qquad \text{(D.2)}$$

Nevertheless, we present the full set of mean characteristics for stayers, movers, and leavers in Tables D.13 and D.14 to complement Tables VI.1 and VI.2 and provide additional context.

Figure D.3. Impacts on Teacher Retention in the District After Three Years by District: Two-Year Districts

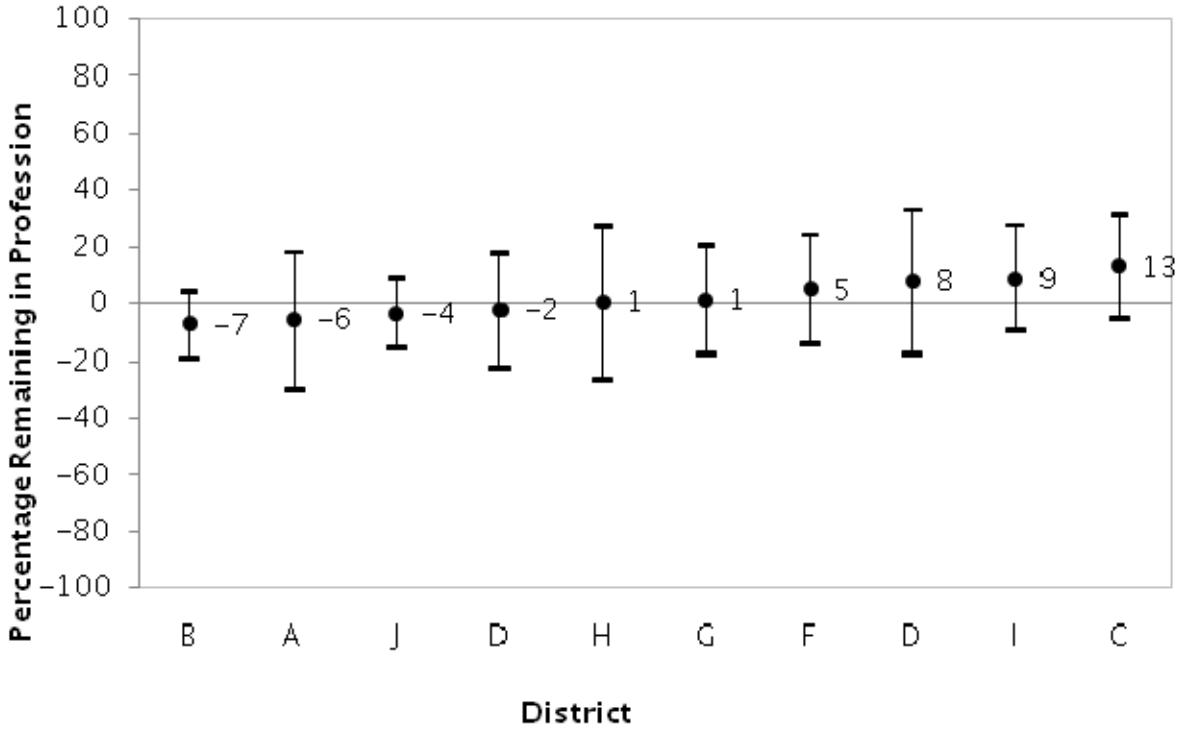


Source: Mathematica Teacher Background Survey administered in fall 2005 and Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes K through Q are arbitrary. Districts are ordered according to the size of the impact. N=375 teachers.

None of the differences is statistically significant at the 0.05 level.

Figure D.4. Impacts on Teacher Retention in the Profession After Three Years by District: One-Year Districts

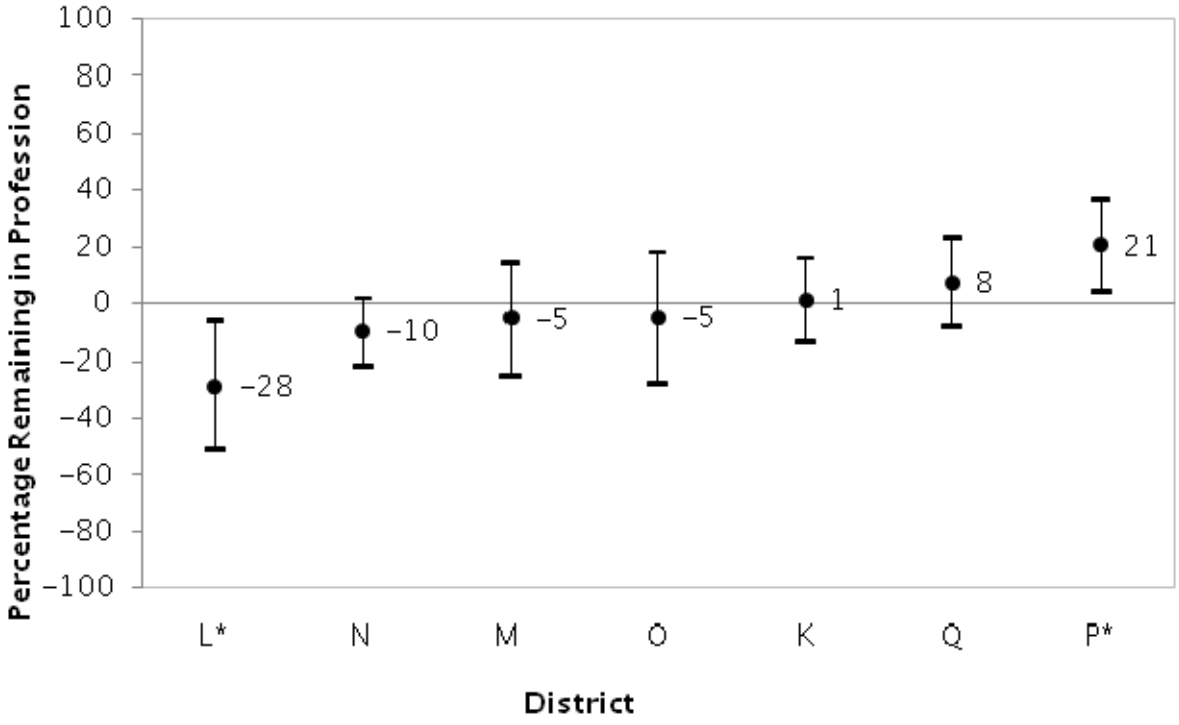


Source: Mathematica Teacher Background Survey administered in fall 2005 and Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

Note: Data pertain to teachers in one-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes A through J are arbitrary. Districts are ordered according to the size of the impact. N=464 teachers.

None of the differences is statistically significant at the 0.05 level.

Figure D.5. Impacts on Teacher Retention in the Profession after Three Years by District: Two-Year Districts



Source: Mathematica Teacher Background Survey administered in fall 2005 and Third Teacher Mobility Survey administered in fall 2008 to all study teachers.

Note: Data pertain to teachers in two-year districts participating in the study. Data are weighted and regression adjusted using ordinary least squares to account for differences in benchmark covariates and robust variance estimation to account for study design and the clustering of teachers within schools. Plot symbols represent the difference between regression-adjusted treatment and control mean within each district, and the vertical lines show the 95 percent confidence interval around each point. District codes K through Q are arbitrary. Districts are ordered according to the size of the impact. N=375 teachers.

*Significantly different from zero at the 0.05 level. (No adjustment is applied for multiple comparisons.)

Table D.13. Characteristics of District Stayers, Movers, and Leavers After Three Years, by Treatment Status (Percentages Except Where Noted): One-Year Districts

	Stayers			Movers			Leavers		
Teacher Characteristic	Treat-ment	Control	Differ-ence	Treat-ment	Control	Differ-ence	Treat-ment	Control	Differ-ence
Characteristic									
College entrance exam scores (SAT combined score or equivalent)	1040	1013	27	1030	1033	-3	1010	1083	-73
Attended highly selective college	27.5	27.2	0.3	26.6	39.1	-12.4	44.3	31.2	13.1
Major or minor in education	78.7	80.9	-2.1	69.4	75.8	-6.4	83.5	60.2	23.3*
Student teaching experience (weeks)	15.8	15.4	0.4	13.9	14.3	-0.4	18.3	11.6	6.7*
Highest degree is master's or doctorate	22.4	28.2	-5.8	19.2	26.7	-7.5	29.5	31.8	-2.3
Entered the profession through traditional four-year program	67.6	58.9	8.7	66.6	61.0	5.6	48.3	34.6	13.6
Certified (regular or probationary)	94.7	94.7	0.0	94.8	96.3	-1.5	90.6	96.8	-6.2
Sample Size (Teachers)	**148**	**139**		**39**	**31**		**29**	**31**	
Sample Size (Schools)	**88**	**84**		**30**	**28**		**22**	**28**	
Year 1 Classroom Observation Score (on 1 to 5 scale)									
Content of a literacy lesson	2.3	2.6	-0.3*	2.1	2.1	0.1	2.8	2.1	0.6*
Implementation of a literacy lesson	2.6	2.8	-0.2	2.6	2.4	0.1	2.9	2.6	0.3
Classroom culture	3.0	3.1	-0.1	2.9	2.8	0.1	3.5	3.0	0.5
Sample Size (Teachers)	**100**	**94**		**26**	**23**		**17**	**18**	
Sample Size (Schools)	**71**	**65**		**20**	**20**		**14**	**18**	

Source: Mathematica calculations using data from the College Board and ACT, Inc.; Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data are weighted to account for the study design. Sample sizes vary due to item nonresponse. The analysis of college entrance exam scores relied on a smaller sample of teachers (84/27/14 treatment stayers/movers/leavers and 86/23/18 control stayers/movers/leavers) and schools (61/22/12 treatment and 62/21/17 control). Stayer: retained in the same school district. Mover: retained in the teaching profession, but not in the same school district. Leaver: no longer teaching.

None of the differences between treatment and control stayers, between treatment and control movers, or between treatment and control leavers is statistically significant at the 0.05 level. p-values are suppressed to make the table easier to read.

Table D.14. Characteristics of District Stayers, Movers, and Leavers After Three Years, by Treatment Status (Percentages Except Where Noted): Two-Year Districts

	Stayers			Movers			Leavers		
Teacher Characteristic	Treat-ment	Control	Differ-ence	Treat-ment	Control	Differ-ence	Treat-ment	Control	Differ-ence
Characteristic									
College entrance exam scores (SAT combined score or equivalent)	905	935	-30	1022	1027	-5	1048	1124	-76
Attended highly selective college	23.7	21.4	2.3	37.5	38.5	-1.0	42.3	49.8	-7.5
Major or minor in education	67.8	66.6	1.2	64.3	74.8	-10.5	60.3	67.6	-7.2
Student teaching experience (weeks)	12.3	12.3	0.1	11.4	14.1	-2.7	9.9	13.2	-3.3
Highest degree is master's or doctorate	16.7	10.2	6.5	10.2	27.2	-17.0*	25.4	0.0	25.4*
Entered the profession through traditional four-year program	61.1	66.4	-5.4	61.9	68.0	-6.2	35.4	65.0	-29.7*
Certified (regular or probationary)	95.8	92.9	2.9	91.4	94.2	-2.8	81.5	84.3	-2.8
Sample Size (Teachers)	**124**	**93**		**35**	**36**		**31**	**26**	
Sample Size (Schools)	**67**	**52**		**28**	**27**		**22**	**21**	
Year 1 Classroom Observation Score (on 1 to 5 scale)									
Content of a literacy lesson	2.4	2.4	0.0	2.2	2.4	-0.1	2.3	2.3	0.0
Implementation of a literacy lesson	2.7	2.6	0.1	2.6	2.7	-0.2	2.7	2.3	0.3
Classroom culture	3.1	3.1	0.1	2.7	3.0	-0.3	2.9	2.8	0.2
Sample Size (Teachers)	**87**	**62**		**28**	**27**		**24**	**16**	
Sample Size (Schools)	**50**	**41**		**22**	**21**		**18**	**12**	

Source: Mathematica calculations using data from the College Board and ACT, Inc.; Mathematica Third Teacher Mobility Survey administered in fall 2008 to all study teachers; Mathematica classroom observations conducted in spring 2006.

Note: Data are weighted to account for the study design. Sample sizes vary due to item nonresponse. The analysis of college entrance exam scores relied on a smaller sample of teachers (56/20/21 treatment stayers/movers/leavers and 47/17/16 control stayers/movers/leavers) and schools (40/17/17 treatment and 35/13/14 control). Stayer: retained in the same school district. Mover: retained in the teaching profession, but not in the same school district. Leaver: no longer teaching.

None of the differences between treatment and control stayers, between treatment and control movers, or between treatment and control leavers is statistically significant at the 0.05 level. p-values are suppressed to make the table easier to read.


DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA