NCEE 2009-4072

U.S. DEPARTMENT OF EDUCATION

Impacts of Comprehensive Teacher Induction

Results from the Second Year of a Randomized Controlled Study

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

Impacts of Comprehensive Teacher Induction

Results from the Second Year of a Randomized Controlled Study

August 2009

Eric Isenberg
Steven Glazerman
Martha Bleeker
Amy Johnson
Julieta Lugo-Gil
Mary Grider
Sarah Dolfin
Mathematica Policy Research

Edward Britton
WestEd

Melanie Ali
Project Officer
Institute of Education Sciences

NCEE 2009-4072
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

National Center for Education Evaluation and Regional Assistance
Phoebe Cottingham
Commissioner

August 2009

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 was the product of many people's efforts beyond the authors of this report. 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, 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. Megan Hague Angus, Andrea Mraz Esposito, Kristin Hallgren, Samina Sattar, Kathy Sonnenfeld, and Ama Takyi-Laryea coordinated with districts at several stages. Matthew Jacobus, assisted by Dominic Harris and Norma Altshuler, provided expert programming assistance. Korantema Kaleem provided project management. Neil Seftor carefully reviewed drafts of the report and provided helpful comments. The report was edited by Laura Bernstein and was produced by Donna Dorsey.

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

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.

CONTENTS

Chapter — **Page**

Chapter **Page**

Chapter **Page**

Chapter **Page**

TABLES

Table **Page**

Table **Page**

Table **Page**

Table **Page**

Table **Page**

Table **Page**

Table **Page**

FIGURES

Figure **Page**

EXECUTIVE SUMMARY

One of the main policy responses to the problems of turnover and inadequate preparation among beginning teachers is to support them with a formal, comprehensive induction program. 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 formative assessment (Berry et al. 2002).

In practice, teacher induction is common, but induction that is intensive, comprehensive, structured, and sequentially delivered in response to teachers' emerging pedagogical needs is not (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.

There is little empirical evidence on whether investing 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), little of the research on teacher induction to date has been conclusive or rigorous. Research based on federal statistics (for example, 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 to the extent 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, Youngs 2002; Fuller 2003; Rockoff 2008) 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.

Congressional interest in formal, comprehensive teacher induction has grown in recent years. 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 a 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 highlight the need to conduct rigorous research to determine whether comprehensive teacher induction programs produce a measurable impact on teacher retention and other positive outcomes for teachers and students.

The National Center for Education Evaluation and Regional Assistance within the U.S. Department of Education's Institute of Education Sciences (IES) contracted with Mathematica Policy Research (MPR) to address this issue by evaluating the impact of structured and intensive teacher induction programs over a three year time period, beginning when teachers first enter the teaching profession. An earlier report (Glazerman et al. 2008) presented results from the first year of the evaluation. The current report presents findings from the second year of the evaluation and a future report will present findings from the third and final year.

Throughout the report, we refer to the more formal, structured programs as "comprehensive" induction. The study examines whether comprehensive teacher induction programs lead to higher teacher retention rates and other positive teacher and student outcomes as compared to prevailing, generally less comprehensive approaches to supporting new teachers. More specifically, the study is designed to address five research questions on the impacts of comprehensive teacher induction:

1. What is the effect of comprehensive teacher induction on the types and intensity of induction services teachers receive compared to the services they receive from the districts' current induction programs?

2. What are the impacts on teachers' classroom practices?[2]

3. What are the impacts on student achievement?

4. What are the impacts on teacher retention?

5. What is the impact on the composition of the district's teaching workforce?

To operationalize the concept of comprehensive teacher induction, we issued a Request for Proposals (RFP) in 2004 to select a comprehensive induction program and program provider for the study. The RFP specified that the induction program should include several

[2] As Glazerman et al. (2008) reports, there was no impact of comprehensive teacher induction on classroom practices in the first year of implementation. Because we did not return to observe classrooms during the second year of the evaluation, we do not re-visit the question about classroom practices in the current report.

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 submitted by Educational Testing Service of Princeton, New Jersey (ETS) and the New Teacher Center at the University of California-Santa Cruz (NTC) as most closely meeting 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; and
- Outreach to district and school-based administrators to educate them about program goals and to garner their systemic support for the program.

MPR contracted with both providers to deliver comprehensive induction services to the districts in the study, with one-half of the districts assigned to ETS, the remaining half to NTC. Researchers from WestEd, a subcontractor to MPR, monitored the implementation of the comprehensive induction services to help the providers ensure there was fidelity to the core service model and to identify and help address any implementation challenges that arose.

STUDY DESIGN

The centerpiece of the study design is the use of random assignment to create a group of teachers exposed to comprehensive teacher induction (treatment) and an equivalent group exposed to the district's usual set of induction services (control). The study design allows us to measure and compare outcomes for these two groups to estimate the impacts of comprehensive induction relative to the services teachers receive from their district's prevailing induction program. We used surveys and school records to measure the background of the study teachers, their receipt of induction services and alternative support services, their attitudes, and the key outcomes of student achievement and teacher mobility.

We selected 17 school districts to participate in the study. District selection was based upon factors such as district size and poverty, whether the district was already implementing a comprehensive teacher induction program, and district willingness to participate in the evaluation. The selected districts, which were spread across 13 states, served low-income students, with every district in the study having more than 50 percent of its students

qualifying for the federal School Lunch Program. We then assigned each district to one of the two providers of comprehensive induction, either ETS or NTC, based primarily on district preferences. Nine districts participated in the ETS program; eight districts participated in the NTC program. 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.

IES later expanded 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 (during the 2005-2006 school year), and the other for districts that received two years (during the 2005-2006 and 2006-2007 school years). In the two-year districts, teachers who had been assigned to the treatment group were offered continued services for a second year. The goal of this expansion was to enable the study to address its main research questions separately for one-year and two-year comprehensive induction programs. Policymakers are interested in both models of service delivery because they are both viable policy options for future implementation.

We used convenience sampling to select the districts to receive a second year of the treatment; we selected the districts based upon 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 districts that received two years of treatment, but instead allows us to investigate the effectiveness of one-year programs separately from that of two-year programs.

In this Year 2 impact report, unlike the Year 1 impact report (Glazerman et al. 2008), we present findings separately for the set of 10 districts that received one year of treatment ("one-year districts") and the other set of 7 districts that received two years of treatment ("two-year districts"). Both sets of findings are based on data collected through two years of the study. When appropriate, however, we compare outcomes from the first year of the study to outcomes from the second year of the study within the one-year districts and within the two-year districts.

Within each district, a subset of elementary schools participated in the study. As noted above, we randomly assigned these elementary schools to either a treatment group, which was offered comprehensive teacher induction, or a control group, which took part in the district's usual teacher induction program. The final sample size included 418 schools across the 17 districts.

Within each study school, we selected all eligible teachers, defined as beginning teachers who met certain criteria: taught in an elementary grade (K-6); were new to the profession; and were not already receiving induction support from a teacher preparation or certification program. Under these criteria, the 252 schools in the one-year districts contained 561 eligible teachers, and the 166 schools in the two-year districts contained 448 eligible teachers. For the student achievement analysis, we limited the collection of student test score data to

teachers meeting another set of eligibility criteria, including teaching a self-contained classroom in a tested grade and subject. This resulted in the collection of reading test scores for 139 teachers and math scores for 123 teachers in the one-year districts, and of reading scores for 96 teachers and math scores for 95 teachers in the two-year districts.[3]

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 of the control group. Therefore, we assigned all eligible teachers to treatment or control status based on the school where they were expected to teach at the point of random assignment.

METHODS AND DATA

We used a model-based approach to estimate program impacts. The statistical model explicitly acknowledges the hierarchical structure of the data—for example, the nesting of teachers within schools—an approach that is sometimes referred to as a hierarchical linear model (HLM). Accordingly, we can properly specify the units of analysis (teachers and schools) and devise unbiased estimates of the standard errors that we used to conduct hypothesis tests. The model also allows us to control for the effects of a range of teacher and school characteristics on the outcomes of interest to increase the precision of the estimates of treatment effects.

For each outcome, we use a different set of control variables (covariates), described in the discussion of key study findings. The control variables used in the body of the report are called the benchmark control variables; in sensitivity analyses presented in appendices to the report, we alter the control variables to test the robustness of the results. These sensitivity tests included re-estimation of the study's main impacts with different sets of covariates, using different samples or sample weights, and different statistical model assumptions.

Data for the study were collected from a variety of sources. In fall 2005 we surveyed mentors participating in the comprehensive induction programs on their background characteristics and reviewed program documents from ETS and NTC. We administered a baseline survey of beginning teachers in fall 2005, at which time we also requested teachers' permission to obtain their college entrance examination scores (SAT or ACT). The baseline survey asked teachers about their formal education, professional training, current teaching assignment, and personal background. We surveyed teachers twice during the 2005-2006 school year on the induction activities in which they participated, including questions about duration and intensity of mentoring and professional development as well as questions about satisfaction with different aspects of their current teaching position. During the 2006-2007 school year, we surveyed teachers in the two-year districts twice and teachers in the one-year

[3] The standard errors of test score impact estimates were in the range of 0.05 to 0.08, meaning that an impact in effect size units of 0.10 to 0.16 would be statistically significant. The study was originally designed to detect test score impacts of 0.10 to 0.22 (Glazerman et al. 2005).

districts once on the induction activities in which they participated and on their job satisfaction.

For the report's core outcomes measuring the impacts of comprehensive teacher induction, we collected districts' student records data at the end of the 2006-2007 school year and conducted the second of three mobility surveys in fall 2007 to learn about teacher retention. We measured student achievement outcomes using district-administered test score data from the spring 2007 (posttest) for students taught by study teachers in the 2006-2007 school year and students' linked scores from the prior grade in spring 2006 (pretest).[4] We conducted all treatment-control comparisons within grade and within district to ensure that treatment status was not confounded with properties of the test. Response rates on teacher surveys ranged from 88 percent to 97 percent for the treatment group and 78 percent to 92 percent for the control group. We used nonresponse adjustment weights and sensitivity analyses to address the differential response rates in the analysis of teacher mobility.

THE TREATMENT: COMPREHENSIVE INDUCTION SERVICES

Treatment teachers in each district were given the opportunity (but were not required) to participate in the comprehensive induction program implemented there. The comprehensive induction program components included carefully selected and trained full-time mentors; a curriculum of intensive and structured support for beginning teachers; 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.

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

[4] For three districts that tested at least some students in the fall, we used a fall 2006 test as a pretest and/or a fall 2007 test as a posttest.

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).

In Year 1, the providers brought their respective mentors together for 10 to 12 days of training. The training was spread across four sessions of 2 to 3 days, with the first session held during the summer of 2005 and the rest taking place throughout the 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. During Year 2, ETS and NTC continued intensive training of their respective mentors in the seven districts that were selected to continue program implementation. ETS brought mentors together for a total of 8 days over 3 sessions. NTC did so over 10 days and 4 sessions. The providers devoted 1.5 to 2.5 days per session. All mentors participated in the trainings, which reflected a focus similar to Year 1. In sum, in two-year districts ETS mentors participated in 18 days of training; NTC mentors participated in 22 days.

Practitioners and policymakers should be aware of two issues related to program implementation. The first is the voluntary nature of teachers' participation in the treatment services. The program models that were implemented did not necessarily require teachers to participate but rather made services available to them, so not all teachers attended every professional development session provided.

The second issue for practitioners and policymakers to be aware of is 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, we aimed for consistent implementation of each program, with a high level of fidelity to the program design and a quick response to any implementation issues. Second, the providers adapted their program for the study to ensure that the required components were included in a one-year curriculum. Once it was decided to add a second year, the programs made additional modifications and adaptations to extend the curriculum another year. Finally, each provider organized off-site mentor training sessions, bringing together the mentors from all of the provider's study districts. For district-wide implementation with a larger number of mentors, training typically occurs within the district, rather than off-site together with mentors from other districts.

THE COUNTERFACTUAL: PREVAILING INDUCTION SERVICES

We designed the study to compare teachers who were exposed to comprehensive teacher induction services (treatment) to an equivalent group that was exposed to the induction services normally offered by the districts (control). We purposefully selected districts whose schools were not already working with ETS or NTC on induction projects, were not using the providers' induction materials, were not spending more than $1,000 per teacher on induction, and did not assign full-time release mentors to work with beginning teachers.

SUMMARY OF FINDINGS AFTER ONE YEAR: ONE-YEAR AND TWO-YEAR DISTRICTS COMBINED

An earlier report (Glazerman et al. 2008) presented findings after the first year of implementation of the comprehensive induction program within study districts. That report showed that teachers assigned to the treatment group reported significantly more induction support, but also that the additional support did not translate into positive impacts on key outcomes after one year.[5] The additional induction support amounted to a greater likelihood of having a mentor formally assigned to beginning teachers (93 versus 75 percent), more time spent in meetings with the mentor (95 versus 74 minutes per week), and greater likelihood of receiving "a moderate amount" or "a lot" of assistance from mentors in areas such as classroom management (65 versus 40 percent), reviewing student work (55 versus 30 percent), and communicating with parents (38 versus 31 percent). There were no positive impacts on classroom practices, student achievement, teacher retention, or the composition of the district's teaching workforce after one year. Nor did we find any evidence of positive impacts on teachers' satisfaction or feelings of preparedness.

[5] All references to "significance" in this report refer to statistical significance. A difference is deemed statistically significant in this report if the probability that it was observed by chance is less than 5 percent. The term "statistically insignificant" does not imply irrelevance for policymakers and similarly the term "statistically significant" does not necessarily mean "large" or meaningful for policy.

SUMMARY OF FINDINGS AFTER TWO YEARS: TREATMENT-CONTROL DIFFERENCES IN ONE-YEAR DISTRICTS

Induction Services Received

Within one-year districts, during Year 1—the year in which comprehensive teacher induction was implemented—we found statistically significant differences between the treatment and control group; the treatment group reported receiving more induction support than the control group across a broad range of measures of the amount, types, and content of supports.

In Year 2—the year in which treatment teachers no longer received comprehensive teacher induction supports—the percentage of teachers with an assigned mentor and the weekly minutes spent with that mentor declined from Year 1 to Year 2 (differences with a p-value of 0.000) for both the treatment and control groups. During this second year, we found statistically significant negative impacts on these and other measures of support, as described below.

Because teachers in one-year districts were not surveyed in the spring of Year 2, we focus the discussion on findings for the fall of each year.[6] Estimates were computed using an ordinary least squares model with district and grade assignment fixed effects that accounted for clustering of teachers within schools; weights were applied to adjust for survey nonresponse and the study design.[7]

Amount of Mentoring. In Year 1, we found statistically significant differences in the likelihood of teachers reporting having a mentor assigned to them and having a full-time mentor. As part of the intervention, every treatment teacher was assigned a mentor by ETS or NTC, but that did not guarantee that all teachers would work with their mentor or acknowledge having had one assigned to them. Still, treatment teachers were more likely than control teachers to report having a mentor assigned to them (90 versus 70 percent) and to report having a full-time mentor (74 versus 8 percent). We found statistically significant differences in teachers' likelihood of having a mentor who was another teacher and in the amount of time teachers reported spending with a mentor during the most recent full week of teaching. Treatment teachers were less likely than control teachers to report having a mentor who was another teacher (25 versus 64 percent). In addition, treatment teachers reported spending an average of 87 minutes per week in mentor meetings compared to 67 minutes for control teachers, with the 20-minute difference attributable entirely to differences in the duration of scheduled meetings, as opposed to informal meetings.

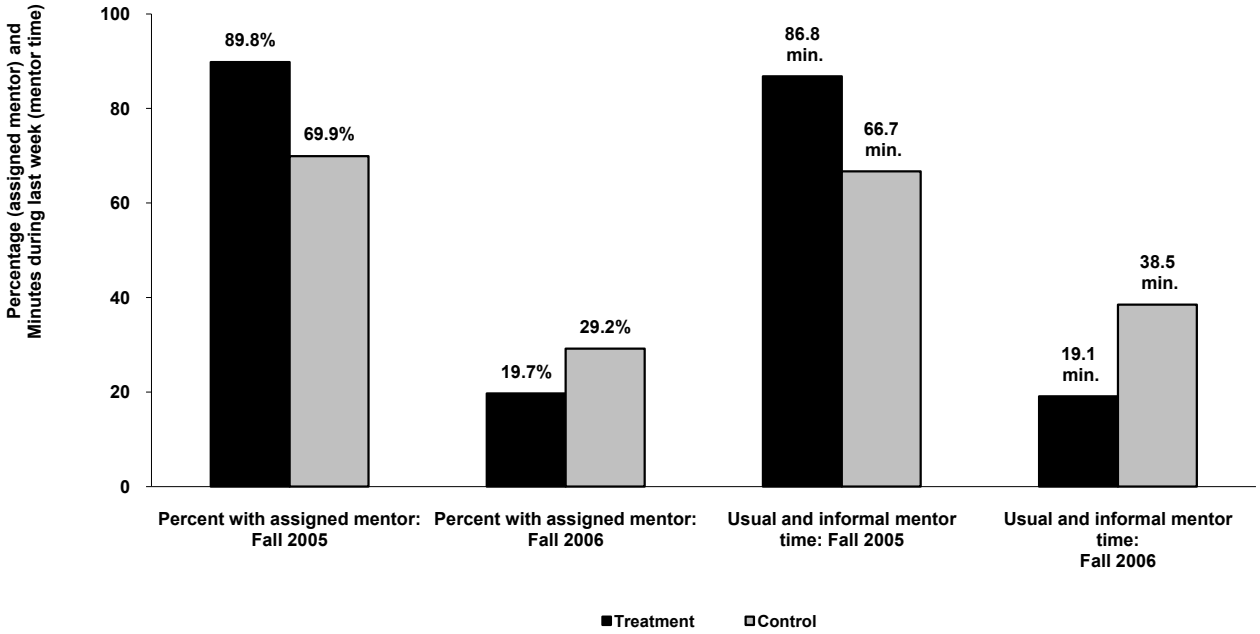
In Year 2, we found statistically significant differences in the prevalence of and time spent in mentoring. Treatment teachers were *less* likely than control teachers to report having

[6] Findings from the fall of Year 1 can be compared to findings from the spring of Year 1, which are shown in Appendix C.

[7] Across all outcomes, the same methods were used in the analysis of two-year districts.

a mentor assigned to them (20 versus 29 percent). Treatment teachers were also less likely than control teachers to report having a mentor who was another teacher (21 versus 31 percent). Treatment teachers spent less time in mentor meetings than control teachers (19 versus 39 minutes per week). Figure ES.1 shows treatment-control differences for having an assigned mentor and time in mentor meetings in Year 1 and Year 2.

Figure ES.1. Treatment-Control Differences in Percent Assigned a Mentor and Total Minutes Spent in Mentoring Per Week: One-Year Districts, Fall 2005 and Fall 2006



Note: All treatment-control differences are significantly different from zero at the 0.05 level, two-tailed test (N=503 teachers in fall 2005 and 472 teachers in fall 2006).

Mentor Activities and Assistance. In Year 1, treatment and control teachers' reports showed statistically significant differences in the amounts of time in various mentor activities and the kinds of assistance received from their mentors. Treatment teachers reported spending more time during the most recent full week of teaching being observed by mentors (34 versus 10 minutes), meeting one-on-one with mentors (34 versus 23 minutes), meeting with mentors together with other first-year teachers (29 versus 9 minutes), and having mentors model lessons (9 versus 6 minutes). During the most recent full week of teaching, treatment teachers were 14 to 27 percentage points more likely than control teachers to report having received mentors' assistance in a variety of topic areas, such as receiving suggestions to improve practice (77 versus 53 percent) and discussing instructional goals (73 versus 48 percent).

By Year 2, we found statistically significant differences in the amount of time teachers reported being observed by mentors during the most recent full week of teaching in fall 2006. Treatment teachers reported less time in a list of six common mentoring activities (22 versus 36 minutes per week) including less time being observed by mentors than control

teachers (2 versus 6 minutes). No statistically significant differences were found between treatment and control group teachers on their reported time spent in any of the other five activities covered by the survey. During the most recent full week of teaching in fall 2006, treatment and control teachers' reports showed statistically significant differences in the likelihood of receiving mentors' assistance in each of the topic areas covered by the survey. Treatment teachers were less likely than control teachers to report receiving mentors' assistance in each topic area, with effects ranging from 8 to 14 percentage points, including, for example, impacts on receiving suggestions to improve practice (15 versus 27 percent) and discussing instructional goals (14 versus 24 percent).

Professional Development. We did not find statistically significant differences between treatment and control teachers in their reported attendance in professional development, except in certain areas. Of the 12 professional development topics covered by the survey, treatment teachers were less likely than control teachers to report having attended professional development sessions in two areas in fall 2005 (Year 1): content area knowledge (61 versus 72 percent) and preparing students for standardized testing (30 versus 41 percent). We did not find statistically significant differences between treatment and control teachers in their reported attendance in any of the 12 professional development activities in fall 2006 (Year 2).

Student Achievement

In Year 2 (school year 2006-2007), we found no statistically significant impacts on reading or math scores in the one-year districts. We compared the test scores for students of treatment teachers to those of control teachers using post-test scores measured in 2007 adjusted for pre-test scores measured in 2006. The test score analysis was based on standardized achievement tests that the district normally conducts.[8] Though 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 most important and worthy of assessing. 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. We kept two broad subject areas, math and reading, distinct. The benchmark model accounts for the nesting of students within schools, using the normalized student pretest score and district-by-grade fixed effects as covariates.

The benchmark impacts on math and reading scores in Year 2 were not significantly different from zero (see Table ES.1). We confirmed that the impact on math and reading in the second year was not statistically significant when the impacts were re-estimated using different samples, sets of covariates, or estimation techniques.

[8] The specific test differs from district to district, and in some cases by grade within district. However, all treatment-control comparisons were made using a common set of tests within grade within district.

Table ES.1. Impacts on Test Scores: One-Year Districts, 2006-2007 School Year

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Subject	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
Reading	0.05	0.01	0.04	0.04	0.380	2,245	135	9
Math	0.05	-0.02	0.08	0.08	0.367	1,995	117	9

Source: MPR analysis of data from 2005-2006 and 2006-2007 school years provided by participating school districts.

Notes: Data are regression-adjusted to account for pretest, district-by-grade fixed effects, and clustering of students within schools. For Reading, there were 1,193 students and 72 teachers in the treatment group, and 1,052 students and 63 teachers in the control group. For Math, there were 994 students and 57 teachers in the treatment group, and 1,001 students and 60 teachers in the control group.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Teacher Retention

We found that comprehensive teacher induction had no statistically significant impact on teacher retention after two years. We measured teacher retention in terms of the percentage of teachers who remained in their originally assigned school, their district, and the teaching profession. Table ES.2 shows the result of the three hypothesis tests specifically focused on retention in the school, in the district, and in the profession as binary outcomes. For each of the outcomes, there was no statistically significant impact. The same result was obtained when we expanded the number of outcomes to differentiate between moving to a school in another public school district and moving to a private, parochial, or other school, and expanded the outcomes for leaving to include leaving to stay at home, leaving to attend school or take a new job, and other reasons for leaving.

Table ES.2. Impacts on Teacher Retention Rates after Two Years (Percentages): One-Year Districts

Outcome	All Teachers	Treatment	Control	Difference	P-value
Retained in the same school	62.5	60.3	64.7	-4.5	0.280
Retained in the same district	79.5	78.6	80.3	-1.7	0.619
Retained in the teaching profession	90.1	90.4	89.8	0.7	0.789
Unweighted Sample Size (Teachers)	476	244	232		
Unweighted Sample Size (Schools)	227	114	113		

Source: MPR Mobility Survey administered in 2007-2008 and Teacher Background Survey administered in 2005-2006 to all study teachers.

Note: 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, two-tailed test.

We also examined the reasons that teachers who left their districts (movers) or left the teaching profession (leavers) gave for leaving and found no statistically significant impacts of treatment. When we asked leavers whether they expected to return and if so, when they would do so, we did not find evidence of a treatment-control difference. In addition, we found that treatment teachers did not report feeling more satisfied with their jobs than control teachers.

Composition of District Teaching Force

The last major research question concerned the impact of comprehensive teacher induction on the composition of the teaching workforce in the district. As shown below, we found no statistically significant impacts on the composition of the district teaching force in one-year districts after two years.

For comprehensive teacher induction to affect the composition of the district's teaching workforce, it has to produce a difference in the types of teachers who decide to remain in the district. As teachers leave the district, the average qualifications of the teachers who remain in the district begin to change, perhaps differentially between the treatment and control groups. We tested this hypothesis by comparing the characteristics of district stayers between the treatment and control groups along two dimensions: (1) their impact on student achievement; and (2) their professional characteristics such as SAT/ACT scores and advanced degrees. The student achievement outcome is regression-adjusted using the same model used in the main analysis.

We found that the treatment had no statistically significant impacts on the student achievement or professional background characteristics of district stayers. Table ES.3 presents the impacts on student achievement outcomes for district stayers. Table ES.4 shows the background characteristics of teachers by mobility status.

Table ES.3. Impacts on Test Scores, District Stayers Only: One-Year Districts, 2005-2006 School Year

Outcome	Treatment	Control	Difference	Effect Size	P-value
Reading scores (all grades)	0.02	-0.03	0.05	0.05	0.331
Unweighted Sample Size (Students)	975	942	1,917		
Unweighted Sample Size (Teachers)	53	56	109		
Unweighted Sample Size (Schools)	47	41	88		
Math scores (all grades)	0.01	-0.02	0.03	0.03	0.629
Unweighted Sample Size (Students)	826	857	1,683		
Unweighted Sample Size (Teachers)	47	52	99		
Unweighted Sample Size (Schools)	43	38	81		

Source: MPR analysis of data from 2004-2005 and 2005-2006 school years provided by participating school districts; MPR Second Mobility Survey administered in 2007-2008 to all study teachers.

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

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Table ES.4. Characteristics of District Stayers, Movers, and Leavers after Two Years by Treatment Status (Percentages Except Where Noted): One-Year Districts

	Treatment			Control			Difference		
Teacher characteristic	Stayers	Movers	Leavers	Stayers	Movers	Leavers	Stayers	Movers	Leavers
College entrance exam scores (SAT combined score or equivalent)	1,026	1,029	1,082	1,021	984	1,080	4	45	2
Attended highly selective college	30.3	27.3	46.0	27.2	50.5	33.3	3.1	-23.2	12.7
Major or minor in education	79.8	65.5	76.1	81.1	65.9	67.2	-1.3	-0.4	8.9
Student teaching experience (Weeks)	16.5	13.9	14.2	15.1	13.5	12.4	1.5	0.4	1.8
Entered the profession through traditional four-year program	64.4	61.0	45.8	60.3	58.7	30.8	4.1	2.4	15.0
Unweighted Sample Size (Teachers)	191	29	24	187	23	22			
Unweighted Sample Size (Schools)	100	25	18	104	22	21			

Source: MPR calculations using data from the College Board and ACT, Inc.; MPR Teacher Background Survey administered in 2005-2006, MPR Second Mobility Survey administered in 2007-2008; MPR First and Second Induction Activities Surveys administered in fall/winter 2005-2006 and spring 2006 to all study teachers.

Notes: 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 (191/29/24 treatment stayers/movers/leavers and 187/23/22 control stayers/movers/leavers) and schools (100/25/18 treatment and 104/22/21 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, two-tailed test. P-values are suppressed to make the table easier to read.

SUMMARY OF FINDINGS AFTER TWO YEARS: TREATMENT-CONTROL DIFFERENCES IN TWO-YEAR DISTRICTS

Induction Services Received

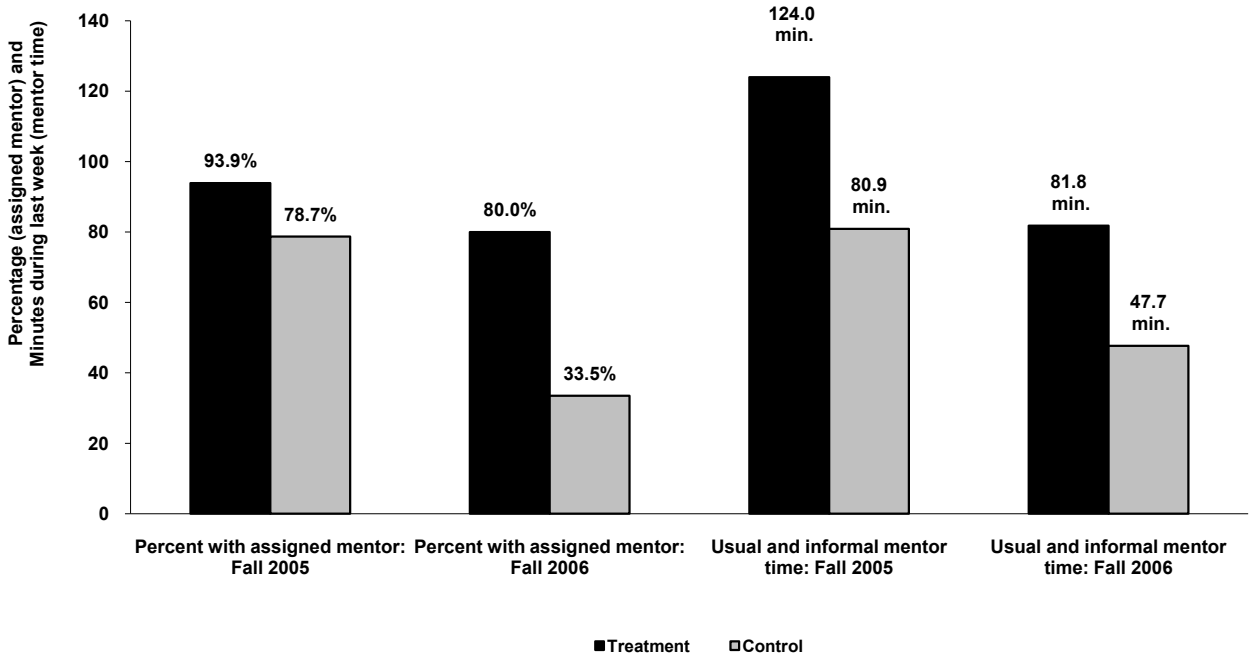
During Year 1 and Year 2, both years in which comprehensive teacher induction services were offered to the treatment group in the two-year districts, treatment and control teachers' reports showed statistically significant differences favoring the treatment group on many measures of the amount, types, or content of supports. For consistency with the way in which results are reported for one-year districts, we report on findings for the fall of each year.[9]

Amount of Mentoring. We found statistically significant differences between the treatment and control teachers with regard to the likelihood of teachers reporting having a mentor assigned to them, having a full-time mentor, and having a mentor who was another teacher. Treatment teachers were more likely than control teachers to report having a mentor assigned to them (94 versus 79 percent in Year 1; 80 versus 34 percent in Year 2), and to report having a full-time mentor (72 versus 16 percent in Year 1; 64 versus 7 percent in Year 2). Treatment teachers were less likely than control teachers to report having a mentor who was another teacher (38 versus 62 percent in Year 1; 12 versus 27 percent in Year 2). We also found statistically significant differences in the amount of time teachers reported spending with their mentors. Treatment teachers reported spending more time working with their mentors than control teachers did during the most recent full week of teaching. Treatment teachers reported spending more time on average in mentor meetings (124 minutes per week versus 81 minutes in Year 1; 82 minutes versus 48 minutes in Year 2). In both years, the differences were attributable primarily to differences in the duration of scheduled meetings. Figure ES.2 shows treatment-control differences for having an assigned mentor and time in mentor meetings in Year 1 and Year 2.

Mentor Activities and Assistance. Treatment and control teachers' reports showed statistically significant differences in the amount of time in various mentor activities and in the kinds of assistance teachers reported receiving from their mentors. Treatment teachers reported spending more time being observed by mentors (38 versus 17 minutes in Year 1; 22 versus 7 minutes in Year 2), meeting one-on-one with mentors (43 versus 23 minutes in Year 1; 25 versus 12 minutes in Year 2), meeting together with mentors and other first-year teachers (38 versus 11 minutes in Year 1; 25 versus 6 minutes in Year 2), and having mentors model lessons (16 versus 10 minutes in Year 1; 12 versus 5 minutes in Year 2). During the most recent full week of teaching, treatment teachers were more likely than control teachers to report receiving mentors' assistance in each of the topic areas covered by the survey: effects ranged from 14 to 28 percentage points in Year 1 and 28 to 44 percent in Year 2.

[9] For two-year districts, findings from spring of Year 1 were consistent with the findings from fall of Year 1. Likewise, findings from spring of Year 2 were consistent with the findings from fall of Year 2.

Figure ES.2. Treatment-Control Differences in Percent Assigned a Mentor and Total Minutes Spent in Mentoring Per Week: Two-Year Districts, Fall 2005 and Fall 2006



Note: All treatment-control differences are significantly different from zero at the 0.05 level, two-tailed test (N=395 teachers in fall 2005 and 360 teachers in fall 2006).

Professional Development. We did not find statistically significant differences between treatment and control teachers' reported attendance in professional development, except that treatment teachers were more likely than control teachers to report having attended sessions focused on classroom management techniques (61 versus 48 percent) in fall 2005 (Year 1).

Student Achievement

We found no evidence of statistically significant impacts on student test scores in two-year districts. The benchmark impacts on math and reading scores in the second year of the study were not significantly different from zero (Table ES.5). The data confirm that the impacts on reading and math in the second year were not statistically significant when we re-estimated the impacts using different samples, different sets of covariates, or different estimation techniques.

Table ES.5. Impacts on Test Scores: Two-Year Districts, 2006-2007 School Year

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Subject	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
Reading	0.00	0.00	0.00	0.00	0.967	1,732	100	7
Math	-0.03	-0.01	-0.02	-0.02	0.746	1,736	99	7

Source: MPR analysis of data from 2005-2006 and 2006-2007 school years provided by participating school districts.

Notes: Data are regression-adjusted to account for pretest, district-by-grade fixed effects, and clustering of students within schools. For Reading, there were 856 students and 52 teachers in the treatment group, and 876 students and 48 teachers in the control group. For Math, there were 780 students and 50 teachers in the treatment group, and 956 students and 49 teachers in the control group.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Teacher Retention

We found that comprehensive teacher induction had no statistically significant impact on teacher retention after two years. Table ES.6 shows the result of the three hypothesis tests specifically focused on retention in the school, in the district, and in the profession as binary outcomes. For each of the outcomes, there was no statistically significant impact. The same result was obtained when we expanded the number of outcomes to differentiate between moving to a school in another public school district and moving to a private, parochial, or other school, and expanded the outcomes for leaving to include leaving to stay at home, leaving to attend school or take a new job, and other reasons for leaving.

We also examined the reasons that teachers who left their districts (movers) or left the teaching profession (leavers) gave for leaving and found no statistically significant impacts of treatment. When we asked leavers whether they expected to return and if so, when they would do so, we did not find evidence of a treatment-control difference. In addition, we found that treatment teachers did not report feeling more satisfied with or prepared for their jobs than control teachers.

Table ES.6. Impacts on Teacher Retention Rates after Two Years (Percentages): Two-Year Districts

Outcome	All Teachers	Treatment	Control	Difference	P-value
Retained in the same school	64.1	62.2	66.2	-4.0	0.386
Retained in the same district	72.3	69.6	75.3	-5.7	0.208
Retained in the teaching profession	88.8	86.9	90.8	-3.9	0.241
Unweighted Sample Size (Teachers)	364	203	161		
Unweighted Sample Size (Schools)	151	81	70		

Source: MPR Second Mobility Survey administered in 2007-2008 and Teacher Background Survey administered in 2005-2006 to all study teachers.

Note: 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, two-tailed test.

Composition of the District Teaching Force

We found that the treatment had no statistically significant impacts on the student achievement outcomes or professional background characteristics of district stayers. Table ES.7 presents the impacts on student achievement outcomes for district stayers. Table ES.8 shows the background characteristics of teachers by mobility status.

Table ES.7. Impacts on Test Scores, District Stayers Only: Two-Year Districts, 2005-2006 School Year

Outcome	Treatment	Control	Difference	Effect Size	P-value
Reading scores (all grades)	0.03	-0.03	0.06	0.06	0.591
Unweighted Sample Size (Students)	745	558	1,303		
Unweighted Sample Size (Teachers)	45	30	75		
Unweighted Sample Size (Schools)	31	24	55		
Math scores (all grades)	-0.04	0.07	-0.11	-0.11	0.162
Unweighted Sample Size (Students)	693	549	1,242		
Unweighted Sample Size (Teachers)	43	30	73		
Unweighted Sample Size (Schools)	29	24	53		

Source: MPR analysis of data from 2004-2005 and 2005-2006 school years provided by participating school districts; MPR Second Mobility Survey administered in 2007-2008 to all study teachers.

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

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Table ES.8. Characteristics of District Stayers, Movers, and Leavers after Two Years by Treatment Status (Percentages Except Where Noted): Two-Year Districts

	Treatment			Control			Difference		
Teacher Characteristic	Stayers	Movers	Leavers	Stayers	Movers	Leavers	Stayers	Movers	Leavers
College entrance exam scores (SAT combined score or equivalent)	916	1,006	1,095	967	1,040	1,081	-51	-34	14
Attended highly selective college	23.4	28.6	59.9	25.1	37.1	52.4	-1.7	-8.5	7.5
Major or minor in education	67.0	70.9	38.9	66.6	70.8	74.7	0.4	0.0	-35.8
Student teaching experience (weeks)	12.2	14.1	6.2	11.9	11.7	9.3	0.3	2.4	-3.1
Entered the profession through traditional four-year program	61.5	76.8	25.2	66.0	61.3	56.1	-4.5	15.5	-30.9
Unweighted Sample Size (Teachers)	143	35	25	121	25	15			
Unweighted Sample Size (Schools)	71	28	20	62	21	13			

Source: MPR calculations using data from the College Board and ACT, Inc.; MPR Teacher Background Survey administered in 2005-2006, MPR Second Mobility Survey administered in 2007-2008; MPR First and Second Induction Activities Surveys administered in fall/winter 2005-2006 and spring 2006 to all study teachers.

Notes: 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 (143/35/25 treatment stayers/movers/leavers and 121/25/15 control stayers/movers/leavers) and schools (71/28/20 treatment and 62/21/13 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 different from zero. P-values are suppressed to make the table easier to read.

CORRELATIONAL ANALYSES

Given the prevalence of supports reported by control teachers, we explored the relationship between induction supports and outcomes independent of group assignment (treatment or control) and district type (one-year or two-year). Using data from the first three Induction Activities surveys, we created a variable that reflects the number of years (0, 1, or 2) the beginning teacher had an assigned mentor and constructed three other new measures[10]:

- The Induction Services Index measuring breadth of services received by the beginning teacher,
- The Instructional Support Index measuring suggestions, guidance, and feedback on teaching, and
- The Induction Intensity Index measuring program duration and intensity.

The analyses use the same methods as the experimental analyses, but instead of assignment to treatment status, which was randomly determined, the key explanatory variables are the number of years the beginning teacher had an assigned mentor and the three indices, included jointly in a regression model. The results should be interpreted with caution because the analyses are correlational and not causal. In particular, a nonexperimental estimate of the relationship 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.

Overall, we found that induction measures were not significantly related to math test scores (p-value of F-test = 0.068) or reading scores (p-value of F-test = 0.651). However, we found that the association between the years the beginning teacher had a mentor and math test scores was statistically significant (regression coefficient = 0.12, p-value = 0.015). For measures of teacher retention, there was a statistically significant relationship between the induction activities variables and retention (p-value of F-test = 0.016 for remaining in the

[10] The variable that reflects the number of years the beginning teacher had an assigned mentor is constructed using three items: the indicator variables at fall 2005, spring 2006, and fall 2006, on whether the beginning teacher had an assigned mentor. This variable has the values 0, 1, and 2 years. The Induction Services Index is the sum of nine indicator variables at fall 2005, spring 2006, and fall 2006, on whether the beginning teacher: (1) met with a literacy or math coach, (2) met with a study group, and (3) observed others teaching. The Induction Services Index has values in the range 0 to 9. The Instructional Support Index is constructed similarly using eight indicator variables on whether the beginning teacher received: (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. The Instructional Support Index has values in the range 0 to 8. The Induction Intensity Index is the sum of the average number of hours per week at fall 2005, spring 2006, and fall 2006 (3 items) 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. The Induction Intensity Index has values in the range 0 to 20.8.

district; p-value of F-test = 0.001 for remaining in teaching). One measure—the Induction Services Index—was positively related and no measures were negatively related to teacher mobility for both remaining in the district and remaining in teaching. The estimate of the regression coefficient on the Induction Services Index for remaining in the district was 0.02; for remaining in teaching, it was 0.01. This implies that, for example, if the retention rate in a district were 80 percent, then an additional induction service, such as meeting with a study group in one semester, would be associated with a district retention rate of 82 percent, all else equal. All results were robust to alternate methods of constructing the indices and alternate model specifications.

SUMMARY OF FINDINGS

The report presents findings from an experimental test of the impact of comprehensive teacher induction on student achievement in beginning teachers' classrooms and on the teachers' retention rates in urban elementary schools. In ten of the study districts, a comprehensive induction program was implemented during beginning teachers' first year in the classroom. In the remaining seven study districts, comprehensive induction was implemented during beginning teachers' first two years in the classroom. This design does not allow for and should not be used to make direct comparisons between the districts that received one year of treatment and districts that received two years of treatment, but instead allows us to investigate the effectiveness of one-year programs separately from that of two-year programs. The main findings are summarized below.

- During their first year in the classroom, in both one- and two-year districts, treatment and control teachers' reports showed statistically significant differences in the amount and types of support received. Treatment teachers were more likely than control teachers to report having an assigned mentor (90 versus 70 percent of teachers reported having an assigned mentor in one-year districts; 94 versus 79 percent in two-year districts) and reported spending more time per week with a mentor (87 versus 67 minutes in one-year districts; 124 versus 81 minutes in two-year districts). Treatment teachers reported spending more time being observed by mentors (34 versus 10 minutes during the most recent full week of teaching in one-year districts; 38 versus 17 minutes in two-year districts) and meeting with mentors together with other first-year teachers (29 versus 9 minutes in one-year districts; 38 versus 11 minutes in two-year districts).

- During their second year in the classroom, treatment teachers in one-year districts received less support than did control teachers. During Year 2, we found a statistically significant difference favoring the control group in teachers' likelihood of having an assigned mentor and in the amount of time teachers spent per week with a mentor. Treatment teachers were less likely than control teachers to report having an assigned mentor (20 versus 29 percent) and reported spending less time per week with a mentor (19 versus 39 minutes).

- During their second year in the classroom, treatment teachers in two-year districts received more support than did control teachers. During Year 2, we found a statistically significant difference favoring the treatment group in teachers' likelihood of having an assigned mentor and in the amount of time teachers spent per week with a mentor. Treatment teachers were more likely than control teachers to report having an assigned mentor (80 versus 34 percent) and reported spending more time per week with a mentor (82 versus 48 minutes).

- No impacts of comprehensive teacher induction were found on student achievement during teachers' second year in the classroom. In both one- and two-year districts, we did not find statistically significant impacts on student achievement across all elementary grade levels in reading or math during the teachers' second year.

- No impacts of comprehensive teacher induction were found on teacher retention rates after two years. There was also no evidence that comprehensive teacher induction induced a change in the kind of teachers retained within the district. In both one- and two-year districts, we did not find statistically significant impacts of comprehensive teacher induction on teacher retention rates in the school, district or profession after two years. In both one- and two-year districts, we did not find statistically significant impacts on the composition of the district teaching workforce after two years, whether measured by district stayers' impacts on student achievement or by their professional background characteristics (for example, SAT/ACT scores or whether the teacher attended a highly selective college).

- In a correlational (nonexperimental) analysis of induction and student test scores, the relationship between four composite induction measures (considered jointly) and test scores was statistically insignificant for both math and reading. When we tested the variables individually, one of the four measures of beginning teacher support (years had a mentor) was positively related to math scores (coefficient = 0.12, p-value = 0.015) and none were related to student achievement in reading. The significant result can be interpreted as a student scoring 12 percent of a standard deviation higher on the math test for each year the beginning teacher had a mentor. The nonexperimental results should be interpreted with caution because the analyses are correlational and not causal.

- In the correlational analysis of induction and teacher mobility, there was a positive relationship between the four composite induction measures and retention that was statistically significant for both retention in the district (p-value=0.016) and retention in the profession (p-value=0.001). When we tested the induction indices one at a time, one of the four explanatory variables was positively related to retention in the district, none were positively related to retention in the profession, and none were negatively related to either type of teacher retention. The estimate of the regression coefficient on the Induction

Services Index for remaining in the district was 0.02. This implies that, for example, if the retention rate in a district were 80 percent, then an additional induction service, such as meeting with a study group in one semester, would be associated with a district retention rate of 82 percent, all else equal. As mentioned above, the nonexperimental results should always be interpreted with caution because the analyses are correlational and not causal.

FUTURE RESEARCH

This report focused on the second year of findings, updating an earlier report (Glazerman et al. 2008) that presented results after one year of implementation for one-year and two-year districts combined. The research team is conducting a follow-up analysis that will include a third and final year of test score and teacher mobility data in one-year and two-year districts.

CHAPTER 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).

Even those 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 of the main policy responses to the problems of high turnover and inadequate preparation among beginning teachers is to support them with a formal, comprehensive induction program. 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 formative assessment (constructive feedback). While most districts use some form of teacher induction or mentoring, they often do so 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, while teacher induction is common, induction that is intensive, comprehensive, 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 is that comprehensive teacher induction is expensive (Villar and Strong 2007; Alliance for Excellent Education 2004). Costs of 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).[11] Moreover, there is little empirical evidence on whether investing more 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), studies of teacher induction to date have been neither conclusive nor rigorous. Research based on federal statistics (e.g., 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) 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. 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

[11] These reports note costs for five programs, four of which are two-year programs and one of which is a 1-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 information on the number of vacancies, this measure does not provide sufficient information to be helpful to districts considering whether to adopt an induction program. See National Commission on Teaching and America's Future (2007), Barnes et al. (2007), Milanowski and Odden (2007), and Fuller (2000).

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 the 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

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 Institute of Education Sciences (IES) contracted with Mathematica Policy Research (MPR), to conduct the Evaluation of the Impact of Teacher Induction Programs. The study examines whether comprehensive teacher induction programs lead to higher teacher retention rates and other positive teacher and student outcomes as compared to prevailing approaches to supporting new teachers that are generally less intensive, formal, or comprehensive. More specifically, the analysis is designed to address five research questions on the impacts of teacher induction services:

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 are the impacts on teachers' classroom practices?

3. What are the impacts on student achievement?

4. What are the impacts on teacher retention?

5. What is the impact on the composition of the district's teaching workforce?

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 ratings from an expert review committee. The providers are Educational Testing Service of Princeton, New Jersey (ETS) and the New Teacher Center at the University of California-Santa Cruz (NTC). MPR 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; NTC operated in 8 districts.

IES later expanded 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 in the one-year districts started in fall 2005 and received induction services in the 2005-06 school year; the teachers in two-year districts also started in fall 2005 but received services in the 2005-06 and 2006-07

school years. We used convenience sampling to select the districts to receive a second year of the treatment; we selected the districts based upon 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 districts that received two years of treatment, but instead 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. In this report, we emphasize the findings from the second year of the study. 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. When appropriate, however, we compare outcomes from the first year of the study to outcomes from the second year of the study within the one-year districts and within the two-year districts.

Researchers from WestEd, a subcontractor to MPR, 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.

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 either from ETS or NTC (depending on the district), or to a control group, which took part in the district's usual teacher induction program. We assigned 418 elementary schools with 1,009 eligible beginning teachers across the 17 districts. While 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 focuses on elementary schools only.

B. FINDINGS AFTER ONE YEAR

The Year 1 report (Glazerman et al. 2008) found that teachers assigned to the treatment group reported more induction support, but also found that the additional support did not translate into positive impacts on key outcomes after one year.[12] The additional induction support amounted to a greater likelihood of having a mentor formally assigned to beginning teachers (93 versus 75 percent), 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 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

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

student work, and communicating with parents). There were no positive impacts on classroom practices, student achievement, teacher retention, or the composition of the district's teaching workforce after one year. Nor did we find any evidence of positive impacts on teachers' satisfaction or feelings of preparedness. The current report re-visits four of the five research questions listed above using an additional year of data and reports on one-year districts and two-year districts separately. Because we did not return to observe classrooms, we did not re-visit the question about classroom practices.

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

Context. The structure and functioning of an induction program is likely to be influenced by the characteristics of the local area, the school, the beginning teacher's classroom, and the teacher (Box A, Figure I.1). 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.

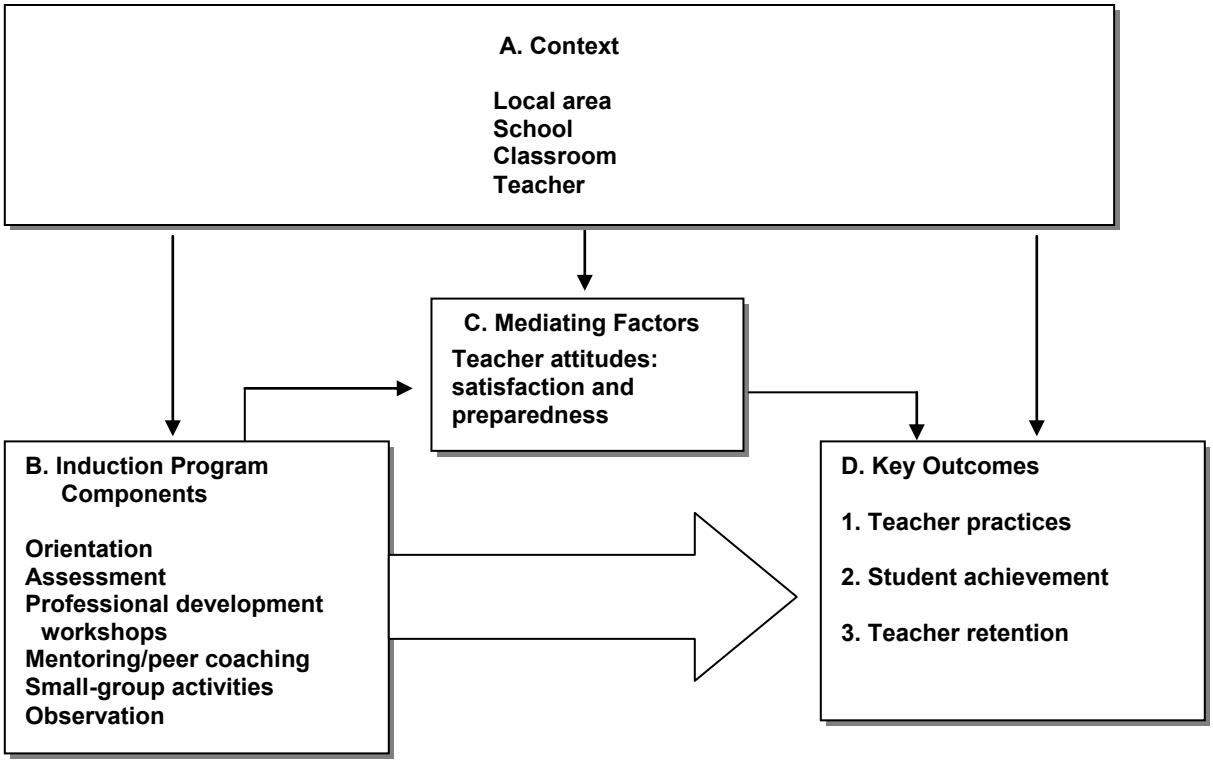
Induction Program Components. Induction programs may include a variety of possible components (Figure I.1, Box B). There is no one-size-fits-all model of teacher induction in either theory or practice: different programs emphasize different approaches. For instance, programs may stress to a greater or 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, we experimentally varied the nature of induction support by packaging induction services into specially selected comprehensive programs (treatment group), comparing the outcomes for the teachers in this group with the outcomes for teachers in the prevailing, less structured induction programs in their districts (control group).

Outcomes for Beginning Teachers. Induction may improve teaching in two ways: by strengthening beginning teachers' attachment to the profession (reflected in mobility patterns) and by improving teaching practices (Figure I.1, Box D). Improving teacher practices is not only a key outcome for teachers but also would help explain possible impacts on retention and student achievement.

Induction may affect several intermediate factors (Figure I.1, Box C) that may help explain changes in 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.

Student Outcomes. The ultimate goal of induction programs is to improve students' academic outcomes (Figure I.1, Box D). Improvements in the teaching workforce achieved through induction may also lead to other positive effects on students, such as a reduction in behavioral problems, improved attendance, and reduced tardiness and disciplinary incidents.

Figure I.1. Conceptual Framework for the Effects of Teacher Induction Programs on Teacher and Student Outcomes



D. Organization and Content of This Report

The rest of this report presents the findings and the methods and data used to generate the findings. Chapter II presents the study design, sample characteristics, and estimation approach. Chapter III discusses the data collection process, including response rates. The report then outlines the interventions under study, both the ETS and NTC models of teacher induction support, as well as the counterfactual condition of prevailing teacher induction programs (Chapter IV). Next, we present findings from the impact analysis for the districts whose treatment groups received one year of intervention (Chapter V) and those whose treatment groups received two years of intervention (Chapter VI), followed by

correlational analyses conducted to add context to the main experimental findings (Chapter VII).

This report presents findings on induction services reported by teachers, student achievement growth, and teacher retention through the first two years of the study, based on data collected in multiple years. A future report will update this one with longer-term follow-up covering the study teachers' third year.

CHAPTER 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 that was exposed to the induction services normally offered by the districts (control). This chapter documents the study design and discusses the methods for selecting districts, schools, and teachers for inclusion in the study, and describes the data analysis methods. Figure II.1 provides an overview of the sample selection process. Although we undertook a purposive selection of districts and schools, the schools within each district were randomly assigned to a treatment or control group.

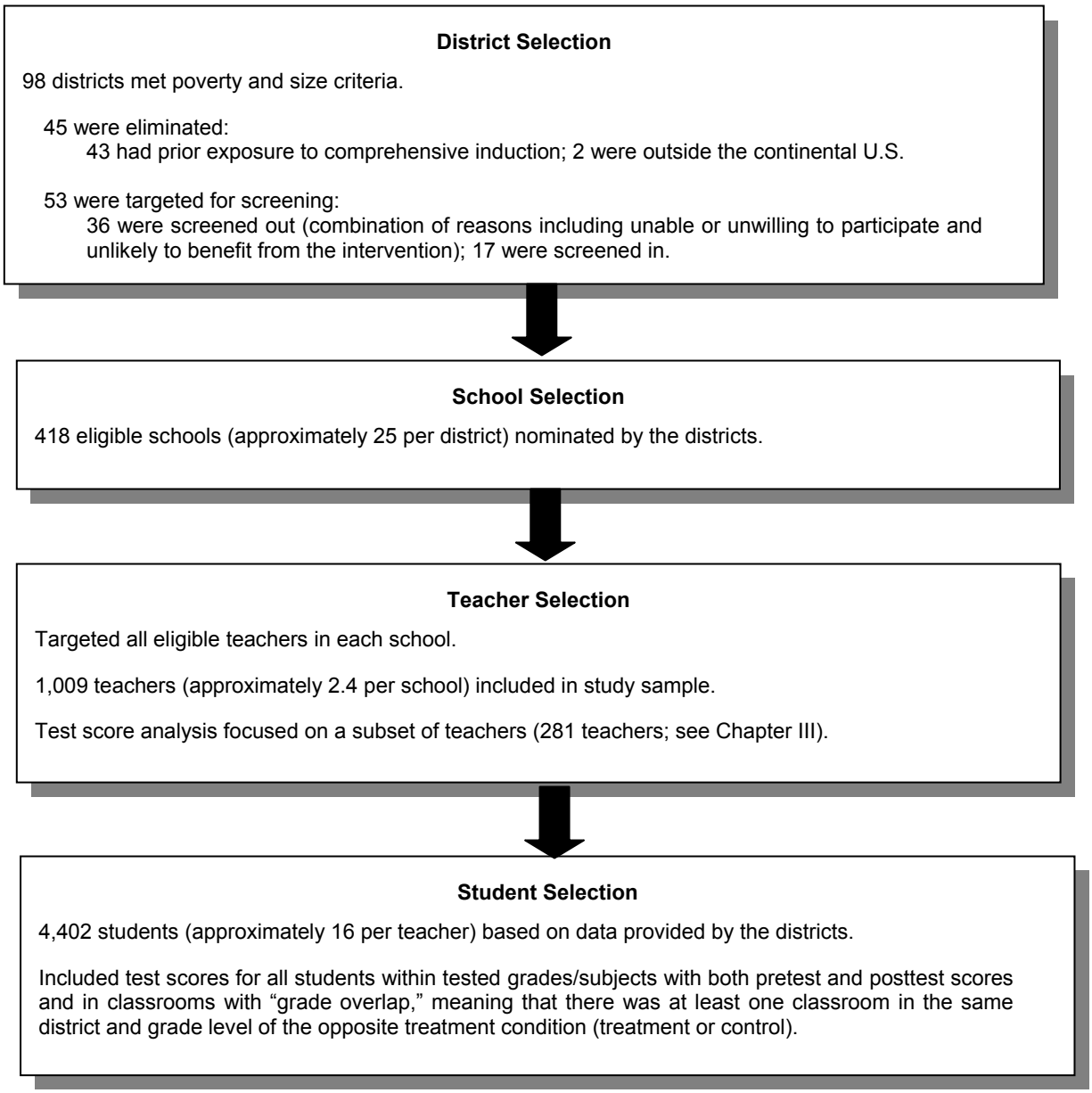
A. SELECTION OF DISTRICTS

The initial list of targeted districts was selected according to size and poverty 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 50 percent of students eligible for free or reduced-price meals under the federal government's National School Lunch Program. We developed these size and poverty targets in consultation with IES, based on 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. MPR 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, MPR and CEL eliminated 2 districts that were outside the continental U.S. 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, Texas, Ohio, or Louisiana, 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.

Figure II.1. Sample Selection Flow Chart



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

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 whose leaders 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 best 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 outcomes may be due to the programs or the set of districts each provider works with; those effects cannot be separated.

Similarly, the decision of which districts would receive a second year of intervention was preference-based. We used convenience sampling to select the districts to receive a second year of the treatment. We ensured a balance of ETS and NTC districts in the two-year group. The self-selection of districts means that they differ in unobserved ways beyond just their having had 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.

Table II.1 shows the characteristics of districts included in the study. The districts served low-income students, with more than 50 percent of students in each district qualifying for the National School Lunch Program. The study included districts serving mostly African American students (7 of the 17 districts), Hispanics (2 of 17), and white students (3 of 17), and 5 diverse districts without a racial/ethnic majority. The districts were located throughout the South (which extends from Delaware to Texas), Northeast, and Midwest and were all urban; 9 of 17 districts enrolled more than 50,000 students, and 11 of 17 included more than 50 elementary schools.

Table II.1 also shows the characteristics for 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 out of seven of the two-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.

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

District Characteristics	Number of Districts: One-Year	Number of Districts: Two-Year	Number of Districts: All	Percent
Demographics				
Low Income (Percent Eligible for NSLP)				
<65	4	2	6	35.3
65–70	2	0	2	11.8
70–75	2	1	3	17.6
75–80	2	3	5	29.4
80–85	0	0	0	0.0
>85	0	0	0	0.0
Unknown (data not available)	0	1	1	5.9
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
Census 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
Unweighted Sample Size (Districts)	10	7	17	100.0

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

Note: NSLP = National School Lunch Program.

B. SELECTION OF SCHOOLS AND TEACHERS

Within each district, a fixed set of elementary schools was selected for 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 that met the following eligibility criteria:

- ***Elementary Grade.*** Teachers in K-6 were considered elementary. We excluded teachers of part-day pre-kindergarten classes or those in middle schools with departmentalized teaching. 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.

- ***Not Already Receiving Support.*** Some alternative teacher preparation or certification programs continue to support teachers during their first year of teaching. While 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. Table II.2 and Table 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 (85 percent and 72 percent) in both types of districts employed one, two, or three eligible beginning teachers.

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 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 availability of comprehensive induction services, ruling out all other confounding factors.

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 or K–5	64.4	65.5	63.4	2.1	
Pre-K or K–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	
Unweighted Sample Size (Schools)	252	124	128		

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

Notes: NSLP = National School Lunch Program; 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, two-tailed test.

1. Method of Random Assignment

Random assignment at the school level was the most feasible approach. Eligible teachers in a school were either all exposed or all not exposed to treatment, a method known as cluster random assignment. Given that varying the types of induction services available in the same school building could result in contamination between services, the cluster random assignment was necessary. Therefore, we assigned all eligible teachers to treatment or control status based on the school where they were expected to teach at the point of random assignment (baseline).

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 or K–5	81.3	84.0	78.6	5.4	
Pre-K or K–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	
Unweighted Sample Size (Schools)	166	86	80		

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

Notes: NSLP = National School Lunch Program; 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, two-tailed test.

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 between 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).[13] 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–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. Table II.4 presents demographic characteristics of the teachers in the study from one-year districts. Of 532 teachers 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). Table II.5 presents demographic characteristics of the teachers in the study from two-year districts. Of 421 teachers 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).

Table II.6 describes the professional backgrounds of teachers for the one-year districts. Similar percentages of treatment and control teachers had advanced degrees (24 and 29 percent), earned bachelor's degrees from highly selective colleges[14] (31 and 30 percent), had an education major or minor (77 and 79 percent), and entered the profession with no student teaching (12 and 16 percent). There was a statistically significant difference in how the teachers entered the profession with a higher percentage of treatment teachers coming from a traditional four-year program (62 percent versus 56 percent) and a lower percentage of treatment teachers entering through an alternative preparation program (13 percent versus 22 percent). There was also a statistically significant 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).

Table II.8 describes the teachers' professional backgrounds for the two-year districts. Similar percentages of treatment and control teachers had advanced degrees (16 percent), earned bachelor's degrees from highly selective colleges (30 and 28 percent), had an education major or minor (64 and 66 percent), entered teaching through a traditional four-year college route (59 and 64 percent), held a regular teaching certificate (50 and 51 percent),

[13] If the admissible allocations are defined independently of treatment status, as they were in this study, then every school and every teacher had a 50 percent probability of assignment to the treatment group.

[14] 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*.

and entered the profession with no student teaching (31 and 26 percent). For those teachers who gave us permission to obtain their SAT or ACT scores and for whom scores were available, we found no statistically significant differences in scores between the treatment and control teachers (Table 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 more	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 under 5 years old	10.4	11.5	9.3	2.2	
One or more children, none under 5 years old	15.1	14.6	15.6	-1.0	
Unweighted Sample Size (Teachers)	532	267	265		

Source: MPR Background Survey administered in 2005-2006 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, two-tailed test.

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 more	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 under 5 years old	19.7	19.8	19.7	0.1	
One or more children, none under 5 years old	15.7	14.6	16.9	-2.3	
Unweighted Sample Size (Teachers)	421	222	199		

Source: MPR Background Survey administered in 2005-2006 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, two-tailed test.

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

Teacher Characteristics	All Teachers	Treatment	Control	Difference	P-value
Has Masters 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	
Unweighted Sample Size (Teachers)	532	267	265		

Source: MPR Background Survey administered in 2005-2006 to all study teachers.

Notes: 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, two-tailed test.

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
Unweighted Sample Size (All Teachers)	561	275	286		
Unweighted Sample Size (Teachers with Usable ACT or SAT Scores)	327	164	163		

Source: MPR calculations 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, two-tailed test.

There were statistically significant differences 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 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 include those who taught a single core subject like math or science as well as those who taught subjects like 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 are 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 because we were interested in these outcomes for all teachers whom districts might have targeted in a real-world implementation and who could have been affected by treatment. The findings were robust to the inclusion or exclusion of special subject teachers.

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	
Unweighted Sample Size (Teachers)	421	222	199		

Source: MPR Background Survey administered in 2005-2006 to all study teachers.

Notes: 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, two-tailed test.

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
Unweighted Sample Size (All Teachers)	448	231	217		
Unweighted Sample Size (Teachers with usable ACT or SAT Scores)	221	117	104		

Source: MPR calculations 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, two-tailed test.

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 one	15.2	14.2	16.2	-1.9	
Grade two	14.4	16.9	11.8	5.0	
Grade three	13.2	15.3	10.9	4.4	
Grade four	12.9	14.5	11.1	3.4	
Grade five	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
Unweighted Sample Size (Teachers)	532	267	265		

Source: MPR Teacher Background Survey administered in 2005-2006 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 .05 level, two-tailed test.

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 one	14.4	14.4	14.4	0.0	
Grade two	16.3	17.4	15.1	2.2	
Grade three	13.6	13.7	13.5	0.2	
Grade four	9.9	9.8	10.1	-0.3	
Grade five	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
Unweighted Sample Size (Teachers)	421	222	199		

Source: MPR Teacher Background Survey administered in 2005-2006 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, two-tailed test.

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.

a. Noncompliance

Noncompliance with treatment assignment—a concern in randomized experiments where 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 MPR, 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 two teachers out of more than 1,000 who transferred from a control to a treatment school and received services. Of those, one could not be included in the analysis due to her failure to complete the surveys.

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 where the induction services would not be available (for example, if they left the district). Nine schools representing 12 teachers in one district and 3 teachers in another district refused to implement the treatment. The 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.

b. 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 the induction study, response rates exceeded 87 percent for the full sample on all major surveys in Year 1 of the study and exceeded 83 percent in Year 2 (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 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 compare 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 nonindependence of observations explicitly, it also allows us to exploit the data on 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 teacher and school variables, 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. While 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 were 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 AND THE MULTIPLE COMPARISON PROBLEM

To interpret the impact estimates, this report relies on conventional notions of statistical significance. That is, the treatment is hypothesized to have no impact (the "null hypothesis") unless we find sufficient evidence to the contrary. In order 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. If the probability of observing a difference (the "p-value") in the absence of a true impact is less than five percent, then we say that there is sufficient evidence to reject the null hypothesis and the effect is deemed statistically significant. If the probability of having observed the difference is five percent or greater, then we assume there is *not* enough evidence to reject the null hypothesis and conclude that the treatment did not cause the observed difference. Maintaining the five percent significance level, there is still a five percent chance that we will reject the null hypothesis and declare a finding to be statistically significant when the treatment was not responsible for the effect. This is called a Type I error. For all of the observed differences with an associated p-value of five percent or larger, we run the risk of failing to attribute that difference to the treatment. This is called a Type II error.

Using these rules, the probability of committing a Type I error is always five 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 multiple comparison problem is the risk that readers will consider one or two statistically significant results as true impacts and ignore the non-significant results. The danger of taking significant findings out of context like this is that it creates a false sense of confidence in the conclusion.

There are many solutions to this problem, but we discuss two here. One solution 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. This approach of contextualizing the significant findings has been followed throughout this report.

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. The second solution 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.[15]

In the report we did not present any adjustments based on the Benjamini-Hochberg (BH) method of addressing multiple comparison inferences because they were unnecessary or inappropriate. For the 62 hypothesis tests that formed the main set of impact analyses (discussed below), the method was unnecessary because there were no significant findings and hence no possibility of Type I error. For the 238 hypothesis tests conducted as part of the sensitivity analysis, 6 tests (3 percent) were rejected and none of those was an appropriate situation for a multiple comparison adjustment.

As mentioned above, the multiple comparison adjustment is unnecessary in cases where there are no significant impacts and hence no risk of Type I error (or of false discoveries). This is the case with all of the impact estimates related to outcomes (teacher attitudes, student achievement, and teacher mobility) presented in Chapters V and VI. The one test that was rejected was an ancillary result, presented in Table V.8, which examined the change in impacts on test scores from one year to the next using a common sample of teachers. The Year 1 and Year 2 impact estimates are presented for reference only, as the focus is on the difference between the two. The Year 1 impact estimate was negative and significant for this sample, a result that is not used to form any conclusions since the more comprehensive analysis of Year 1 impacts for the full sample was presented in an earlier report (Glazerman et al. 2008).

In other cases, the method is inappropriate because the assumption made by Benjamini and Hochberg that the tests being grouped together are independent is violated. One example of such a violation is a sensitivity analysis, where one hypothesis test is typically repeated several times with same data, same outcomes, and same explanatory variables, with small changes in the underlying assumption or sample restrictions in each run. In such cases,

[15] 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.

the statistical significance of the result is not used to draw a conclusion about the particular relationship. Rather, the entire set of results is used to draw a conclusion about the robustness of the main result. The analysis is designed to provide context for, not overturn, the main result. Hence, the analysis does not carry the same elevated risk of Type I error as a traditional analysis. This point applies to the appendices to this report.

Another case in which the BH method is unnecessary is when it is possible to conduct a joint significance test of all of the hypotheses in a group or to reduce the number of tests by aggregating data or measures. By conducting a joint test, one can render an overall judgment about the significance of the collection of treatment-control contrasts. This is the case in Chapter VII, where we test the significance of the relationships (expressed as regression coefficients) between different induction support variables and the study's main outcomes.

CHAPTER III

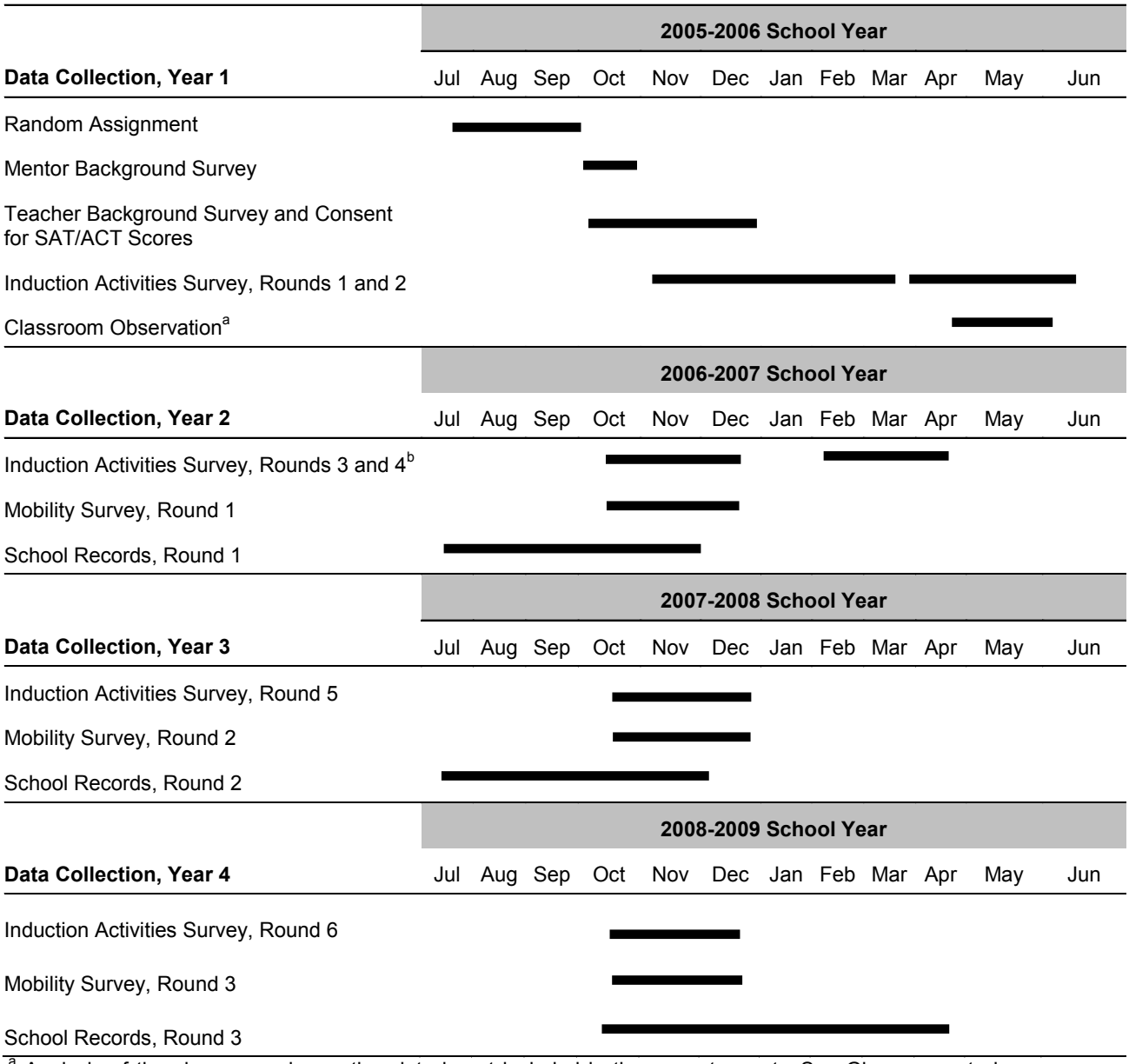
DATA

In accordance with the conceptual framework presented in Chapter I, we collected detailed data on teacher induction services, outcomes, and contextual factors that may have influenced the induction outcomes. 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). We surveyed mentors on their background characteristics and reviewed program documents from ETS and NTC in fall 2005. Surveys of teacher induction activities were administered to both treatment and control teachers during the 2005-2006, 2006-2007, and 2007-2008 school years. Teachers in the seven districts that received two years of comprehensive teacher induction (two-year districts) were surveyed an additional time during spring 2007 to gather more in-depth information about their induction activities during that second year.

For the study's core outcomes, we observed classrooms in spring 2006, collected the districts' student records data following the 2005-2006 and 2006-2007 school years, and conducted teacher mobility surveys in fall 2006 and fall 2007 to learn about teacher retention. Future plans include collection of another year of student records data and, to help track mobility patterns, we are following study teachers with a mobility survey administered in fall and winter 2008. In addition, a final round of the teacher induction activities survey was administered to study teachers beginning in fall 2008.

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. Figure III.1 shows a timeline for the data collection activities. The current report presents the findings pertaining to the first and second years of the study (2005-2006 and 2006-2007), both for the set of districts that received one year of treatment and for those that received two years of treatment. A brief description of each data collection activity is provided below. Copies of the survey instruments may be found in Glazerman et al. (2005).

Figure III.1. Data Collection Schedule



[a] Analysis of the classroom observation data is not included in the current report. See Glazerman et al. (2008) for the classroom practices findings.

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

Figures III.2 and III.3 present flow diagrams of sample members that explain how we derived our analysis samples from the pool of originally identified teachers in one- and two-year districts, respectively.

The test score analysis pertains only to the subset of teachers in tested grades and subjects. Specifically, the eligible sample included teachers who had been assigned to grades and subjects for which their students took a test that year (posttest) and in the prior grade (pretest). State assessment systems under No Child Left Behind focus on grades 3 through 8, which means that only teachers in grades 4 and 5 in K-5 elementary schools routinely have students with both a post-test and a pre-test score. Across one-year and two-year districts for treatment and control groups, the teachers in non-tested grades or subjects represent about 620 teachers or 61 percent of all teachers in the study for the reading analysis (63 percent for the math analysis).

Once the eligible sample for test score analysis was identified, we excluded teachers from the test score analysis if they did not meet certain data conditions, as follows:

(a) 53 teachers were linked to an implausibly high or low number of students to be a regular classroom teacher (see Appendix A for details),

(b) 61 teachers could not be linked by the district to any students

(c) 40 teachers were teaching in grade levels for which a treatment-control comparison could not be made within their district.

These exclusions from the reading score analysis amount to 15 percent of all teachers. For the math score analysis, the same categories of exclusions represent 16 percent of all teachers. As a result, the teachers in the test score analysis sample represent 23 and 22 percent of all teachers in the study for reading and math, respectively. The resulting standard errors of test score impact estimates were in the range of 0.05 to 0.08, meaning that an impact in effect size units of 0.10 to 0.16 would be statistically significant. The study was originally designed to detect test score impacts of 0.10 to 0.22 (Glazerman et al. 2005).

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

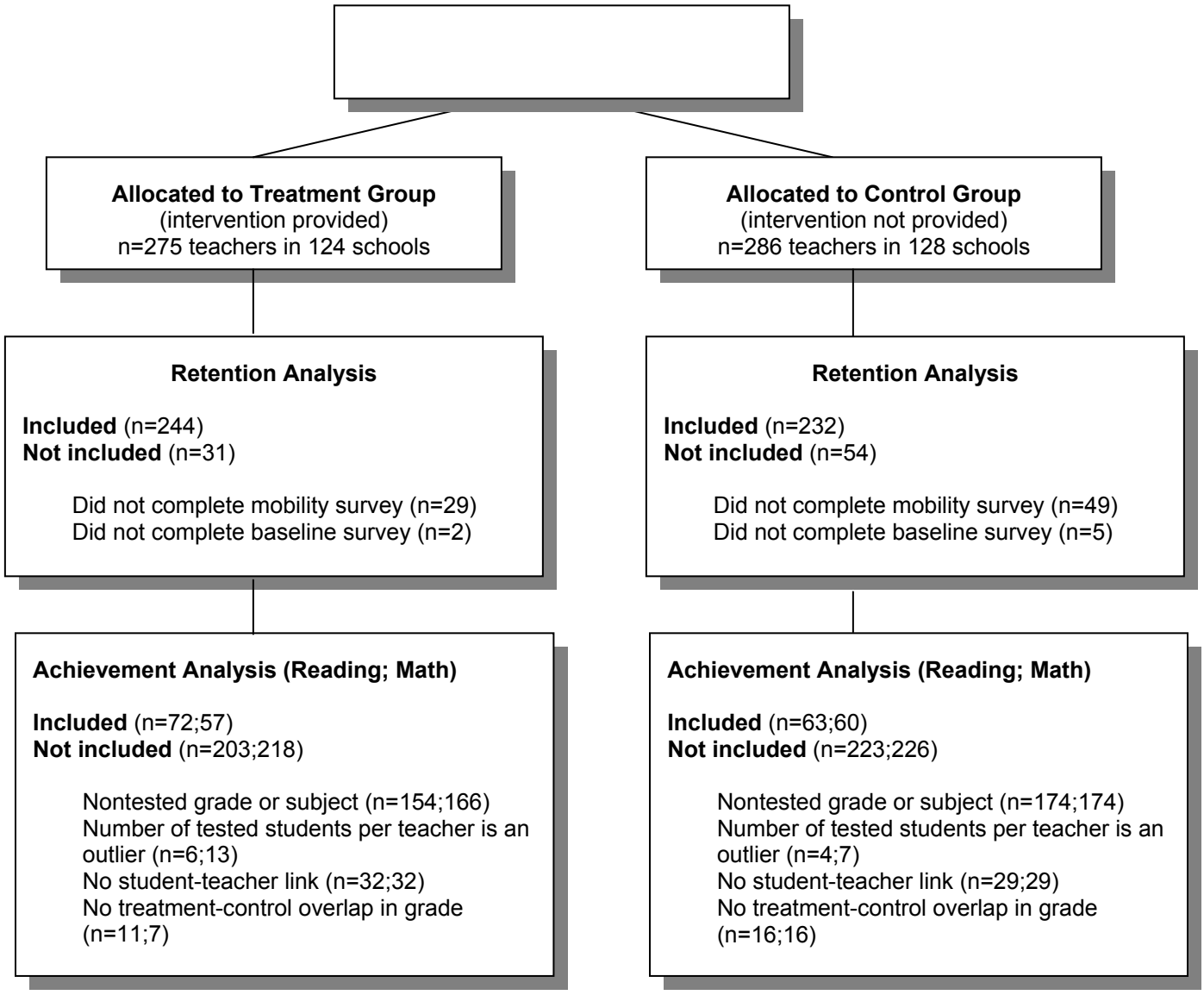
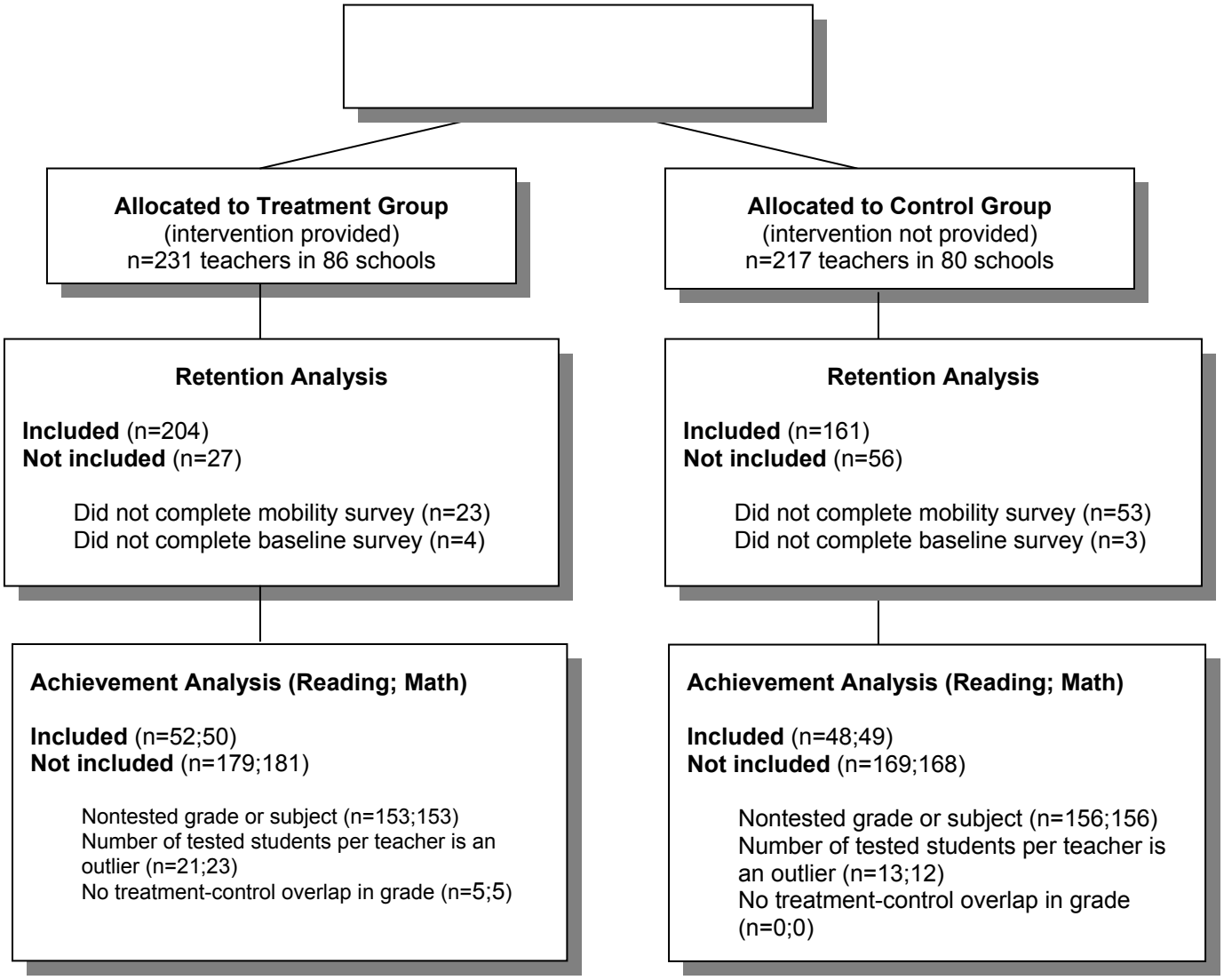


Figure III.3. Flow of Teachers Through the Study in Two-Year Districts



Response rates on teacher surveys ranged from 88 percent to 97 percent for the treatment group and 78 percent to 92 percent for the control group (Table III.1). Table III.2 shows the 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).

Table III.1. Response Rates 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
Teacher Mobility Survey				
Fall 2006*	1,009	88.7	91.5	85.9
Fall 2007*	1,009	85.3	90.2	80.2

Source: MPR teacher induction survey management system.

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

[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, two-tailed test.

n.a. = not applicable.

To reduce any possible bias that nonresponse may cause, we conducted a nonresponse analysis and created nonresponse adjustment weights (see Appendix A). This allowed us to place greater weight on respondents who are most similar to nonrespondents so that the former may stand in for their missing counterparts. For dichotomous outcomes, such as teacher retention, we conducted sensitivity analyses that allowed us to place upper and lower bounds on the effect of nonresponse (including differential nonresponse) on the findings (see Chapters V and VI).

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. We surveyed mentors in order to learn about their professional backgrounds, information that can be used to understand program implementation.

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

	Response Rate (Percentages)							
	Teacher Background Survey, Fall 2005		Induction Activities/Mobility Survey, Fall 2006		Induction Activities, Spring 2007		Induction Activities/Mobility Survey, Fall 2007	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
District Type (Years of Implementation)								
One Year	97.1	92.7	92.7	87.2	n.a.	n.a.	89.5	82.9
Two Year	96.1	91.7	88.8	82.0	87.9	77.9	90.0	75.6
Grade Level								
K or Pre-K	96.3	97.2	94.7	91.3	93.2	86.2	91.3	80.6
1	98.6	97.2	95.4	89.7	83.9	81.5	95.9	88.7
2	97.6	91.0	91.0	89.0	92.1	82.9	89.3	76.9
3	97.5	94.7	89.7	84.3	91.2	77.8	86.4	84.2
4	96.7	91.7	91.1	84.5	85.0	78.3	85.0	73.3
5	100.0	96.2	93.0	91.1	83.3	82.4	91.3	84.6
Other/multiple	91.5	84.1	83.5	72.9	82.5	67.8	89.0	74.3
School Type (Percent in Free Lunch Program)								
Unknown	100.0	100.0	83.3	66.7	100.0	66.7	100.0	66.7
0–24.9%	100.0	92.3	96.7	87.0	100.0	66.7	90.3	92.3
25–49.9%	95.9	91.4	90.2	85.6	90.9	77.6	89.8	80.5
50–74.9%	97.1	92.1	91.8	85.2	86.0	78.2	89.7	78.9
75–100%	90.0	96.6	79.3	79.3	89.3	79.3	86.7	75.9

Source: MPR teacher induction survey management system; MPR 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.

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: MPR calculations 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, two-tailed test.

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 practices 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. While 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

more than 90 percent (89 percent of control group teachers and 96 percent of treatment group teachers).

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 provide an objective measure of a teacher's cognitive ability before he or she 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.

2. Induction Activities Survey

It was important to understand the differences in the services delivered by the comprehensive and prevailing district programs, and to investigate teachers' participation in induction activities after treatment has ended. 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 and fall 2007.[16] 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 may 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 in fall 2005, spring 2006, fall 2006, and spring 2007. Findings in the main report pertain to the fall surveys. Results from the spring surveys are presented in the appendices. We focus the discussion on the fall results for two reasons: the spring results for 2007 exclude the one-year districts, and the choice of fall versus spring results did not change the discussion because the findings are consistent.

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.

[16] 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.

3. Teacher Mobility Survey

We sent mobility surveys to all teachers in fall 2006 and fall 2007 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 status and plans for returning (if applicable). For example, we asked about job responsibilities and salary for those who had changed jobs. We intend to repeat the mobility survey in fall 2008 to identify teachers who moved or left teaching after three years on the job. 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.

C. STUDENT RECORDS

To gauge whether comprehensive teacher induction has any impact on student achievement, we collected student records data from all 17 districts for students in both treatment and control classrooms.[17] The data included scores from standardized tests administered by the districts during spring 2006 (pretest) and spring 2007 (posttest), as well as student background data such as race/ethnicity, date of birth (to determine if a student was over age for grade), eligibility for free or reduced-price meals under the federal School Lunch Program, and disability status.[18]

As shown in Figures III.2 and III.3, we excluded some teachers from the sample based on an examination of the student records data. This exclusion pertains to any teachers who were not linked to individual student test scores in reading or math. We also excluded teachers who were linked with so many or so few students that it was implausible that the teacher was primarily responsible for student achievement in one or both of these subjects. See Appendix A for the details of how we used data on the number of tested students per teacher to determine which students were unlikely to be full-time reading and math students of a particular teacher. We further excluded teachers who lacked a counterpart because there were only treatment teachers or only control teachers in a particular grade within a district. One additional data edit was to replace student test score values that were more than three standard deviations above average with a top-coded score of three and to replace student test score values that were three or more standard deviations below average with a bottom-coded score of negative three. These implausible scores are believed to be outliers and the result of data errors. To test whether this edit made a difference we re-estimated the impacts with the scores included as they originally appeared in the data. The results, shown in Appendices C and D, suggest that the main study findings are robust to this data edit.

[17] The student records data provided by one of the districts could not be used in the impact analysis. This district provided student records data that could not be linked to teachers participating in the evaluation study.

[18] For three districts that tested at least some students in the fall, we used a fall 2006 test as a pretest and/or a fall 2007 test as a posttest.

Aggregating test score data across multiple districts and grades posed a serious challenge, but we made treatment-control comparisons within grades and within districts. Therefore, it was only necessary for the data to come from tests that had been standardized and administered under common testing conditions within each grade within district. Scores were scaled scores, normal curve equivalents, or percentile rankings. We rescaled all tests to have a common mean (0) and variance (1) within each district-grade combination. Further details on aggregation are provided in the impact findings presented in Appendix A.

D. OTHER SUPPORTING DATA

To interpret the impact findings, we needed to understand how the comprehensive teacher induction program was delivered and how it compared to the existing array of services. The induction activities surveys described above represent 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 teacher induction activities surveys. The materials, which provide the basis for the detailed description of program support (see Chapter IV), include documents such as training agenda and materials, curriculum guides, and assessment tools.

CHAPTER IV

PROGRAM IMPLEMENTATION

The Evaluation of the Impact of Teacher Induction Programs set out to study comprehensive teacher induction, an intervention that combines orientation, professional development, and ongoing mentoring services to support new teachers as they begin their careers. The word "comprehensive" is intended to underscore the contrast with the services typically offered to beginning teachers in high-need districts. 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 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.

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, MPR 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 also increases the ability to generalize findings about the impacts 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 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, reviewing 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 and, in the spring, either visited again or conducted semi-structured telephone interviews.[19] 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 beginning teachers whom they served. 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 was 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.

[19] Four of the nine ETS districts (44 percent) and three of the eight NTC districts (38 percent) received a visit. The others received a telephone call.

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 for 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 below. For district-wide implementation with a larger number of mentors, training typically occurs within the district, rather than off site together with mentors from other districts.

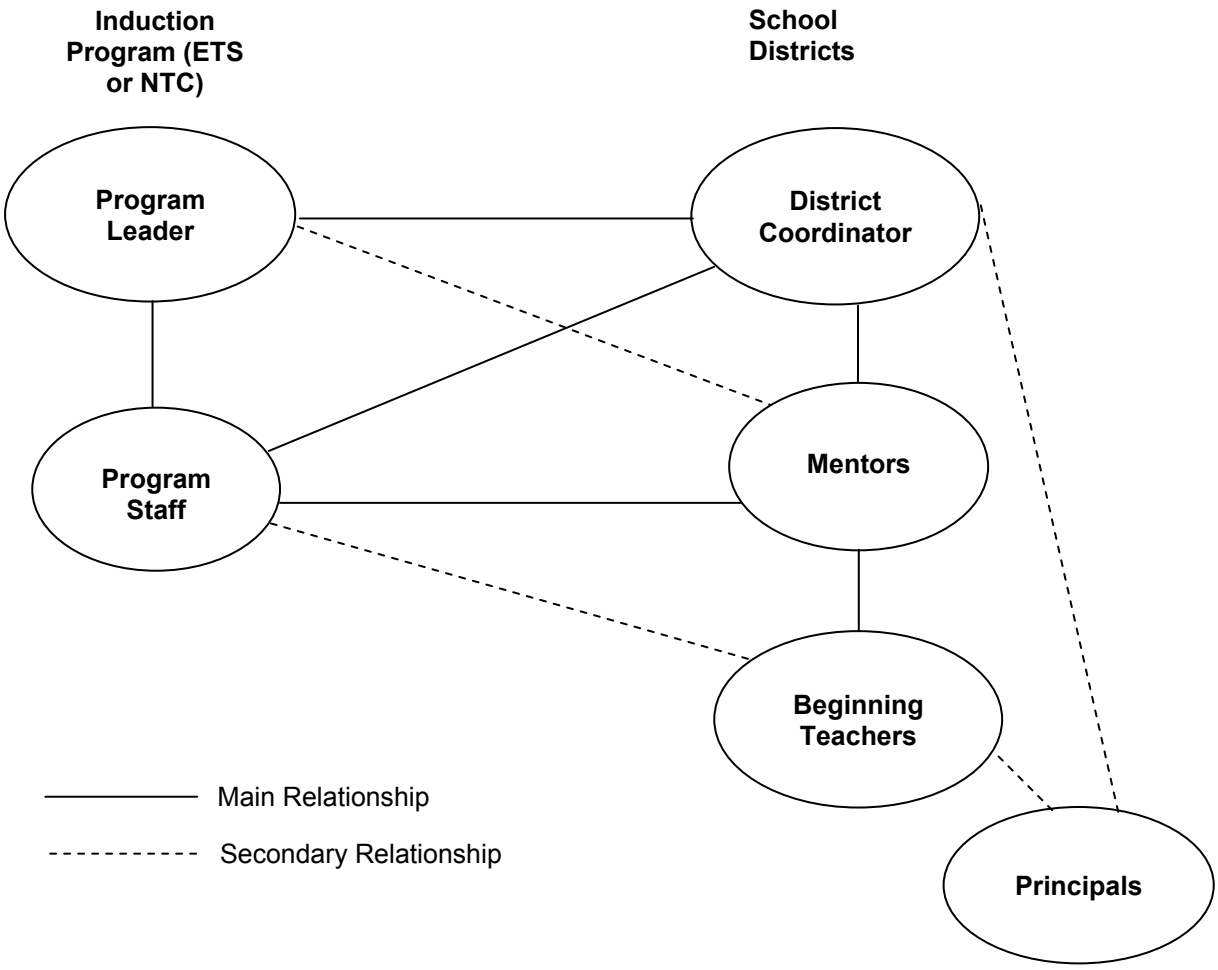
B. ADMINISTRATIVE SUPPORT STRUCTURE

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.[20] 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

[20] 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.

productive relationships among various staff members. In Year 2, an ETS co-leader left the study and was replaced by one of the mentors, while the NTC leader continued in her role.[21]

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



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.[22] 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

[21] The ETS co-leader for the study, who had served under the program leader in Year 1, left due to 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. While the NTC leader continued in this role, this person also served as program staff for one of the districts in Year 2.

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

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.

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, key functions were to help establish district positions for mentors and recruit candidates, establish procedures for job reporting and evaluation, create functional working conditions for mentors by locating office space, and set up email and telephone access. They also helped to 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 coordinator 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 from 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 rooms to serve as meeting spaces or mentor offices. 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.[23]

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.[24] 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.[25] 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.

C. 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), though 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,

[23] 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.

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

[25] 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 briefing.

programs sought individuals 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). In each district, candidates were interviewed by a committee that included the district coordinator for the study and other participants such as representatives from human resources, the teachers' union, and professional development; an assistant superintendent for instruction; other experienced mentors; and/or school administrators. Program leaders traveled to the interviews or conducted telephone consultations with the district coordinator about the finalists, but districts made the final mentor selections. In all but three districts, there were two or more applicants per mentor position. There was one instance of turnover among mentors 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.[26]

Since our analysis is not designed to compare one-year and two-year districts directly, the characteristics of study mentors serving these two types of districts are presented separately. Table IV.1 describes the background of the 25 mentors selected to deliver the comprehensive induction services in the one-year districts. These data are taken from a survey administered to mentors at the outset of program implementation in Year 1. In one-year districts, all mentors reported at least 5 years of teaching experience, with an average of 16.7 years. Forty percent had worked in non-teaching positions in education and all held at least a bachelor's degree; 76 percent had a master's degree. The average age of these mentors was 42 years old in 2005. Mentors were overwhelmingly female (95 percent across both types of districts, not shown in the tables) and 63 percent were white non-Hispanic. While the mentors were implementing the particular program under study for the first time during the 2005-2006 school year, 76 percent reported having prior mentoring experience—6.5 years on average. Ninety percent of these individuals had attended mentor training in the past. The most commonly reported areas of training addressed classroom management, the delivery of effective feedback, and mentor roles (at least 85 percent for each area).

Table IV.2 describes the background of the 19 mentors in the two-year districts based on data from the same source. All mentors in these districts reported at least 5 years of teaching experience, with an average of 19.5 years. Fifty-three percent had worked in non-teaching positions in education. All mentors in these districts had earned a master's degree and 36 percent were certified through the National Board of Professional Teaching Standards. Mentors were aged 44 years old on average in 2005 and 35 percent were white non-Hispanic. Seventy-nine percent reported having prior mentoring experience—5.8 years on average—and 55 percent had previously attended mentor training. The most commonly

[26] Half-way 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.

reported areas of training addressed classroom management, the delivery of effective feedback, and mentor roles (at least 85 percent for each area).

Table IV.1. Mentor Background: One-Year Districts

Characteristics	Percentage	
Race/Ethnicity: Percent White, Non-Hispanic	62.5	
Education: Has Master's Degree	76.0	
Certified Through National Board of Professional Teaching Standards (NBPTS)	[a]	
Teaching Experience		
Last position before mentoring was as a classroom teacher	84.0	
Ever worked in nonteaching position(s) within education	40.0	
Mentoring Background		
Any mentoring experience	76.0	
Any previous mentoring training (if have mentoring experience)	89.5	
Areas of Mentor Training (If Received Mentor Training)		
Classroom management	82.4	
Giving effective feedback	88.2	
Mentor roles	88.2	
Coaching strategies	82.4	
Lesson planning	76.5	
Classroom observations	64.7	
Helping adult learners set goals	47.1	
Analyzing student work	47.1	
Leading study groups	35.3	
Coaching in literacy/language or math	35.3	
	Average	Range (Min., Max.)
Age in 2005 (Years)	42.1	(28, 61)
Teaching Experience (Years)	16.7	(5, 35)
Experience in Nonteaching Position(s) Within Education (Years)	1.2	(0, 4.6)
Years of Mentoring Experience (If Have Mentoring Experience)	6.5	(1, 30)
Caseload (Number of Beginning Teachers)	11.4	(9, 14)
Unweighted Sample Size (Mentors)	25	

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

[a]Exact value suppressed to protect respondent confidentiality.

Table IV.2. Mentor Background: Two-Year Districts

Characteristics	Percentage	
Race/Ethnicity: Percent White, Non-Hispanic	35.3	
Education: Has Master's Degree	100.0	
Certified Through National Board of Professional Teaching Standards (NBPTS)	36.3	
Teaching Experience		
Last position before mentoring was as a classroom teacher	78.9	
Ever worked in nonteaching position(s) within education	52.6	
Mentoring Background		
Any mentoring experience	78.9	
Any previous mentoring training (if have mentoring experience)	55.3	
Areas of Mentor Training (If Received Mentor Training)		
Classroom management	100.0	
Giving effective feedback	85.7	
Mentor roles	85.7	
Coaching strategies	75.0	
Lesson planning	85.7	
Classroom observations	66.7	
Helping adult learners set goals	66.7	
Analyzing student work	57.1	
Leading study groups	50.0	
Coaching in literacy/language or math	62.5	
	Average	Range (Min., Max.)
Age in 2005 (Years)	44.2	(32, 54)
Teaching Experience (Years)	19.5	(10, 32)
Experience in Nonteaching Position(s) Within Education (Years)	1.7	(0, 6.8)
Years of Mentoring Experience (If Have Mentoring Experience)	5.8	(2, 20)
Caseload (Number of Beginning Teachers)	12.1	(8, 14)
Unweighted Sample Size (Mentors)	19	

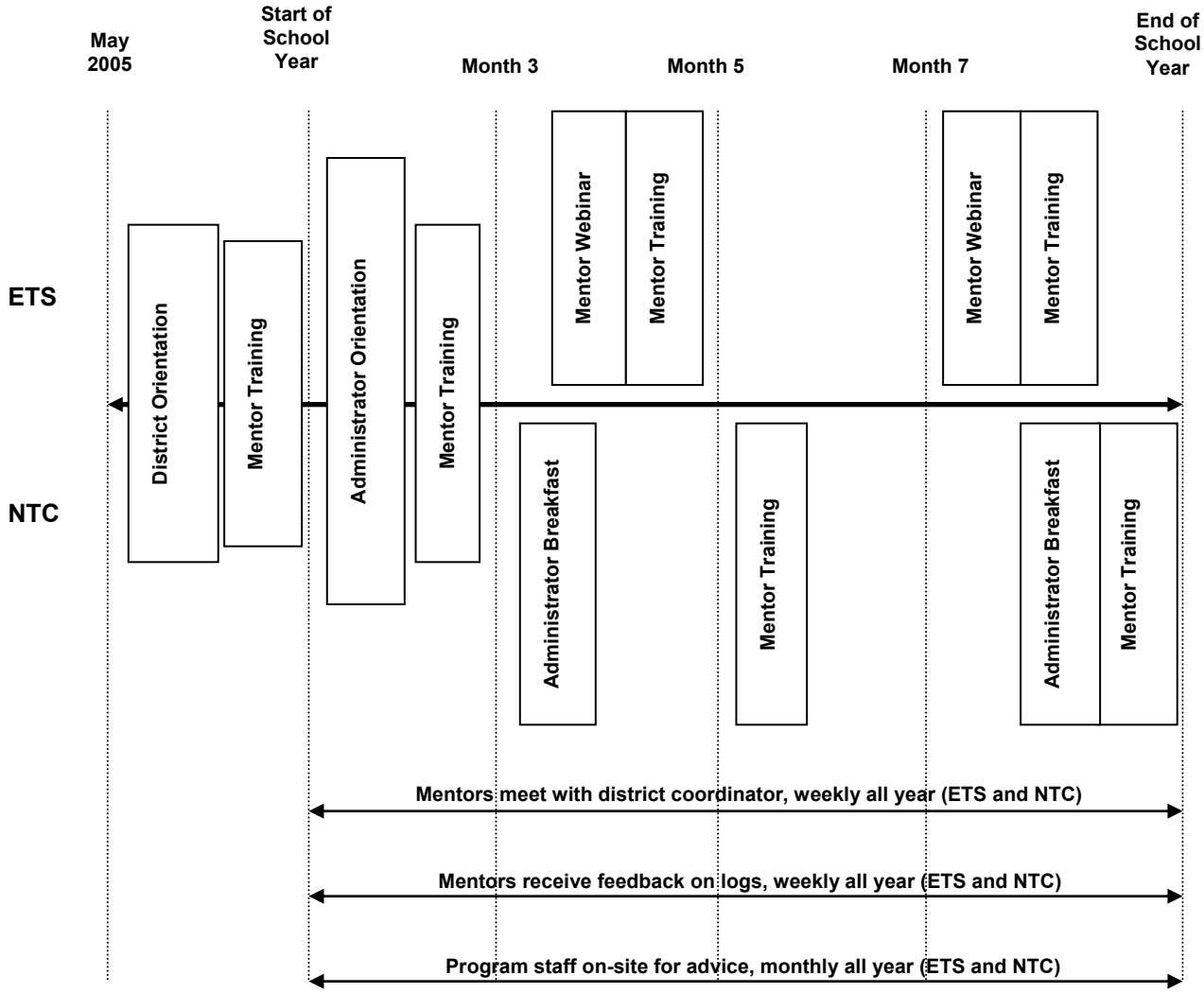
Source: MPR Mentor Survey administered in fall 2005 to all study mentors.

Once mentors were selected for program participation, during the first year of program implementation both ETS and NTC trained their respective mentors in four training sessions 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 caused by reasons such as a death in the family or serious illness. 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' mid-year 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.[27]

[27] Examples of survival and more general needs are how to interact with your principal, teachers' own 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



Notes: 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 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 gaining the support of building administrators. At the outset of the school year, district coordinators provided substantive advice during weekly mentor meetings and three-

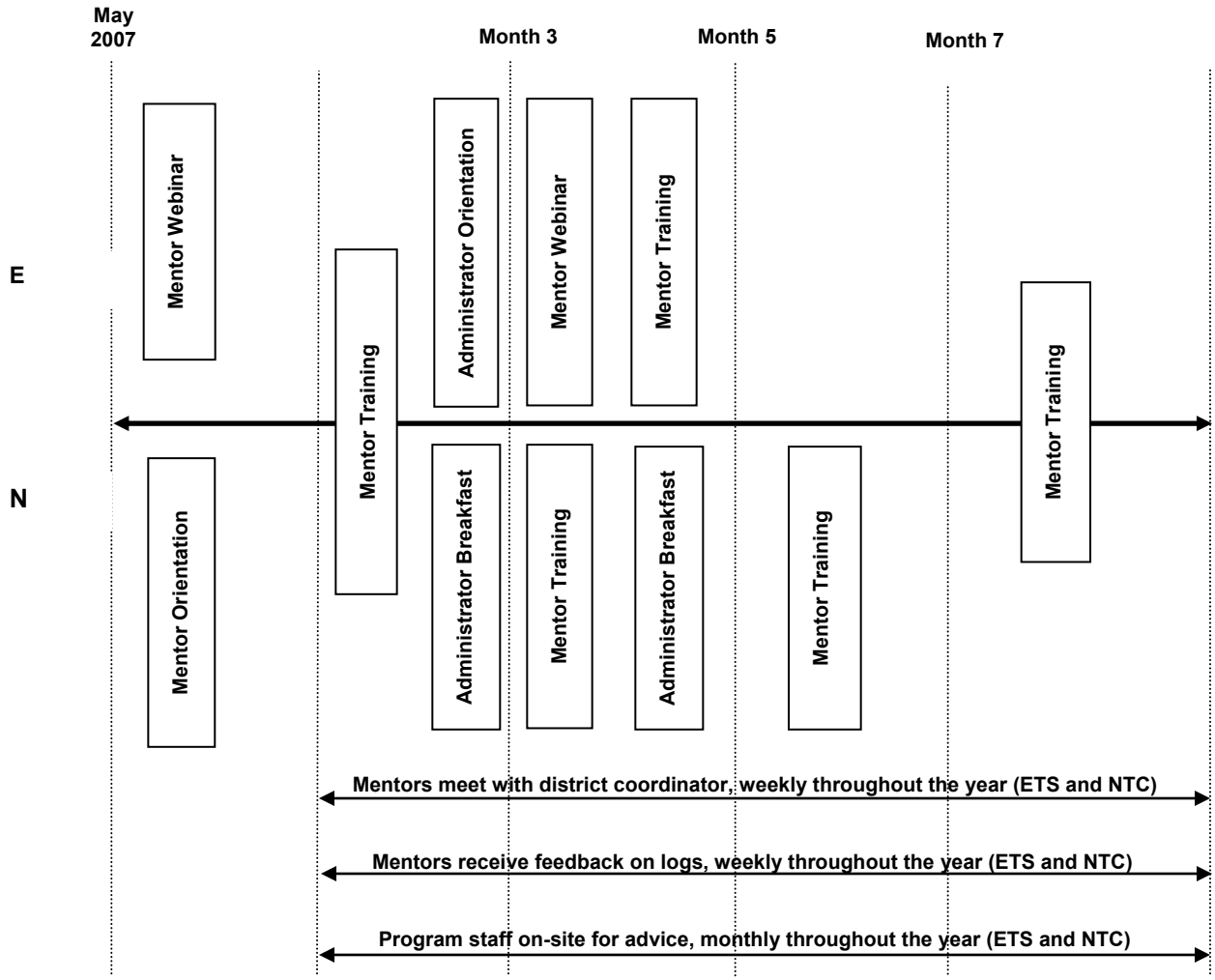
quarters of them continued to join mentor meetings throughout the year. Program staff and district coordinators regularly responded to telephone or email inquiries from mentors, and the ETS program held two one-hour webinars for mentors and district coordinators. The fall webinar helped mentors shift from providing the types 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.)

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 3 and 4 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, while 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.

All mentors participated in the trainings, which reflected a focus similar to 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 also 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 below.

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



Notes: 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.

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 above for Year 1.

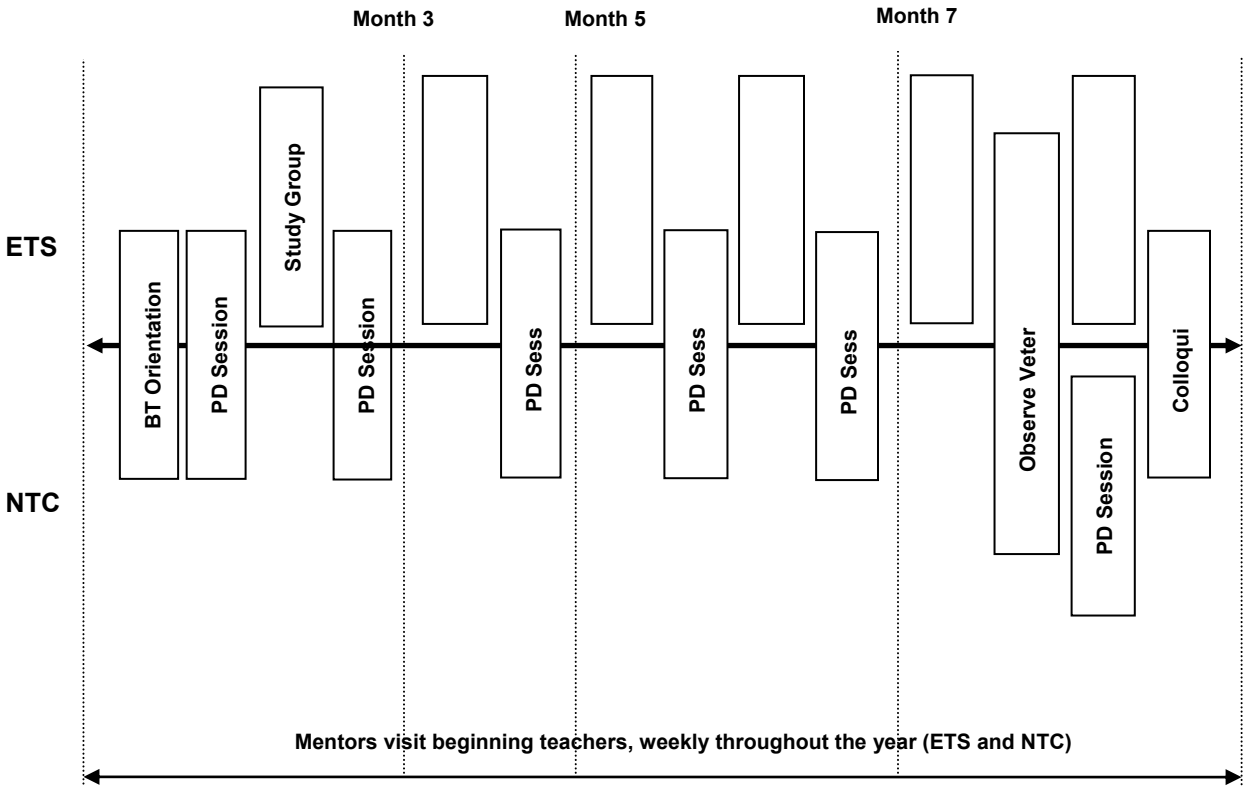
D. PROGRAM SERVICES AND ACTIVITIES

1. 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 (Figure IV.4). 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.[28] 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.[29]

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.

[28] The primary obstacle to holding these early meetings was the delay in district staff 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 where 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.

[29] 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 service did not take time away from focusing on instruction for those teachers who were on track in their development.

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 year-long curriculum for beginning teachers that focuses on effective teaching (Table IV.3). 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.[30] The NTC induction model defines effective teaching in terms of six Professional Teaching Standards.[31] 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.4).

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.

[30] The ETS program derives its content from *Enhancing Professional Practice: A Framework for Teaching* (Danielson 1996).

[31] 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. 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 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 *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).

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 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, though 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.3) and the continuum of performance (Table IV.4). 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 mid-year review to assess progress in meeting goals.

Table IV.4. 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.[32] Mentors were expected to spend some of that time every week meeting with beginning 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 enable mentors to assist teachers in understanding their students' learning challenges.

Monthly Professional Development Sessions.[33] During the 2005-2006 school year, both ETS and NTC held monthly, two-hour professional development sessions (Table IV.5),[34] which complemented the interactions between mentors and beginning teachers as described in the seven ETS events and NTC's FAS. On average, the professional development sessions drew 72 and 65 percent of the beginning teachers (ETS and NTC, respectively, as shown in Tables IV.6 and IV.7). 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 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 84 to 63 percent.

[32] 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.

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

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

Table IV.5. 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.6. 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[a]	High	Low	Teachers Missing No More Than 1 Session	Teachers Missing 3 or More Sessions
Orientation*	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*	87	96	75	n.a.	n.a.

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

*Data not available for orientations. Data available from four of nine districts for end-of-year colloquia.

[a]BT = beginning teacher.
[b]Average of district averages across all five sessions.

n.a. = not applicable.

Table IV.7. Teacher Attendance at NTC Induction Activities (Percentages): 2005-2006 School Year

Activity	Average Attendance of BTs[a]	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)[b]	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]BT = beginning teacher.
[b]Average of district averages across all six sessions.

n.a. = not applicable.

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 interest to 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 from them.[35]

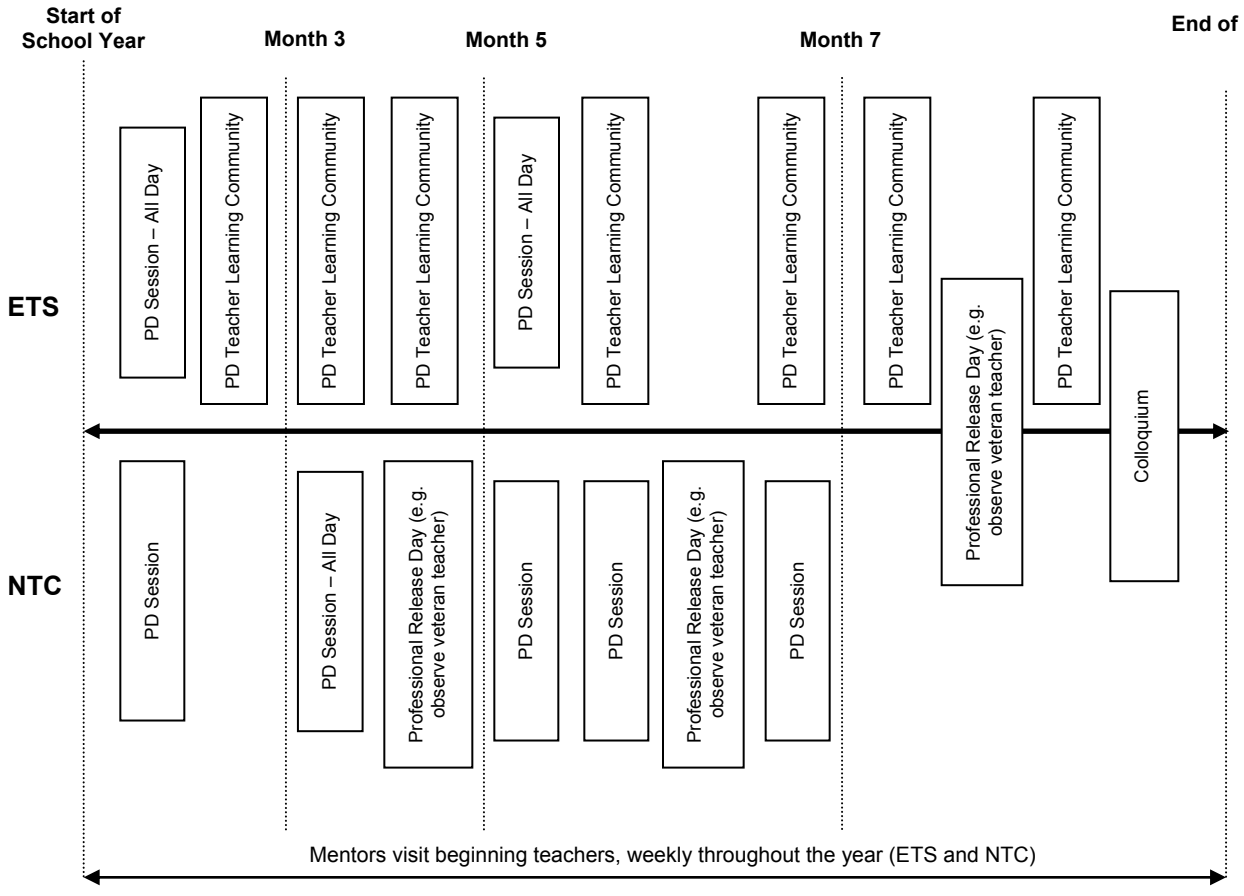
End-of-Year Colloquium. The two- to three-hour colloquium in each district focused on celebrating the first year's successes and teachers' professional growth. It also encouraged teachers to set goals for improved instruction for the year ahead. Attendance at the end-of-year colloquia was similar to 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 higher and lower levels in some districts (ranging from 96 to 46 percent).

[35] 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 them 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.

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

As in Year 1, **mentoring** of beginning teachers (those who were randomly assigned to treatment in Year 1 and are 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 also 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 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



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

Mentoring. Mentoring in the second year was very 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, engaged in the same kinds of mentor/novice interactions described for Year 1. The framework for ETS mentors was again Pathwise Induction Events, while NTC mentors again used the FAS.

Professional Development. The ETS and NTC programs included between 35 to 40 hours of professional development for beginning teachers in Year 2.[36] In ETS districts, a total of eight two-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 two-hour sessions throughout the year, and three release days for observation of other teachers, or individual work with their mentors.[37] As in Year 1, topics of sessions continued to be related to the mentors' weekly work with their beginning teachers.

Programs changed the content and conduct of the professional development sessions during this second year to reflect the growth of mentors and beginning teachers as well as the evolution of their circumstances and needs. While program staff of both programs traveled to districts to conduct or lead 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 (e.g., differentiation in instruction) and designed activities to reflect local needs, in consultation with the program leader and their 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 local 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 and 58 percent of the beginning teachers over the course of the year, for ETS and NTC respectively (Table IV.8).[38,39] 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 and 79 percent for the first ETS and NTC all-day sessions, and 55 percent for the second ETS all-day session.

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

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

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

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

Thirty-eight 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 (over 50 percent) of the sessions (36 and 35 percent of ETS and NTC teachers, respectively).

Table IV.8. ETS and NTC Teacher Attendance: Professional Development Sessions and Colloquia (Percentages): 2006-2007 School Year

		Range of Average Attendance Across Districts		Regularity of Attendance	
Activity	BT Average Attendance[a]	High (Percent)	Low (Percent)	BTs Attending Most Sessions (Percent)	BTs Missing Most Sessions (Percent)
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.

[a]BT = beginning teacher.

n.a. = not applicable.

Table IV.9 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.9, was introduced in the two all-day professional development sessions; during their monthly TLC meetings, teachers then discussed the topics and the 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, e.g., developing more elementary-school examples than the standard product contained.

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 interest to 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.

Table IV.9. 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.3).
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 its data from Year 1 participant survey.

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. It 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 and 60 percent of teachers, ETS and NTC, respectively), with notably higher and lower levels among individual districts (ranging from 96 to 29 percent).

CHAPTER V

IMPACT FINDINGS: ONE-YEAR DISTRICTS

The main goal of this study is to estimate the impact of comprehensive teacher induction on teacher and student outcomes. In this chapter, we present findings from the impact analysis for the ten school districts whose treatment groups received one year of intervention and subsequently returned to the prevailing district induction and professional development services received by the control groups. The first section of the chapter compares the induction experiences of teachers in the treatment group with the experiences of those in the control group, both in Year 1 of the study (during implementation) and Year 2 (after implementation). The gap in services, or service contrast, represents the effect of offering treatment during the first year on the types and intensity of induction services received in both the first and second years of the study. This contrast in services is an important precursor to impacts on desirable outcomes such as student test scores and teacher retention.

The second section of the chapter presents the impact estimates for teacher attitudes, student achievement, and teacher retention. Readers may refer to Appendix A for a detailed description of analytic methods. For each outcome, we present a summary of methods, findings, and sensitivity tests. Despite the simplicity of analysis under a randomized design, some aspects of the study design and outcome measurement required decisions on the part of the researcher that could affect either the impact estimates or the hypothesis tests. For example, each outcome was regression-adjusted using a set of covariates specific to that outcome, a specification known as the "benchmark analysis" for the outcome. We conducted a series of sensitivity analyses to demonstrate the robustness of the findings using alternate samples or specifications of covariates for each outcome.

A. TREATMENT-CONTROL DIFFERENCES IN TEACHER INDUCTION SERVICES

This study does not compare comprehensive teacher induction to the absence of any support services for new teachers; rather, it compares comprehensive teacher induction to the prevailing level of induction services in the selected districts. We use 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 derive estimates of the service contrast, which provides the necessary context for understanding the impacts on teacher and student outcomes. Estimates were computed using an ordinary least squares regression model with district and grade fixed effects. The computation of standard errors accounted for clustering of teachers within schools; weights were applied to adjust for survey nonresponse and the study design.

The data, drawn from the induction activities surveys that were administered during fall 2005 and fall 2006, characterize the induction services received by the treatment and control groups during the fall of Year 1 and the fall of Year 2. We focus on these two time points to illustrate the difference in services received by treatment and control teachers both before and after the comprehensive induction services had ended.[40] Although treatment teachers were offered the same usual district services as control teachers in Year 2, the examination of service usage in this year is important. Analysis of the services received by control teachers in Year 2 provides a description of typical district induction support during teachers' second year in the classroom. Moreover, our analysis can show whether the intervention in Year 1 induced changes in treatment teachers' usage of these services in Year 2 beyond what it would have been in the absence of the intervention.

1. Mentor Assignments

During the first year of the study, in fall 2005, treatment teachers were significantly more likely than control teachers to report having a mentor (93 versus 78 percent, Table V.1). The survey asks teachers if they have a mentor and if the mentor was assigned. Mentors could have been assigned by a teacher's district or principal or by a teacher preparation program. Treatment teachers also reported having an assigned mentor at higher rates, 90 versus 70 percent. One year later, treatment teachers were significantly *less* likely than control teachers to report having a mentor (25 versus 38 percent) or having an assigned mentor (20 versus 29 percent). There are no data to explain why treatment teachers in one-year districts received significantly less support in Year 2 than control teachers. Districts provided a mix of one-year and two-year induction programs to teachers in the control schools, although data are not available to indicate which control schools had which types of programs.

[40] For ease of exposition, the presentation in this chapter excludes results from the induction activities survey administered in spring 2006, which are described in Glazerman et al. 2008 and can be found in Appendix C of this report, Tables C.1-C.5.

Table V.1. Teacher Reports on Professional Support and Duties (Percentages): One-Year Districts

	Fall 2005				Fall 2006			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
BT[a] 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
Unweighted Sample Size (Teachers)	258	245	503		241	231	472	

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

Notes: 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]BT = beginning teacher.

*Significantly different from zero at the .05 level, two-tailed test.

2. Number and Types of Mentors

Table V.2 presents estimates of treatment-control differences in mentor assignments and mentor profiles in fall 2005 and fall 2006. Treatment teachers were significantly more likely than control teachers to report having multiple mentors (25 versus 15 percent), having two mentors assigned to them (19 versus 7 percent), and having a full-time mentor (74 versus 8 percent) in fall 2005. Treatment teachers were significantly less likely than control teachers to report having a mentor who was another teacher (25 versus 64 percent).

In fall 2006, after the comprehensive induction services had ended, treatment teachers were significantly less likely than control teachers to report having two assigned mentors (2 versus 6 percent, Table V.2). Treatment teachers were also significantly less likely than control teachers to report having a mentor who was another teacher (21 versus 31 percent).

Given that some teachers had more than one mentor, Tables V.2–V.5 report on the induction services received by teachers for all of a teacher's mentors. For example, under "Mentor Positions" in Table V.2, the row labeled "Full-time mentor" indicates the percentages of teachers reporting any full-time mentor.

3. Meetings with Mentors

Table V.3 presents estimates of treatment-control differences in mentor meetings and activities in fall 2005 and fall 2006. Combining usual scheduled time and informal time during the most recent full week of teaching, we find treatment teachers spent an average of 87 minutes in mentor meetings compared to 67 minutes for control teachers in fall 2005. Since total meeting time is not reported directly but must be constructed from reports of the frequency and duration of usual scheduled meetings and the time spent in informal meetings, we cannot determine precisely whether treatment teachers met with their study mentors for two hours per week as the ETS and NTC programs expected. The reported meeting time includes all mentors, which may capture time spent with mentors that were not part of the experimental intervention. Thus 87 minutes (Year 1) and 19 minutes (Year 2) represent upper bound estimates of time that treatment teachers spent with mentors assigned through the ETS or NTC programs.

The statistically significant 21-minute difference is attributable entirely to differences in the duration of the usual scheduled meetings (56 versus 34 minutes). Treatment teachers reported spending significantly more time meeting with full-time mentors than did control teachers (60 versus 4 minutes) during the most recent week of teaching, but reported significantly less time than control teachers with mentors who were also teachers (23 versus 60 minutes).

Table V.2. Impacts on Teacher-Reported Mentor Profiles (Percentages): One-Year Districts

	Fall 2005				Fall 2006			
Mentoring Characteristic	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Number of Mentors								
Multiple Mentors (More Than One)	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
Unweighted Sample Size (Teachers)	258	245	503		241	231	472	

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

Notes: 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 .05 level, two-tailed test.

Table V.3. Impacts on Teacher-Reported Mentor Services Received in Most Recent Full Week of Teaching: 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[c] 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 something BT requested[d]	71.9	50.7	21.1*	n.a.	0.000	12.0	20.5	-8.6*	n.a.	0.015
Unweighted Sample Size (Teachers)	258	245	503			241	231	472		

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

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]BT = beginning teacher.

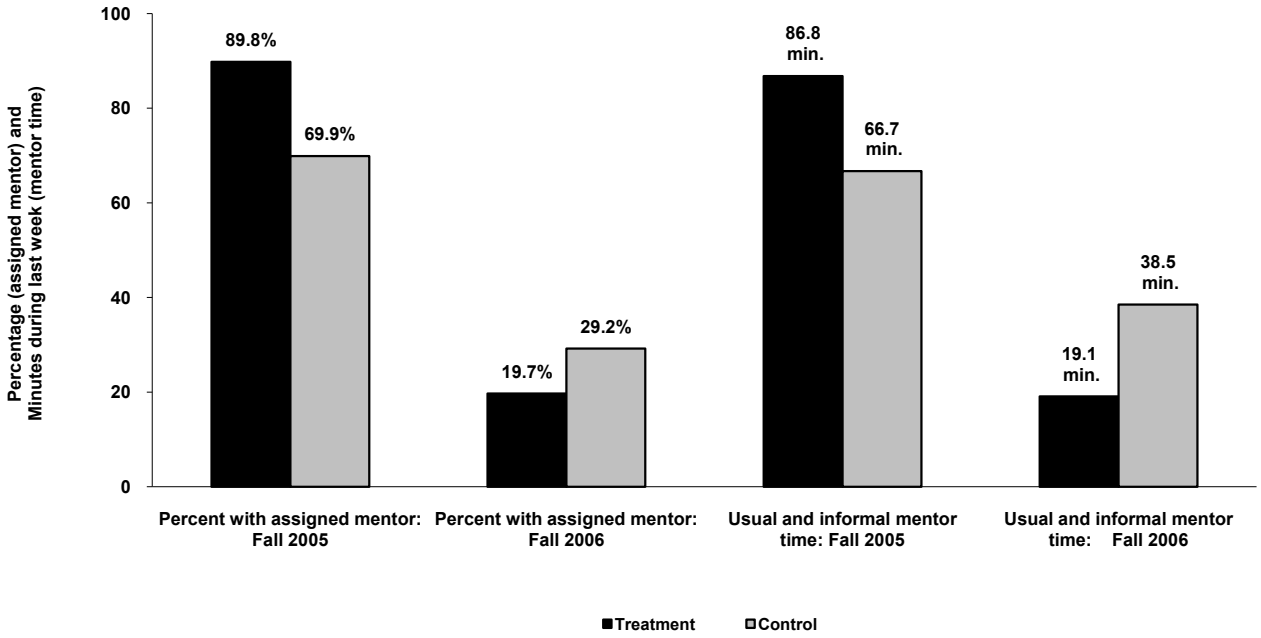
[d]Total sample size is 396 in fall 2005; 441 in fall 2006. The question did not apply to teachers who did not make a request to their mentors.

*Significantly different from zero at the .05 level, two-tailed test.

n.a. = not applicable.

In fall 2006, combining the usual scheduled time and informal time during the most recent full week of teaching, on average treatment teachers spent significantly less time in mentor meetings than control teachers (19 versus 39 minutes), which resulted from spending less time both in scheduled meetings (10 versus 18 minutes) and in informal meetings with mentors (9 versus 20 minutes). Treatment teachers also reported spending significantly less time with mentors who were teachers (17 versus 33 minutes). Figure V.1 shows treatment-control differences for having an assigned mentor and time in mentor meetings in Year 1 and Year 2. The declines in these two key measures of services from Year 1 to Year 2 are statistically significant (p-value=0.000) for both treatment and control teachers.[41] Estimates of the treatment-control difference in time spent with mentors are shown separately by district in Appendix B, Figure B.1.

Figure V.1. Treatment-Control Differences in Percent Assigned a Mentor and Total Minutes Spent in Mentoring Per Week: One-Year Districts, Fall 2005 and Fall 2006



Note: All treatment-control differences are significantly different from zero at the 0.05 level, two-tailed test (N=503 teachers in fall 2005 and 472 teachers in fall 2006).

[41] We did not test the differences in the declines between treatment and control teachers for statistical significance because we did not have a hypothesis regarding the sign of this difference.

4. Mentor Activities and Assistance

In addition to reporting spending more time meeting with mentors during Year 1, Table V.3 shows that treatment teachers reported spending significantly more time than control teachers in specific types of mentoring activities during the most recent full week of teaching in fall 2005. These activities included being observed by mentors (34 versus 10 minutes), meeting one-on-one with mentors (34 versus 23 minutes), meeting with mentors together with other first-year teachers (29 versus 9 minutes), and having mentors model lessons (9 versus 6 minutes). The total time spent in the six types of activities covered by the survey averaged 130 minutes per week for treatment teachers and 67 minutes per week for control teachers, a significant difference of 63 minutes per week.

In contrast, treatment teachers in Year 2 reported significantly less time being observed by mentors than control teachers (2 versus 6 minutes) during the most recent full week of teaching in fall 2006 but did not differ significantly on their reported time spent in any of the other five activities covered by the survey. Treatment teachers averaged less total time than control teachers in the six types of activities covered by the survey (22 minutes per week for treatment teachers versus 36 minutes per week for control teachers).

In Year 1, treatment teachers were significantly more likely than control teachers to report receipt of a wide range of types of mentor assistance. The bottom panel of Table V.3 shows that, during the most recent full week of teaching in fall 2005, treatment teachers were significantly more likely than control teachers to report receiving mentors' assistance in 9 out of 10 topic areas covered by the survey, with effects ranging from 14 to 27 percentage points, and significant differences above 20 percentage points on receiving suggestions to improve practice (77 versus 53 percent), receiving encouragement or moral support (87 versus 66 percent), having opportunities to raise issues and discuss concerns (86 versus 65 percent), receiving help on identifying teaching challenges and solutions (82 versus 55 percent), discussing instructional goals (73 versus 48 percent), and receiving help that the beginning teacher requested (72 versus 51 percent). Among treatment teachers, the percentage reporting each type of assistance ranged from 56 percent sharing lesson plans, assignments, and other instructional activities, to 87 percent receiving encouragement or moral support. Among control teachers, the percentage reporting each type of assistance ranged from 44 percent receiving guidance on how to assess students to 66 percent receiving encouragement or moral support.

In Year 2, treatment teachers were significantly less likely than control teachers to report receipt of a wide range of types of mentor assistance. Table V.3 shows that during the most recent full week of teaching in fall 2006, treatment teachers were significantly less likely than control teachers to report receiving mentors' assistance in each of the topic areas covered by the survey, with effects ranging from 8 to 14 percentage points, and significant differences above 10 percentage points on receiving suggestions to improve practice (15 versus 27 percent), receiving encouragement or moral support (21 versus 33 percent), having an opportunity to raise issues or discuss concerns (18 versus 32 percent), receiving help with administrative/logistical issues (12 versus 25 percent), discussing instructional goals (14 versus 24 percent), and receiving guidance on how to assess students (11 versus 21 percent). Among treatment teachers, the percentage reporting each type of assistance ranged from 11 percent receiving guidance on how to assess

students to 21 percent receiving encouragement or moral support. Among control teachers, the percentage reporting each type of assistance ranged from 19 percent receiving help teaching to state standards to 33 percent receiving encouragement or moral support.

5. Professional Development

Table V.4 presents estimates of treatment-control differences in professional development activities in fall 2005 and fall 2006. During the three months prior to the fall 2005 survey, treatment teachers were significantly more likely than control teachers to report working with study groups of new teachers (66 versus 34 percent) and observing others teaching in their classrooms (61 versus 44 percent). Treatment teachers were significantly less likely than control teachers to report meeting with a resource specialist to discuss needs of a particular student (66 versus 77 percent). Compared to control teachers, treatment teachers were also significantly more frequently observed by mentors during the three months prior to the fall survey (4.0 versus 1.5 times) and more frequently given feedback on teaching not as part of a formal evaluation (3.2 versus 2.4 times) during this period.

In contrast, during the three months prior to the fall 2006 survey, treatment teachers were significantly less likely than control teachers to report working with a study group of new teachers (11 versus 21 percent) and were significantly less likely to be observed by a mentor (0.3 times versus 0.6 times).

Nearly all study teachers reported having been offered professional development sessions in fall 2005 (99.4 percent) and fall 2006 (97.4 percent); differences between treatment and control teachers were not statistically significant (p-values 0.639 and 0.430, respectively). Treatment and control teachers did not differ significantly in their reported attendance in professional development, except in certain areas. See Table V.5 for the fall 2005 and fall 2006 service contrast estimates for professional development topic sessions attended by teachers during the past three months. Of the 12 professional development topics covered by the survey, treatment teachers were significantly less likely than control teachers to report having attended professional development sessions in two areas in fall 2005: content area knowledge (61 versus 72 percent) and preparing students for standardized testing (30 versus 41 percent). Treatment and control teachers did not differ significantly in attendance in any of the 12 professional development areas in fall 2006, as shown in Table V.5.

Table V.4. Impacts on Teacher-Reported Professional Development Activities During Past Three Months: One-Year Districts

Aspect of Professional Development	Fall 2005					Fall 2006				
	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 3 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
Unweighted Sample Size (Teachers)	258	245	503			241	231	472		

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

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 .05 level, two-tailed test.

[a]Effect sizes are reported for continuous measures, but are not indicated for dichotomous variables that are reported as percentages.

n.a. = not applicable.

Table V.5. Impacts on Teacher-Reported Areas of Professional Development During the Past Three Months (Percentages): 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
Unweighted Sample Size (Teachers)	258	245	503		241	231	472	

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

Notes: 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, two-tailed test.

B. IMPACT FINDINGS: TEACHER SATISFACTION

The impact of teacher induction on teacher attitudes was not one of the study's central research questions, but it can nonetheless be viewed as an important early signal of whether the program is generating its intended effect—an intermediate step on the way to improving teaching and encouraging retention. The induction activities surveys allowed us to examine whether comprehensive teacher induction made teachers feel more satisfied with their jobs. The survey results indicated that this was not the case. As shown below, there were no statistically significant impacts of treatment on teacher satisfaction in fall 2005 or fall 2006.

1. Methods

Using items from the induction activities surveys, we measured teachers' feelings of satisfaction in 19 areas. Factor analysis suggested that teacher satisfaction consisted of three categories: (1) school, (2) class, and (3) career (details are given in Appendix A). The constructed scales for each of these three categories exhibited internal consistency ranging from 0.73 to 0.91, as tested by the Cronbach's alpha coefficient. Psychometric properties for each scale are given in Appendix A, Table A.4.

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

2. Impact Estimates

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 both fall 2005 and fall 2006. Out of the six differences examined (three measures at two points in time), none were statistically significant (Table V.6).[42] As a sensitivity analysis, we recoded the teacher satisfaction data into two categories and examined individual survey items separately. The results show no statistically significant differences with regard to teachers' reports of satisfaction in fall 2005, fall 2006, or spring 2006.[43] See Appendix C (Tables C.7-C.8) for details.

[42] The spring 2006 impact analysis is presented in Appendix C, Table C.6. We reached the same general conclusion of no statistically significant positive impacts of treatment on teacher satisfaction in spring 2006.

[43] Teacher attitudes were not measured in one-year districts in spring 2007.

Table V.6. Impacts on Teacher Satisfaction (Scores on a Four-Point Scale): One-Year Districts

	Fall 2005					Fall 2006				
	Treatment	Control	Difference	Effect Size	P-value	Treatment	Control	Difference	Effect Size	P-value
Feel Satisfied with:										
School	3.1	3.1	0.0	0.0	0.751	3.2	3.1	0.0	0.0	0.843
Class	3.0	3.0	0.1	0.1	0.339	3.1	3.1	0.0	0.0	0.812
Teaching career	3.0	3.0	-0.1	-0.1	0.290	3.0	3.0	0.0	-0.1	0.615
Unweighted Sample Size (Teachers)	258	245	503			241	231	472		

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

Notes: Data pertain to teachers in all 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, two-tailed test.

C. IMPACT FINDINGS: STUDENT TEST SCORES

We compared the test scores for students of treatment teachers to those of control teachers, adjusted for pretest scores. Though 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 most important and worthy of assessing.

We focused on results from the teachers' second year of teaching but also compared results from the second year to results from the first year of teaching. Although comprehensive teacher induction services ended after the 2005-06 school year, we hypothesize that there can be delayed impacts of induction programs because teachers may not be able to implement the advice they have been given immediately. We found no overall impacts for math or reading in the second year. We checked the findings using different methods of aggregation, model specification, and model estimation.

1. Methods

Estimating impacts on student achievement posed a challenge, requiring careful use of test score data from nine districts, which administered different tests under different conditions and followed different recordkeeping practices. Although ten one-year districts participated in the study, one of these districts was unable to match teachers in the study with student test scores.

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. We kept two broad subject areas, math and reading, distinct. The benchmark model was a hierarchical linear model, which accounts for the nesting of students within schools. As shown in Table A.1 in Appendix A, the normalized student pretest score and district-by-grade fixed effects are covariates in the benchmark model. Appendix A describes in more detail the aggregation method, treatment of missing data, regression model, and estimation strategies

2. Impact Estimates

The benchmark impacts on math and reading scores in the second year of teaching were not significantly different from zero (see Table V.7).

Table V.7. Impacts on Test Scores: One-Year Districts, 2006-2007 School Year

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Subject	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
Reading	0.05	0.01	0.04	0.04	0.380	2,245	135	9
Math	0.05	-0.02	0.08	0.08	0.367	1,995	117	9

Source: MPR analysis of data from 2005-2006 and 2006-2007 school years provided by participating school districts.

Notes: Data are regression-adjusted to account for pretest, district-by-grade fixed effects, and clustering of students within schools. Treatment and control group sample sizes are shown in Appendix Table C.13.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Another way to analyze these data is to consider how test score impacts in Year 2 may have differed from impacts in Year 1. If there were no treatment/control differences in Year 1 or Year 2 but a significant gain in impacts of the treatment teachers relative to the control teachers from one year to the next, this might indicate that the effect of comprehensive teacher induction on outcomes has a delayed impact. This may result if teachers need time to assimilate the advice they were given in Year 1. It might also suggest that further gains of treatment teachers relative to control teachers may be possible in Year 3.

We focus on the subsample of teachers who had students with valid test score data in both years, the "common sample." There are two reasons why this subsample may differ from the entire sample in a single year. First, test score data were available in a different set of district-grade combinations in the two years. Second, some teachers left teaching or changed assignments (out of tested grades and subjects) before the end of Year 2. An added benefit of the common sample analysis is that, by including only teachers who are in the sample in both years, we isolate the productivity effect of teacher induction on student achievement separate from the composition effect.

The impacts on reading and math for the common sample of teachers, shown in Table V.8, indicate no significant improvement for reading or math test scores.

In addition to the common sample analysis, we conducted other sensitivity tests using the benchmark sample and model. We confirmed that the impacts on reading and math scores in Year 2 were not statistically significant when the impacts were re-estimated using different samples, sets of covariates, or estimation techniques. First, the results are disaggregated by grade, with each grade considered individually and with the sample restricted to students from grades 3–5 (the grades typically covered by state assessments). Second, we use the original data, without forcing outliers to have minimum values of -3 and

Table V.8. Impacts on Test Scores in Year 1 and Year 2, Common Sample of Teachers: One-Year Districts

Subject	Adjusted Mean Test Scores		Difference	Effect Size	P-value	Unweighted Sample Sizes		
	Treatment	Control				Students	Teachers	Districts
Reading								
Year 1	0.07	0.10	-0.03	-0.03	0.553	1,519	82	7
Year 2	0.06	-0.01	0.07	0.07	0.236	1,458	82	7
Year 2–Year 1	-0.01	-0.11	0.10	0.10	0.231			
Math								
Year 1	0.08	0.05	0.03	0.03	0.667	1,274	73	6
Year 2	0.02	0.01	0.01	0.01	0.832	1,266	73	6
Year 2–Year 1	-0.06	-0.05	-0.01	-0.01	0.867			

Source: MPR analysis of data from 2004–2005, 2005–2006, and 2006–2007 school years provided by participating school districts.

Notes: Data are regression-adjusted to account for pretest, district-by-grade fixed effects, and clustering of students within schools. Treatment and control group sample sizes are shown in Appendix Table C.14.

The common sample is the subsample of teachers who had students with valid test score data in Year 1 and Year 2.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

maximum values of 3. Third, we add student demographic covariates. Fourth, we use student and teacher covariates. The covariates used in these models are given in Appendix A, Table A.1. Fifth, we use ordinary least squares rather than hierarchical linear modeling, and account for correlation of outcomes for students in the same school using robust standard errors. Sixth, we estimate impacts without controlling for a pretest. Seventh, we estimate a model in which the math pretest is used as an instrumental variable to control for measurement error in the reading pretest. See Appendix C (Tables C.9-C.12) for details. Figures B.2 and B.3 in Appendix B show estimates of the impacts on reading and math scores separately by district. There is one outlier district with a statistically significant negative impact on each subject's scores, but the exclusion of this district did not alter the findings from the benchmark model.

D. IMPACT FINDINGS: TEACHER RETENTION

An often-cited goal of comprehensive teacher induction is the increase in retention of beginning teachers, who are presumed to be at greatest risk of leaving the profession in the first five years of their teaching career (Kapadia et al. 2007). To address the question of turnover, the effect of comprehensive induction programs on the retention of new teachers was examined.

We are interested not only in the rate of retention overall, but also in the effects of such retention on the composition of the teaching force in the district. Although staff turnover can be disruptive and costly, 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. The random assignment design allowed us to test directly the effects of comprehensive teacher induction on the composition of the teaching force by comparing the characteristics of treatment teachers who stayed in the district in subsequent years to control teachers who did so. Under random assignment, the treatment and control teachers are equivalent, on average, prior to the intervention. At the end of two years of teaching, after some teachers have left the district (or teaching), the average quality and qualifications of both groups of teachers may change. We examined the impacts for Year 1 test score performance on teachers who stayed in the same district as well as differential attrition by teacher qualifications like advanced degrees and certification status. We found no evidence of a retention impact or composition effect after two years.

1. Methods

Teachers' mobility status can be defined in a variety of ways but most commonly it falls into three categories: (1) stayers—teachers who stay at their original school; (2) movers—teachers who move to another school either within the same district or to another district; and (3) leavers—teachers who leave the teaching profession. Sometimes it is useful to redefine stayers and movers in terms of whether the teacher remains in the district rather than in the school. Many teachers may change schools but remain in the district, especially newer teachers who may be involuntarily transferred to help the district match staffing to student enrollment patterns. Thus, mobility rates are always higher at the school level than at the district level. We use the district perspective here unless otherwise noted because adoption of a comprehensive induction program, such as the ones under study, is a district-level policy decision. A teacher's mobility status can vary over time; unless otherwise stated, we report mobility as of fall 2007, which indicates whether the teacher returned to the district for a third year.

The impact estimates are derived from a logistic regression model that mimics the models used for teacher satisfaction and student achievement, except that the outcome variable is binary. The model is described in Appendix A and the covariates are listed in Table A.1. As part of the sensitivity tests, we estimated the model with other assumptions such as a linear probability model and multinomial logit model (one that models staying/moving/leaving as a categorical outcome).

To estimate the impacts of comprehensive induction on the composition of the district's teaching force, we re-estimated the impacts on student achievement but included only the district stayers in the analysis. If comprehensive teacher induction is to improve the composition of the district's teaching force, then one would expect the teachers with more credentials to be more highly represented among those who remained in the district after movers and leavers are accounted for. Similarly, a positive composition effect would imply that the teachers who had produced greater achievement gains would be more highly represented among the stayers. We assume that the average quality and qualifications of

replacement teachers are unaffected by treatment status and, thus, there can be no difference in the composition of the teaching force without having made a difference between the two groups of stayers.

2. Impact Estimates

After two years, 63 percent of study teachers returned to the same schools (see Table V.9). Another 17 percent had changed schools since fall 2005 but remained in the same district. An additional 11 percent stayed in teaching but changed districts or left the public sector. The remaining 10 percent of teachers left the profession altogether. The regression-adjusted district retention rate was 80 percent and the total retention rate in teaching (including movers) was 90 percent.

No impacts of treatment were found on this pattern of teacher mobility after two years. The control group's teacher mobility pattern was statistically indistinguishable from that of the treatment group. Table V.9 shows the result of the three hypothesis tests specifically focused on retention in the school, in the district, and in the profession as binary outcomes. For each of the outcomes, there was no statistically significant impact.

Table V.9. Impacts on Teacher Retention Rates After Two Years (Percentages): One-Year Districts

Outcome	All Teachers	Treatment	Control	Difference	P-value
Retained in the same school	62.5	60.3	64.7	-4.5	0.280
Retained in the same district	79.5	78.6	80.3	-1.7	0.619
Retained in the teaching profession	90.1	90.4	89.8	0.7	0.789
Unweighted Sample Size (Teachers)	476	244	232		
Unweighted Sample Size (Schools)	227	114	113		

Source: MPR Mobility Survey administered in 2007-2008 and Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: 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, two-tailed test.

We also examined movers' and leavers' self-reported reasons for leaving their schools and found no statistically significant impacts of treatment. The two possible reasons for a lack of statistical significance are that the sample size is too small to detect a relationship—about 10 percent of the sample members were leavers and 10 percent were movers—or that there may in fact be no relationship between comprehensive induction and the reasons for moving or leaving. We do not present tabulations for reasons for moving out of one's original school to protect respondent confidentiality. The reasons for leaving are not presented because there were too few cases to draw meaningful inferences. When we asked leavers whether they expected to return and, if so, when they would do so, we found no evidence of a treatment-control difference.

The treatment did not result in the retention, after Year 2, of teachers who had produced higher Year 1 test scores than control teachers. The observed differences between test scores of treatment and control stayers were not statistically significant. Table V.10 presents the impacts on Year 1 student achievement outcomes for those who returned to teach in the same district for the 2007–2008 school year.

Table V.10. Impacts on Test Scores, District Stayers Only: One-Year Districts, 2005-2006 School Year

Outcome	Treatment	Control	Difference	Effect Size	P-value
Reading Scores (All Grades)	0.02	-0.03	0.05	0.05	0.331
Unweighted Sample Size (Students)	975	942	1,917		
Unweighted Sample Size (Teachers)	53	56	109		
Unweighted Sample Size (Schools)	47	41	88		
Math Scores (All Grades)	0.01	-0.02	0.03	0.03	0.629
Unweighted Sample Size (Students)	826	857	1,683		
Unweighted Sample Size (Teachers)	47	52	99		
Unweighted Sample Size (Schools)	43	38	81		

Source: MPR analysis of data from 2004-2005 and 2005-2006 school years provided by participating school districts; MPR Second Mobility Survey administered in 2007-2008 to all study teachers.

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

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Table V.11 shows the background characteristics of teachers by mobility status. We also looked at certification (regular or probationary), highest degree earned, and whether the teacher was a career changer, but do not present these tabulations to protect respondent confidentiality. Across a wide variety of characteristics we found no differences between the treatment and control group stayers nor were there significant treatment-control differences between movers or between leavers, suggesting that comprehensive teacher induction did not induce a change in the mix of teachers who remained in the districts under study.

We examined the robustness of the teacher retention findings with respect to different sample inclusion/exclusion criteria, definitions of mobility, and modeling assumptions and, in each case, reached the same conclusion. In addition, Figure B.4 in Appendix B shows estimates of impacts on teacher retention separately by district.

Table V.11. Characteristics of District Stayers, Movers, and Leavers After Two Years by Treatment Status (Percentages Except Where Noted): One-Year Districts

	Treatment			Control			Difference		
Teacher Characteristic	Stayers	Movers	Leavers	Stayers	Movers	Leavers	Stayers	Movers	Leavers
College entrance exam scores (SAT combined score or equivalent)	1,026	1,029	1,082	1,021	984	1,080	4	45	2
Attended highly selective college	30.3	27.3	46.0	27.2	50.5	33.3	3.1	-23.2	12.7
Major or minor in education	79.8	65.5	76.1	81.1	65.9	67.2	-1.3	-0.4	8.9
Student teaching experience (weeks)	16.5	13.9	14.2	15.1	13.5	12.4	1.5	0.4	1.8
Entered the profession through traditional four-year program	64.4	61.0	45.8	60.3	58.7	30.8	4.1	2.4	15.0
Unweighted Sample Size (Teachers)	191	29	24	187	23	22			
Unweighted Sample Size (Schools)	100	25	18	104	22	21			

Source: MPR calculations using data from the College Board and ACT, Inc.; MPR Second Mobility Survey administered in 2007-2008; MPR First and Second Induction Activities Surveys administered in fall/winter 2005-2006 and spring 2006 to all study teachers.

Notes: 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 (191/29/24 treatment stayers/movers/leavers and 187/23/22 control stayers/movers/leavers) and schools (100/25/18 treatment and 104/22/21 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, two-tailed test. P-values are suppressed to make the table easier to read.

Finally, we considered nonresponse to the mobility survey. Though the overall response rate to this survey was 85 percent, the response rates for treatment and control groups differed (90 and 80 percent, respectively). If nonrespondents differed from respondents in characteristics related to outcomes, then differential nonresponse could bias the impact estimates. To test this, we re-estimated impacts under alternate assumptions about nonrespondents, and found no impacts of treatment except under the most extreme and implausible assumptions. See Appendix C (Table C.19) for details.

CHAPTER VI

IMPACT FINDINGS: TWO-YEAR DISTRICTS

This chapter presents the impact analysis for the seven school districts whose treatment groups were offered two years of comprehensive teacher induction. The organization of this chapter parallels Chapter V, which reports outcomes for one-year districts. The first section of the chapter compares the induction experiences of teachers in the treatment group with the experiences of those in the control group, both in Year 1 and Year 2 of the study. The second section of the chapter presents the impact estimates for teacher attitudes, student achievement, and teacher retention. The basic methodological issues are discussed in Chapters II and V. Readers may refer to Appendix A for a detailed description of analytic methods.

A. TREATMENT-CONTROL DIFFERENCES IN TEACHER INDUCTION SERVICES

Consistent with the analysis of one-year districts, we compare differences in induction service receipt between the treatment and control groups in the two-year districts in fall 2005 and fall 2006, the study teachers' first and second years of teaching, respectively. This analysis characterizes the two years of comprehensive induction services received by the treatment teachers, as well as the district and school services received by the control teachers over the same two-year period.

1. Mentor Assignments

During the first year of the study, in fall 2005, treatment teachers were significantly more likely than control teachers to report having a mentor (98 versus 86 percent, Table VI.1) or having an assigned mentor (94 versus 79 percent). During the second year of the study, in fall 2006, treatment teachers were still significantly more likely than control teachers to report having a mentor (80 versus 41 percent) or having an assigned mentor (80 versus 34 percent).

Table VI.1. Teacher Reports on Professional Support and Duties (Percentages): Two-Year Districts

	Fall 2005				Fall 2006			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
BT[a] 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
Unweighted Sample Size (Teachers)	213	182	395		191	169	360	

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

Notes: 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]BT = beginning teacher.

*Significantly different from zero at the .05 level, two-tailed test.

2. Number and Types of Mentors

Treatment teachers were significantly more likely than control teachers to report having multiple mentors (38 versus 23 percent), having two mentors assigned to them (31 versus 13 percent), and having a full-time mentor (72 versus 16 percent) in fall 2005 (see Table VI.2). Treatment teachers were significantly less likely than control teachers to report having a mentor who was another teacher (38 versus 62 percent).

In fall 2006, treatment teachers were no longer significantly more likely than control teachers to report having multiple mentors or two assigned mentors, but were significantly more likely than control teachers to have one assigned mentor (73 versus 27 percent). Treatment teachers were still significantly more likely than control teachers to report having a full-time mentor (64 versus 7 percent) and significantly less likely than control teachers to report having a mentor who was another teacher (12 versus 27 percent).

3. Meetings with Mentors

Table VI.3 presents estimates of treatment-control differences in mentor meetings and activities in fall 2005 and fall 2006. Taking usual scheduled time and informal time during the most recent full week of teaching together, treatment teachers spent an average of 124 minutes in mentor meetings compared to 81 minutes for control teachers in fall 2005. The statistically significant 43-minute difference is attributable primarily to disparities in the duration of the usual scheduled meetings (79 versus 43 minutes). Treatment teachers also reported spending significantly more time meeting with full-time mentors than did control teachers (75 versus 6 minutes) during the most recent week of teaching, but reported significantly less time than control teachers with mentors who were also teachers (39 versus 70 minutes).

In fall 2006, during the second year of comprehensive induction services, taking usual scheduled time and informal time during the most recent full week of teaching together, on average, treatment teachers spent significantly more time in mentor meetings than control teachers (82 versus 48 minutes), mostly attributable to spending more time in scheduled meetings with mentors (55 versus 30 minutes). Treatment teachers also reported spending significantly more time with full-time mentors (59 versus 2 minutes) and significantly less time with those who were teachers (14 versus 42 minutes). Estimates of the treatment-control difference in time spent with mentors are shown separately by district in Appendix B, Figure B.5.

In both fall 2005 and fall 2006, we cannot determine precisely whether treatment teachers met with their study mentors for two hours per week as specified by the ETS and NTC program models. This is because total meeting time is not reported directly but must be constructed from reports of the frequency and duration of usual scheduled meetings and the time spent in informal meetings. The reported meeting time includes all mentors, which may capture time spent with mentors that were not part of the experimental intervention. Thus 124 minutes (Year 1) and 82 minutes (Year 2) represent upper bound estimates of time that treatment teachers spent with mentors assigned through the ETS or NTC programs.

Table VI.2. Impacts on Teacher-Reported Mentor Profiles (Percentages): 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								
Full-time mentor	71.5	15.8	55.7*	0.000	63.6	6.5	57.1*	0.000
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
Unweighted Sample Size (Teachers)	213	182	395		191	169	360	

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

Notes: 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 .05 level, two-tailed test.

Table VI.3. Impacts on Teacher-Reported Mentor Services Received in Most Recent Full Week of Teaching: Two-Year Districts

Mentor Service	Fall 2005 Treatment	Fall 2005 Control	Fall 2005 Difference	Fall 2005 Effect Size[a]	Fall 2005 P-value	Fall 2006 Treatment	Fall 2006 Control	Fall 2006 Difference	Fall 2006 Effect Size[b]	Fall 2006 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[c] 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 (Percentage)										
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 something BT requested[d]	77.9	50.0	27.9*	n.a.	0.000	59.7	23.0	36.7*	n.a.	0.000
Unweighted Sample Size (Teachers)	213	182	395			191	169	360		

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

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]BT = beginning teacher.

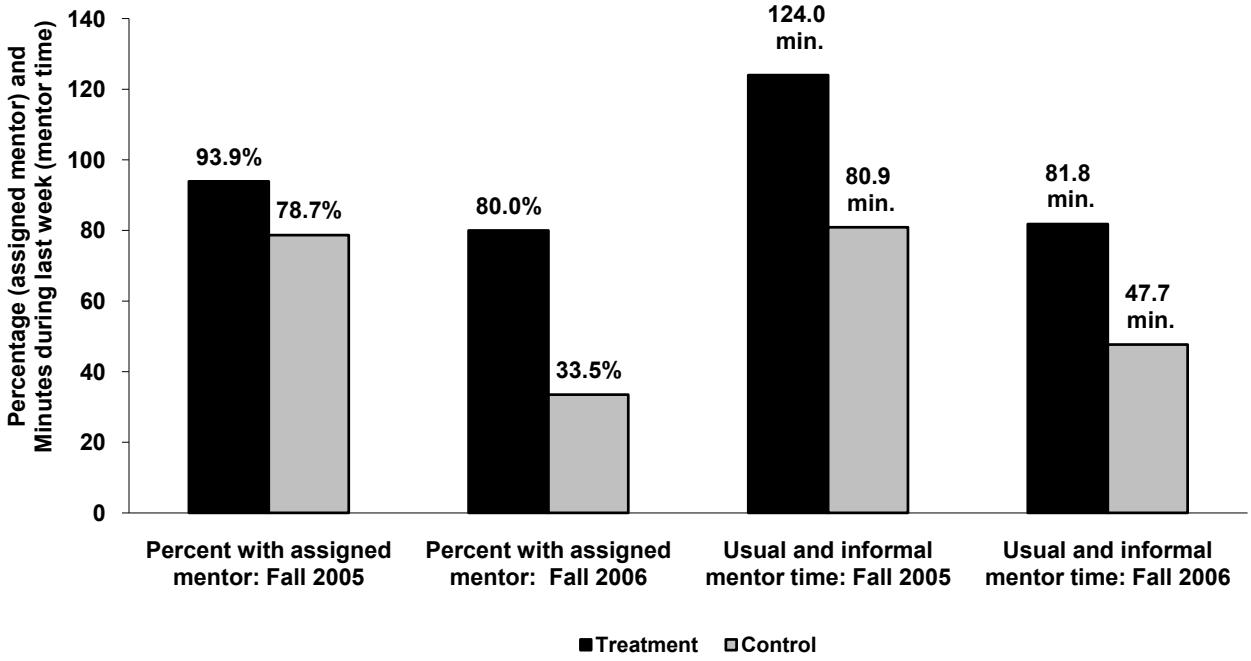
[d]Total sample size is 315 in fall 2005; 313 in fall 2006. The question did not apply to teachers who did not make a request to their mentors.

*Significantly different from zero at the .05 level, two-tailed test.

n.a. = not applicable.

Figure VI.1 shows treatment-control differences for having an assigned mentor and time in mentor meetings in Year 1 and Year 2. The declines in these two key measures of services from Year 1 to Year 2 are statistically significant for both treatment and control teachers.[44] However, while the usual scheduled and informal time that treatment teachers spent with all mentors showed a statistically significant decline, the time they spent with their study mentors did not show a statistically significant decline. (Treatment teachers' time with study mentors was 77 minutes per week in Year 1 and 65 minutes per week in Year 2; the p-value of the difference is 0.177.) This indicates that the decline in mentor time is due to a decline in time spent with non-study mentors.

Figure VI.1. Treatment-Control Differences in Percent Assigned a Mentor and Total Minutes Spent in Mentoring Per Week: Two-Year Districts, Fall 2005 and Fall 2006



Note: All treatment-control differences are significantly different from zero at the 0.05 level, two-tailed test (N=395 teachers in fall 2005 and 360 teachers in fall 2006).

[44] The declines in the percentages of treatment and control teachers with an assigned mentor are both statistically significant with p-values of 0.000. The decline in minutes spent with mentors is statistically significant with a p-value of 0.001 for treatment teachers and 0.027 for control teachers.

4. Mentor Activities and Assistance

In addition to spending more time meeting with mentors during Year 1, Table VI.3 shows that treatment teachers reported spending significantly more time than control teachers in specific types of mentoring activities during the most recent full week of teaching in fall 2005. These activities included being observed by mentors (38 versus 17 minutes), meeting one-on-one with mentors (43 versus 23 minutes), meeting together with mentors and other first-year teachers (38 versus 11 minutes), and having mentors model lessons (16 versus 10 minutes). The total time spent in the six types of activities surveyed in fall 2005 averaged 170 minutes per week for treatment teachers and 87 minutes per week for control teachers, a significant difference of 83 minutes per week.

Treatment teachers in Year 2 continued to report spending significantly more time being observed by mentors than control teachers (22 versus 7 minutes), meeting one-on-one with mentors (25 versus 12 minutes), meeting together with mentors and other first-year teachers (25 versus 6 minutes), and having mentors model lessons (12 versus 5 minutes) during the most recent full week of teaching in fall 2006. Treatment teachers averaged more total time than control teachers in the six types of activities surveyed (106 minutes versus 44 minutes per week).

In Year 1, treatment teachers were significantly more likely than control teachers to report receipt of a wide range of mentor assistance. The bottom panel of Table VI.3 shows that, during the most recent full week of teaching in fall 2005, treatment teachers were significantly more likely than control teachers to report receiving mentors' assistance in all 10 topic areas surveyed by 14 to 28 percentage points, with significant differences above 20 percentage points on having opportunities to raise issues and discuss concerns (90 versus 69 percent), receiving help on identifying teaching challenges and solutions (82 versus 58 percent), discussing instructional goals (75 versus 48 percent), and receiving help that the beginning teachers requested (78 versus 50 percent). Among treatment teachers, the percentage reporting each type of assistance ranged from 66 percent on receiving guidance on how to assess students to 92 percent on receiving encouragement or moral support. Among control teachers, the percentage reporting each type of assistance ranged from 48 percent receiving guidance on how to assess students to 73 percent receiving encouragement or moral support.

In Year 2, treatment teachers were still significantly more likely than control teachers to report receiving mentors' assistance in each of the topic areas surveyed by 28 to 44 percentage points. Significant differences above 35 percentage points are found for: receiving suggestions to improve practice (62 versus 23 percent), receiving encouragement or moral support (72 versus 30 percent), having opportunities to raise issues or discuss concerns (72 versus 28 percent), and receiving help with administrative/logistical issues (63 versus 24 percent). Among treatment teachers, the percentage reporting each type of assistance ranged from 50 percent receiving guidance on how to assess students to 72 percent receiving encouragement or moral support. Among control teachers, the percentage reporting each type of assistance ranged from 21 percent receiving guidance on how to assess students to 30 percent receiving encouragement or moral support.

5. Professional Development

Table VI.4 presents estimates of treatment-control differences in professional development activities in fall 2005 and fall 2006. During the three months prior to the fall 2005 survey, treatment teachers were significantly more likely than control teachers to report working with study groups of new teachers (67 versus 24 percent) and being observed by mentors (3.4 versus 2.1 times). During the three months prior to the fall 2006 survey, treatment teachers were significantly more likely than control teachers to report working with a study group of new teachers (42 versus 19 percent) and working with a study group of new and experienced teachers (54 versus 40 percent). Treatment teachers were also significantly more likely than control teachers to be observed by mentors (2.3 versus 0.8 times) and receive feedback on teaching not as part of a formal evaluation (1.9 versus 1.5 times).

Nearly all teachers reported having been offered professional development services in fall 2005 (98.6 percent) and fall 2006 (96.1 percent); differences between treatment and control teachers were not statistically significant (p-values 0.523 and 0.341, respectively). Table VI.5 presents estimates of treatment-control differences in teachers' attendance at professional development activities. Treatment and control teachers did not differ significantly in their reported attendance in professional development, except that treatment teachers were significantly more likely than control teachers to report having attended sessions focused on classroom management techniques (61 versus 48 percent) in fall 2005.

B. IMPACT FINDINGS: 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 both fall 2005 and fall 2006. Out of the six differences examined (three measures at two points in time), none were statistically significant (Table VI.6).[45, 46]

As a sensitivity analysis, we recoded the teacher satisfaction data into two categories and examined individual survey items separately. There were two statistically significant differences with regard to teachers' reports of satisfaction out the 76 tests conducted for fall 2005, spring 2006, fall 2006, or spring 2007: treatment teachers reported feeling more satisfied than control teachers with opportunities for professional development in fall 2006 and spring 2007.[47] See Appendix D (Tables D.7-D.8) for details.

[45] The spring 2006 and spring 2007 impact analysis is presented in Table D.6 in Appendix D. We reached the same general conclusions of no statistically significant positive impacts of treatment on teacher satisfaction in spring 2006 and spring 2007.

[46] A similar analysis of teachers' feelings of preparedness is conducted in Appendix E. It shows that there were no statistically significant impacts of treatment on teacher preparedness in spring 2006 or spring 2007.

[47] See Chapter II for a discussion of multiple comparisons and false discoveries that is relevant to the interpretation of these findings.

Table VI.4. Impacts on Teacher-Reported Professional Development Activities During Past Three Months: 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 3 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
Unweighted Sample Size (Teachers)	213	182	395			191	169	360		

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

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.

PD = professional development.

[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 .05 level, two-tailed test.

n.a. = not applicable.

Table VI.5. Impacts on Teacher-Reported Areas of Professional Development During Past Three Months (Percentages): 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
Unweighted Sample Size (Teachers)	213	182	395		191	169	360	

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

Notes: 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, two-tailed test.

Table VI.6. Impacts on Teacher Satisfaction (Scores on a Four-Point Scale): Two-Year Districts

	Fall 2005					Fall 2006				
	Treatment	Control	Difference	Effect Size	P-value	Treatment	Control	Difference	Effect Size	P-value
Feel Satisfied with:										
School	3.1	3.1	0.0	0.0	0.908	3.1	3.2	0.0	0.0	0.793
Class	3.1	3.1	0.0	0.0	0.895	3.2	3.1	0.1	0.1	0.280
Teaching career	3.0	3.1	-0.1	-0.2	0.127	3.0	3.0	0.0	0.0	0.999
Unweighted Sample Size (Teachers)	213	182	395			191	169	360		

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

Notes: 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, two-tailed test.

C. IMPACT FINDINGS: STUDENT TEST SCORES

The benchmark impacts on math and reading scores in the second year of the study were not significantly different from zero (Table VI.7). The same finding holds when we focus on tests that were administered as part of state accountability systems, those in grades three and above (detailed results shown in Appendix D, Tables D.9 and D.11).

Table VI.7. Impacts on Test Scores: Two-Year Districts, 2006-2007 School Year

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Subject	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
Reading	0.00	0.00	0.00	0.00	0.967	1,732	100	7
Math	-0.03	-0.01	-0.02	-0.02	0.746	1,736	99	7

Source: MPR analysis of data from 2005-2006 and 2006-2007 school years provided by participating school districts.

Notes: Data are regression-adjusted to account for pretest, district-by-grade fixed effects, and clustering of students within schools. Treatment and control group sample sizes are shown in Appendix Table D.13.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

As discussed in Chapter V, another way to analyze these data is to consider how test score impacts in Year 2 may have differed from Year 1. Using the same approach as with one-year districts, we conducted a formal test of the difference in treatment effects from Year 1 and Year 2 for the subsample of teachers who had students with valid test score data in both years. We found that the estimated changes in impacts on reading and math for the common sample of teachers were statistically insignificant (Table VI.8).

Finally, we re-estimated the Year 2 impacts using different samples, different sets of covariates, and different estimation techniques. The estimated impacts on reading and math in the second year under these alternative models were not statistically significant. See Appendix D (Tables D.9 to D.12) for details. Figure B.6 in Appendix B shows estimates of impacts on reading scores separately by district and Figure B.7 shows these estimates for math scores.

D. IMPACT FINDINGS: TEACHER RETENTION

After the two study years, 64 percent of study teachers had returned to the same schools (see Table VI.9). Another 8 percent had changed schools since fall 2005 but remained in the same district. An additional 16 percent stayed in teaching but changed districts or left the public sector. The remaining 11 percent of teachers had left teaching altogether. The regression-adjusted district retention rate was 72 percent and the total retention rate in teaching (including movers) was 89 percent.

No impacts of treatment were found on this pattern of teacher mobility after two years. The control group's mobility pattern was statistically indistinguishable from that of the treatment group. Table VI.9 shows the result of the three hypothesis tests specifically focused on retention in the school, in the district, and in the profession as binary outcomes. For each of the outcomes, there was no statistically significant impact.

Table VI.8. Impacts on Test Scores, Year 1 and Year 2 Common Sample: Two-Year Districts

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Subject	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
Reading								
Year 1	0.08	0.08	0.00	0.00	0.957	1,280	76	6
Year 2	-0.02	0.03	-0.05	-0.05	0.478	1,344	76	6
Year 2 – Year 1	-0.10	-0.05	-0.05	-0.05	0.521			
Math								
Year 1	0.02	0.17	-0.15*	-0.15	0.041	1,241	74	6
Year 2	-0.02	0.03	-0.05	-0.05	0.467	1,323	74	6
Year 2 – Year 1	-0.04	-0.13	0.10	0.10	0.292			

Source: MPR analysis of data from 2004–2005, 2005–2006, and 2006–2007 school years provided by participating school districts.

Notes: Data are regression-adjusted to account for pretest, district-by-grade fixed effects, and clustering of students within schools. Treatment and control group sample sizes are shown in Appendix Table D.14.

The common sample is the subsample of teachers who had students with valid test score data in Year 1 and Year 2.

*Significantly different from zero at the 0.05 level, two-tailed test.

Table VI.9. Impacts on Teacher Retention Rates after Two Years (Percentages): Two-Year Districts

Outcome	All Teachers	Treatment	Control	Difference	P-value
Retained in the same school	64.1	62.2	66.2	-4.0	0.386
Retained in the same district	72.3	69.6	75.3	-5.7	0.208
Retained in the teaching profession	88.8	86.9	90.8	-3.9	0.241
Unweighted Sample Size (Teachers)	364	203	161		
Unweighted Sample Size (Schools)	151	81	70		

Source: MPR Second Mobility Survey administered in 2007-2008 and Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: 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, two-tailed test.

We also examined movers' and leavers' self-reported reasons for leaving their schools and found no statistically significant impacts of treatment. To protect respondent confidentiality, we do not present the reasons for moving out of one's original school. The reasons for leaving are not presented because there were too few cases to draw meaningful inferences. When we asked leavers whether they expected to return and, if so, when they would do so, we did not find evidence of a treatment-control difference. The two possible reasons for a lack of statistical significance are that the sample size is too small to detect a relationship – about 12 percent of the sample members were leavers and 18 percent were movers – or that there may in fact be no relationship between comprehensive induction and the reasons for moving or leaving.

The reasons for moving provide some insight into the problem that teacher induction was meant to address. Dissatisfaction with administrative support was the most commonly cited single reason for treatment group movers (16 percent) and involuntary transfer was most commonly cited by control group teachers (28 percent), although there were a variety of reasons given by teachers in both groups.

The treatment did not result in the retention, after Year 2, of teachers who had produced higher Year 1 test scores than control teachers. In other words, we did not find evidence for a beneficial composition effect. We used Year 1 test scores to estimate composition effects because Year 2 scores already include the effects of learning on the job and the possible effects of the second year of induction services on the quality of teaching. The observed differences between test scores of treatment and control stayers were not statistically significant. Table VI.10 presents the impacts on Year 1 student achievement outcomes for those who returned to teach in the same district for the 2007–2008 school year.

Table VI.11 shows the background characteristics of teachers by mobility status. We also looked at certification (regular or probationary), highest degree earned, and whether the teacher was a career changer, but do not present these tabulations to protect respondent confidentiality. Across a wide variety of characteristics, no differences were found between the treatment and control group stayers nor were there significant treatment-control differences between movers or leavers, suggesting that comprehensive teacher induction did not induce a change in the mix of teachers who remained in the districts under study.

We examined the robustness of the teacher retention findings with respect to different sample inclusion/exclusion criteria, different definitions of mobility, and different modeling assumptions and, in each case, reached the same conclusion. In addition, Figure B.8 in Appendix B shows estimates of impacts on teacher retention separately by district.

Finally, we considered nonresponse to the mobility survey. Though the overall response rate was 85 percent, the response rates for treatment and control groups differed (90 and 80 percent, respectively). If nonrespondents differed from respondents in characteristics related to outcomes, then differential nonresponse could bias the impact estimates. To test this, we re-estimated impacts under alternate assumptions about nonrespondents, and found no impacts of treatment except under the most extreme and implausible assumptions. See Appendix D (Table D.19) for details.

Table VI.10. Impacts on Test Scores, District Stayers Only: Two-Year Districts, 2005-2006 School Year

Outcome	Treatment	Control	Difference	Effect Size	P-value
Reading scores (all grades)	0.03	-0.03	0.06	0.06	0.591
Unweighted Sample Size (Students)	745	558	1,303		
Unweighted Sample Size (Teachers)	45	30	75		
Unweighted Sample Size (Schools)	31	24	55		
Math scores (all grades)	-0.04	0.07	-0.11	-0.11	0.162
Unweighted Sample Size (Students)	693	549	1,242		
Unweighted Sample Size (Teachers)	43	30	73		
Unweighted Sample Size (Schools)	29	24	53		

Source: MPR analysis of data from 2004-2005 and 2005-2006 school years provided by participating school districts; MPR Second Mobility Survey administered in 2007-2008 to all study teachers.

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

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Table VI.11. Characteristics of District Stayers, Movers, and Leavers After Two Years by Treatment Status (Percentages Except Where Noted): Two-Year Districts

	Treatment			Control			Difference		
Teacher Characteristic	Stayers	Movers	Leavers	Stayers	Movers	Leavers	Stayers	Movers	Leavers
College entrance exam scores (SAT combined score or equivalent)	916	1,006	1,095	967	1,040	1,081	51	34	14
Attended highly selective college	23.4	28.6	59.9	25.1	37.1	52.4	-1.7	-8.5	7.5
Major or minor in education	67.0	70.9	38.9	66.6	70.8	74.7	0.4	0.0	-35.8
Student teaching experience (weeks)	12.2	14.1	6.2	11.9	11.7	9.3	0.3	2.4	-3.1
Entered the profession through traditional four-year program	61.5	76.8	25.2	66.0	61.3	56.1	-4.5	15.5	-30.9
Unweighted Sample Size (Teachers)	143	35	25	121	25	15			
Unweighted Sample Size (Schools)	71	28	20	62	21	13			

Source: MPR calculations using data from the College Board and ACT, Inc.; MPR Second Mobility Survey administered in 2007-2008; MPR First and Second Induction Activities Surveys administered in fall/winter 2005-2006 and spring 2006 to all study teachers.

Notes: 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 (143/35/25 treatment stayers/movers/leavers and 121/25/15 control stayers/movers/leavers) and schools (71/28/20 treatment and 62/21/13 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, two-tailed test. P-values are suppressed to make the table easier to read.

CHAPTER 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 (Chapters V and VI). We also showed, however, that the comprehensive induction services did not translate into a robust finding of positive impacts as hypothesized in the conceptual framework in Figure I.1. By the end of Year 2, there were no statistically significant positive impacts on teacher attitudes, retention, or test scores in one-year districts or two-year districts (Chapters V and VI).

This chapter attempts to answer a new set of questions raised by these findings. The overall question is: if there are no impacts associated with a particular increment in comprehensive induction services (the experimental contrast) might there still be a relationship between induction services more generally and outcomes? We report on correlational (nonexperimental) analyses of how variation in induction activities, both within and between treatment arms of the experiment, was related to student test scores and teacher retention. Test scores include math and reading test scores for the 2006-2007 school year in 16 districts[48] and teacher retention was measured in fall 2007, which would be the start of teachers' third year, in all 17 districts.

The results presented in this chapter should be interpreted with caution because the analyses are correlational and not causal. In particular, a nonexperimental estimate of the relationship 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 the principal's decision to help weak, struggling teachers who would likely have poor outcomes anyway.

[48] We could not use data for all initial districts because it was not possible to link student-level information to teacher-level information for one of the initial districts.

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

We analyzed a set of key measures of the induction services received by both treatment and control teachers. Three primary dimensions on which teacher induction programs can vary are the (1) breadth of services teachers receive, (2) instructional focus of the services, and (3) duration and intensity of services (Ingersoll and Kralik 2004).[49] This analysis focuses on induction supports that were considered important in the teacher induction literature (Portner 2005) and/or that ETS and NTC emphasized in their comprehensive induction programs (see Chapter IV).

The breadth of services received by the beginning teacher is measured by four indicator (yes or no) variables which inform whether the beginning teacher:

- Was assigned a mentor
- Met with a literacy or math coach in the prior three months
- Worked with a study group (with new or both new and experienced teachers) during the prior three months
- Observed others teaching during the prior three months

We used the indicator on 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. The mean and standard deviation of this variable vary, respectively, from 1.12 to 1.14 and 0.57 to 0.60, for the three samples analyzed in this chapter (sample used in the student math test scores analyses, sample used in the student reading test scores analyses, and sample used in the teacher retention analyses). The properties of this variable are presented in Table A.4 in Appendix A.

Using the other three measures of induction services, we constructed a new measure called the **Induction Services Index**, which was the sum of their values at three points in time: fall 2005, spring 2006, and fall 2006. Thus, the index is the sum of the values of 9 items and takes on values of 0 (never received any of the supports) to 9 (reported receiving all three supports at each of the three time points). For the samples used in the analyses of math test scores, reading test scores, and teacher retention, the mean of the index varies from 5.24 to 5.69, the standard deviation from 1.95 to 2.23, and the alpha from 0.39 to 0.54. The properties of the index are shown in Table A.4 in Appendix A.

[49] 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.

We also constructed an **Instructional Support Index** by examining another set of indicator variables at the same three time points. 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[50]
- Feedback on teaching, whether or not as part of a formal evaluation, during the prior three months

Because the question on subject area guidance (the second item above) was not included in the fall 2006 survey, the index is the sum of the 8 items and takes on values from 0 to 8 (not 9). The index can be interpreted as measuring the strength of instructional support received by beginning teachers. For the samples analyzed in this chapter, the mean of the index varies from 4.82 to 5.04, the standard deviation from 1.76 to 1.95, and the alpha from 0.34 to 0.64. The properties of the index are presented in Table A.4 in Appendix A.

For program duration and intensity, we constructed an **Induction Intensity Index** by averaging the number of hours per week[51] that beginning teachers reported spending in the following activities in the fall 2005, spring 2006, and fall 2006:

- Mentoring sessions (both scheduled and informal)
- Being observed teaching by 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

[50] 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 mobility.

[51] 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.

Since the questions on time spent on professional development activities were not included in the fall 2006 survey, the index includes 10 items and takes on values from 0 to 20.8. For the samples analyzed in this chapter, the mean of the index varies from 1.61 to 1.79, the standard deviation from 1.49 to 1.98, and the alpha from 0.29 to 0.48. The properties of the index are presented in Table A.4 in Appendix A.

The analyses use the same methods as the experimental analyses discussed in Chapter V, but instead of assignment to treatment status, which was randomly determined, the key explanatory variables are the (1) number of years the teacher had an assigned mentor, (2) Induction Services Index, (3) Instructional Support Index, and (4) Induction Intensity Index.[52] For each outcome, these four measures were included jointly in a regression model, along with the same set of covariates used in the corresponding experimental analyses. Including the four measures of induction services jointly, as opposed to including each measure individually in regression models without the other three measures, allows us to investigate how student achievement and teacher retention are associated with the multiple dimensions—the breadth of services received by beginning teachers, the extent of instructional support, and the duration and intensity of induction services—that characterize the induction services and support received by beginning teachers. To address concerns about multicollinearity, a problem described in more detail below, obscuring the effects of any individual measure, we also estimated regression models in which each induction services measure is included individually without the other three.

Thus, for the regression models in which the four measures are included jointly, if more induction services and more intense services are associated with better teacher and student outcomes, the induction measures should be positively related to each outcome. For each of the four measures, the reported coefficient represents the relationship between the outcome and the measure, holding all other measures equal. For instance, the coefficients on induction services, reported in Tables VII.1 and VII.2, measure the effect of receiving more services while leaving other information unchanged, including whether the teacher had an assigned mentor, how much time was spent being mentored, or the amount of instructional support the teacher received. See Appendix A for details of the statistical model. We conducted a number of sensitivity analyses using alternate constructions of the indices and specifications of the regression model.

B. NONEXPERIMENTAL RESULTS

The nonexperimental findings can be summarized as follows:

1. For student achievement, we found that one of the four measures of beginning teacher support was positively related to math scores and none were related to student achievement in reading. The four explanatory variables considered

[52] The results presented in this chapter should be interpreted with caution because of the reliability coefficients (ranging from 0.29 to 0.64) of the Induction Services Index, Instructional Support Index, and Induction Intensity Index.

collectively were not jointly related to student achievement in either subject (Table VII.1).

2. For teacher mobility, we found that one of the four explanatory variables was positively related to retention in the district, none were positively related to retention in the profession, and none were negatively related to either type of teacher retention. The four explanatory variables considered collectively were jointly related to teacher retention using both measures (Table VII.2).

Table VII.1. Association Between Beginning Teacher Support and Test Scores

	Math[a]		Reading[a]	
Induction Measure	Coefficient	P-value	Coefficient	P-value
Years BT had an assigned mentor	0.12*	0.015	0.00	0.971
Induction Services Index	-0.01	0.704	0.01	0.457
Instructional Support Index	0.02	0.334	0.01	0.297
Induction Intensity Index	-0.03	0.083	-0.01	0.448
Unweighted Sample Size (Districts)	16		16	
Unweighted Sample Size (Schools)	152		159	
Unweighted Sample Size (Teachers)	202		220	
Unweighted Sample Size (Students)	3,476		3,693	

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

Notes: BT = beginning teacher. The variable "years BT had an assigned mentor" has the following values: 0, 1, and 2 years. The Induction Services Index is the sum of the indicator variables at fall 2005, spring 2006, and fall 2006, on whether the beginning teacher: (1) met with a literacy or math coach, (2) met with a study group, and (3) observed others teaching (range: 0-9). The Instructional Support Index is constructed similarly using the indicator variables on whether the beginning teacher received: (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 (range 0-8). The Induction Intensity Index is the sum of 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.

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, two-tailed test.

[a]The following variables are not jointly significant: years BT had an assigned mentor, Induction Services Index, Instructional Support Index, and Induction Intensity Index (p-value = 0.063 for math, 0.542 for reading).

Table VII.2. Association Between Beginning Teacher Support and Teacher Mobility

	Remains in District[a]		Remains in Teaching[a]	
Induction Measure	Coefficient	P-value	Coefficient	P-value
Years BT had an assigned mentor	-0.04	0.166	0.00	0.624
Induction Services Index	0.02*	0.002	0.01*	0.003
Instructional Support Index	-0.00	0.956	0.00	0.822
Induction Intensity Index	0.01	0.424	0.00	0.439
Unweighted Sample Size (Teachers)	786		786	

Source: MPR Mobility Survey administered in 2007-2008; MPR Teacher Background Survey administered in 2005-2006; and First, Second, and Third Induction Activities Surveys administered in fall/winter 2005-2006, spring 2006, and fall/winter 2006-2007 to all study teachers.

Notes: BT = beginning teacher. The variable "years BT had an assigned mentor" has the following values: 0, 1, and 2 years. The Induction Services Index is the sum of the indicator variables at fall 2005, spring 2006, and fall 2006, on whether the beginning teacher: (1) met with a literacy or math coach, (2) met with a study group, and (3) observed others teaching (range: 0-9). The Instructional Support Index is constructed similarly using the indicator variables on whether the beginning teacher received: (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 (range 0-8). The Induction Intensity Index is the sum of 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.

Data are regression-adjusted using a logit model with robust standard errors to account for baseline characteristics and clustering of teachers within schools.

*Significantly different from zero at the 0.05 level, two-tailed test.

[a]The following variables are not jointly significant: years BT had an assigned mentor, Induction Services Index, Instructional Support Index, and Induction Intensity Index (p-value = 0.063 for math, 0.542 for reading).

1. Student Achievement

Overall, we found that induction measures were not significantly related to math test scores (p-value = 0.068) or reading scores (p-value = 0.651). These inferences are based on a test of whether the regression coefficients for the four induction measures are jointly equal to zero. The associations between each test score measure and each of the four individual induction measures are shown in Table VII.1. Each estimate in Table VII.1 is stated in terms of a standard unit of test scores. Because 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. For example, the regression coefficient suggests that students scored 12 percent of a standard deviation higher on the math test for each year the beginning teacher had a mentor. The coefficient on years the beginning teacher had a mentor is not statistically significant for reading test scores. The Induction Services Index, the Instructional Support

Index and the Induction Intensity index, were not significantly related to math or reading test scores.

In an earlier report from this study, we found that observing others teaching was negatively associated with some outcome measures using a similar correlational analysis (Glazerman et al. 2008), so we repeated the analyses omitting this measure from the Induction Services Index and found that the relationship between the years the beginning teacher had a mentor and math test scores remained statistically significant (regression coefficient = 0.12, p-value = 0.016). The relationship between each of the other three constructed induction indices and the math and reading tests scores remained statistically insignificant. The results from this analysis are presented in Table F.1 in Appendix F.

We were concerned that the similarity of the four induction services measures to each other would make it difficult to identify their overall effects, a problem known as multicollinearity. To address this concern, we estimated regression models in which each induction services measure is entered without the other three measures. Under this approach, the relationship between the years the beginning teacher had a mentor and math test scores remained statistically significant (regression coefficient = 0.09, p-value = 0.046). The associations between the other three induction services measures and math and reading test scores remained statistically insignificant. These results are presented in Table F.2 in Appendix F.

Another concern, raised at the beginning of this chapter, was that nonexperimental results reported here do not support a causal interpretation. In other words, even if induction services appear positively correlated with beneficial outcomes, it does not mean that the outcomes were caused by the services. In an attempt to address this concern, we conducted an instrumental variables (IV) analysis suggested by Angrist, Imbens and Rubin (1996). The approach exploits the fact that random assignment status helps explain the degree to which beginning teachers received support. If treatment-induced variation in service receipt is related to outcomes, then it would suggest that such services may indeed produce the outcomes in question.

The IV results were obtained in two stages. In the first stage, the indicator of treatment status[53] is used as the explanatory variable in regression models with each of the four indices of beginning teacher supports as outcome variables. The estimated coefficients obtained from those regressions are then used to calculate estimated values of these four indices. In the second stage, each of the estimated index variables is included individually (without the other three) in a regression model in which the outcome is the student math or reading test scores.

[53] Randomization to treatment or control status in the study was done at the school level. Thus, students taught by beginner teachers in treatment schools are considered treatment students, and students taught by beginner teachers in control schools are considered control students.

We found that with the IV approach, the associations between all the induction services measures and math and reading test scores were statistically insignificant. These results are presented in Table F.3 in Appendix F. If the assumption underlying the instrumental variable approach is valid, and there is a causal relationship, then we conclude from this analysis that the relationship is not strong enough to be detected with the available data.

2. Teacher Mobility

We rejected the hypothesis of no relationship between the induction activities variables and teacher retention. The p-value for this joint test was 0.016 for remaining in the district and 0.001 for remaining in teaching, meaning that there was an overall relationship. The estimates for each individual measure, shown in Table VII.2, are measured in changes in the estimated probability of a teacher remaining in the school district or the teaching profession after two years.[54] One measure—the Induction Services Index—was positively related and no measures were negatively related to teacher mobility for both remaining in the district and remaining in teaching. The estimate on the Induction Services Index for remaining in the district was 0.02 (p-value = 0.002); for remaining in teaching, it was 0.01 (p-value = 0.003). This implies that, for example, if the retention rate in a district were 80 percent, then an additional induction service, such as meeting with a study group in one semester, would be associated with a district retention rate of 82 percent, all else equal. The other variables—assignment to a mentor, the Instructional Support Index, and the Induction Intensity Index—were not significantly related to teacher retention.

As we did above, we repeated the analysis using an alternate Induction Services Index that omits the measure of observing others teaching and found that the association between the alternate index and the likelihood of remaining in the district is 0.03 and it is statistically significant (p-value = 0.000). The associations between the likelihood of remaining in the district and the other indices are not statistically significant (regression coefficients and p-values are -0.04 (p-value = 0.166), 0.00 (p-value = 0.988), and 0.01 (p-value = 0.412) for the years the beginning teacher had an assigned mentor, the Instructional Support Index, and the Induction Intensity Index, respectively). These results are presented in Table F.4 in Appendix F.

In order to avoid the problem of multicollinearity among the indices, we conducted separate analyses for each of the four induction services measures. We found that the association between the Induction Services Index and the likelihood of remaining in the district was 0.03 and it was statistically significant (p-value = 0.000) for a regression model in which the Induction Services Index is entered without the other indices. The associations between the likelihood of remaining in the district and the other induction services indices were not statistically significant for this specification of the model (regression coefficients and p-values are -0.01 (p-value = 0.600), 0.01 (p-value = 0.154), and 0.01 (p-value = 0.221)

[54] Similar to the experimental analysis, the retention effects are estimated using a logit model. The results presented are marginal effects predicted by the logit model with the covariates set at the mean values for the full sample.

for the years the beginning teacher had an assigned mentor, the Instructional Support Index, and the Induction Intensity Index, respectively). These results are presented in Table F.5 in Appendix F.

When we substituted a measure of whether the teacher remained in teaching as an alternate outcome measure, repeating the analysis using an alternate Induction Services Index that omits the measure of observing others teaching, we found that the coefficient on the alternate index remains statistically significant for teaching in the profession (regression coefficient = 0.01, p-value = 0.001). The associations between the likelihood of remaining teaching and the other indices are not statistically significant (regression coefficients and p-values are -0.00 (p-value = 0.557), 0.00 (p-value = 0.789), and 0.00 (p-value = 0.413) for the years the beginning teacher had an assigned mentor, the Instructional Support Index, and the Induction Intensity Index, respectively). These results are presented in Table F.4 in Appendix F.

As we did for the likelihood of remaining in the district, and to avoid the problem of multicollinearity among the indices, we conducted separate analyses for the associations of each of the four induction services measures and the likelihood of remaining teaching. We found that the association between each of the four induction services measures and the likelihood of remaining teaching was positive and statistically significant for a regression model in which each induction services measure is entered without the other three measures. The associations between the likelihood of remaining teaching and (1) the years the beginning teacher had an assigned mentor is 0.01 (p-value = 0.050); (2) the Induction Services Index is 0.01 (p-value = 0.000); (3) the Instructional Support Index is 0.01 (p-value = 0.004); and (4) the Induction Intensity Index is 0.01 (p-value = 0.030). These results are presented in Table F.5 in Appendix F.

As discussed above, we conducted an IV analysis with teacher mobility, using randomization status as the instrument, and again found that the IV analysis produced statistically insignificant estimates of the relationship. The results from this approach are presented in Table F.6 in Appendix F. If the assumption underlying the instrumental variable approach is valid, and there is a causal relationship, then we conclude from this analysis that the relationship is not strong enough to be detected with the available data.

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.

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.

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.

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, Institute of Education Sciences, National Center for Educational Evaluation and Regional Assistance, 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, Inc., September 2005.

Good, Roland H., and Ruth A. Kaminski (eds.) *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement, 2002.

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." *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. "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.

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.

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 [www.nber.org/papers/w13868]. 2008.

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.

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.

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.

White, H. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica,* vol. 48, 1980, pp. 817–830.

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.

APPENDIX A

ANALYTIC METHODS

This appendix provides technical details of the impact estimation method, analysis weights, and constructed variables used in the analysis.

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 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_{ij} 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 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 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 third 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 similar. Equations (A.4) and (A.5) express the basic student achievement model, with the unified model expressed by Equation (A.6):

$$Y_{ij,t} = c_j + \lambda Y_{ij,t-1} + \gamma' D_{ij} + e_{ij} \quad \text{(A.4)}$$

$$c_j = \mu + \delta T_j + u_j \quad \text{(A.5)}$$

$$Y_{ij,t} = \mu + \delta T_j + \lambda Y_{ij,t-1} + \gamma' D_{ij} + [u_j + e_{ij}] \quad \text{(A.6)}$$

Equation (A.6) differs from Equation (A.3) in two main ways. First, the *i* subscript models the student level rather than the teacher level. Second, Equation (A.6) is a value-added model in which student achievement at the end of the year $Y_{ij,t}$ (measured by the posttest) depends on student achievement at the beginning of the year $Y_{ij,t-1}$ (measured by the pretest) as well as random assignment to the treatment group T_j and a set of district-by-grade fixed effects D_{ij}. We substitute data for both math and reading test scores in place of Y_{ij}.

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.

Equations (A.3) and (A.6) can collectively be thought of as a mixed model or hierarchical model. They are "mixed" because they contain fixed effects (represented by μ, δ, β, γ, and λ) as well as random effects (represented by e and u). It is hierarchical because it embeds a school-level model (indexed by j) within an individual-level model (indexed by i). Several techniques are available for estimating such a model, including ordinary least squares (OLS) with robust standard errors (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. Unless noted otherwise, we report findings based on GLS estimates of 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, for the teacher retention analysis, we include school-level variables (the *Z* vector) from the Common Core of Data (CCD) of the National Center for Education Statistics.[55] For the student achievement analyses, districts provided student pretest scores and student demographic characteristics that could be included.

We used a separate set of covariates for each type of outcome we analyzed. Table A.1 presents the lists by analysis type. The benchmark analysis of teacher attitudes (Tables V.6, VI.6, and E.2) had district and grade fixed effects and no other covariates. The student achievement benchmark analyses (Tables V.7, V.8, VI.7, and VI.8) had normalized student pretest score and district-by-grade fixed effects. In the sensitivity analysis, we also include an *X* vector of student characteristics, teacher personal characteristics, and teacher professional characteristics. Finally, the benchmark teacher retention analysis (Tables V.9 and VI.9) included teacher personal characteristics, teacher professional characteristics, teacher neighborhood characteristics, school characteristics, and district and grade fixed effects.

Instrumental Variable Estimation to Correct for Measurement Error. As a specification check of the main student achievement results, we estimated a regression model using the method of instrumental variables to correct for measurement error in the pretest coefficient. 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, we included these results to account for the possibility that in the second year of teaching, principals may have assigned students to treatment and control teachers differently. We therefore estimate this system of equations:

$$Y_{ij,t-1} = \alpha + \theta W_{ij,t-1} + \varphi_1 T_j + \varphi_2 ' X_{ij} + \varphi_3 ' D_j + v_{ij} \quad \text{(A.7)}$$

$$Y_{ij,t} = \mu + \lambda \hat{Y}_{ij,t-1} + \delta T_j + \beta ' X_{ij} + \gamma ' D_j + e_{ij} \quad \text{(A.8)}$$

[55] 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
Teacher attitudes	V.6, VI.6, C.6, C.7, C.8, D.6, D.7, D.8, E.1, E.2	District fixed effects Grade fixed effects
Student achievement (benchmark model)	V.7, V.8, V.10, VI.7, VI.8, VI.10, C.9, C.10 (rows 1-2, 5-7), C.11, C.12 (rows 1-2, 5-7), D.9, D.10 (rows 1-2, 5-7), D.11, D.12 (rows 1-2, 5-7)	Normalized pretest score District-by-grade fixed effects
Student achievement (alternate model 1)	C.10 (row 3), C.12 (row 3), D.10 (row 3), D.12 (row 3)	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 (alternate model 2)	C.10 (row 4), C.12 (row 4), D.10 (row 4), D.12 (row 4)	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 • Gender • Race/ethnicity • Teacher race/ethnicity matches that of a majority of students Teacher professional characteristics: • Months of relevant teaching experience • Route into teaching • Certification status • 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 non-teaching job for five or more years. Normalized pretest score District-by-grade fixed effects
Student achievement (alternate model 3)	C.10 (row 6), C.12 (row 6), D.10 (row 6), D.12 (row 6)	District-by-grade fixed effects

Table A.1 *(continued)*

Analysis	Tables	Covariates included in the impact estimation model
Teacher mobility	V.9, VI.9	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 non-teaching 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

In this instrumental variables model, the pretest $Y_{ij,t-1}$ may be measured with error. Therefore, we ran a "first stage" regression model (A.7) in which we estimate the value of $Y_{ij,t-1}$ by regressing this variable on all of the other independent variables from the main equation (A.8) plus an instrumental variable, the opposite-subject pretest, $W_{ij,t-1}$ (that is, we use the math pretest as an instrument for the reading pretest and vice versa). In the main, or "second stage" regression model (A.8), $Y_{ij,t-1}$ is replaced by its predicted value, which is generated from equation (A.7) by setting the error term v_{ij} to zero. In equation (A.8), we use robust standard errors to account for correlation in outcomes for students clustered within schools. Instrumental variable results are reported in row 7 of Tables C.10, C.12, D.10, and D.12.

Difference-in-Differences Analysis of the Change in the Treatment Effect for Student Achievement from Year 1 to Year 2. To measure the improvement of treatment teachers relative to control teachers from Year 1 to Year 2, we employ a difference-in-differences estimator. That is, we compare the difference in student outcomes between

treatment and control teachers in Year 2 to the corresponding differences in Year 1. Because teachers typically teach the same grade in Year 1 and Year 2, the students taught by any given teacher will change between years. We pool data on all students taught by the common sample of teachers in the data from both years and estimate the following model:

$$Y_{ij,t} = \mu + \pi C_{ij} + \delta_1 T_j + \delta_2 (T_j * C_{ij}) + \gamma_1 ' Z_j + \gamma_2 '(Z_j * C_{ij}) + \lambda_1 Y_{ij,t-1} + \lambda_2 (Y_{ij,t-1} * C_{ij}) + [u_j + e_{ij}] \quad \text{(A.9)}$$

In this model, the student posttest is regressed on an indicator variable for cohort C_{ij} (that is, the Year 1 or Year 2 cohort of students), assignment to the treatment group T_j, the interaction of cohort and treatment status, district-by-grade fixed effects, the interaction of district-by-grade fixed effects with cohort, student pretest, and the interaction of the student pretest with the cohort.

Students in the Year 1 cohort are assigned weights in order to make the sum of the weights for a teacher equal across cohorts. For example, if a teacher has 20 students in cohort one and 10 students in cohort two, each student in cohort one will receive a weight of 0.5 so that the total weight for that teacher in cohort one is 10 (since 20*0.5 = 10). Conversely, for a teacher with 10 students in cohort one and 20 students in cohort two, each student in cohort one 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 between Year 1 and Year 2 of the treatment/control contrast in teacher effect on student test scores. We use robust standard errors to account for correlation in outcomes for students clustered within schools.

Nonexperimental Analysis. Chapter VII presents findings from nonexperimental 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. The result, presented in Equation (A.10), is an extension of the retention analysis. The student achievement model (not shown) is analogous.

$$Y_{ij} = \mu + \theta ' Q_{ij} + \beta ' X_{ij} + \gamma ' Z_j + [u_j + e_{ij}] \quad \text{(A.10)}$$

where Q_{ij}, representing a vector of indices of the level or intensity of induction services, replaces T, the indicator variable for assignment to the treatment group in Equation (A.3). Each coefficient in the θ vector captures the relationship between an induction index and the outcome Y. We estimated the relationships between the induction indices and the two main outcomes of interest—student achievement and teacher mobility—by substituting measures of the outcomes for Y_{ij}. The same vector of X and Z variables used in the experimental section is used here. The regressions are unweighted. If more induction services and more intense services are associated with better teacher and student outcomes, our measures of the level or intensity of

services provided should be positively related to each outcome. Psychometric properties of the indices we use are given in Table A.4.

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 such an outcome, 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 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.11):

$$WEIGHT_{ikm} \propto (1/\hat{p}_i) * \frac{1}{n_{km}} \frac{(n_{kT} + n_{kC})}{2}, \quad \text{(A.11)}$$

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. 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.2 and for teacher preparedness items in Table A.3.[56] The constructed scales for each of these categories exhibited good internal consistency (ranging from 0.73 to 0.92), as tested by the Cronbach's alpha coefficient. Psychometric properties for each scale are given in Table A.4.

2. Test Score Data

Aggregation of Test Scores across Grades, Subjects, and Districts. We observed considerable variation across districts and even across grades within some districts with respect to types of tests administered. Aggregating test scores across different tests posed a serious challenge for the analysis. In expectation 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, reducing the challenge from aggregating 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.

[56] The impact analysis for teacher preparedness data is presented in Appendix E.

Table A.2. 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: MPR First Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006.

Notes: Data pertain to teachers in all 17 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.

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. The resulting score can be interpreted as the distance from the average score as a fraction of a standard deviation; therefore, a z-score of –0.5, for example, means that the score was one-half of a standard deviation below the mean. We used the mean and standard deviation of the control group within each grade-district combination at each time point, thereby permitting us to interpret the z-scores as performance relative to that reference group.

Table A.3. 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: MPR First Induction Activities Survey administered to all study teachers in fall/winter 2005-2006.

Notes: Data pertain to teachers in all 17 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.

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.[57] 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 above the mean at the end of grade five. These would be considered equivalent, as both sets of students moved up one-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 present the findings separately for those two subjects. We dropped tests in early grades from three districts because they were scored by the school rather than by a test publisher. We excluded other subjects from the main impact analysis, such as foreign languages, social studies, or science, which are not available in

[57] Seattle and Cleveland are listed as hypothetical examples. They are not in the study.

enough districts to yield meaningful findings. Psychometric properties of the test score measures are given in Table A.4.

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 could have been exempt from testing, be missing a test score because of repeated absence, or not have been enrolled during the test period. These problems can result in a missing pretest or posttest score, each of which was required for the value-added analysis. Though we were better able to account for missing cases in some districts than in others, they appeared to be restricted to a small percentage of students and applied equally to the treatment and control groups. Because the difference in the percentages of students who had valid scores in treatment versus control schools was 4.3 percentage points for reading and 3.9 percentage points for math, we assumed that the data were missing at random.

Restrictions. Based on the data provided by school districts, we excluded some students from the model if it appeared implausible that the teacher linked to them was their full-time teacher for one or both subjects. We used four criteria. First, 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. Second, we excluded all students from a teacher who was linked to 40 or more students and indicated on a survey that she was responsible for both reading and math outcomes. Third, we excluded students of teachers who were linked to fewer than 7 students in a subject-grade combination unless at least 80 percent of the teacher's students were in special education. This restriction did not necessarily exclude all of a teacher's students if the teacher was linked to students in more than one subject-grade. Fourth, we excluded students of teachers who reported teaching small groups or a mixture of small groups and regular classes on the Teacher Background Survey, who taught at least 3 different grades, and who were linked to no more than 10 students in any one grade.

Because we relied on the pretest from the prior year, we excluded the youngest grade at which testing begins. For example, in districts that test in grades three through eight and operate elementary schools that include kindergarten through grade five (the most common case), we were able to estimate impacts on achievement for grades four and five. It is important to note, therefore, that the test score analysis pertains only to these tested grades and subjects. 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.4. 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
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
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
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
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
Preparedness to work with students								
Fall 2005	4	2.73	2.75	0.59	1.00	4.00	895	0.78
Fall 2005	4	2.84	2.75	0.61	1.00	4.00	876	0.83
Fall 2005	4	2.99	3.00	0.57	1.00	4.00	371	0.75
Student Achievement								
Reading Test Scores, 2007	1	0.00	0.03	1.00	-4.64	3.49	4,551	n/a
Math Test Scores, 2007	1	0.00	-0.01	1.00	-4.35	3.56	3,897	n/a
Induction Support								
Full sample of teachers								
Years BT had an assigned mentor	3	1.14	1.00	0.60	0.00	2.00	965	n/a
Induction Services Index	9	5.24	5.00	2.23	0.00	9.00	957	n/a
Fall 2005	3							0.47
Spring 2006	3							0.54
Fall 2006	3							0.54
Instructional Support Index	8	4.89	5.00	1.95	0.00	8.00	936	n/a
Fall 2005	3							0.61
Spring 2006	3							0.64
Fall 2006	2							0.43
Induction Intensity Index	10	1.61	1.30	1.49	0.00	20.81	877	n/a
Fall 2005	4							0.30
Spring 2006	4							0.30
Fall 2006	2							0.43

Table A.4 *(continued)*

Outcome	Number of Items	Mean	Median	SD	Minimum Value	Maximum Value	Sample Size	Cronbach's Alpha
Sample of teachers in student math test scores analyses								
Years BT had an assigned mentor	3	1.13	1.00	0.58	0.00	2.00	223	n/a
Induction Services Index	9	5.67	6.00	1.95	0.00	9.00	223	n/a
Fall 2005	3							0.49
Spring 2006	3							0.45
Fall 2006	3							0.40
Instructional Support Index	8	4.82	5.00	1.76	0.00	8.00	220	n/a
Fall 2005	3							0.55
Spring 2006	3							0.51
Fall 2006	2							0.34
Induction Intensity Index	10	1.79	1.41	1.98	0.00	20.81	211	n/a
Fall 2005	4							0.30
Spring 2006	4							0.29
Fall 2006	2							0.48
Sample of teachers in student reading test scores analyses								
Years BT had an assigned mentor	3	1.12	1.00	0.57	0.00	2.00	259	n/a
Induction Services Index	9	5.69	6.00	2.06	0.00	9.00	259	n/a
Fall 2005	3							0.51
Spring 2006	3							0.51
Fall 2006	3							0.39
Instructional Support Index	8	5.04	5.00	1.81	0.00	8.00	254	n/a
Fall 2005	3							0.59
Spring 2006	3							0.59
Fall 2006	2							0.36
Induction Intensity Index	10	1.77	1.41	1.88	0.00	20.81	241	n/a
Fall 2005	4							0.29
Spring 2006	4							0.30
Fall 2006	2							0.47

Source: MPR First, Second, Third, and Fourth Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006, spring 2006, fall/winter 2006-2007, and spring 2007; MPR analysis of data from 2006-2007 school year provided by participating school districts.

Note: Cronbach's alpha was calculated separately for variables within each time point for the Induction Services Index, Instructional Support Index, and Induction Intensity Index. The indices used in the correlational analyses are the sum of values from all three time points.

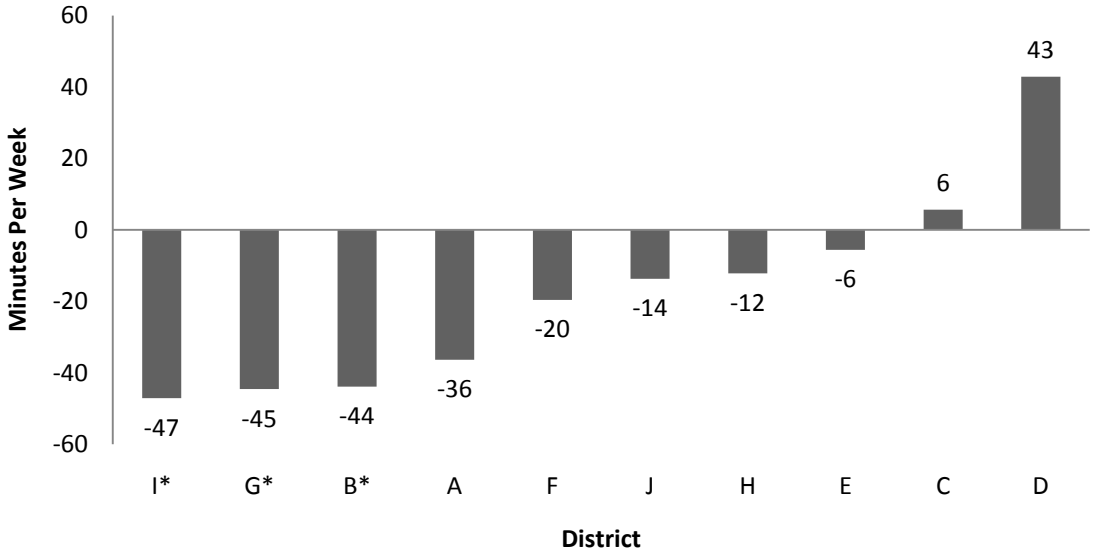
[a]BT = beginning teacher.

APPENDIX B

SUPPLEMENTARY FIGURES: IMPACTS BY DISTRICT

It is possible with an intervention like comprehensive teacher induction that the impacts may vary by school district because so much of what defines the counterfactual, that is, the induction support that teachers would have received the absence of intervention, is partly determined at the school district level. For example, the prevailing level of teacher induction supports is influenced by district policies and budgets. Also, each district (or state) has its own curriculum and assessment regime, which affects test score results and its own local labor market, which affects teacher mobility. The study was never designed to be able to detect impacts at the individual school district level, but nevertheless it is instructive to examine the *distribution* of impact estimates across districts. The remainder of this appendix shows figures with these distributions for a variety of intermediate and final outcomes. One goal of this analysis is to illustrate the degree of heterogeneity in the magnitude of the service contrast. The other is to show the extent to which the main study findings may be a reflection of any district outliers. All district identities have been masked by an arbitrary district identifier used in the figures.

Figure B.1. Impacts on Total Minutes Spent in Mentoring Per Week by District: One-Year Districts, Fall 2006

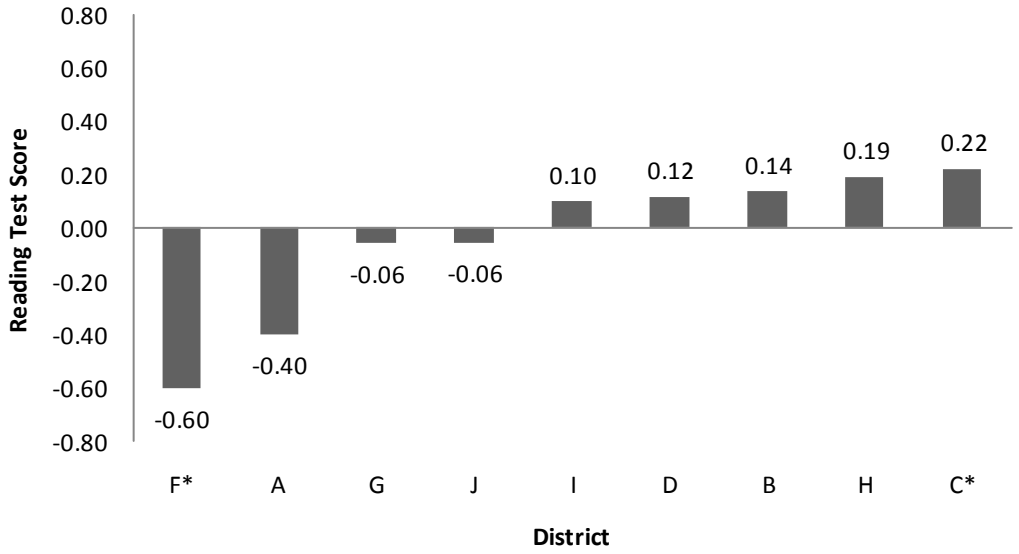


Source: MPR Third Induction Activities Survey administered in fall/winter 2006-2007 to all study teachers.

Notes: Vertical bars represent the regression-adjusted treatment group mean minus the regression-adjusted control group mean within each district. A negative impact estimate is shown as a bar that extends below the horizontal axis. District codes A through J are arbitrary.

*District-specific impact estimate is statistically significant at 0.05 level, two-tailed test. (No adjustment is applied for multiple comparisons.)

Figure B.2. Impacts on Reading Test Scores by District: One-Year Districts, 2006-2007 School Year



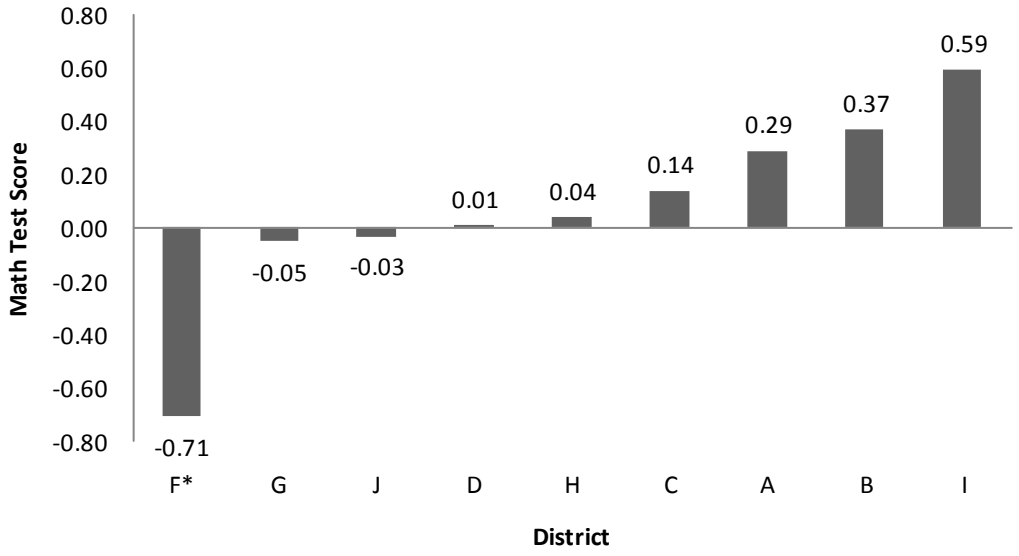
Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Notes: Vertical bars represent the regression-adjusted treatment group mean minus the regression-adjusted control group mean within each district. A negative impact estimate is shown as a bar that extends below the horizontal axis. District codes A through J are arbitrary.

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.

*District-specific impact estimate is statistically significant at 0.05 level, two-tailed test. (No adjustment is applied for multiple comparisons.)

Figure B.3. Impacts on Math Test Scores by District: One-Year Districts, 2006-2007 School Year



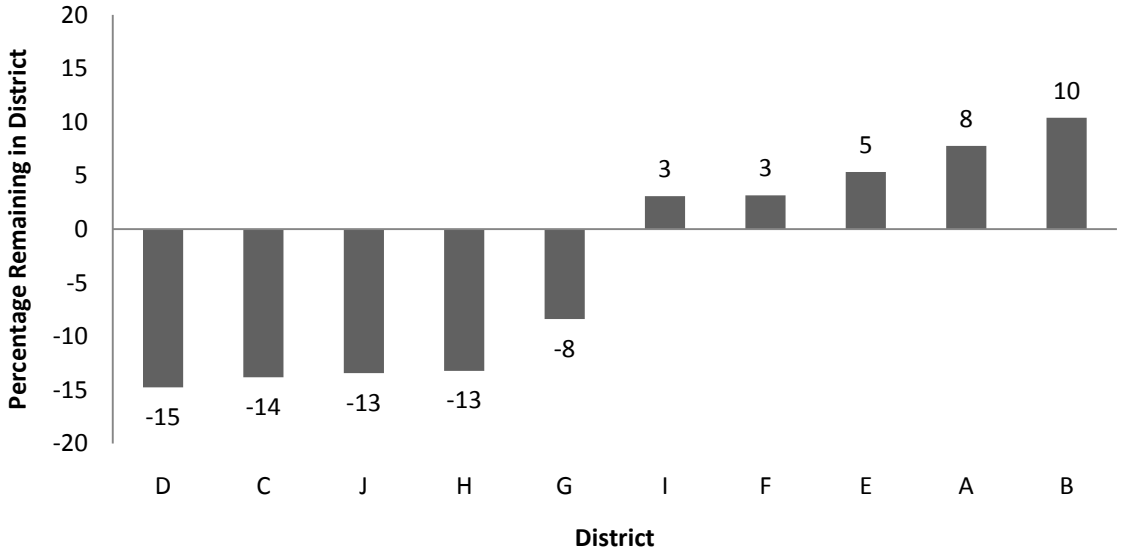
Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Notes: Vertical bars represent the regression-adjusted treatment group mean minus the regression-adjusted control group mean within each district. A negative impact estimate is shown as a bar that extends below the horizontal axis. District codes A through J are arbitrary.

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.

*District-specific impact estimate is statistically significant at 0.05 level, two-tailed test. (No adjustment is applied for multiple comparisons.)

Figure B.4. Impacts on Teacher Retention by District After Two Years: One-Year Districts

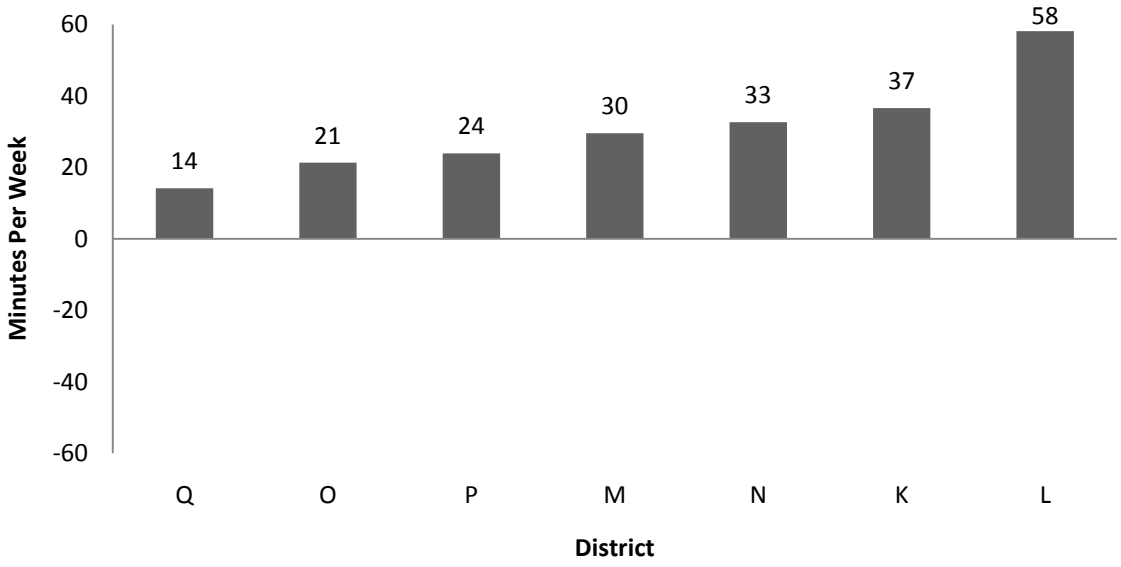


Source: MPR Second Mobility Survey administered in 2007-2008 and Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Vertical bars represent the regression-adjusted treatment group mean minus the regression-adjusted control group mean within each district. A negative impact estimate is shown as a bar that extends below the horizontal axis. District codes A through J are arbitrary.

None of the district-specific impact estimates are statistically significant at 0.05 level, two-tailed test.

Figure B.5. Impacts on Total Minutes Spent in Mentoring Per Week by District: Two-Year Districts, Fall 2006

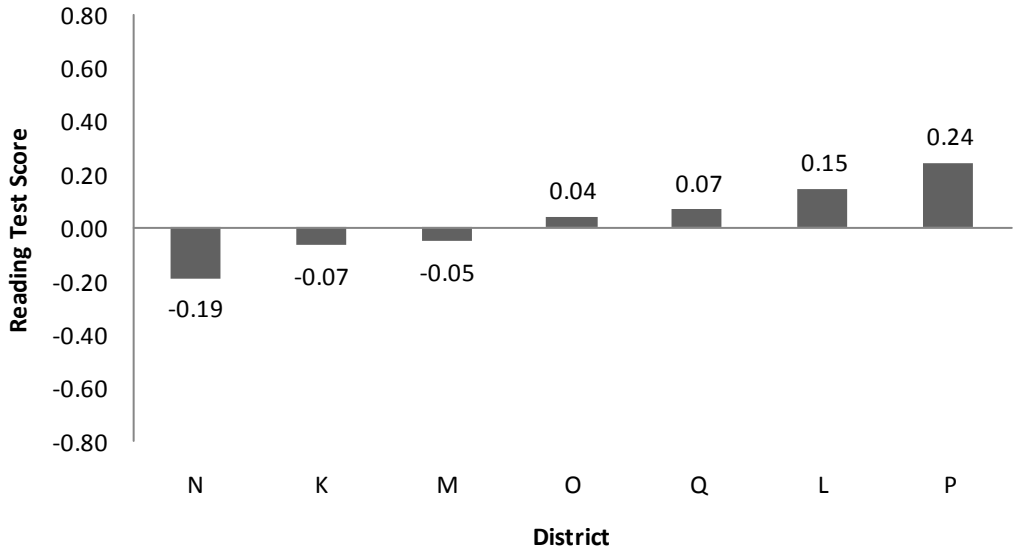


Source: MPR Third Induction Activities Survey administered in fall/winter 2006-2007 to all study teachers.

Notes: Vertical bars represent the regression-adjusted treatment group mean minus the regression-adjusted control group mean within each district. A negative impact estimate is shown as a bar that extends below the horizontal axis. District codes K through Q are arbitrary.

None of the district-specific impact estimates are statistically significant at 0.05 level, two-tailed test.

Figure B.6. Impacts on Reading Test Scores by District: Two-Year Districts, 2006-2007 School Year



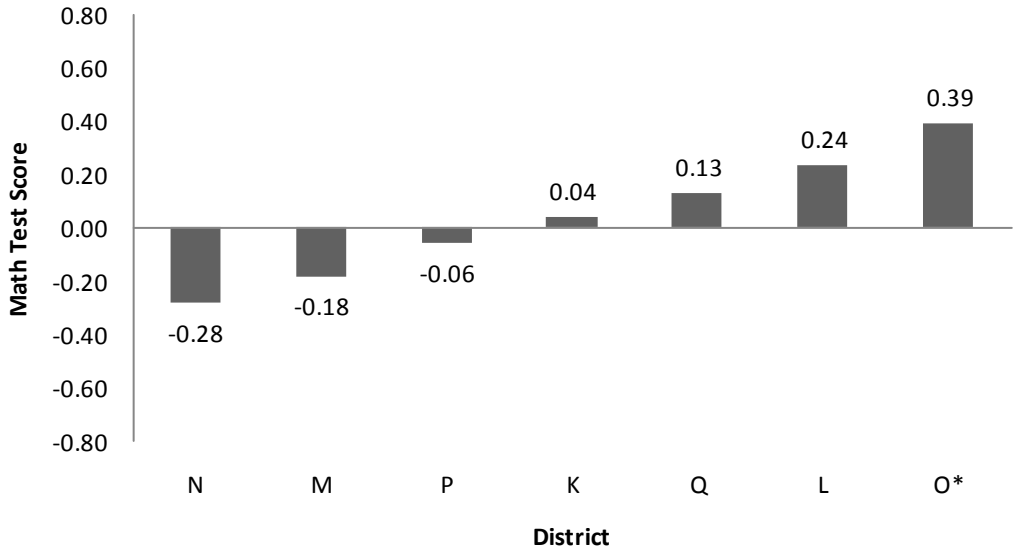
Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Notes: Vertical bars represent the regression-adjusted treatment group mean minus the regression-adjusted control group mean within each district. A negative impact estimate is shown as a bar that extends below the horizontal axis. District codes K through Q are arbitrary.

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.

None of the district-specific impact estimates are statistically significant at 0.05 level, two-tailed test.

Figure B.7. Impacts on Math Test Scores by District: Two-Year Districts, 2006-2007 School Year



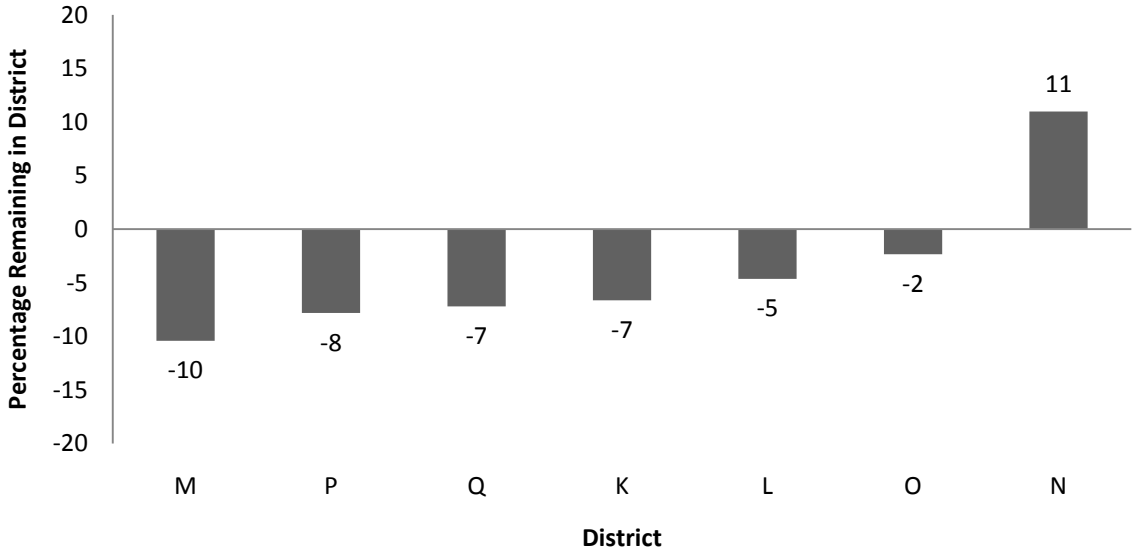
Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Notes: Vertical bars represent the regression-adjusted treatment group mean minus the regression-adjusted control group mean within each district. A negative impact estimate is shown as a bar that extends below the horizontal axis. District codes K through Q are arbitrary.

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.

*District-specific impact estimate is statistically significant at 0.05 level, two-tailed test. (No adjustment is applied for multiple comparisons.)

Figure B.8. Impacts on Teacher Retention by District After Two Years: Two-Year Districts



Source: MPR Second Mobility Survey administered in 2007-2008 and Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Vertical bars represent the regression-adjusted treatment group mean minus the regression-adjusted control group mean within each district. A negative impact estimate is shown as a bar that extends below the horizontal axis. District codes K through Q are arbitrary.

None of the district-specific impact estimates are statistically significant at 0.05 level, two-tailed test.

APPENDIX C

SENSITIVITY ANALYSES AND SUPPLEMENTAL TABLES FOR CHAPTER V

A. SUPPLEMENTARY TABLES FOR TEACHER INDUCTION SERVICES

The appendix tables present additional data with which to judge the impact of comprehensive teacher induction on service receipt. Tables C.1 to C.5 present results for teacher induction services for one-year districts based on the Second Induction Activities Survey administered in spring 2006. The corresponding tables in the main report (Tables V.1-V.5) present results based on the First Induction Activities Survey (administered in fall 2005) and the Third Induction Activities Survey (administered in fall 2006). The conclusions do not change when we examine data from spring 2006.

Table C.1. Teacher Reports on Professional Support and Duties (Percentages): One-Year Districts

	Spring 2006			
	Treatment	Control	Difference	P-value
BT[a] has a mentor	90.3	80.3	10.0*	0.001
BT has an assigned mentor	89.7	72.2	17.5*	0.000
Unweighted Sample Size (Teachers)	258	241	499	

Source: MPR Second Induction Activities Survey administered to all study teachers in spring 2006.

Notes: 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]BT = beginning teacher.

*Significantly different from zero at the .05 level, two-tailed test.

Table C.2. Impacts on Teacher-Reported Mentor Profiles (Percentages): 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
Unweighted Sample Size (Teachers)	258	241	499	

Source: MPR Second Induction Activities Survey administered to all study teachers in spring 2006.

Notes: 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 .05 level, two-tailed test.

Table C.3. Impacts on Teacher-Reported Mentor Services Received in Most Recent Full Week of Teaching: 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	28.8	33.7	-4.9	-0.12	0.197
Total time (minutes)					
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[c] 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 something BT requested[d]	61.2	46.3	14.8*	n.a.	0.002
Unweighted Sample Size (Teachers)	258	241	499		

Source: MPR Second Induction Activities Survey administered to all study teachers 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 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]BT = beginning teacher.

[d]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 .05 level, two-tailed test.

n.a. = not applicable.

Table C.4. Impacts on Teacher-Reported Professional Development Activities During Past Three Months: 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
Unweighted Sample Size (Teachers)	258	241	499		

Source: MPR Second Induction Activities Survey administered to all study teachers 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 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 .05 level, two-tailed test.

n.a. = not applicable.

Table C.5. Impacts on Teacher-Reported Areas of Professional Development During the Past Three Months (Percentages): 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
Unweighted Sample Size (Teachers)	258	241	499	

Source: MPR Second Induction Activities Survey administered to all study teachers in spring 2006.

Notes: 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, two-tailed test.

B. SUPPLEMENTARY TABLE AND SENSITIVITY ANALYSIS FOR TEACHER SATISFACTION

Table C.6 presents results for teacher satisfaction for one-year districts based on the Second Induction Activities Survey administered in spring 2006.[58] The corresponding table in the main report (Table V.6) presents results based on the First Induction Activities Survey (administered in fall 2005) and the Third Induction Activities Survey (administered in fall 2006).

Table C.6. Impacts on Teacher Satisfaction (Scores on a Four-Point Scale): One-Year Districts

	Spring 2006			
	Treatment	Control	Difference	P-value
Feel Satisfied with:				
School	3.0	3.0	0.0	0.927
Class	3.0	3.0	0.0	0.720
Teaching career	2.9	2.9	-0.1	0.201
Unweighted Sample Size (Teachers)	258	241	499	

Source: MPR Second Induction Activities Survey administered to all study teachers in spring 2006.

Notes: 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, two-tailed test.

One concern with the analysis of teacher satisfaction 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 this dichotomous variable. The results for one-year districts show no statistically significant differences with regard to teachers' reports of satisfaction in fall 2005 and fall 2006 (Table C.7) or spring 2006 (Table C.8).

[58] Teacher attitudes were not measured in one-year districts in spring 2007.

Table C.7. Impacts on Teacher Satisfaction (Percent "Somewhat Satisfied" or "Very Satisfied"): One-Year Districts, Fall 2005 and Fall 2006

	Fall 2005					Fall 2006				
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	74.1	76.3	-2.3	-0.05	0.572	71.4	74.6	-3.2	-0.07	0.408
Availability of resources and materials/equipment for your classroom	66.4	66.5	-0.1	0.00	0.985	67.1	66.3	0.8	0.02	0.858
Input into school policies and practices	68.6	72.2	-3.7	-0.08	0.395	71.4	72.2	-0.8	-0.02	0.838
Opportunities for professional development	84.9	84.5	0.4	0.01	0.915	81.1	79.8	1.3	0.03	0.675
Principal's leadership and vision	77.6	75.9	1.7	0.04	0.680	73.6	72.2	1.4	0.03	0.745
Professional caliber of colleagues	80.4	85.3	-4.9	-0.13	0.152	77.4	76.6	0.8	0.02	0.811
Supportive atmosphere among faculty/collaboration with colleagues	84.0	81.2	2.8	0.07	0.456	80.1	76.6	3.5	0.09	0.358
School facilities such as the building or grounds	76.5	73.5	3.0	0.07	0.491	70.3	72.6	-2.3	-0.05	0.565
School policies	79.9	78.4	1.6	0.04	0.667	82.3	75.4	6.8	0.17	0.054
Satisfaction with Students										
Autonomy or control over own classroom	86.8	86.5	0.3	0.01	0.929	81.7	82.9	-1.3	-0.03	0.662
Student motivation to learn	73.5	70.6	2.9	0.07	0.464	69.0	72.2	-3.2	-0.07	0.395
Student discipline and behavior	65.4	58.8	6.7	0.14	0.147	61.7	63.9	-2.2	-0.04	0.614
Parental involvement in the school	45.0	44.9	0.1	0.00	0.978	47.6	42.5	5.1	0.10	0.271
Grade assignment	89.2	87.8	1.4	0.04	0.619	86.6	85.7	0.9	0.03	0.683
Students assigned	83.9	84.9	-1.0	-0.03	0.758	81.6	79.4	2.2	0.06	0.433
Satisfaction with Teaching Career										
Salary and benefits	78.3	78.0	0.4	0.01	0.917	67.5	68.7	-1.2	-0.03	0.748
Professional prestige	82.2	83.3	-1.0	-0.03	0.766	75.8	73.4	2.3	0.05	0.542
Intellectual challenge	85.7	89.8	-4.1	-0.13	0.160	85.3	84.1	1.2	0.03	0.631
Workload	52.2	55.9	-3.7	-0.07	0.422	51.6	54.0	-2.4	-0.05	0.595
Unweighted Sample Size (Teachers)	258	245	503			241	231	472		

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

Notes: 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, two-tailed test.

Table C.8. Impacts on Teacher Satisfaction (Percent "Somewhat Satisfied" or "Very Satisfied"): One-Year Districts, Spring 2006

	Spring 2006				
Area of Satisfaction	Treatment	Control	Difference	Effect Size	P-value
Satisfaction with School					
Administration support for beginning teachers	69.8	65.2	4.6	0.10	0.280
Availability of resources and materials/equipment for your classroom	65.1	63.9	1.2	0.02	0.791
Input into school policies and practices	66.2	67.6	-1.4	-0.03	0.737
Opportunities for professional development	81.2	78.8	2.3	0.06	0.520
Principal's leadership and vision	67.7	66.8	0.9	0.02	0.845
Professional caliber of colleagues	78.7	80.9	-2.2	-0.06	0.567
Supportive atmosphere among faculty/ collaboration with colleagues	80.2	79.3	1.0	0.02	0.809
School facilities such as the building or grounds	70.3	71.8	-1.4	-0.03	0.729
School policies	73.6	73.0	0.5	0.01	0.903
Satisfaction with Students					
Autonomy or control over own classroom	87.8	87.6	0.3	0.01	0.923
Student motivation to learn	70.1	66.8	3.3	0.07	0.493
Student discipline and behavior	57.9	52.7	5.2	0.10	0.302
Parental involvement in the school	41.0	36.1	4.9	0.10	0.331
Grade assignment	88.8	90.9	-2.0	-0.07	0.464
Students assigned	81.3	83.8	-2.5	-0.07	0.477
Satisfaction with Teaching Career					
Salary and benefits	71.8	76.8	-4.9	-0.11	0.198
Professional prestige	76.9	74.3	2.6	0.06	0.554
Intellectual challenge	86.1	89.2	-3.1	-0.09	0.299
Workload	54.4	56.8	-2.4	-0.05	0.587
Unweighted Sample Size (Teachers)	258	241	499		

Source: MPR Second Induction Activities Survey administered to all study teachers in spring 2006.

Notes: 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, two-tailed test.

C. SUPPLEMENTARY TABLES AND SENSITIVITY ANALYSIS FOR STUDENT ACHIEVEMENT

We performed a sensitivity analysis for one-year districts by re-estimating the impacts on 2006-2007 test scores using different samples, sets of covariates, and estimation techniques:

- Disaggregating results by grade
- Omitting test scores from the lower grades and presenting results using only test scores from grades 3 and above
- Using the original value of outliers rather than truncating them to a range from –3 to +3
- Adding student demographic covariates as control variables (see Table A.1 in Appendix A for a list of control variables)
- Adding student demographic covariates and teacher covariates as control variables (see Table A.1 in Appendix A for a list of control variables)
- Estimating impacts using ordinary least squares with robust standard errors;
- Estimating impacts without controlling for a pretest
- Estimating impacts using the opposite-subject pretest as an instrumental variable to control for measurement error in the pretest

All of these alternate models showed statistically insignificant impacts for reading and math.

When the reading results are disaggregated by grade, impact estimates at the individual grade levels are not significantly different from zero; neither are they significant in a sample composed of data pooled from grades 3 and above (in the case of one-year districts, grades 3–5). Results are shown in Table C.9. 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. We present results with the sample restricted to students from grades 3–5 for two reasons. First, paper-and-pencil tests for older students may be more reliable than those given to younger students. Second, because grades 3 and above are subject to the federal No Child Left Behind Act, teachers in these grades may feel more pressure to raise test scores than teachers of grades 1 and 2. In this way, test scores may be a more accurate indicator of teacher quality for teachers in grades 3 and above. As Table C.9 indicates, however, the impact of comprehensive teacher induction on reading scores in these upper grades is not statistically significant.

Table C.9. Impacts on Reading Test Scores by Grade: One-Year Districts, 2006-2007 School Year

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
2	0.02	0.02	0.00	0.00	0.981	473	27	3
3	0.10	0.01	0.08	0.08	0.499	478	37	5
4	0.10	-0.01	0.11	0.11	0.254	873	51	8
5	-0.01	0.00	-0.01	-0.01	0.885	421	24	5
All Grades	0.05	0.01	0.04	0.04	0.380	2,245	135	9
Grades 3–5	0.06	0.01	0.06	0.06	0.269	1,772	109	9

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Notes: Data are regression-adjusted to account for pretest, district-by-grade fixed effects, and clustering of students within schools. Sample sizes for treatment and control groups are shown in Appendix Table C.15.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

A set of additional specification checks, shown in Table C.10, confirms that there is no statistically significant effect of treatment on reading in the second year of teaching. The top row of Table C.10 repeats the results of the benchmark analysis for reference. The second row uses the original data, without forcing outliers to have minimum values of -3 and maximum values of 3. This does not change the result. The third and fourth rows are estimated with additional covariates. For the model shown in the third row, student demographic covariates have been added but the impact estimate remains not significantly different from zero. This is also true when covariates for teacher characteristics are added as well. See Appendix Table A.1 for a list of student and teacher covariates used in these models.

Another set of models change the estimation strategy rather than include extra covariates, but none shows a statistically significant impact. The fifth row of Table C.10 shows results of a model that uses ordinary least squares rather than hierarchical linear modeling, and accounts for correlation of outcomes for students in the same school using robust standard errors. The result is unchanged.

The sixth row shows results from estimating impacts without controlling for a pretest. Some students in our sample were missing pretest data. We excluded from the main analysis any student with missing test scores. This decision risked excluding mobile students and students in lower grades in some districts, who could have experienced a different impact of treatment than the students with both a posttest and pretest. As shown on the sixth row of Table C.10, the impact is not significantly different from zero.

Table C.10. Impacts on Reading Test Scores, Alternate Model Specifications: One-Year Districts, 2006-2007 School Year

Model	Adjusted Mean Test Scores		Difference	Effect Size	P-value	Unweighted Sample Sizes		
	Treatment	Control				Students	Teachers	Districts
Benchmark	0.05	0.01	0.04	0.04	0.380	2,245	135	9
With outliers	0.05	0.01	0.04	0.04	0.474	2,245	135	9
Student covariates	0.05	0.02	0.03	0.03	0.498	2,245	135	9
Student, teacher covariates	0.07	0.00	0.07	0.07	0.201	2,245	135	9
Robust standard errors	0.06	0.00	0.06	0.06	0.195	2,245	135	9
No pretest	0.02	-0.02	0.04	0.04	0.468	3,213	169	9
Instrumental variables	0.06	0.00	0.05	0.05	0.272	1,946	122	9

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts; MPR Teacher Background Survey administered in 2005–2006 to all study teachers.

Notes: Data are regression-adjusted to account for district-by-grade fixed effects and clustering of students within schools. See Appendix Table A.1 for a list of other covariates used in these models. Sample sizes for treatment and control groups are shown in Appendix Table C.16.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

The seventh row shows results from a regression model in which the math pretest is used as an instrumental variable to control for measurement error in the reading pretest. (This also alters the sample since students who lack a math pretest are excluded.) In non-experimental 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 Year 2. For example, if principals believed that comprehensive teacher induction gave teachers in their second year 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. Using the instrumental variables model, however, did not change our findings.

For the math results for one-year districts, Table C.11 shows that grade-by-grade impacts are not statistically significant, nor is the aggregate impact using only data from grades 3–5. Using all grades for math tests, Table C.12 shows that the initial finding of no impact is robust to changes in covariates or specification. Table C.12 indicates no statistically significant impact, whether allowing outliers to have their original values (line 2), altering the

covariates (lines 3 and 4), using an ordinary least squares model with robust standard errors (line 5), excluding the pretest from the model and thereby expanding the sample size (line 6), or using an instrumental variables approach (line 7).

Table C.11. Impacts on Math Test Scores by Grade: One-Year Districts, 2006-2007 School Year

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
2	-0.09	0.12	-0.20	-0.20	0.231	332	20	2
3	0.04	0.02	0.03	0.03	0.804	327	24	4
4	0.06	-0.06	0.12	0.12	0.362	914	51	8
5	0.09	-0.08	0.17	0.17	0.135	422	24	5
All Grades	0.05	-0.02	0.08	0.08	0.367	1,995	117	9
Grades 3–5	0.09	-0.06	0.15	0.15	0.080	1,663	97	9

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Notes: Data are regression-adjusted to account for pretest, district-by-grade fixed effects, and clustering of students within schools. Sample sizes for treatment and control groups are shown in Appendix Table C.17.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Table C.12. Impacts on Math Test Scores, Alternate Model Specifications: One-Year Districts, 2006-2007 School Year

Model	Adjusted Mean Test Scores		Difference	Effect Size	P-value	Unweighted Sample Sizes		
	Treatment	Control				Students	Teachers	Districts
Benchmark	0.05	-0.02	0.08	0.08	0.367	1,995	117	9
With outliers	0.05	-0.03	0.08	0.08	0.419	1,995	117	9
Student covariates	0.05	-0.01	0.06	0.06	0.475	1,995	117	9
Student, teacher covariates	0.04	-0.04	0.08	0.08	0.345	1,995	117	9
Robust standard errors	0.04	-0.02	0.05	0.05	0.284	1,995	117	9
No pretest	-0.01	0.01	-0.01	-0.01	0.829	2,885	148	9
Instrumental variables	0.06	-0.02	0.08	0.08	0.143	1,966	117	9

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts; MPR Teacher Background Survey administered in 2005–2006 to all study teachers.

Notes: Data are regression-adjusted to account for district-by-grade fixed effects and clustering of students within schools. See Appendix Table A.1 for a list of other covariates used in these models. Sample sizes for treatment and control groups are shown in Appendix Table C.18.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

The remaining appendix tables for student achievement—Tables C.13 to C.18—show treatment and control sample sizes for models corresponding to Tables V.7-V.8, and C.9-C.12.

Table C.13. Treatment and Control Sample Sizes for Impacts on Test Scores (Benchmark Model): One-Year Districts, 2006-2007 School Year

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Subject	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Reading	1,193	72	52	9	1,052	63	47	9
Math	994	57	46	9	1,001	60	45	9

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Table C.14. Treatment and Control Sample Sizes for Impacts on Test Scores (Year 1 and Year 2 Common Sample): One-Year Districts, 2006-2007 School Year

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Subject: Year	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Reading: Year 1	870	45	38	7	649	37	29	7
Reading: Year 2	814	45	38	7	644	37	29	7
Math: Year 1	687	38	31	6	587	35	27	6
Math: Year 2	653	38	31	6	613	35	27	6

Source: MPR analysis of data from 2004–2005, 2005–2006, and 2006–2007 school years provided by participating school districts.

Table C.15. Treatment and Control Sample Sizes for Impacts on Reading Test Scores, by Grade Level: One-Year Districts, 2006-2007 School Year

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
2	220	12	12	3	253	15	15	3
3	320	23	17	5	158	14	13	5
4	456	26	25	8	417	25	24	8
5	197	11	10	5	224	13	11	5
All Grades	1,193	72	52	9	1,052	63	47	9
Grades 3–5	973	60	43	9	799	49	37	9

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Table C.16. Treatment and Control Sample Sizes for Impacts on Reading Test Scores, Alternate Model Specifications: One-Year Districts, 2006-2007 School Year

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Model	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Benchmark	1,193	72	52	9	1,052	63	47	9
With outliers	1,193	72	52	9	1,052	63	47	9
Student covariates	1,193	72	52	9	1,052	63	47	9
Student, teacher covariates	1,193	72	52	9	1,052	63	47	9
Robust standard errors	1,193	72	52	9	1,052	63	47	9
No pretest	1,739	90	57	9	1,474	79	55	9
Instrumental variables	981	60	49	9	965	62	46	9

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Table C.17. Treatment and Control Sample Sizes for Impacts on Math Test Scores by Grade Level: One-Year Districts, 2006-2007 School Year

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
2	156	9	9	2	176	11	11	2
3	175	12	12	4	152	12	11	4
4	465	25	24	8	449	26	25	8
5	198	11	10	5	224	13	11	5
All Grades	994	57	46	9	1,001	60	45	9
Grades 3–5	838	48	40	9	825	49	37	9

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Table C.18. Treatment and Control Sample Sizes for Impacts on Math Test Scores, Alternate Model Specifications: One-Year Districts, 2006-2007 School Year

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Model	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Benchmark	994	57	46	9	1,001	60	45	9
With outliers	994	57	46	9	1,001	60	45	9
Student covariates	994	57	46	9	1,001	60	45	9
Student, teacher covariates	994	57	46	9	1,001	60	45	9
Robust standard errors	994	57	46	9	1,001	60	45	9
No pretest	1,472	74	51	9	1,413	74	51	9
Instrumental variables	973	60	46	9	993	60	45	9

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

D. SENSITIVITY ANALYSIS FOR TEACHER RETENTION

For the teacher retention analysis using one-year districts, the conclusions did not change when we expanded the number of outcomes to differentiate between moving to a school in another public school district and moving to a private, parochial, or other school, and expanded the outcomes for leaving to include leaving to stay at home, leaving to attend school or take a new job, and other reasons for leaving. When we re-estimated the models using a linear probability model or a multinomial logit model, we reached the same conclusions as when we used a binary logit model.

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 (Table C.19).[59] 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.

The results did not change when we assumed that all nonrespondents were stayers or all were leavers. The only exceptions were the most extreme assumptions, in which we first assumed that *all* of the treatment group nonrespondents were stayers and *all* of the control group nonrespondents were movers or leavers, which gave an upper bound on the impact estimate, and then assumed the reverse to derive a lower bound estimate. The impact estimates based on all other assumptions were not statistically significant.

[59] 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 they attended a selective college/university, whether they majored in an education-related field, and the amount of student-teaching experience.

Table C.19. Mobility Impacts After Two 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)	75.9	81.5	-5.6
No weights	78.5	80.5	-2.1
Enhanced weights	78.6	80.1	-1.5
Respondents and Nonrespondents			
Assume 100% of treatment nonrespondents are movers, 0% of controls	71.7	82.6	-10.9 *
Assume 0% of nonrespondents are movers	80.0	82.8	-2.9
Assume 25% of nonrespondents are movers	78.0	79.1	-1.1
Assume 50% of nonrespondents are movers	75.7	77.7	-1.9
Assume 100% of nonrespondents are movers	71.4	70.3	1.1
Assume 0% of treatment nonrespondents are movers, 100% of controls	79.8	70.7	9.0*
Respondents and Selected Nonrespondents			
Recode selected nonrespondents from other data sources	79.0	81.4	-2.4
Recode selected nonrespondents and assume 100% of other nonrespondents are movers	75.0	74.3	0.7
Retention in the Teaching Profession			
Respondents			
Benchmark weights (benchmark estimates)	89.5	88.8	0.7
No weights	90.5	89.7	0.8
Enhanced weights	90.4	89.3	1.1
Respondents and Nonrespondents			
Assume 100% of treatment nonrespondents are leavers, 0% of controls	83.2	90.8	-7.7 *
Assume 0% nonrespondents are leavers	91.2	91.1	0.1
Assume 25% of nonrespondents are leavers	89.3	87.4	1.9
Assume 50% of nonrespondents are leavers	87.1	85.8	1.2
Assume 100% of nonrespondents are leavers	82.8	78.5	4.3
Assume 0% of treatment nonrespondents are leavers, 100% of controls	91.1	79.0	12.0*
Respondents and Selected Nonrespondents			
Recode selected nonrespondents from other data sources	90.8	90.3	0.5
Recode selected nonrespondents and assume 100% of other nonrespondents are leavers	86.3	82.4	3.8
Sample Size (Teachers)			
Respondents	197	195	392
Respondents and Selected Nonrespondents	253	243	496
Respondents and Nonrespondents	267	265	532

Source: MPR Mobility Survey administered to all study teachers in 2007–2008.

*Significantly different from zero at the 0.05 level, two-tailed test.

APPENDIX D

SENSITIVITY ANALYSES AND SUPPLEMENTAL TABLES FOR CHAPTER VI

A. SUPPLEMENTARY TABLES FOR TEACHER INDUCTION SERVICES

The tables in this appendix present additional data with which to judge the impact of comprehensive teacher induction on service receipt. Tables D.1 to D.5 present results for teacher induction services for two-year districts based on the Second Induction Activities Survey (administered in spring 2006) and Fourth Induction Activities Survey (administered in spring 2007). The corresponding tables in the main report (Tables VI.1-VI.5) present results based on the First Induction Activities Survey (administered in fall 2005) and the Third Induction Activities Survey (administered in fall 2006). The conclusions do not change when we examine data from spring 2006 and spring 2007.

Table D.1. Teacher Reports on Professional Support and Duties (Percentages): Two-Year Districts

	Spring 2006				Spring 2007			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
BT[a] 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
Unweighted Sample Size (Teachers)	210	176	386		203	169	372	

Source: MPR Second and Fourth Induction Activities Surveys administered to all study teachers in spring 2006 and spring 2007.

Notes: 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]BT = beginning teacher.

*Significantly different from zero at the .05 level, two-tailed test.

Table D.2. Impacts on Teacher-Reported Mentor Profiles (Percentages): 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
Unweighted Sample Size (Teachers)	210	176	386		203	169	372	

Source: MPR Second and Fourth Induction Activities Surveys administered to all study teachers in spring 2006 and spring 2007.

Notes: 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 .05 level, two-tailed test.

Table D.3. Impacts on Teacher-Reported Mentor Services Received in Most Recent Full Week of Teaching: Two-Year Districts

Mentor Service	Spring 2006					Spring 2007				
	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[c] 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 something BT requested[d]	75.9	54.2	21.7*	n.a.	0.000	60.4	23.5	37.0*	n.a.	0.000
Unweighted Sample Size (Teachers)	210	176	386			203	169	372		

Source: MPR Second and Fourth Induction Activities Surveys administered to all study teachers in spring 2006 and spring 2007.

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]BT = beginning teacher.

[d]Total sample size is 306 in spring 2006; 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 .05 level, two-tailed test.

n.a. = not applicable.

Table D.4. Impacts on Teacher-Reported Professional Development Activities During Past Three Months: 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
Unweighted Sample Size (Teachers)	210	176	386			203	169	372		

Source: MPR Second and Fourth Induction Activities Surveys administered to all study teachers in spring 2006 and spring 2007.

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 .05 level, two-tailed test.

n.a. = not applicable.

Table D.5. Impacts on Teacher-Reported Areas of Professional Development During the Past Three Months (Percentages): 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
Unweighted Sample Size (Teachers)	210	176	386		203	169	372	

Source: MPR Second and Fourth Induction Activities Surveys administered to all study teachers in spring 2006 and spring 2007.

Notes: 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, two-tailed test.

B. SUPPLEMENTARY TABLE AND SENSITIVITY ANALYSIS FOR TEACHER SATISFACTION

Table D.6 presents results for teacher satisfaction for two-year districts based on the Second Induction Activities Survey administered in spring 2006 and the Fourth Induction Activities Survey administered in spring 2007. The corresponding table in the main report (Table VI.6) presents results based on the First Induction Activities Survey (administered in fall 2005) and the Third Induction Activities Survey (administered in fall 2006).

The results of the sensitivity analysis show no statistically significant differences with regard to teachers' reports of satisfaction in fall 2005 (Table D.7) or spring 2006 (Table D.8). In fall 2006, treatment teachers were significantly more likely than control teachers to report satisfaction with opportunities for professional development. Treatment teachers were also significantly more likely than control teachers to report satisfaction with opportunities for professional development in spring 2007.

Table D.6. Impacts on Teacher Satisfaction (Scores on a Four-Point Scale): Two-Year Districts

	Spring 2006				Spring 2007			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Feel Satisfied with:								
School	3.1	3.0	0.0	0.596	3.0	2.9	0.1	0.207
Class	3.0	3.0	0.0	0.977	3.1	3.1	0.0	0.797
Teaching career	2.9	3.0	-0.1	0.286	2.8	2.9	-0.1	0.214
Unweighted Sample Size (Teachers)	210	176	386		203	169	372	

Source: MPR Second and Fourth Induction Activities Surveys administered to all study teachers in spring 2006 and spring 2007.

Notes: 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, two-tailed test.

Table D.7. Impacts on Teacher Satisfaction (Percent "Somewhat Satisfied" or "Very Satisfied"): Two-Year Districts, Fall 2005 and Fall 2006

	Fall 2005					Fall 2006				
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	77.7	74.2	3.5	0.08	0.448	75.2	75.0	0.2	0.01	0.956
Availability of resources and materials/equipment for your classroom	67.7	68.1	-0.4	-0.01	0.937	68.4	71.1	-2.7	-0.06	0.552
Input into school policies and practices	65.4	69.2	-3.8	-0.08	0.429	69.3	70.6	-1.2	-0.03	0.770
Opportunities for professional development	85.1	81.9	3.2	0.09	0.387	85.0	77.8	7.2*	0.19	0.025
Principal's leadership and vision	83.9	79.1	4.7	0.12	0.268	72.6	77.2	-4.6	-0.11	0.248
Professional caliber of colleagues	81.1	84.6	-3.5	-0.09	0.364	72.2	79.4	-7.2	-0.17	0.050
Supportive atmosphere among faculty/collaboration with colleagues	82.3	81.3	1.0	0.03	0.830	78.8	80.6	-1.8	-0.04	0.624
School facilities such as the building or grounds	77.3	76.9	0.4	0.01	0.934	73.6	65.0	8.6	0.19	0.112
School policies	83.0	80.2	2.7	0.07	0.570	75.9	80.6	-4.6	-0.11	0.206
Satisfaction with Students										
Autonomy or control over own classroom	84.9	85.7	-0.8	-0.02	0.807	80.3	82.8	-2.5	-0.06	0.454
Student motivation to learn	77.0	75.3	1.8	0.04	0.701	73.3	70.6	2.7	0.06	0.511
Student discipline and behavior	67.3	65.4	1.9	0.04	0.681	67.0	57.8	9.2	0.19	0.067
Parental involvement in the school	46.0	45.6	0.4	0.01	0.939	54.8	50.6	4.3	0.09	0.438
Grade assignment	88.9	86.8	2.1	0.06	0.530	87.4	88.9	-1.5	-0.05	0.529
Students assigned	82.2	82.4	-0.3	-0.01	0.947	81.5	80.6	0.9	0.02	0.782
Satisfaction with Teaching Career										
Salary and benefits	73.3	78.6	-5.3	-0.12	0.218	62.8	63.3	-0.6	-0.01	0.905
Professional prestige	80.5	81.3	-0.8	-0.02	0.847	72.5	75.0	-2.5	-0.06	0.534
Intellectual challenge	90.3	89.0	1.3	0.04	0.707	86.2	85.0	1.2	0.03	0.665
Workload	57.3	63.2	-5.9	-0.12	0.215	49.8	51.7	-1.9	-0.04	0.693
Unweighted Sample Size (Teachers)	213	182	395			191	169	360		

Source: MPR First and Third Induction Activities Surveys administered to all study teachers in fall/winter 2005-2006 and fall/winter 2006-2007.

Notes: 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, two-tailed test.

Table D.8. Impacts on Teacher Satisfaction (Percent "Somewhat Satisfied" or "Very Satisfied"): Two-Year Districts, Spring 2006 and Spring 2007

	Spring 2006					Spring 2007				
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	75.0	72.2	2.8	0.06	0.560	79.9	72.8	7.1	0.16	0.131
Availability of resources and materials/equipment for your classroom	71.8	70.4	1.3	0.03	0.791	74.9	65.7	9.2	0.20	0.093
Input into school policies and practices	60.9	65.3	-4.5	-0.09	0.401	67.7	67.5	0.2	0.01	0.961
Opportunities for professional development	84.2	79.6	4.7	0.12	0.240	86.9	69.2	17.7*	0.42	0.000
Principal's leadership and vision	76.4	77.3	-0.8	-0.02	0.852	80.2	76.3	3.9	0.09	0.377
Professional caliber of colleagues	75.0	80.1	-5.1	-0.12	0.301	77.7	76.3	1.3	0.03	0.776
Supportive atmosphere among faculty/collaboration with colleagues	77.1	82.4	-5.3	-0.13	0.251	73.8	75.7	-1.9	-0.04	0.676
School facilities such as the building or grounds	76.9	75.6	1.3	0.03	0.803	75.6	65.7	9.9	0.22	0.071
School policies	78.5	81.3	-2.7	-0.07	0.541	80.0	75.7	4.3	0.10	0.322
Satisfaction with Students										
Autonomy or control over own classroom	83.2	82.9	0.2	0.01	0.954	88.8	90.5	-1.8	-0.06	0.605
Student motivation to learn	73.3	72.2	1.1	0.03	0.808	73.3	74.6	-1.3	-0.03	0.795
Student discipline and behavior	59.2	60.2	-1.0	-0.02	0.843	64.2	65.7	-1.5	-0.03	0.799
Parental involvement in the school	41.9	43.2	-1.3	-0.03	0.809	49.7	47.9	1.8	0.03	0.755
Grade assignment	91.7	89.8	1.9	0.07	0.527	93.2	90.5	2.7	0.10	0.355
Students assigned	80.7	84.1	-3.4	-0.09	0.389	89.9	87.0	2.9	0.09	0.391
Satisfaction with Teaching Career										
Salary and benefits	65.3	73.3	-8.0	-0.17	0.080	61.3	65.1	-3.7	-0.08	0.474
Professional prestige	73.5	77.3	-3.7	-0.09	0.410	75.5	74.0	1.6	0.04	0.763
Intellectual challenge	86.5	87.5	-1.0	-0.03	0.768	86.9	88.8	-1.9	-0.06	0.594
Workload	57.5	65.3	-7.8	-0.16	0.115	58.6	63.9	-5.3	-0.11	0.335
Unweighted Sample Size (Teachers)	210	176	386			203	169	372		

Source: MPR Second and Fourth Induction Activities Surveys administered to all study teachers in spring 2006 ad spring 2007.

Notes: 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, two-tailed test.

C. Supplementary Tables and Sensitivity Analysis for Student Achievement

The results from alternate estimations of student achievement results for two-year districts are shown in Tables D.9 through D.12. The same specification tests that were discussed for the one-year districts in Appendix C are discussed in this section.

When the results are disaggregated by grade, test score impacts at individual grade levels are not significantly different from zero; nor are they significant using data pooled from grades 3 and above (in the case of two-year districts, grades 3–6). Results are shown in Table D.9 for reading and D.11 for math. Additional specification checks, shown in Table D.10 for reading and D.12 for math, confirm that there is no statistically significant effect of treatment on either subject in Year 2 of teaching. This is true whether allowing outliers to have their original values (line 2), altering the covariates (lines 3 and 4), using an ordinary least squares model with robust standard errors instead of a random effects specification (line 5), excluding the pretest from the model and thereby expanding the sample size (line 6), or using an instrumental variables approach (line 7).

The remaining appendix tables for student achievement—Tables D.13 to D.18—show treatment and control sample sizes for models corresponding to Tables VI.7-VI.8, and D.9-D.12.

Table D.9. Impacts on Reading Test Scores by Grade: Two-Year Districts, 2006-2007 School Year

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
2	0.01	0.02	-0.01	-0.01	0.964	156	12	1
3	0.03	-0.09	0.11	0.11	0.631	469	29	2
4	0.01	0.06	-0.05	-0.05	0.665	705	41	7
5	-0.10	-0.08	-0.02	-0.02	0.858	350	20	4
6	-0.03	0.07	-0.10	-0.10	0.824	52	4	1
All Grades	0.00	0.00	0.00	0.00	0.967	1,732	100	7
Grades 3–6	-0.01	0.00	0.00	0.00	0.969	1,576	90	7

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Notes: Data are regression-adjusted to account for pretest, district-by-grade fixed effects, and clustering of students within schools. Sample sizes for treatment and control groups are shown in Appendix Table D.15.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Table D.10. Impacts on Reading Test Scores, Alternate Model Specifications: Two-Year Districts, 2006-2007 School Year

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Model	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
Benchmark	0.00	0.00	0.00	0.00	0.967	1,732	100	7
With outliers	-0.01	0.00	-0.01	-0.01	0.906	1,732	100	7
Student covariates	0.02	-0.01	0.03	0.03	0.695	1,732	100	7
Student, teacher covariates	0.01	-0.02	0.02	0.02	0.757	1,732	100	7
Robust standard errors	0.00	0.01	-0.02	-0.02	0.747	1,732	100	7
No pretest	0.01	0.00	0.01	0.01	0.904	2,500	136	7
Instrumental variables	0.01	0.00	0.00	0.00	0.956	1,725	100	7

Sources: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts; MPR Teacher Background Survey administered to all study teachers in 2005–2006.

Notes: Data are regression-adjusted to account for district-by-grade fixed effects and clustering of students within schools. See Appendix Table A.1 for a list of other covariates used in these models. Sample sizes of treatment and control groups are shown in Appendix Table D.16.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Table D.11. Impacts on Math Test Scores by Grade: Two-Year Districts, 2006-2007 School Year

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
2	-0.04	0.03	-0.07	-0.07	0.617	154	12	1
3	0.04	-0.09	0.12	0.12	0.496	426	28	2
4	0.01	-0.01	0.01	0.01	0.930	676	40	7
5	-0.09	0.00	-0.09	-0.09	0.406	428	21	4
6	0.09	-0.06	0.15	0.15	0.500	52	4	1
All Grades	-0.03	-0.01	-0.02	-0.02	0.746	1,736	99	7
Grades 3–6	-0.04	-0.01	-0.03	-0.03	0.704	1,582	89	7

Sources: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Notes: Data are regression-adjusted to account for pretest, district-by-grade fixed effects and clustering of students within schools. Sample sizes of treatment and control groups are shown in Appendix Table D.17.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Table D.12. Impacts on Math Test Scores, Alternate Model Specifications: Two-Year Districts, 2006-2007 School Year

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Model	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
Benchmark	-0.03	-0.01	-0.02	-0.02	0.746	1,736	99	7
With outliers	-0.04	-0.01	-0.03	-0.03	0.706	1,736	99	7
Student covariates	-0.01	-0.02	0.01	0.01	0.903	1,736	99	7
Student, teacher covariates	-0.02	0.00	-0.02	-0.02	0.820	1,736	99	7
Robust standard errors	0.00	0.00	0.00	0.00	0.968	1,736	99	7
No pretest	-0.01	0.00	-0.01	-0.01	0.893	2,525	134	7
Instrumental variables	0.02	0.00	0.02	0.02	0.724	1,729	99	7

Sources: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts; MPR Teacher Background Survey administered to all study teachers in 2005–2006.

Notes: Data are regression-adjusted to account for district-by-grade fixed effects and clustering of students within schools. See Appendix Table A.1 for a list of other covariates used in these models. Sample sizes for treatment and control groups are shown in Appendix Table D.18.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Table D.13. Treatment and Control Sample Sizes for Impacts on Test Scores (Benchmark Model): Two-Year Districts, 2006-2007 School Year

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Subject	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Reading	856	52	36	7	876	48	34	7
Math	780	50	36	7	956	49	35	7

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Table D.14. Treatment and Control Sample Sizes for Impacts on Test Scores (Year 1 and Year 2 Common Sample): Two-Year Districts, 2006-2007 School Year

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Subject: Year	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Reading: year 1	663	40	30	6	617	36	27	6
Reading: year 2	671	40	30	6	673	36	27	6
Math: year 1	560	37	29	6	681	37	28	6
Math: year 2	577	37	29	6	746	37	28	6

Source: MPR analysis of data from 2004–2005, 2005–2006, and 2006–2007 school years provided by participating school districts.

Table D.15. Treatment and Control Sample Sizes for Impacts on Reading Test Scores by Grade Level: Two-Year Districts, 2006-2007 School Year

	Treatment				Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
2	59	5	5	1	97	7	5	1
3	274	17	16	2	195	12	10	2
4	327	21	20	7	378	20	18	7
Grades 5-6	196	12	11	4	206	11	11	4
All Grades	856	52	36	7	876	48	34	7
Grades 3–6	797	49	35	7	779	41	31	7

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Table D.16. Treatment and Control Sample Sizes for Impacts on Reading Test Scores, Alternate Model Specifications: Two-Year Districts, 2006-2007 School Year

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Model	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Benchmark	856	52	36	7	876	48	34	7
With outliers	856	52	36	7	876	48	34	7
Student covariates	856	52	36	7	876	48	34	7
Student, teacher covariates	856	52	36	7	876	48	34	7
Robust standard errors	856	52	36	7	876	48	34	7
No pretest	1,250	68	44	7	1,250	68	41	7
Instrumental variables	854	52	36	7	871	48	34	7

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Table D.17. Treatment and Control Sample Sizes for Impacts on Math Test Scores by Grade Level: Two-Year Districts, 2006-2007 School Year

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
2	59	5	5	1	95	7	5	1
3	231	16	15	2	195	12	10	2
4	325	21	20	7	351	19	17	7
5 and 6	165	11	10	4	315	13	13	4
All Grades	780	50	36	7	956	49	35	7
Grades 3–6	721	47	35	7	861	42	32	7

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

Table D.18. Treatment and Control Sample Sizes for Impacts on Math Test Scores, Alternate Model Specifications: Two-Year Districts, 2006-2007 School Year

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Model	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Benchmark	780	50	36	7	956	49	35	7
With outliers	780	50	36	7	956	49	35	7
Student covariates	780	50	36	7	956	49	35	7
Student, teacher covariates	780	50	36	7	956	49	35	7
Robust standard errors	780	50	36	7	956	49	35	7
No pretest	1,178	69	43	7	1,347	69	42	7
Instrumental variables	779	49	36	7	950	49	35	7

Source: MPR analysis of data from 2005–2006 and 2006–2007 school years provided by participating school districts.

D. SENSITIVITY ANALYSIS FOR TEACHER RETENTION

For the teacher retention analysis using two-year districts, the conclusions did not change when we expanded the number of outcomes to differentiate between moving to a school in another public school district and moving to a private, parochial, or other school, and expanded the outcomes for leaving to include leaving to stay at home, leaving to attend school or take a new job, and other reasons for leaving. When we re-estimated the models using a linear probability model or a multinomial logit model, we reached the same conclusions as when we used a binary logit model.

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 (Table D.19). 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.

The results did not change when we assumed that all nonrespondents were stayers or all were leavers. The only exceptions were the most extreme assumptions, in which we first assumed that *all* of the treatment group nonrespondents were stayers and *all* of the control group nonrespondents were movers or leavers, which gave an upper bound on the impact

estimate, and then assumed the reverse to derive a lower bound estimate. The impact estimates based on all other assumptions were not statistically significant.

Table D.19. Mobility Impacts After Two 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)	68.0	72.0	-4.0
No weights	69.6	75.2	-5.6
Enhanced weights	69.5	75.5	-6.0
Respondents and Nonrespondents			
Assume 100% of treatment nonrespondents are movers, 0% of controls	63.8	80.5	-16.8*
Assume 0% of nonrespondents are movers	72.2	80.2	-8.0
Assume 25% of nonrespondents are movers	71.5	75.1	-3.6
Assume 50% of nonrespondents are movers	68.5	70.8	-2.3
Assume 100% of nonrespondents are movers	63.9	61.4	2.5
Assume 0% of treatment nonrespondents are movers, 100% of controls	72.2	61.1	11.1*
Respondents and Selected Nonrespondents			
Recode selected nonrespondents from other data sources	70.4	77.8	-7.4
Recode selected nonrespondents and assume 100% of other nonrespondents are movers	66.9	70.5	-3.6
Retention in the Teaching Profession			
Respondents			
Benchmark weights (benchmark estimates)	86.9	90.8	-3.9
No weights	87.3	90.6	-3.3
Enhanced weights	86.7	90.9	-4.2
Respondents and Nonrespondents			
Assume 100% of treatment nonrespondents are leavers, 0% of controls	79.6	92.8	-13.2*
Assume 0% of nonrespondents are leavers	88.4	92.3	-3.9
Assume 25% of nonrespondents are leavers	87.4	87.4	0.0
Assume 50% of nonrespondents are leavers	84.5	83.1	1.3
Assume 100% of nonrespondents are leavers	79.8	73.8	6.0
Assume 0% of treatment nonrespondents are leavers, 100% of controls	88.3	73.4	14.9*
Respondents and Selected Nonrespondents			
Recode selected nonrespondents from other data sources	87.8	91.5	-3.7
Recode selected nonrespondents and assume 100% of other nonrespondents are leavers	82.9	82.9	0.1
Sample Size (Teachers)			
Respondents	164	117	281
Respondents and Selected Nonrespondents	210	179	389
Respondents and Nonrespondents	222	199	421

Source: MPR Mobility Survey administered to all study teachers in 2007–2008.

*Significantly different from zero at the 0.05 level, two-tailed test.

APPENDIX E

IMPACTS ON TEACHER PREPAREDNESS (TWO-YEAR DISTRICTS)

An extra Induction Activities Survey administered in spring 2007 in two-year districts allowed us to examine whether comprehensive teacher induction made teachers feel more prepared to do their jobs than control teachers in those districts. The survey results indicated that this was not the case. There were no statistically significant impacts of treatment on teacher preparedness in spring 2006 or spring 2007.

A. METHODS

Using items from the induction activities surveys, we measured teachers' feelings of preparedness in 13 areas. Factor analysis suggested that teacher preparedness consisted of three categories: (1) instruction, (2) working with students, and (3) working with others (details are given in Appendix B). Benchmark estimates are based on a regression model that has district and grade fixed effects and no other covariates. The results did not vary according to estimation method or the set of control variables used.

B. IMPACT ESTIMATES

Overall, teachers from the treatment and control groups reported feelings of preparedness that differed by 0.10 or less on a four-point scale, in both spring 2006 and spring 2007. Out of the six differences we examined (three measures at two points in time), none were statistically significant (Table E.1).

C. SENSITIVITY ANALYSIS

One concern with this analysis 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 differences between the first and second categories may be larger than those between the third and fourth. We recoded teacher preparedness into two categories—(1) "not at all prepared" or "somewhat prepared" or (2) "well prepared" or "very well prepared"—and found no change in the conclusions. We then examined item-specific impacts on the outcome defined as a dichotomous variable and found no change in

the conclusions. The results of changing both assumptions (Table E.2) show that treatment teachers were significantly less likely than control teachers to report preparedness with working effectively with parents in spring 2006.

Table E.1. Impacts on Teacher Preparedness (Scores on a Four-Point Scale): Two-Year Districts

	Spring 2006				Spring 2007			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Feel Prepared to:								
Instruct	3.0	3.0	0.0	0.703	3.2	3.1	0.0	0.869
Work with students	2.9	2.8	0.1	0.472	3.0	3.0	0.0	0.614
Work with other school staff	3.0	3.0	-0.1	0.338	3.1	3.1	0.0	0.933
Unweighted Sample Size (Teachers)	210	176	386		203	169	372	

Source: MPR Second and Fourth Induction Activities Surveys administered to all study teachers in spring 2006 and spring 2007.

Notes: 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. 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.

None of the differences is statistically significant at the 0.05 level, two-tailed test.

Table E.2. Impacts on Teacher Preparedness (Percent "Somewhat Prepared" or "Very Prepared"): Two-Year Districts

	Spring 2006					Spring 2007				
Area of Preparedness	Treatment	Control	Difference	Effect Size	P-value	Treatment	Control	Difference	Effect Size	P-value
Prepared to Instruct										
Managing classroom activities, transitions, and routines	83.6	81.8	1.8	0.05	0.713	88.3	88.8	-0.4	-0.01	0.906
Using variety of instructional methods	73.5	74.4	-1.0	-0.02	0.857	81.8	84.6	-2.8	-0.08	0.480
Assessing your students	69.5	72.2	-2.6	-0.06	0.580	85.4	84.6	0.8	0.02	0.849
Selecting and adapting instructional materials	66.1	65.3	0.8	0.02	0.882	75.4	76.9	-1.6	-0.04	0.733
Planning effective lessons	75.8	82.9	-7.1	-0.17	0.143	86.3	85.8	0.5	0.01	0.902
Being an effective teacher	79.6	79.6	0.0	0.00	0.999	85.6	90.5	-4.9	-0.15	0.156
Addressing needs of a diversity of learners	73.9	72.2	1.8	0.04	0.756	79.7	76.3	3.4	0.08	0.487
Prepared to Work with Students										
Handling range of classroom behavior or discipline situations	75.1	72.2	2.9	0.07	0.542	84.5	83.4	1.1	0.03	0.788
Motivating students	78.7	79.6	-0.9	-0.02	0.844	83.5	89.9	-6.4	-0.18	0.082
Working effectively with parents	64.1	73.9	-9.7*	-0.21	0.047	75.5	77.5	-2.0	-0.05	0.655
Working with students with special challenges	50.4	42.0	8.3	0.17	0.110	53.9	46.8	7.1	0.14	0.166
Prepared to Work with Other School Staff										
Working with other teachers to plan instruction	73.1	80.1	-7.0	-0.17	0.113	85.1	84.0	1.1	0.03	0.779
Working with the principal or other instructional leaders	74.7	73.9	0.8	0.02	0.854	79.8	78.7	1.1	0.03	0.796
Unweighted Sample Size (Teachers)	210	176	386			203	169	372		

Source: MPR Second and Fourth Induction Activities Surveys administered to all study teachers in spring 2006 and spring 2007.

Notes: 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, two-tailed test.

APPENDIX F

SENSITIVITY ANALYSES FOR CHAPTER VII

A. SENSITIVITY ANALYSES FOR STUDENT ACHIEVEMENT

We performed three different sensitivity analyses to test the robustness of the correlational analyses presented in the main report (Table VII.1). The first sensitivity analysis replaces the Induction Services Index with an alternative that omits the measure of observing others teaching. The results from this analysis, shown in Table F.1, indicate that the association between the years the beginning teacher had an assigned mentor and math test scores is positive and statistically significant (regression coefficient = 0.12, p-value = 0.016). In this analysis, the relationship between the alternate Induction Services Index and math and reading test scores was not statistically significant. The associations between the other induction services measures (Instructional Support Index and Induction Intensity Index) and math and reading test scores are also statistically insignificant.

In order to rule out any concern that the similarity of each induction services measure to each other makes it difficult to identify their overall effects, a problem known as multicollinearity, we conducted separate analyses for each of the four induction services measures individually. Table F.2 shows the results from regression models in which each induction services measure is entered without the other three measures. Under this approach, the association between the years the beginning teacher had an assigned mentor and math test scores is positive and statistically significant (regression coefficient = 0.09, p-value = 0.046). The associations between each of the other three induction services measures and math and reading test scores remained statistically insignificant.

To further explore the robustness of the association measures, we used an Instrumental Variables approach (Angrist, Imbens and Rubin 1996). Under this approach, the randomization indicator (that is, an indicator for whether the student was taught in a treatment or a control school) is used as an instrument for each of the induction services measures. We then estimated regression models in which each instrumented services measure is entered without the other three measures. The results from this approach, as presented in Table F.3, show that the associations between the beginning teacher support indices and student math and reading test scores are not statistically significant.

Table F.1. Association Between Beginning Teacher Support and Test Scores (Induction Services Index Excludes Observing Others Teaching)

	Math[a]		Reading[a]	
Induction Measure	Coefficient	P-value	Coefficient	P-value
Years BT had an assigned mentor	0.12*	0.016	0.00	0.992
Induction Services Index (Excludes Observing Others Teaching)	0.01	0.510	0.01	0.276
Instructional Support Index	0.01	0.502	0.01	0.307
Induction Intensity Index	-0.03	0.098	-0.01	0.453
Unweighted Sample Size (Districts)	16		16	
Unweighted Sample Size (Schools)	152		159	
Unweighted Sample Size (Teachers)	202		220	
Unweighted Sample Size (Students)	3,476		3,693	

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

Notes: BT = beginning teacher. The variable "years BT had an assigned mentor" has the following values: 0, 1, and 2 years. The Induction Services Index is the sum of the indicator variables at fall 2005, spring 2006, and fall 2006, on whether the beginning teacher: (1) met with a literacy or math coach, and (2) met with a study group (range: 0-6). The Instructional Support Index is constructed similarly using the indicator variables on whether the beginning teacher received: (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 (range 0-8). The Induction Intensity Index is the sum of 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.

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, two-tailed test.

[a]The following variables are not jointly significant: years BT had an assigned mentor, Induction Services Index, Instructional Support Index, and Induction Intensity Index (p-value = 0.063 for math, 0.542 for reading).

Table F.2. Association Between Beginning Teacher Support and Test Scores (One Regression Model per Induction Measure)

Outcome	Years BT Had an Assigned Mentor	Induction Services Index	Instructional Support Index	Induction Intensity Index
Math				
Coefficient	0.09*	0.00	0.01	-0.01
P-value	0.046	0.749	0.448	0.576
Unweighted Sample Size (Districts)	16	16	16	16
Unweighted Sample Size (Schools)	161	161	158	152
Unweighted Sample Size (Teachers)	214	214	211	202
Unweighted Sample Size (Students)	3,705	3,705	3,645	3,476
Reading				
Coefficient	0.01	0.01	0.01	0.00
P-value	0.726	0.351	0.425	0.785
Unweighted Sample Size (Districts)	16	16	16	16
Unweighted Sample Size (Schools)	168	168	165	159
Unweighted Sample Size (Teachers)	233	233	229	220
Unweighted Sample Size (Students)	3,952	3,952	3,864	3,693

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

Notes: BT = beginning teacher. The variable "years BT had an assigned mentor" has the following values: 0, 1, and 2 years. The Induction Services Index is the sum of the indicator variables at fall 2005, spring 2006, and fall 2006, on whether the beginning teacher: (1) met with a literacy or math coach, and (2) met with a study group (range: 0-6). The Instructional Support Index is constructed similarly using the indicator variables on whether the beginning teacher received: (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 (range 0-8). The Induction Intensity Index is the sum of 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.

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, two-tailed test.

Table F.3. Association Between Beginning Teacher Support and Test Scores (Instrumental Variables Analyses)

Outcome	Years BT Had an Assigned Mentor	Induction Services Index	Instructional Support Index	Induction Intensity Index
Math[a]				
Coefficient	0.04	0.02	0.02	0.06
P-value	0.605	0.600	0.564	0.375
Unweighted Sample Size (Districts)	16	16	16	16
Unweighted Sample Size (Schools)	161	161	158	152
Unweighted Sample Size (Teachers)	214	214	211	202
Unweighted Sample Size (Students)	3,705	3,705	3,645	3,476
Reading[a]				
Coefficient	0.05	0.02	0.02	0.07
P-value	0.489	0.484	0.586	0.371
Unweighted Sample Size (Districts)	16	16	16	16
Unweighted Sample Size (Schools)	168	168	165	159
Unweighted Sample Size (Teachers)	233	233	229	220
Unweighted Sample Size (Students)	3,952	3,952	3,864	3,693

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

Notes: BT = beginning teacher. The variable "years BT had an assigned mentor" has the following values: 0, 1, and 2 years. The Induction Services Index is the sum of the indicator variables at fall 2005, spring 2006, and fall 2006, on whether the beginning teacher: (1) met with a literacy or math coach, (2) met with a study group, and (3) observed others teaching (range: 0-9). The Instructional Support Index is constructed similarly using the indicator variables on whether the beginning teacher received: (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 (range 0-8). The Induction Intensity Index is the sum of 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.

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

None of the differences is statistically significant at the 0.05 level, two-tailed test.

[a]These are results for regression models in which each induction services measure has been entered without the other three induction measures. The randomization indicator (indicator of whether student was taught by a treatment or a control teacher) was used as an instrument for each of the induction services measures in these models.

B. SENSITIVITY ANALYSES FOR TEACHER MOBILITY

We performed three different sensitivity analyses to test the robustness of the correlational analyses presented in the main report (Table VII.2). In the first sensitivity analysis, we used an alternate Induction Services Index that omits the measure of observing others teaching. As shown in Table F.4, the association between the alternate index and the likelihood of remaining in the district is positive and statistically significant. The associations between the likelihood of remaining in the district and the variable on years teacher had an assigned mentor, the Instructional Support Index, and the Induction Intensity Index, are statistically insignificant. Similarly, we found that the alternate index is statistically significantly associated with the likelihood of remaining teaching. The associations between the other induction services measures and the likelihood of remaining teaching are not statistically significant.

In the interest of avoiding the possible multicollinearity among the induction services measures, we conducted separate analyses for each of the four induction services measures. We found that the association between the Induction Services Index and the likelihood of remaining in the district is positive and statistically significant, as shown in Table F.5, for a regression model in which the Induction Services Index is entered without the other indices. The associations between the likelihood of remaining in the district and the other induction services measures are not statistically significant for this specification of the model. Table F.5 also shows that the association between each of the four induction services measures and the likelihood of remaining in teaching is positive and statistically significant for a regression model in which each induction services measure is entered without the other three measures.

The results from the IV analysis (which was discussed in Chapter VII), are presented in Table F.6. They show that the associations between the induction services measures and the likelihood of remaining in the district and the likelihood of remaining teaching are not statistically significant.

Table F.4. Association Between Beginning Teacher Support and Teacher Mobility (Induction Services Index Excludes Observing Others Teaching)

	Remains in District[a]		Remains in Teaching[a]	
Induction Measure	Coefficient	P-value	Coefficient	P-value
Years BT had an assigned mentor	-0.04	0.166	0.00	0.557
Induction Services Index (excludes observing others teaching)	0.03*	0.000	0.01*	0.001
Instructional Support Index	0.00	0.988	0.00	0.789
Induction Intensity Index	0.01	0.412	0.00	0.413
Unweighted Sample Size (Teachers)	786		786	

Source: MPR Mobility Survey administered in 2007-2008; MPR Teacher Background Survey administered in 2005-2006; and First, Second, and Third Induction Activities Surveys administered in fall/winter 2005-2006, spring 2006, and fall/winter 2006-2007 to all study teachers.

Notes: BT = beginning teacher. The variable "years BT had an assigned mentor" has the following values: 0, 1, and 2 years. The Induction Services Index is the sum of the indicator variables at fall 2005, spring 2006, and fall 2006, on whether the beginning teacher: (1) met with a literacy or math coach, and (2) met with a study group (range: 0-6). The Instructional Support Index is constructed similarly using the indicator variables on whether the beginning teacher received: (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 (range 0-8). The Induction Intensity Index is the sum of 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.

Data are regression-adjusted using a logit model with robust standard errors to account for baseline characteristics and clustering of teachers within schools.

*Significantly different from zero at the 0.05 level, two-tailed test.

[a]The following variables are not jointly significant: years BT had an assigned mentor, Induction Services Index, Instructional Support Index, and Induction Intensity Index (p-value = 0.063 for math, 0.542 for reading).

Table F.5. Association Between Beginning Teacher Support and Teacher Mobility (One Regression Model per Induction Measure)

Outcome	Years BT had an Assigned Mentor	Induction Services Index	Instructional Support Index	Induction Intensity Index
Remains in District				
Coefficient	-0.01	0.03*	0.01	0.01
P-value	0.600	0.000	0.154	0.221
Unweighted Sample Size (Teachers)	840	836	826	786
Remains in Teaching				
Coefficient	0.01*	0.01*	0.01*	0.01*
P-value	0.050	0.000	0.004	0.030
Unweighted Sample Size (Teachers)	840	836	826	786

Source: MPR Mobility Survey administered in 2007-2008; MPR Teacher Background Survey administered in 2005-2006; and First, Second, and Third Induction Activities Surveys administered in fall/winter 2005-2006, spring 2006, and fall/winter 2006-2007 to all study teachers.

Notes: BT = beginning teacher. The variable "years BT had an assigned mentor" has the following values: 0, 1, and 2 years. The Induction Services Index is the sum of the indicator variables at fall 2005, spring 2006, and fall 2006, on whether the beginning teacher: ((1) met with a literacy or math coach, (2) met with a study group, and (3) observed others teaching (range: 0-9). The Instructional Support Index is constructed similarly using the indicator variables on whether the beginning teacher received: (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 (range 0-8). The Induction Intensity Index is the sum of 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.

Data are regression-adjusted using a logit model with robust standard errors to account for baseline characteristics and clustering of teachers within schools.

*Significantly different from zero at the 0.05 level, two-tailed test.

Table F.6. Association Between Beginning Teacher Support and Teacher Mobility (Instrumental Variables Analyses)

Outcome	Years BT Had an Assigned Mentor	Induction Services Index	Instructional Support Index	Induction Intensity Index
Remains in District[a]				
Coefficient	-0.11	-0.04	-0.04	-0.06
P-value	0.280	0.221	0.259	0.280
Unweighted Sample Size (Teachers)	840	836	826	786
Remains in Teaching[a]				
Coefficient	-0.04	-0.02	-0.01	-0.03
P-value	0.371	0.281	0.395	0.247
Unweighted Sample Size (Teachers)	840	836	826	786

Source: MPR Mobility Survey administered in 2007-2008; MPR Teacher Background Survey administered in 2005-2006; and First, Second, and Third Induction Activities Surveys administered in fall/winter 2005-2006, spring 2006, and fall/winter 2006-2007 to all study teachers.

Notes: BT = beginning teacher. The variable "years BT had an assigned mentor" has the following values: 0, 1, and 2 years. The Induction Services Index is the sum of the indicator variables at fall 2005, spring 2006, and fall 2006, on whether the beginning teacher: (1) met with a literacy or math coach, (2) met with a study group, and (3) observed others teaching (range: 0-9). The Instructional Support Index is constructed similarly using the indicator variables on whether the beginning teacher received: (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 (range 0-8). The Induction Intensity Index is the sum of 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.

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, two-tailed test.

[a]These are results for regression models in which each induction services measure has been entered without the other three induction measures. The randomization indicator (indicator of whether teacher is a treatment or a control teacher) was used as an instrument for each of the induction services measures in these models.