NCEE 2009-4034

U.S. DEPARTMENT OF EDUCATION

Impacts of Comprehensive Teacher Induction

Results From the First 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 First Year of a Randomized Controlled Study

OCTOBER 2008

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

Edward Britton
WestEd

Melanie Ali
Project Officer
Institute of Education Sciences

NCEE 2009-4034
U.S. DEPARTMENT OF EDUCATION

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

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Evaluation and Regional Assistance
Phoebe Cottingham
Commissioner

October 2008

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 and Nakis Evgeniou. Amy Raduzycki, Phyllis Schulman, and Susan Golden played critical roles in achieving high response rates to the surveys. Renée Nogales was instrumental and very persistent in gathering test score records from the many districts that participated in the study. Gail Baxter, Kristin Hallgren, and Kathy Sonnenfeld coordinated with districts at several stages. Matthew Jacobus and Ama Takyi-Laryea, assisted by Dominic Harris and Elizabeth Seif, provided expert programming assistance. Roberto Agodini carefully reviewed drafts of the report and provided helpful comments. The report was produced by Donna Dorsey. Mustafa Menai and Mindy Hu provided additional production support.

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, Inc., 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**

TABLES

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 less so (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 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), 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, Fuller 2003; Youngs 2002) has yielded more detailed descriptions of teacher supports. Like the national studies, however, it has relied on evaluation designs that leave doubt about whether the inferences are causal.

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 (ED) Institute of Education Sciences (IES) contracted with Mathematica Policy Research, Inc. (MPR), to evaluate the impact of structured and intensive teacher induction programs. Throughout this 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 teacher induction services:

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?

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?

In 2004, we issued a Request for Proposals (RFP) to implement a comprehensive induction program as part of the study. The RFP specified that the induction program should include several components that earlier research and professional wisdom gleaned from practice had suggested were important features of successful teacher induction programs (Alliance for Excellent Education 2004; Ingersoll and Smith 2004; Smith and Ingersoll 2004; Kelly 2004; Serpell and Bozeman 2000).

The components that constitute comprehensive teacher induction include the following: 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.

A group of outside expert reviewers read and scored the proposals we received in response to the RFP. Among the proposals received, those submitted by Educational Testing Service of Princeton, New Jersey (ETS) and the New Teacher Center at the University of California-Santa Cruz (NTC) stood out as most closely meeting the study's specified requirements. The two programs included the required components and were roughly comparable in structure. MPR therefore contracted with both providers to deliver one year of 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. As discussed below, we used surveys, classroom observations, and school records to measure the background of the study teachers, their receipt of induction services and alternative support services, their attitudes, and their outcomes related to the study's main research questions: classroom practices, student achievement, and teacher mobility.

We recruited 17 school districts to participate in the study. The 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. 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.

Within each district, a subset of elementary schools participated in the study. The study used an experimental design in which we randomly assigned elementary schools within each of the 17 participating districts to either a treatment group, which received comprehensive teacher induction—from ETS or NTC, depending on the district—or a control group, which took part in the district's usual teacher induction program. Districts nominated approximately 500 schools across the 17 districts. It turned out that some schools that we targeted for random assignment had no eligible teachers, so the final sample sizes included 418 schools: 100 treatment schools and 103 control schools in the 9 ETS districts and 110 treatment and 105 control schools in the 8 NTC districts.

With 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. The 418 schools participating in the study contained 1,009 eligible teachers.

Not all of the 1,009 teachers eligible for the study were eligible for all analyses. We limited the collection of classroom practices data to 698 teachers who met certain eligibility requirements such as teaching English/language arts to a self-contained classroom. Because we focused on reading instruction, it was not appropriate or even possible to include teachers such as music, art, or math specialists who were not responsible for teaching reading. 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 281 teachers and math scores for 261 teachers.

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.

We found that random assignment produced groups that were equivalent on a wide variety of teacher and school characteristics. Of the dozens of baseline attributes we examined, we found statistically significant differences between treatment and control groups in one area: teacher assignments.[2] The control group contained a higher percentage of special subject teachers (such as art and music) than did the treatment group (7 versus 3 percent) and consequently a lower percentage of teachers who taught just a single grade (79 versus 85 percent) and who said they were responsible for math (85 versus 90 percent) or reading outcomes (83 versus 91 percent). Accounting for such differences did not change the study's conclusions.

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. The set of benchmark control variables (covariates), which differs for each outcome, are described in the discussion of key study findings.

To test the robustness of the study findings, we conducted several sensitivity tests. These tests included re-estimation of the study's main impacts with different sets of covariates and sample weights and different statistical model assumptions. We also reported whether the findings would change if we were to use post-hoc adjustments for multiple

[2] All differences discussed in the text are statistically significant at the 0.05 level unless stated otherwise.

comparison errors. Multiple comparison errors are those that arise when researchers report on a large number of hypothesis tests, at least some of which may result in falsely rejecting the null hypothesis. Specifically, we applied a method developed by Benjamini and Hochberg (1995) for reducing the rate of false discoveries.

Findings are pooled across ETS and NTC districts throughout this report because the study was intended to explore the effects of comprehensive teacher induction in general, not the specific impacts of any one program. However, we conducted separate analyses by district type (ETS or NTC) to ensure that the findings were not peculiar to one of the providers.

Data for the study were collected from a variety of sources. We administered a baseline teacher survey 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 and preparedness for different aspects of their current teaching position. We surveyed mentors participating in the comprehensive induction programs on their background characteristics and reviewed program documents from ETS and NTC. Additional detail on these measures is included in the discussion of findings below.

For the study's core outcomes, we observed the teachers teaching a literacy unit in the classroom in the spring of 2006, collected districts' student records data at the end of the 2005-2006 school year, and conducted the first of three mobility surveys in fall 2006 to learn about teacher retention. We achieved response rates of over 85 percent on the teacher surveys and observations, although the rates for the control group (for example, 92 percent on the background survey) were not as high as those for the treatment group (97 percent on the same survey). We used nonresponse adjustment weights and sensitivity analyses to address the differential response rates.

The instrument used to conduct the observations was the Vermont Classroom Observation Tool (VCOT). The VCOT measures the teacher practices that current research suggests are essential to good teaching or that have been linked to student achievement growth (Cawelti 2004). The VCOT also measures instructional practices that closely reflect those recognized by both the ETS and NTC induction programs, particularly for literacy instruction. We observed eligible study teachers once while they were teaching a literacy unit. The observations lasted between one to two hours, with duration dependent on how the district or school structured its class periods. Observers scored teachers in each of three constructs based on a set of items that are believed to be indicators of good practice: implementation of a lesson, content of a lesson, and classroom culture. Implementation was measured with five items that focused on the effectiveness of instruction and learning that occurred during the lesson. Content was measured with four items that assessed the accuracy, importance, level of abstraction, and connections to other concepts. Classroom culture was measured with seven items that assessed the learning environment, the level of

student engagement, the nature of working relationships, and issues of student equity (Saginor and Hyjek 2005). The three domains comprise five, four, and seven items, respectively. Observers rated the extent of evidence of teacher behavior for each item on a five-point scale showing (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence.

We measured student achievement outcomes using district-administered test score data from the spring 2006 (post-test) for students taught by study teachers in the 2005-2006 school year and students' linked scores from the prior grade in spring 2005 (pre-test).[3] We conducted all treatment-control comparisons within grade and within district to ensure that treatment status was not confounded with properties of the test.

THE TREATMENT: COMPREHENSIVE INDUCTION SERVICES

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. The curriculum 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. Treatment teachers were also 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. Treatment teachers also observed veteran teachers once or twice during the year. At the end of the school year, treatment teachers in both ETS and NTC districts participated in a colloquium celebrating the year's successes and teachers' professional growth.

The goal of the study was to assign each mentor to 12 beginning teachers, though mentor caseloads ranged from 8 to 14 teachers over the course of the year. The program providers 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). 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

[3] One district tested students in the fall, so we used data that tracked growth from fall 2005 to fall 2006.

analyzing student work. The trainings focused on improving beginning teachers' instruction, including the use of forms and processes to accomplish this, and 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. Additional support and development opportunities were provided to mentors during the year through weekly meetings of district mentors as well as feedback and advice from district coordinators and program staff.

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.[4] 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.[5]

There are other similar features as well. Under each program, 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.

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, 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. 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.

[4] The four domains are planning and preparation, classroom environment, instruction, and professional responsibilities. As an example, components of the instruction domain include communicating clearly and accurately and using questioning and discussion techniques. The full set of components in each domain is presented in Chapter IV.

[5] The six standards are planning instruction and designing learning experiences, creating/maintaining effective environments, understanding/organizing subject matter, development as a professional educator, engaging/supporting all students in learning, and assessing student learning. As an example, categories of teaching behaviors in the standard of engaging/supporting all students in learning include connecting prior knowledge, life experience, and interests with learning goals and promoting self-directed, reflective learning. The full set of teaching behaviors in each standard is presented in Chapter IV.

SUMMARY OF FINDINGS: POSITIVE IMPACTS ON INDUCTION SUPPORT RECEIVED

We found statistically significant differences between the treatment and control groups in the amount, types, and content of induction support teachers reported having received (see Chapter IV). This finding was similar in the fall and the spring of the intervention year. 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.

Treatment Teachers Reported Receiving More Mentoring Than Did Control Teachers. Treatment teachers were significantly more likely than control teachers to report having any mentor (94 versus 83 percent) and having more than one mentor (29 versus 17 percent). The types of mentors also differed between treatment and control groups. Treatment teachers were significantly more likely than control teachers to report having a mentor assigned to them (93 versus 75 percent) and to report having a full-time mentor (74 versus 13 percent).[6] Treatment teachers reported spending significantly more time working with their mentors than control teachers did during the most recent full week of teaching. Treatment teachers reported spending an average of 95 minutes per week in mentor meetings compared to 74 minutes for control teachers, with the 21-minute difference attributable entirely to differences in the duration of scheduled meetings. For a typical school year of 36 weeks, the treatment-control difference in the total hours of mentor contact time during the year is estimated to be 12.5 hours.

Treatment Teachers Were More Likely Than Control Teachers to Report Participating in Specific Induction Activities. Treatment teachers reported spending significantly more time during the most recent full week of teaching being observed by their mentors (26 versus 11 minutes), observing mentors modeling lessons (11 versus 7 minutes), and meeting one-on-one with a mentor (34 versus 21 minutes) or meeting with mentors and other first-year teachers (27 versus 7 minutes) as compared to control teachers. During the most recent full week of teaching, treatment teachers were 15 to 26 percentage points more likely than control teachers to report having received mentors' assistance in a variety of topic areas, with a difference of more than 20 percentage points in discussing instructional goals and how to achieve them (70 versus 44 percent), receiving suggestions to improve practice (74 versus 52 percent), and receiving guidance on assessing students (62 versus 40 percent). Examining a broader window of three months prior to the spring survey, treatment teachers were a significant 7 to 36 percentage points more likely than control teachers to receive each type of guidance the survey asked about, with a difference of 25 percentage points or more in reflecting on instructional practice (68 versus 33 percent); managing classroom activities, transitions, and routines (65 versus 40 percent); reviewing and assessing student work (55

[6] Although all treatment teachers were assigned a full-time ETS or NTC mentor, not all treatment teachers reported this person as their mentor. In addition, not all treatment teachers reported having a mentor assigned to them (as opposed to being someone the teacher sought out) or reported having a full-time mentor who had been released from teaching. We discuss teacher-reported mentor profiles in detail in Chapter IV.

versus 30 percent); and using student assessments to inform their teaching (54 versus 29 percent).

Treatment Teachers Spent More Time in Certain Professional Activities Than Did Control Teachers During the Three Months Prior to the Spring Survey. During the three months prior to the spring survey, treatment teachers were significantly more likely than control teachers to report having kept written logs (40 versus 28 percent), worked with study groups of new teachers (68 versus 27 percent) and study groups of new and experienced teachers (47 versus 37 percent), and observed others teaching both in their classrooms (70 versus 42 percent) and in the teacher's classroom (47 versus 38 percent). However, treatment and control teachers did not differ significantly in their likelihood to report having engaged in other activities such as keeping a portfolio and analysis of student work or meeting with principals, literacy or mathematics coaches, or resource specialists. Compared to control teachers, treatment teachers were significantly more frequently observed by mentors (3.4 versus 1.5 times), though not by principals, and more frequently given feedback on teaching both as part of a formal evaluation (1.7 versus 1.5 times) and not as part of a formal evaluation (2.5 versus 2.0 times) than control teachers during this period. Of 17 areas of professional development asked about, treatment teachers were significantly more likely than control teachers to report having attended professional development in three areas: lesson planning (38 versus 26 percent), analyzing student work/assessment (56 versus 42 percent) and differentiated instruction (55 versus 46 percent). Treatment teachers reported spending significantly more time in professional development in 4 of the 17 areas: analyzing student work/assessment (58 versus 41 minutes), lesson planning (36 versus 26 minutes), parent and community relations (23 versus 15 minutes), and assigning grades/record keeping (17 minutes versus 10 minutes). Treatment teachers reported spending significantly less time than control teachers in one area: preparing students for standardized testing (43 minutes versus 53 minutes).

SUMMARY OF FINDINGS: NO IMPACTS ON TEACHER PRACTICES IN THE FIRST YEAR

Observers scored teachers on a set of 16 indicators of teaching practice using a five-point scale.[7] The indicators are grouped into three domains: lesson implementation, lesson content, and classroom culture. The analysis included teacher demographic characteristics, teacher's educational and professional background, teaching assignment, school characteristics, and district and grade fixed effects.

We observed no statistically significant differences between treatment and control teachers' performance on any of the three domains of classroom practices (Table 1). We express the impact on each domain of classroom practice as the difference in scores on the five-point scale. An impact of 0.5 point, for example, would suggest that the intervention moves the average teacher from being able to demonstrate "moderate" evidence of good practice in that domain half of the distance to being able to demonstrate "consistent" evidence of good practice if they start at the moderate level.

[7] The instrument used to conduct teacher observations was the Vermont Classroom Observation Tool (Saginor and Hyjek 2005).

Table 1. Impacts on Classroom Practices (Average Score on a 5-Point Scale)

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

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

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

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

SUMMARY OF FINDINGS: NO POSITIVE IMPACTS ON STUDENT TEST SCORES IN THE FIRST YEAR

The test score analysis was based on standardized achievement tests that the district normally conducts.[8] While district-administered test scores may 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. The impact estimates are regression-adjusted using covariates that include the normalized student pre-test score, student characteristics, teacher personal characteristics, teacher professional characteristics, and district-by-grade fixed effects.

The findings, summarized in Tables 2 and 3, show the grade-specific impact estimates to be negative and statistically significant for grade 2 for reading (effect size = -0.22) and for grades 2 and 3 for math (effect size = -0.38 and -0.26, respectively), but the average impacts across all grades were not significantly different from zero for math or reading. The findings were robust to different analysis methods, such as regression with an omitted pre-test or regression with alternative weights or different sets of control variables.

[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). We standardized all test scores to have a mean of zero and a standard deviation of one to facilitate aggregation of impacts across districts and grades.

Table 2. Impacts on Reading Test Scores

	Adjusted Mean Test Scores					Unweighted Student Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Total	Treatment Group	Control Group
2	-0.12	0.10	-0.22*	-0.22	0.034	543	243	300
3	-0.06	0.07	-0.13	-0.13	0.119	1,113	629	484
4	0.02	-0.02	0.04	0.04	0.421	1,679	919	760
5	0.01	-0.01	0.01	0.01	0.843	1,516	707	809
6	--	--	--	--		48	24	24
All Grades	**0.01**	**-0.01**	**0.01**	**0.01**	**0.735**	**4,899**	**2,522**	**2,377**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

None of the differences is significantly different from zero after applying a Benjamini-Hochberg correction.

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

SUMMARY OF FINDINGS: NO IMPACTS ON TEACHER RETENTION AFTER ONE YEAR

We found that comprehensive teacher induction had no statistically significant impact on teacher retention. 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 4 shows the percentages of treatment and control teachers who stayed in the same school, moved within the profession, and left the teaching profession. The difference in mobility patterns between the two groups was not statistically significant. Even when we collapsed the mobility patterns into summary measures, we found no statistically significant differences between the treatment and control groups.

Table 3. Impacts on Math Test Scores

	Adjusted Mean Test Scores					Unweighted Student Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Total	Treatment Group	Control Group
2	-0.20	0.18	-0.38*†	-0.38	0.000	472	226	246
3	-0.11	0.15	-0.26*†	-0.26	0.002	837	469	368
4	0.01	-0.01	0.03	0.03	0.617	1,545	805	740
5	-0.02	0.02	-0.04	-0.04	0.549	1,510	699	811
6	--	--	--	--		48	24	24
All Grades	**-0.03**	**0.03**	**-0.05**	**-0.05**	**0.184**	**4,412**	**2,223**	**2,189**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

†Significantly different from zero at the 0.05 level, two-tailed test, after applying a Benjamini-Hochberg correction.

Table 4. Impacts on Teacher Mobility, by Destination (Percentages)

Outcome	Treatment	Control	Difference
Stayers			
Stayed at original school	75.0	74.6	0.3
Movers			
Moved, same district	11.2	10.6	0.6
Moved, different district	6.3	7.4	-1.1
Moved, private, parochial, or other school	2.4	1.4	1.1
Leavers			
Left, to stay at home	0.8	1.3	-0.5
Left, in school or new job	3.9	4.2	-0.3
Left, other	0.4	0.5	-0.1
Unweighted Sample Size (Teachers)	**470**	**433**	**903**
Unweighted Sample Size (Schools)	**201**	**193**	**394**

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

Note: Data are weighted to account for the study design.

Difference in the distributions is not statistically significant using a design-based F-test (p= 0.890).

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 better prepared for their jobs as teachers than control teachers. We will repeat these analyses in the coming years when we collect additional follow-up data, at which point we expect there to be more teacher mobility to explain.

SUMMARY OF FINDINGS: NO POSITIVE IMPACTS ON COMPOSITION OF THE DISTRICT TEACHING WORKFORCE AFTER ONE YEAR

The last major research question concerned the impact of comprehensive teacher induction on the composition of the teaching workforce in the district. 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 return to 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 three dimensions: (1) their observed classroom practices; (2) their effect on student achievement; and (3) their professional characteristics such as SAT/ACT scores and advanced degrees. Classroom practice and student achievement outcomes are regression-adjusted using the same sets of covariates used in the main analysis.

We found that the treatment had no positive impacts on the classroom practices, no positive impacts on student achievement (and one statistically significant negative impact), and no significant impacts on the professional background characteristics. Table 5 presents the impacts on classroom practices and student achievement outcomes for district stayers. Table 6 shows the background characteristics of teachers by mobility status.

Correlational Analyses Explore Relationships between Induction and Outcomes

Because a majority of both treatment and control teachers reported receiving induction support (93 percent of treatment teachers and 75 percent of control teachers reported having an assigned mentor), we looked at the relationship between the types and intensity of support and our three main outcome measures: classroom practices, student achievement, and teacher retention. These nonexperimental analyses investigate whether there was a relationship between induction support and outcomes, regardless of treatment status. The analyses mimic the experimental analyses discussed above, using the same covariates and model specification, but replacing the indicator for assignment to treatment status with a measure of induction services. We re-ran the model once for each of 12 measures of

Table 5. Impacts on Classroom Practices and Student Achievement, District Stayers Only

Outcome	Treatment	Control	Difference	Effect Size	P-value
Classroom Practices (Average Score on a 5-Point Scale)					
Implementation of literacy lesson	2.7	2.7	0.0	0.02	0.814
Content of literacy lesson	2.4	2.4	0.0	-0.05	0.586
Classroom culture	3.1	3.1	0.0	0.05	0.613
Unweighted Sample Size (Teachers)	**281**	**236**	**517**		
Unweighted Sample Size (Schools)	**153**	**141**	**294**		
Student Achievement (Effect Size)					
Reading scores (all grades)	0.00	0.01	-0.01	-0.01	0.785
Unweighted Sample Size (Students)	**2,100**	**1,750**	**3,850**		
Unweighted Sample Size (Teachers)	**122**	**99**	**221**		
Unweighted Sample Size (Schools)	**94**	**73**	**167**		
Math scores (all grades)	-0.04	0.04	-0.08*	-0.08	0.037
Unweighted Sample Size (Students)	**1,874**	**1,647**	**3,521**		
Unweighted Sample Size (Teachers)	**113**	**95**	**208**		
Unweighted Sample Size (Schools)	**87**	**70**	**157**		

Source: MPR classroom observations conducted in spring 2006; MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 and Mobility Survey administered in 2006-2007 to all study teachers.

Note: Classroom practice means are weighted and regression-adjusted using ordinary least squares to account for differences in baseline characteristics and the study design. Student achievement means are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. The test scores are expressed as z-scores, with mean of 0 and standard deviation of 1 within each district and grade.

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

induction services measured from the fall survey and again using the measures from the spring survey.[9] The results from these analyses should be viewed cautiously. They should be used to generate hypotheses rather than to establish causal inferences because any association may confound effects of the induction services themselves with the pre-existing differences between the types of teachers who receive different levels of services. For example, those who receive the most support may be the most assertive and effective teachers who are most attached to the profession. Due to the number of analyses conducted, we focus upon the relationships that are statistically significant after applying a Benjamini-Hochberg correction for multiple hypothesis testing within each of the three main outcome domains.

After adjusting for multiple hypothesis testing, none of the relationships between the induction variables and classroom practices was statistically significant. Three of the relationships between the induction variables and student test scores and eight of the relationships between the induction variables and retention measures were positive and statistically significant. Specifically, students of teachers who reported meeting with a subject coach in the fall scored higher on math tests by 0.14 of a standard deviation. The students of teachers who reported receiving feedback on teaching during the fall scored higher on both math and reading tests by 0.02 of a standard deviation per instance that the teacher received feedback. Having an assigned mentor in the spring, receiving guidance in math or literacy content in the spring, each hour spent in the fall on professional development related to content area knowledge, and each hour spent in the fall and spring on professional development related to instructional techniques were associated with a 1 to 6 percentage point increase in the likelihood of remaining in the district or in the teaching profession.

FUTURE RESEARCH

This report focused on the first year of findings only. The research team is conducting longer term followup to include additional collection of test score and teacher mobility data. In addition, the intervention was expanded to include a second year of services for treatment teachers in seven of the districts (4 ETS and 3 NTC), selected based on districts' willingness and ability to continue the program. Future reports will therefore provide evidence on the longer-term effects of both a one-year program in 10 districts and a two-year program in 7 districts.

[9] The twelve induction measures were: whether the beginning teacher was assigned a mentor (yes/no), whether the beginning teacher met with a literacy or math coach (yes/no), whether the beginning teacher worked with a study group (yes/no), whether the beginning teacher observed others teaching (yes/no), whether the mentor gave the beginning teacher suggestions to improve his/her practices (yes/no), whether the beginning teacher received a "moderate amount" or "a lot" of guidance in math content (yes/no), whether the beginning teacher received a "moderate amount" or "a lot" of guidance in literacy content (yes/no), the frequency with which the beginning teacher received feedback on his/her teaching (number of times in a three-month period), time the beginning teacher spent in mentoring sessions (hours per week), time the mentor spent observing the beginning teacher teaching (hours per week), time spent on instruction techniques and strategies as part of professional development activities (hours per three-month period), and time spent on content area knowledge as part of professional development activities (hours per three-month period).

Table 6. Characteristics of District Stayers, Movers, and Leavers by Treatment Status (Percentages Except Where Noted)

	Treatment			Control			Difference		
Teacher Characteristic	Stayers	Movers	Leavers	Stayers	Movers	Leavers	Stayers	Movers	Leavers
College entrance exam scores (SAT combined score or equivalent)	1,000	1,009	1,016	1,009	997	1,059	-9	12	-44
Attended highly selective college	29.6	27.6	42.4	27.9	44.8	32.7	1.7	-17.2	9.7
Major or minor in education	73.5	61.8	70.3	74.0	81.8	51.9	-0.5	-20.0	18.4
Student teaching experience (weeks)	14.5	14.5	12.3	13.9	13.6	11.7	0.6	0.8	0.6
Highest degree is master's or doctorate	20.7	20.5	22.7	21.6	30.5	23.2	-0.9	-10.0	-0.6
Entered the profession through traditional four-year program	64.1	61.7	35.7	60.3	62.3	37.6	3.8	-0.6	-1.8
Certified (regular or probationary)	92.1	97.1	95.7	94.2	94.3	87.0	-2.1	2.8	8.7
Career changer	14.7	10.4	21.1	13.4	15.9	25.8	1.2	-5.5	-4.7
Unweighted Sample Size (Teachers)	**394**	**40**	**23**	**361**	**38**	**26**			
Unweighted Sample Size (Schools)	**188**	**34**	**21**	**180**	**36**	**25**			

Source: MPR calculations using data from the College Board and ACT, Inc.; MPR Teacher Background Survey administered in 2005-2006, Mobility Survey administered in fall/winter 2006-2007, and 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 (190/59/13 treatment stayers/movers/leavers and 183/44/11 control stayers/movers/leavers) and schools (111/36/7 treatment and 106/24/6 control). See Table V.6 for a definition of stayers, movers, and leavers.

None of the differences between treatment stayers and control stayers, between treatment movers and control movers, or between treatment leavers and control leavers is significant at the 0.05 level, two-tailed test. P-values are suppressed to make the table easier to read.

CHAPTER I

INTRODUCTION AND BACKGROUND

Policymakers and researchers have 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), though 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 46 percent of beginning teachers leave the classroom within five years (Ingersoll 2003).

High teacher turnover can have several 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 in other ways (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 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. As a result, teacher induction is common, but induction that is intensive, comprehensive, structured, and sequentially delivered in response to teachers' emerging pedagogical needs is less so. 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 estimated in recent literature range from $1,660 to $6,605 per teacher per year (Villar and Strong 2007; Alliance for Excellent Education 2004).[10] 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), little of the research on teacher induction to date has been conclusive or 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 (e.g., Fuller 2003; Youngs 2002) has yielded more detailed descriptions of teacher supports but, like the national studies, has relied on evaluation designs that leave doubt about whether the inferences are causal. For example, 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). These non-experimental approaches cannot be treated as causal estimates of the impact of interest: the retention rate for participants compared to *what it would have been* in the absence of the program.

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 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

[10] These reports note costs for 5 programs, 4 of which are 2-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 3 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).

and teacher mentoring programs, with a call to fund "proven models" to meet these objectives. In addition, the Higher Education Opportunity Act of 2008 authorizes grants that include teacher induction or mentoring programs for new teachers. These initiatives demonstrate 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 (ED) Institute of Education Sciences (IES) contracted with Mathematica Policy Research, Inc. (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 in, respectively, approximately half of the 17 districts participating in the study. IES later expanded the intervention to include a second year of services for seven of the districts (4 ETS and 3 NTC), selected based on the districts' willingness and ability to continue the program.

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 to the implementation process to help ensure that there was fidelity 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 elementary schools within each of 17 participating districts to either a treatment group, which received comprehensive teacher induction—from ETS or NTC, depending on the district—or 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. CONCEPTUAL BACKGROUND FOR THE STUDY

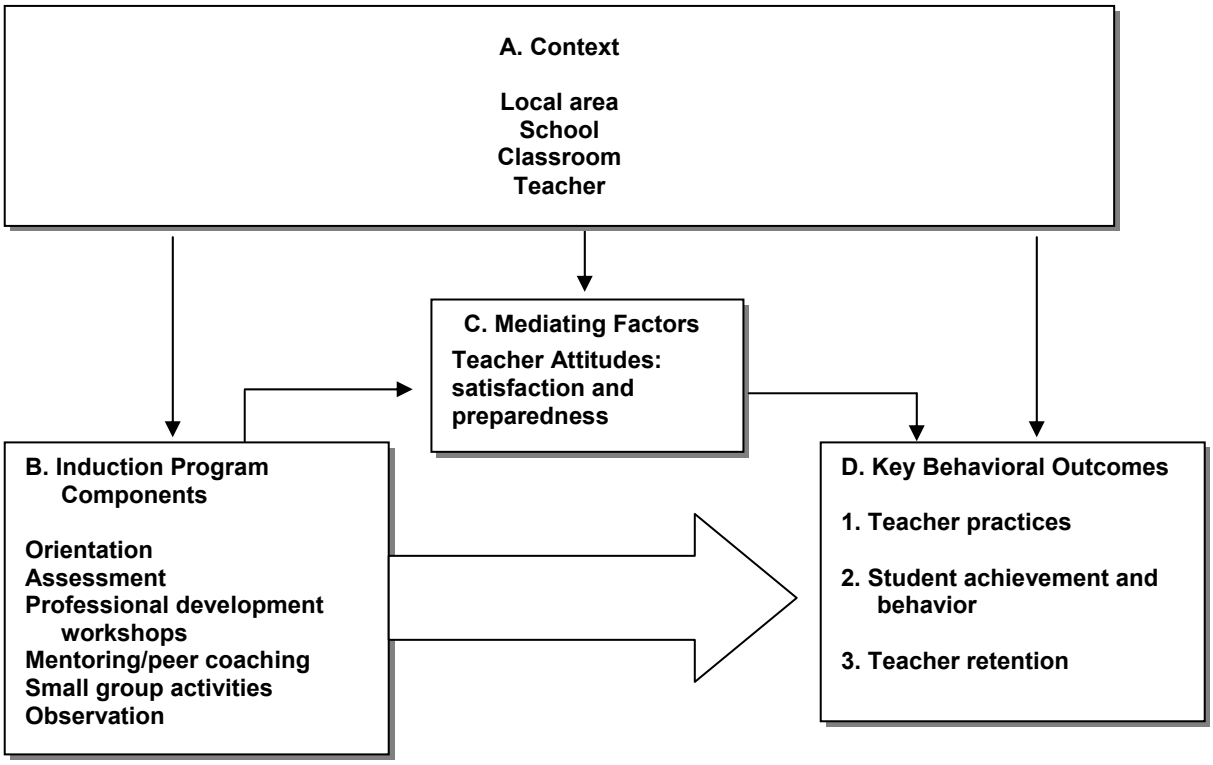
To answer the research questions, we began by identifying the pathways by which teacher induction programs may lead to teacher and student outcomes. Figure I.1 illustrates the pathways and highlights some of the contextual factors that are useful to consider when planning and interpreting analyses. More 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 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 (Box B, Figure I.1). 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. The more intense the emphasis on a given component, the larger is its effect on outcomes—presumably. 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) and compared outcomes for teachers in this group with 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 skills (Box D, Figure I.1). Improving teacher practices is not only a key outcome for teachers but also helps explain possible impacts on retention and student achievement.

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



Induction may affect several intermediate factors (Box C, Figure I.1) that may help explain changes in final outcomes. For instance, a possible precursor to teacher mobility is dissatisfaction or feeling 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 (Box D, Figure I.1). Improvements in the teaching force 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.

C. 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 (Chapter V), followed by correlational analyses conducted to add context to the main experimental findings (Chapter VI).

The report presents findings on classroom practices measured during the spring semester of the teacher's first year in the profession, student achievement growth demonstrated during the teacher's first year, and teacher retention after the first year in the profession. Future reports will update this one with longer term follow-up from additional collection of test score and teacher retention data.

CHAPTER II

STUDY DESIGN AND METHODS

This chapter documents the study design. The centerpiece of the design 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). We discuss the methods for selecting districts, schools, and teachers for inclusion in the study and for constructing the randomly assigned treatment and control groups. We then describe the data analysis methods. Though we undertook a purposeful selection of districts and schools, we then randomly assigned schools to a treatment or control group. Figure II.1 provides an overview of the sample selection process.

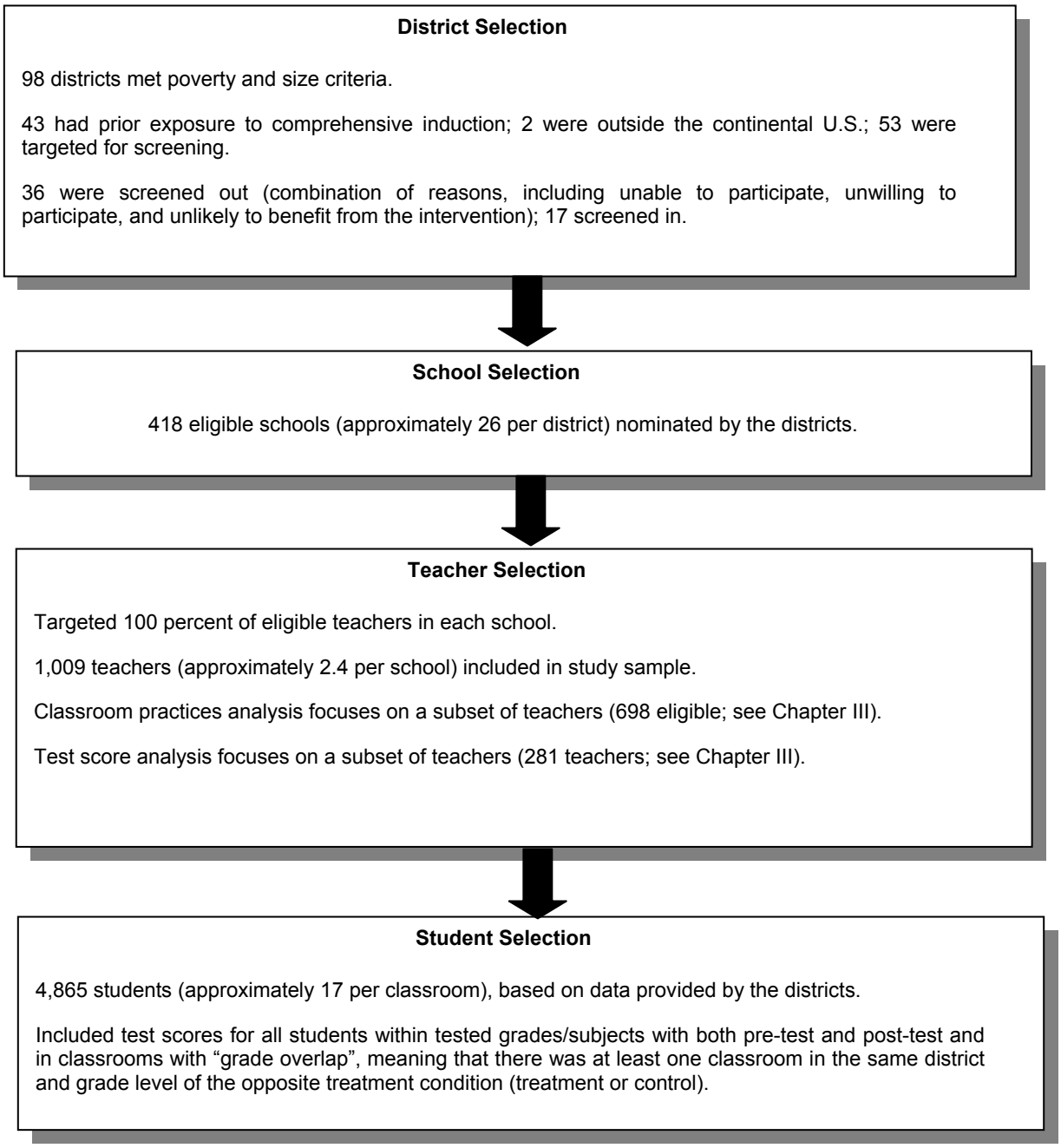
A. SELECTION OF DISTRICTS

We selected the initial list of targeted districts 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 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 School Lunch Program. We developed these size and poverty targets in consultation with IES, based on our earlier feasibility analysis (see Glazerman et al. 2005). Ninety-eight districts nationally 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. We also eliminated 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.

Figure II.1. Sample Selection Flow Chart



At the end of the screening and recruiting process, we had a final sample of 17 school districts in 13 states. By selecting districts that both met our criteria and 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 a study of teacher induction. See Appendix A for an analysis of data from the Schools and Staffing Survey (Tourkin et al 2007) that compares the level of induction services in the study districts to that in the full set of districts that met the criteria for inclusion in the study and to that in a national sample of school districts.

We assigned each district 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. Such comparisons would confound differences in the districts each provider works with, and such differences cannot be guaranteed to cancel out one another.

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 federal School Lunch Program. The study included districts serving mostly African American students (7 of the 17 districts), Hispanic students (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 Texas to Delaware), 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.

It is important to note that the two induction program providers worked in different types of districts. For example, more of the NTC districts than ETS districts (six versus three) had over 50,000 students and NTC districts were more likely to serve mostly African American students (five out of eight districts) compared to the ETS districts (two out of nine). The two types of districts were balanced regionally, with three states including at least one ETS district and at least one NTC district. Throughout this report, we present findings for the entire sample (both types of districts) combined. However, in order to understand whether the impacts are peculiar to one provider, we present results in appendices that show the findings by district type (ETS or NTC).

B. SELECTION OF SCHOOLS AND TEACHERS

Within each district, a fixed set of elementary schools in which to conduct the study was selected. Large districts exercised some discretion over the subset of schools considered for the study. Otherwise, we selected all schools with eligible teachers. We selected all the teachers within those schools that met certain eligibility criteria, as follows:

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

Table II.1. Characteristics of Districts in Teacher Induction Sample, by Induction Provider

District Characteristic	Number of Districts			Percent
	ETS	NTC	All	
Demographics				
Low Income (Percent Eligible for School Lunch Program)				
<65	2	0	2	11.8
65–70	1	0	1	5.9
70–75	2	4	6	35.3
75–80	2	2	4	23.5
80–85	1	2	3	17.6
>85	0	0	0	0.0
Unknown (data not available)	1	0	1	5.9
Race/Ethnicity				
Majority African American	2	5	7	41.2
Majority Hispanic	1	1	2	11.8
Majority white	3	0	3	17.6
No single majority group	3	2	5	29.4
Region				
Northeast	2	2	4	23.5
Midwest	2	2	4	23.5
West	0	0	0	0.0
South	5	4	9	52.9
District Size				
Student Enrollment				
5,000–25,000	1	0	1	5.9
25,000–50,000	5	2	7	41.2
50,000–100,000	2	3	5	29.4
More than 100,000	1	3	4	23.5
Number of Elementary Schools				
Fewer than 50	4	2	6	35.3
50–100	3	2	5	29.4
More than 100	2	4	6	35.3
Study Sample				
Number of Mentors				
2	6	5	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	4	4	8	47.1
50–74	3	3	6	35.3
75–100	2	0	2	11.8
More than 100	0	1	1	5.9
Unweighted Sample Size (Districts)	**9**	**8**	**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.

- ***New to the Profession.*** We encountered some teachers who reported prior teaching experience in some capacity, even if the district did not recognize such experience. The most relevant criteria were (1) that the district considered such teachers as new from the perspective of eligibility for beginning-teacher induction services and (2) that the method for identifying teachers for the study was applied consistently for all schools within each district.

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

We ultimately included 418 elementary schools in the study across the 17 districts. Table II.2 shows the percentages of schools serving poor students, minority students and the grade configurations of the schools. Most of the schools (80 percent) employed one, two, or three eligible beginning teachers. Statistics are shown for ETS and NTC districts separately in Appendix E.

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.

1. Method of Random Assignment

The most feasible approach was random assignment of schools such that 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).

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 accounts for the considerable variation between districts in the policies, student populations, and environments that could affect the study's outcomes.

Table II.2. School Characteristics By Treatment Status (Percentages)

School Characteristic	All Schools	Treatment	Control	Differ-ence	P-value
Percent Eligible for School Lunch Program					0.441
<50%	8.0	9.3	6.7	2.6	
50–75%	20.6	17.7	23.5	-5.8	
75–100%	64.3	66.1	62.5	3.6	
Unknown	7.2	7.0	7.4	-0.4	
Race/Ethnicity					0.476
Majority African American	44.2	43.8	44.5	0.6	
Majority Hispanic	21.7	24.5	18.9	5.5	
Majority white	16.8	16.2	17.4	-1.2	
Majority other	0.2	0.5	0.0	0.5	
Other/mixed	17.1	15.0	19.2	-4.2	
Grade Configuration					0.012*
Pre-K or K–5	71.1	72.9	69.4	3.5	
Pre-K or K–6	2.9	0.4	5.2	-4.8	
Pre-K or K–8	20.5	19.5	21.6	-2.1	
Other	5.5	7.3	3.8	3.5	
Number of Sample Teachers					0.317
1	37.8	35.6	40.0	-4.4	
2	23.9	25.4	22.5	2.9	
3	18.1	20.7	15.5	5.2	
4	8.7	9.5	7.9	1.6	
5	6.1	5.4	6.8	-1.4	
More than 5	5.3	3.4	7.2	-3.8	
Unweighted Sample Size (Schools)	**418**	**210**	**208**		

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

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.

*Difference in distributions is significant at the 0.05 level, two-tailed test.

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 randomly selected one allocation with equal probability. The admissible allocations were those that achieved an appropriate degree of balance between the treatment and control groups in terms of overall number of eligible teachers and teaching

assignment (grade level).[11] 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 through II.6 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 (a majority of students were African American in approximately two-fifths of the schools, a majority Hispanic in one-fifth), as shown in Table II.2. Table II.3 presents demographic characteristics of the teachers in the study. Of 953 teachers responding to our baseline survey, similar percentages of treatment and control group members were white (60 and 62 percent, respectively), female (88 and 89 percent), under age 25 (49 and 48 percent), married (55 and 56 percent), and had no children at home (70 percent). Table II.4 describes the teachers' professional backgrounds. Similar percentages of treatment and control teachers had advanced degrees (20 and 23 percent), earned bachelor's degrees from highly selective colleges[12] (31 and 29 percent), had an education major or minor (71 and 73 percent), entered teaching through a traditional four-year college route (61 and 59 percent), held a regular teaching certificate (61 and 56 percent), and entered the profession with no student teaching (20 percent). For those teachers who gave us permission to obtain their SAT or ACT scores from the test publishers and for whom we found the scores, we found no statistically significant differences in scores between the treatment and control teachers (Table II.5).

There were statistically significant differences between treatment and control groups in one area: teachers' assignments. The control group contained a higher percentage of special subject teachers (such as art and music) than did the treatment group (7 versus 3 percent) and consequently a lower percentage of teachers who taught just a single grade (79 versus 85 percent) and who said they were responsible for math (85 versus 90 percent) or reading outcomes (83 versus 91 percent). See Table II.6. 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 classroom practices and 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.

[11] 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.

[12] 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*.

Table II.3. Teacher Demographic Characteristics by Treatment Status (Percentages)

Teacher Characteristic	All Teachers	Treatment	Control	Difference	P-value
Gender					0.407
Male	11.5	12.4	10.6	1.8	
Female	88.5	87.6	89.4	-1.8	
Race/Ethnicity					0.124
White, non-Hispanic	61.3	60.2	62.4	-2.2	
African American, non-Hispanic	19.2	21.5	16.7	4.8	
Hispanic	15.1	13.1	17.2	-4.1	
Other/mixed	3.5	3.6	3.4	0.2	
Unknown	0.9	1.6	0.2	1.4	
Age (Years)[a]					0.809
20–25	48.6	49.0	48.2	0.8	
26–29	19.5	19.4	19.6	-0.2	
30–39	20.2	19.0	21.5	-2.5	
40–49	9.1	9.8	8.4	1.4	
50 or more	2.6	2.9	2.3	0.6	
Marital Status					0.702
Married or living with a partner	55.5	54.8	56.2	-1.4	
Single, separated, divorced, or widowed	44.5	45.2	43.8	1.4	
Children Living in the Home					0.754
None	70.0	70.2	69.8	0.4	
1 or more children under 5 years old	14.6	15.2	14.0	1.2	
1 or more children, none under 5 years old	15.4	14.6	16.2	-1.6	
Unweighted Sample Size (Teachers)	**953**	**489**	**464**		

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.4. Teacher Professional Background by Treatment Status (Percentages)

Teacher Characteristic	All Teachers	Treatment	Control	Difference	P-value
Highest Degree Earned					0.136
Associate's	0.8	1.3	0.2	1.1	
Bachelor's	77.5	78.2	76.9	1.3	
Master's	21.5	20.0	22.9	-2.9	
Doctoral	0.3	0.5	0.0	0.5	
Earned a Bachelor's Degree from a Highly Selective College	30.0	30.7	29.1	1.6	0.595
Earned a Degree with Education-Related Major or Minor	71.9	70.9	72.8	-1.9	0.586
How Entered the Profession					0.899
Traditional program (4-year)	60.2	61.0	59.3	1.7	
Traditional program (post-baccalaureate)	16.7	16.3	17.2	-0.9	
Teach for America	3.2	3.3	3.0	0.4	
Other alternative preparation program	16.8	15.8	17.9	-2.0	
Other/unknown	3.2	3.5	2.7	0.8	
Career Changer	14.1	14.2	13.9	0.3	0.891
Teaching Certificate					0.174
Regular	58.4	60.8	55.8	5.0	
Probationary	35.0	31.6	38.6	-6.9	
Emergency/waiver	5.5	6.1	5.0	1.1	
Other	1.1	1.5	0.7	0.8	
Weeks of Student Teaching					0.770
Zero	20.3	20.3	20.2	0.1	
1–12	19.2	17.9	20.6	-2.8	
13–16	36.6	36.8	36.4	0.4	
17 or more	23.9	25.0	22.8	2.4	
Unweighted Sample Size (Teachers)	**953**	**489**	**464**		

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.

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

Table II.5. Teacher College Entrance Exams by Treatment Status

Teacher Characteristic	All Teachers	Treatment	Control	Difference	P-value
College Entrance Exam Scores (Percentages)					0.271
Did not take exam	11.3	10.4	12.2	-1.8	
Did not provide consent to obtain scores	20.8	19.7	22.1	-2.4	
Scores not found	11.1	13.1	9.0	4.1	
Scores reported	56.8	56.9	56.7	0.2	
SAT Combined Score (or ACT Equivalent)	1008	1003	1013	-10	0.548
Unweighted Sample Size (All Teachers)	**1,009**	**506**	**503**		
Unweighted Sample Size (Teachers with usable ACT or SAT Scores)	**526**	**271**	**255**		

Source: MPR calculations using data from the College Board and ACT, Inc.

Note: ACT scores were converted to SAT score equivalents using concordance tables found in Dorans et al. (1997).

None of the differences is statistically significant 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 researchers' ability to draw inferences about the effectiveness of the intervention on the people receiving it. 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

Table II.6. Teaching Assignments by Treatment Status (Percentages)

Teacher Characteristic	All Teachers	Treatment	Control	Difference	P-value
Grade Level					0.311
Kindergarten	15.7	15.7	15.7	0.0	
Grade one	14.8	14.3	15.3	-1.0	
Grade two	15.3	17.1	13.3	3.8	
Grade three	13.4	14.6	12.1	2.5	
Grade four	11.6	12.4	10.7	1.7	
Grade five	9.0	8.6	9.5	-0.9	
Grade six	0.7	0.6	0.9	-0.3	
Multiple, other	19.5	16.6	22.6	-6.0	
Responsible for Reading Outcomes	87.1	91.4	82.6	8.7*	0.000
Responsible for Mathematics Outcomes	87.6	90.1	84.9	5.2*	0.024
Subject Specialty[a]					
Teaches only one grade level	82.4	85.3	79.3	6.0*	0.040
Specialist: bilingual, ESL, or ELL	1.2	1.2	1.3	-0.1	0.934
Specialist: special education	6.5	6.1	7.0	-0.9	0.615
Specialist: core academic subjects (e.g., reading, social studies, mathematics, science)	1.8	1.3	2.4	-1.0	0.261
Specialist: other subjects (e.g., computers, foreign language, art, music, gym)	4.6	2.5	6.7	-4.2*	0.003
Teaching in Preferred Grade and Subject	79.6	81.6	77.6	4.0	0.138
Unweighted Sample Size (Teachers)	**953**	**489**	**464**		

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 .05 level, two-tailed test.

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 via field reports that we 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 they would not be mentored (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. Overall, however, noncompliance was relatively rare—the 15 teachers made up 3 percent of the treatment group—and some noncompliance is always to be expected, even in the absence of experimental conditions. We discuss the degree of program dropout in Chapter IV. We included all sample members in the impact analysis regardless of compliance status and classified them 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 (see Chapter III, Table III.1), yet we observed differences in response rates by treatment status that were statistically significant. For example, the largest difference was for the spring 2006 induction activities questionnaire, where the control group response rate was 83 percent and the treatment group response 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, 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. 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 B. 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 the outcomes that *would have been observed* in the absence of the comprehensive program. To that end, we examined whether classroom teaching practices, 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 C details the methods we used for estimating the impacts of the comprehensive induction programs as well as the alternative estimation approaches we used for testing the robustness of the study's findings. We illustrate the effect of alternative approaches by using a benchmark model that imposes the most reasonable set of assumptions and measurement rules and then compares them to a set of alternatives that implement deviations—one at a time—from that benchmark. For example, the benchmark model specifies a set of variables used as covariates for regression adjustment of the impact estimates. The set of benchmark covariates differs for each outcome.

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

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

Therefore, we use a model-based approach to estimate program impacts. The statistical model not only allows us to represent the non-independence of observations explicitly, it also allows us to exploit the data on 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 student achievement, for example, we can reduce the amount of unexplained variation in test scores 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, we can properly specify the units of analysis and devise unbiased estimates of

the standard errors that we used to conduct hypothesis tests. 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 C 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. We computed effect sizes as the impact divided by the standard deviation of the outcome variable. The standard deviation is computed using the full sample (treatment and control group).

CHAPTER III

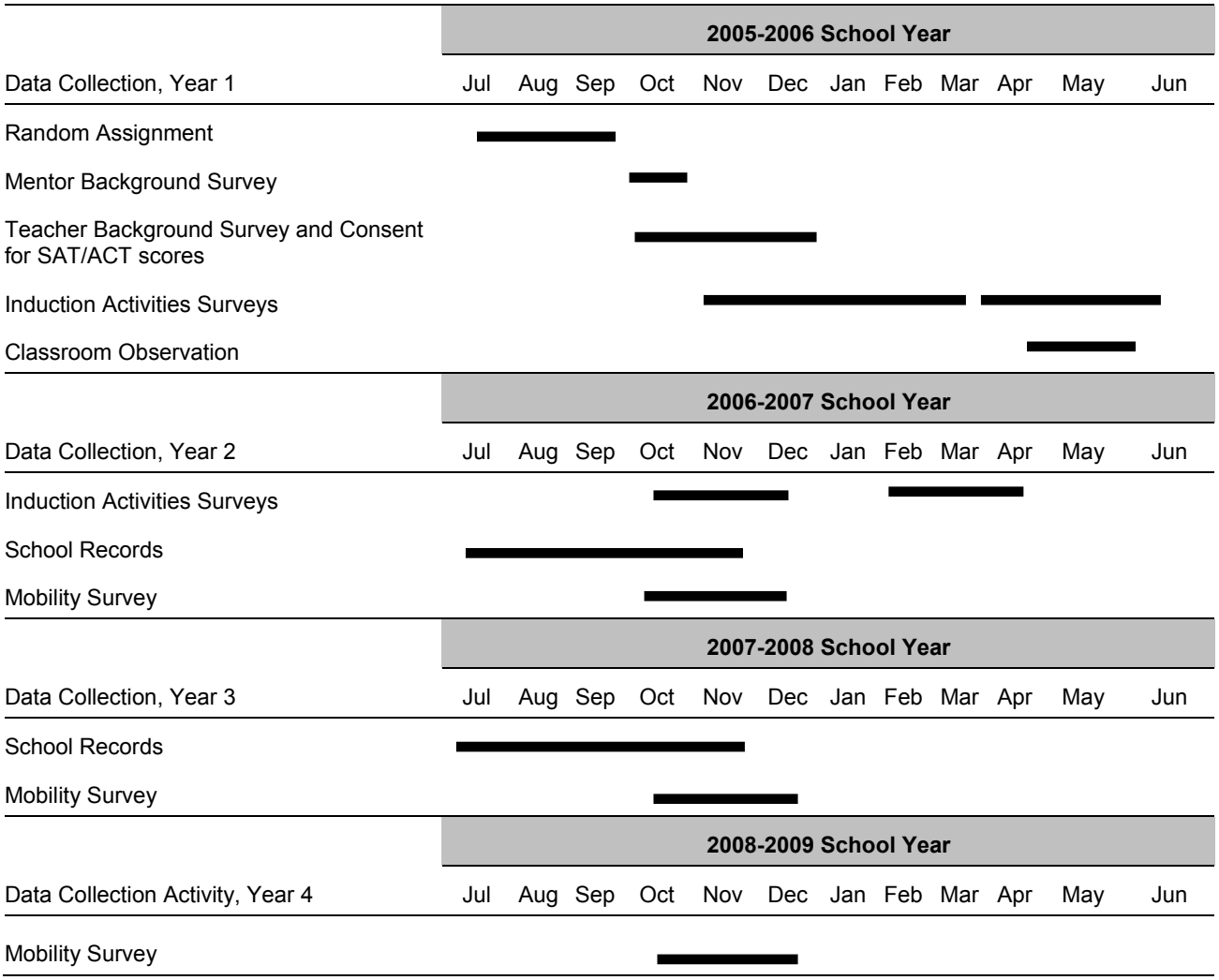
DATA

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 baseline 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 teachers twice during the 2005-2006 school year on the induction activities they participated in, and we surveyed mentors on their background characteristics and reviewed program documents from ETS and NTC.

For the study's core outcomes, we observed the teachers teaching a literacy unit in the classroom, collected districts' student records data at the end of the school year, and conducted the first of three mobility surveys in fall 2006 to learn about teacher retention. In the future, we plan to collect another year of student records data and, to help us track mobility patterns, continue following study teachers with a survey administered in fall 2007 and fall 2008.

The data collection effort was most intense during the 2005-2006 school year, while the induction programs were being implemented in the treatment schools. Figure III.1 shows a time line for the data collection activities. The current report presents the findings from the Year 1 data collection activities plus the first mobility survey and the first round of school records data. 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.2 presents a flow diagram of sample members that explains how we derived our analysis sample from the sample of originally identified teachers.

Figure III.1. Data Collection Schedule



We achieved response rates on teacher surveys and observations that ranged from 92 percent to 97 percent for the treatment group and 83 percent to 92 percent for the control group (Tables III.1 and III.2.). Table III.3 shows the rates for different subgroups. Despite overall response rates above 80 percent, the control group response rates persistently fell below treatment group response rates 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 and the full set of respondents and nonrespondents on observable school characteristics—the only data that are available for respondents and nonrespondents—are not statistically significant (see Table III.4).

Figure III.2. Flow of Teachers Through Study

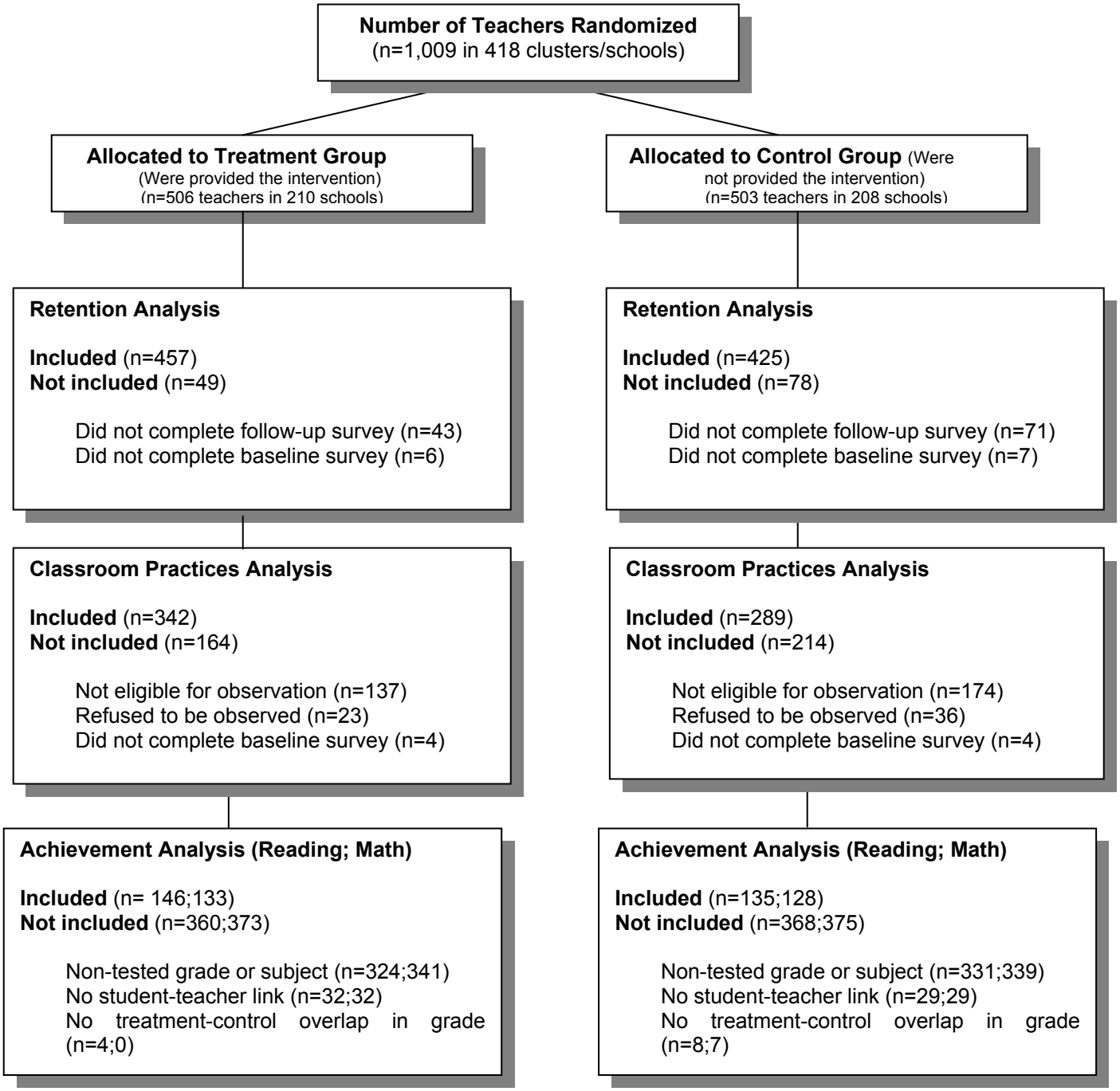


Table III.1. Response Rates by Treatment Status

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
Classroom Observations	698[a]	91.6	93.8	89.1
Teacher Mobility Survey	1,009	88.7	91.5	85.9

Source: MPR teacher induction survey management system.

[a]Teachers who did not teach reading in spring 2006 (e.g. art, music, or math specialists and teachers who left the profession) were not eligible for classroom observation.

n.a. = not applicable.

Table III.2. Response Status to Classroom Observation and Reasons for Nonresponse

Status/Reason	Number of Teachers	Percentage of Teachers	Percent of Eligible Teachers		
			All Eligibles	Treatment	Control
Eligibles					
Completes	639	63.3	91.6	93.8	89.1
Refusals	59	5.6	8.5	6.2	10.9
Ineligibles					
Does not teach reading in classroom setting	175	17.3	n.a.	n.a.	n.a.
Not teaching	64	6.3	n.a.	n.a.	n.a.
Not beginning teachers	70	6.9	n.a.	n.a.	n.a.
Other ineligible	2	0.2	n.a.	n.a.	n.a.
Total	**1,009**	**100.0**	**100.0**	**100.0**	**100.0**

Source: MPR teacher induction survey management system.

n.a. = not applicable

Table III.3. Response Rates to Teacher Surveys, by Subgroup and Treatment Status

	Response Rate (Percentages)							
	Background		Induction Activities 1		Induction Activities 2		Mobility	
	Treatment	Control	Treatment	Control	Treatment	Control	Treatment	Control
District Type (Program Provider)								
ETS	95.8	93.0	93.1	86.3	92.3	83.4	91.5	88.9
NTC	97.6	91.4	93.5	82.8	92.7	82.3	91.5	82.3
Grade Level								
K or Pre-K	96.3	97.2	95.0	90.3	92.5	90.3	95.0	91.7
1	98.6	97.2	95.9	94.4	95.9	87.3	95.9	90.1
2	97.6	91.0	95.2	78.2	89.3	76.9	91.7	89.7
3	97.5	94.7	95.1	86.0	96.3	80.7	90.1	86.0
4	96.7	91.7	95.0	88.3	93.3	86.7	91.7	85.0
5	100.0	96.2	95.7	88.5	97.8	90.4	93.5	92.3
Other/multiple	91.5	84.1	82.9	75.2	85.4	75.2	84.1	74.3
School Type (Percent in Free Lunch Program)								
Unknown	90.0	96.6	83.3	75.9	93.3	79.3	80.0	79.3
0–25%	100.0	100.0	100.0	66.7	100.0	66.7	83.3	66.7
25–50%	100.0	92.3	93.5	73.1	93.5	73.1	96.8	88.5
50–75%	95.9	91.4	92.9	84.4	91.8	81.3	90.8	86.7
75–100%	97.1	92.1	94.1	86.8	92.4	84.9	92.4	86.1

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

To reduce any possible bias that nonresponse may cause, we conducted a nonresponse analysis and created nonresponse adjustment weights (see Appendix B). The weights allow 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 Chapter V).

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 the professional backgrounds of these individuals, information that can be used to understand program implementation.

Table III.4. School Characteristics of Respondents and Nonrespondents

	Respondents Only			
	Background Survey (N=953)	Induction Activities Surveys (N=936)	First Mobility Survey (N=895)	Respondents and Nonrespondents (N=1,009)
Percent Free Lunch in School				
Unknown	5.8	5.7	5.3	5.9
0–50%	6.7	6.1	6.7	6.5
50–75%	22.1	22.0	22.4	22.4
75–100%	65.4	66.2	65.7	65.2
Percent White in School				
Unknown	0.9	1.0	1.0	0.9
0–50%	81.1	81.1	80.8	81.4
50–75%	16.4	16.4	16.5	16.3
75–100%	1.6	1.6	1.7	1.5
Percent Black in School				
Unknown	0.9	1.0	1.0	0.9
0–50%	59.3	59.7	59.8	59.8
50–75%	6.9	6.4	7.4	6.8
75–100%	32.8	32.9	31.8	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) is statistically significant at the 0.05 level, two-tailed test.

During the ETS and NTC mentor training sessions in late summer and 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 the background survey was a consent form asking teachers to permit the research team to obtain their college entrance exam scores, either SAT or ACT scores. The college entrance exam scores provide an objective measure of teachers' cognitive ability before they received any special preparation to become a teacher. 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 is important to understand the differences in the services delivered by the comprehensive and prevailing district programs. To that end, we administered a survey of teacher induction activities to both treatment and control teachers twice: in January 2006 and April 2006.[13] Given that the nature of induction activities may change often during the school year, the administration of two surveys reduced any difficulties teachers may have had in recalling induction activities over the full year, 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. We report on data from the spring survey, and also present findings from the fall survey when they differed significantly from the spring results.

The survey included questions applicable to services delivered by both the comprehensive programs and prevailing programs. The survey asked questions about mentoring from any source, the 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 felt prepared for their first year of teaching and satisfied with various aspects of teaching. We mailed the surveys and followed up by telephone, in some cases using field interviewers to complete the survey in person to achieve a high response rate.

3. Teacher Mobility Survey

We sent surveys to all teachers in fall 2006 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 2007 and fall 2008 to identify teachers who moved or left teaching after two and three years on the job. As with the other teacher

[13] The two induction activities surveys were administered over a period that stretched from November to early March and late March to June, respectively. A large share of the surveys were returned in January and March. 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.

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. CLASSROOM OBSERVATIONS

We observed classrooms of teachers in the treatment and control groups to measure their classroom practices in the area of reading and literacy. We excluded from this data collection any teachers who were responsible for small classes such as special education resource teachers, taught special populations such as bilingual classes, taught mathematics only, were not first-year teachers, or were no longer teaching in the district. Thus, the eligible sample for the classroom observations (698) was smaller than the full study sample (1,009), as shown in Table III.2.

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

The observations focused on pedagogical practices and classroom management. Classroom observers, all of whom had teaching experience and underwent training for this study, visited classrooms in late spring 2006 (toward the end of the treatment year), when differences in teacher practices due to the comprehensive induction program would most likely be evident. They were blind to the treatment status of the classrooms they observed.

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

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

Observers scored teachers in each of three constructs based on a set of items that are believed to be indicators of good practice: implementation of a lesson, content of a lesson, and classroom culture. The three domains comprise five, four, and seven items, respectively. Observers rated the extent of evidence of teacher behavior for each item on a five-point scale showing (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence. Examples of the indicators include the following. For lesson implementation: "The pace of the lesson is appropriate for the developmental level of the students." For literacy content: "Understanding of content and concepts is taught through *close reading of text* and vocabulary instruction." For classroom culture: "Classroom management maximizes learning opportunities." The VCOT provides observers with examples of specific behaviors to look for in assessing the extent of evidence of teacher practice within each item. We found all items within each of the three literacy constructs to be highly correlated with other items in the construct, based on standardized inter-item reliability coefficients.[14] To determine how each item reflected the reliability within each of the three constructs, we calculated an alpha coefficient after deleting each item independently from the construct. Psychometric details are presented as an appendix to Chapter V (Appendix G), where the findings based on the observation measures first appear.

D. STUDENT RECORDS

To gauge whether induction has any impact on student achievement, we collected student records data from all 17 districts for students in both treatment and control classrooms. The data included scores from the standardized tests administered by the districts during spring 2006 (post-test) and 2005 (pre-test) as well as student background data such as race/ethnicity, date of birth (to determine over age for grade), eligibility for free or reduced-price meals under the federal School Lunch Program, and disability status.[15]

Aggregating test score data across many districts and grades poses a serious challenge, but the fact that we made treatment-control comparisons within grade within districts means that it was only necessary to standardize tests and testing conditions within each grade within

[14] MPR calculated the psychometric properties of the VCOT by using preliminary data from a national evaluation of teacher preparation models (see Decker et al. 2005) in which we administered the same instrument under similar circumstances. The standardized inter-item reliability coefficients, Cronbach's alphas, were 0.89 (implementation), 0.75 (content), and 0.91 (culture). Inter-rater reliability indices from the publisher are not available. In the current study, observers were deemed certified to conduct observations based on a comparison of their 16-item scores to the observations of a "gold standard" panel; following certification, however, inter-rater reliability was not measured in the field.

[15] One district tested students in the fall of each year.

each district. Most scores were state assessments that we standardized such that each score described student performance relative to all other students in the same grade and district, a common standard for treatment and control groups. Scores were scaled scores, normal curve equivalents, or percentile rankings. We re-scaled all tests to have a common mean (0) and variance (100). Chapter V further details the aggregation process.

E. 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 first-year 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.

The first section of this chapter describes the intervention provided to the treatment group during the 2005-2006 school year. The second section compares the 2005-2006 induction experiences of teachers in the treatment group with that of the teachers in the control group. The gap in services, or service contrast, represents the impact of offering the treatment on the types and intensity of induction services received. It is an important precursor to impacts on desirable outcomes such as improved classroom practices and student test scores.

A. COMPREHENSIVE TEACHER INDUCTION

To test the hypothesis that a comprehensive teacher induction program would be more effective than the services normally provided to beginning teachers by their schools and districts, we had to identify such a program as well as a provider of program services. Accordingly, we issued a Request for Proposals (RFP) in 2004. The RFP specified that the induction program should include components that earlier research and professional wisdom gleaned from practice had suggested were important features of successful teacher induction programs (Alliance for Excellent Education 2004; Ingersoll and Smith 2004; Smith and Ingersoll 2004; Kelly 2004; Serpell and Bozeman 2000). The components include carefully selected and trained full-time mentors; a curriculum of intensive and structured support for beginning teachers 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.

A group of outside expert reviewers read and scored the proposals we received in response to the RFP. Among the proposals received, those submitted by ETS and NTC stood out as most closely meeting the study's specified requirements. We selected the ETS and NTC programs because we were interested in 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 increases our ability to generalize the findings of the comprehensive induction model, as compared to including just one program. Furthermore, the expert panel that was convened to select the study's intervention rated both the ETS and NTC programs as high in quality, and the panel agreed that both programs were similar enough in goals and structure that including both in the study (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. 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 both interviewed program leaders and staff and received reports from them regularly, from weekly at start-up to monthly later in the school year. For each program, the monitors also observed one initial local orientation for beginning teachers, another one for administrators, and an end-of-year Colloquium for beginning teachers.

WestEd monitors visited every district in the fall and, in the spring, either visited again or conducted semi-structured telephone interviews.[16] 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; 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.

[16] Four of the 9 ETS districts (44 percent) and 3 of the 8 NTC districts (38 percent) received a visit. The others received a telephone call.

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, 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 programs for the study to ensure that the required components were included in a one-year curriculum. Finally, the providers adjusted their usual methods of service delivery to meet the requirements of the study. 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.

1. Delivery of Induction Services

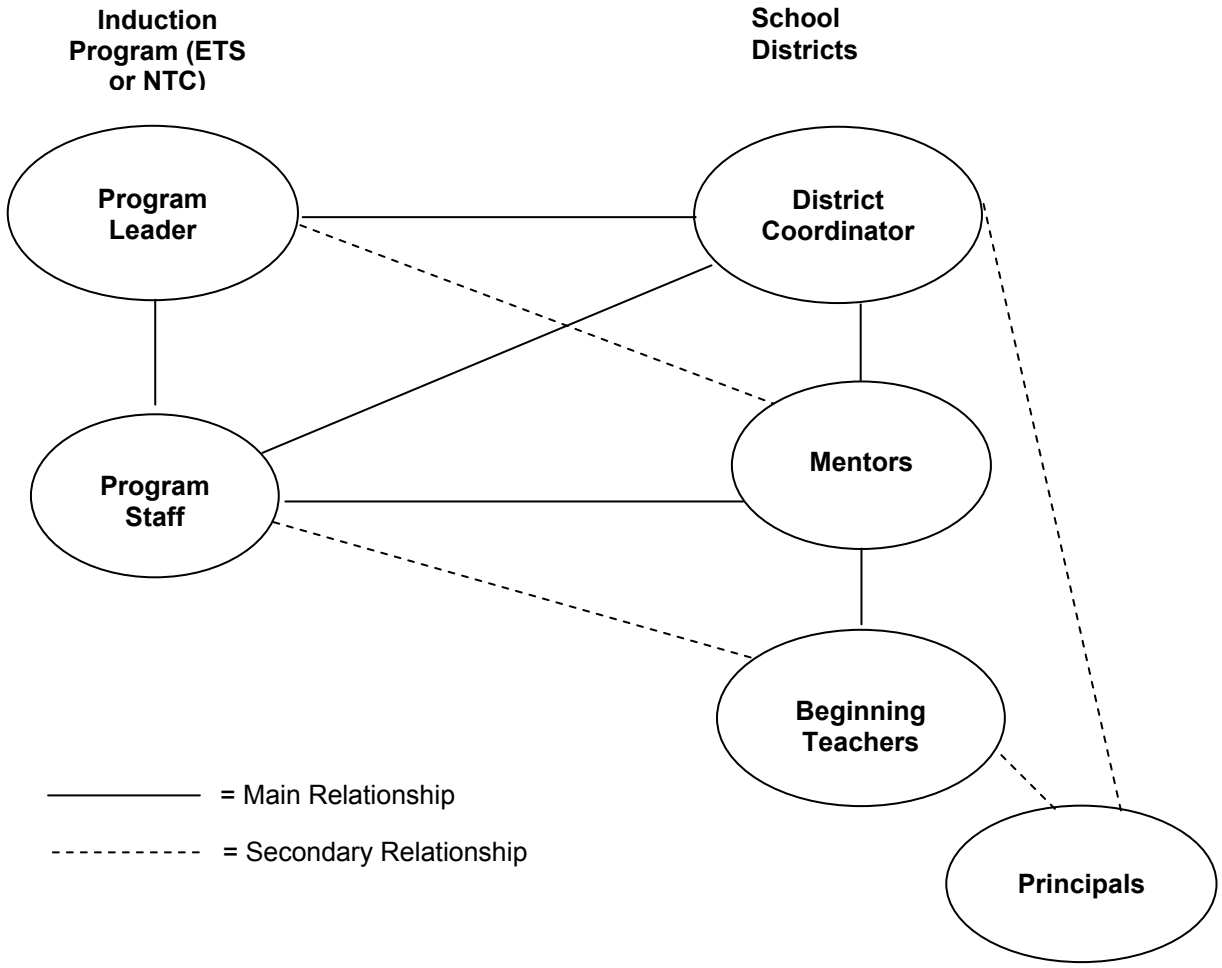
To understand the treatment provided by each program, we begin with an overview of the key roles played by designated staff members in implementing the programs (Figure IV.1). Oversight for implementation of the ETS and NTC programs was the responsibility of a designated staff member from the respective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

In collaboration with the program leaders, designated ETS and NTC **program staff** worked with assigned districts to help implement the program consistently across the districts.[18] 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 any 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, thereby providing these 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.

[17] In addition, WestEd staff provided external oversight of service provision to help address any issues that arose and keep implementation consistent across all sites.

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

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



Districts designated their own staff members to provide local oversight to program implementation. **District coordinators** worked in departments of human resources or professional development. They helped establish district positions for mentors and recruited candidates, established procedures for job reporting and evaluation, created functional working conditions for mentors by locating office space and setting up email and telephone access, facilitated mentors' weekly meetings, and joined mentors at off-site trainings throughout the year. The district coordinators 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. 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. 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).

According to WestEd's monitors' interviews with district coordinators, coordinators 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. In two districts, the coordinator position turned over during the study period. Given that the coordinator role was an addition to a full set of existing responsibilities, coordinators struggled to carve out the time needed for program implementation.[19]

Principals also played an important role in program implementation. Both ETS and NTC asked principals to encourage and support beginning teachers' participation in induction activities, particularly by permitting them to attend professional development sessions and minimizing conflicts that could impede mentors' efforts to schedule time with beginning teachers. Both programs offered an initial orientation for administrators, and NTC held a fall and spring administrator briefing over breakfast. On average, fewer than half of administrators attended the orientations. 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, schools and districts evidenced wide variation in the level of principal support, ranging from principals 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.[20] Five principals out of the 210 treatment schools in the study fell into this latter category. Such resistance abated over the course of the year in response to the intervention of district coordinators, mentors, and program staff. The resistant principals either required beginning teachers to attend school or district events that conflicted with induction program events or imposed heavy restrictions on when mentors could visit teachers. Induction programs encouraged mentors to visit their beginning teachers' principals at least once a month. When program staff shadowed mentors, they met briefly with principals who did not strongly support the induction program.

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

[20] 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.

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 mentor caseloads ranged from 8 to 14 teachers over the course of the year. With mentoring as the largest component of the comprehensive induction programs, mentors necessarily underwent careful selection and training. 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 teacher's union, and professional development; an assistant superintendent for instruction; other experienced mentors; and/or building 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.

Tables IV.1, IV.2, and IV.3 describe the background of the 44 mentors selected to deliver the comprehensive induction services, as reported by mentors in responses to a survey. As noted, all mentors reported at least 5 years of teaching experience, with 18 years on average, and 46 percent had worked in non-teaching positions in education (Table IV.1). All held at least a bachelor's degree, while 86 percent had earned a master's degree. All were certified, 55 percent in several areas and 14 percent through the National Board of Professional Teaching Standards. About one-third of the mentors (33 percent) also reported working toward an advanced degree or additional credits, and 20 percent were working toward additional certification. The average age of the mentors was 43 years old in 2005; mentors were overwhelmingly female (95 percent) and 51 percent white, 32 percent African American, and 17 percent Hispanic. There was one instance of turnover among mentors during the study year covered in this report.

While the mentors were all first-year implementers of the particular program under study, more than three-quarters of them (77 percent) reported having prior mentoring experience—six years on average—and 74 percent of these individuals had attended mentor training in the past, as shown in Table IV.2. The most commonly reported areas of training addressed classroom management, the delivery of effective feedback, and mentor roles (88 percent for each area). Table IV.3 indicates the types of activities in which those with mentoring experience had participated. At least 40 percent reported having helped with classroom management (46 percent), assisted with strategies for effective instruction (43 percent), or helped teachers plan lessons (40 percent) on a weekly or bimonthly basis. Appendix F provides separate mentor profiles for each induction program, with Tables F.1-F.3 presenting characteristics reported by mentors working in ETS districts and Tables F.4-F.6 presenting characteristics reported by mentors working in NTC districts.

Table IV.1. Mentor Characteristics

	Percentage	
Race/ethnicity		
White, non-Hispanic	51.2	
Black or African American, non-Hispanic	31.7	
Hispanic	17.1	
Other/multiple	0.0	
Gender (percent female)	95.5	
Education: Highest Degree Attained		
Bachelor's degree	13.6	
Master's degree	86.4	
Working Toward Advanced Degree or Additional Credits	32.5	
Certification		
Not certified	0.0	
Certified in one area	45.5	
Certified in multiple areas	54.6	
Area of Certification		
General elementary education	90.9	
Bilingual education	11.4	
Special education	15.9	
Special subject area(s)	27.3	
Other area	31.8	
Certified Through National Board of Professional Teaching Standards (NBPTS)	13.6	
Working Toward Additional Certification	19.5	
Working Toward Certification Through NBPTS	a.	
Teaching Experience		
Last position before mentoring was as a classroom teacher	81.8	
Ever worked in nonteaching position(s) within education	45.5	
	Average	Range (Min., Max.)
Age in 2005 (Years)	43.0	(28, 61)
Teaching Experience (Years)	17.9	(5, 35)
Experience in Nonteaching Position(s) Within Education (Years)	1.4	(0, 6.8)
Caseload (Number of Beginning Teachers)	11.7	(8, 14)
Unweighted Sample Size (Mentors)	**44**	

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

Note: a. Values suppressed to protect respondent confidentiality.

Table IV.2. Mentor Experience and Training

Characteristic	Percentage	
Any Mentoring Experience	77.3	
Years of Mentoring Experience		
0	23.3	
1	14.0	
2 or more	62.8	
Types of Teachers Mentored (If Have Mentoring Experience)		
Beginning teachers	38.2	
Veteran teachers	a.	
Both beginning and veteran teachers	58.8	
Any Previous Mentoring Training (If Have Mentoring Experience)	73.5	
Areas of Mentor Training (If Received Mentor Training)		
Classroom management	87.5	
Giving effective feedback	87.5	
Mentor roles	87.5	
Coaching strategies	80.0	
Lesson planning	79.2	
Classroom observations	65.2	
Helping adult learners set goals	52.2	
Analyzing student work	50.0	
Leading study groups	39.1	
Coaching in literacy/language	27.5	
Coaching in math	20.8	
	Average	Range (Min, Max)
Years of Mentoring Experience (If Have Mentoring Experience)	6.2	(1, 30)
Unweighted Sample Size (Mentors)	**44**	

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

Note: a. Values suppressed to protect respondent confidentiality.

Once mentors were selected for program participation, 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) over four sessions, 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

Table IV.3. Previous Mentoring Activities

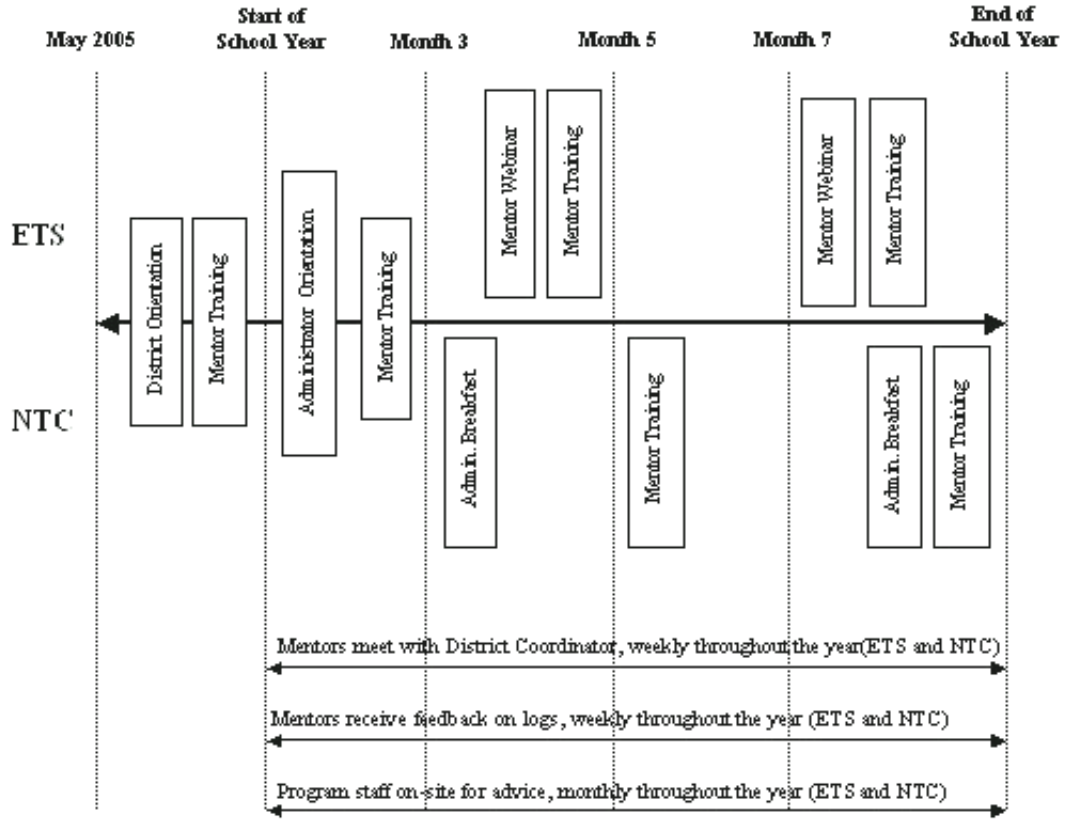
Activity	Frequency of Activity (Percentage)			
	Never	A Few Times a Year or on Request	Monthly	Weekly or Bimonthly
Helped teachers with strategies for effective instruction	22.7	18.2	15.9	43.2
Helped teachers plan lessons	22.7	22.7	13.6	40.2
Helped teachers with classroom management	25.6	23.2	a.	46.5
Observed teachers and provided feedback	27.3	31.8	a.	34.1
Helped teachers set goals to improve practice	31.8	34.1	15.9	18.2
Provided opportunities for teachers to observe others	36.4	36.4	a.	20.5
Reviewed teacher portfolios	53.5	32.6	11.6	a.
Led study groups on teaching	55.8	14.0	a.	23.3
Unweighted Sample Size (Mentors)	**44**			

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

Note: a. Values suppressed to protect respondent confidentiality.

best to address any implementation issues. By holding sessions over the course of the school year, programs 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 begin setting longer-range goals for their development.

Figure IV.2. Comprehensive Induction Program Training for Mentors, District Coordinators, and Administrators



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 district administrators from the central office. The administrator orientation was offered to school building administrators.

Trainings focused on active learning in two main areas: (1) improving beginning teachers' instruction, including the use of forms and processes to accomplish this; 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. ETS trainings conveyed stronger expectations for mentors' extensive use of materials focused on beginning teachers' instruction, and NTC staff gave more attention to mentoring skills. Both programs gave limited attention to a third area: 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.[21]

[21] 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.

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. 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 mentor 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 year, district coordinators provided substantive advice during the 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 their 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 a program tool, and how to stay on schedule with program implementation.

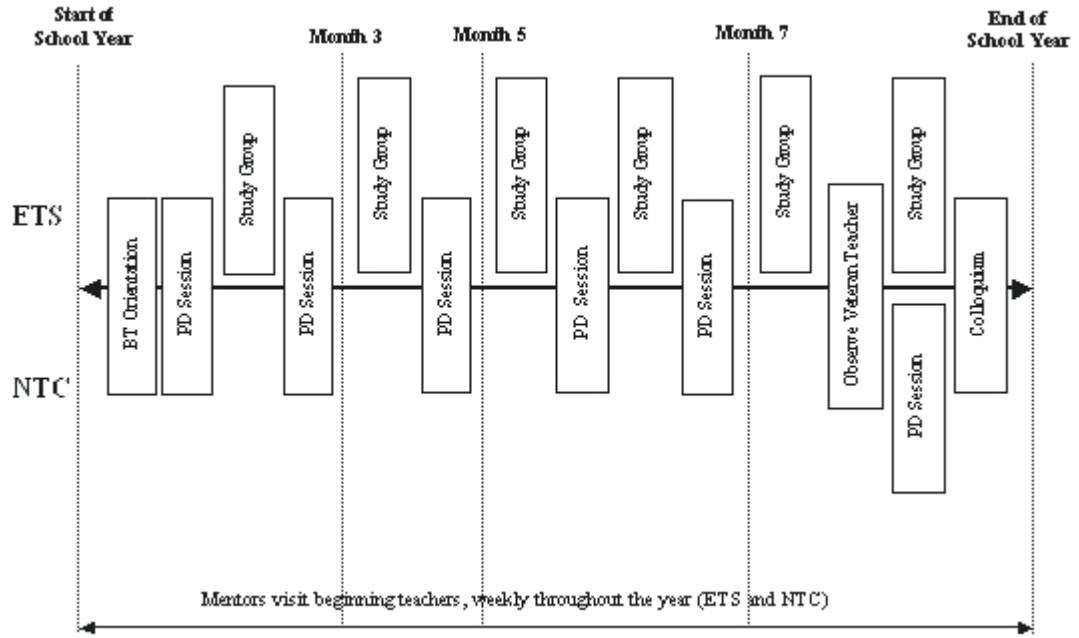
2. Program Services and Activities

When possible, **mentoring** of beginning teachers began during the first week of school, following an orientation session during which teachers were introduced to induction program goals and schedules (Figure IV.3). 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.[22] 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.[23]

[22] The primary obstacle to holding these early meetings was the delay in districts 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 the 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.

[23] Especially in the early part of the 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

Figure IV.3. Comprehensive Induction Program Activities for Beginning Teachers



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.

All beginning treatment teachers in the study 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. 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.[24] The NTC induction

(*continued*)
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.

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

model defines effective teaching in terms of six Professional Teaching Standards.[25] Each standard or domain is broken into a succession of more discretely defined categories of teaching behaviors (Table IV.4).

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.5).

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

The centerpiece of the NTC mentoring model is the *NTC Formative Assessment System* (FAS). FAS involves a series of collaborative processes between the mentor and beginning teacher that aims to collect and analyze a variety of data focused on teacher practices and students' 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 largely 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 six teaching standards (Table IV.4) and the continuum of performance (Table IV.5). 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.

[25] 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.4. ETS and NTC Content: Four Domains and Six Professional Teaching Standards

ETS Domains of Professional Practice		
Four Domains	Example, Subcategories of a Domain (*Instruction*)	Example, Details of a 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		
Six Professional Teaching Standards	Example, Subcategories of a Standard (*Engaging Students in Learning*)	Example, Details of a 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 experience, and interests with learning goals *Promoting self-directed, reflective learning* * 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 own 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* (Association for Supervision and Curriculum Development 1996). The content of the NTC program is based on two documents—*California's Standards for the Teaching Profession* (California Commission on Teacher Credentialing 1997) and *Continuum of Teacher Development* (New Teacher Center 2002).

Table IV.5. 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* (Association for Supervision and Curriculum Development 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 of the profile of the students in the teacher's class. 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, mentors were asked to allocate approximately two hours for contact time each week with every beginning teacher in their caseloads. 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.[26] Both ETS and NTC held monthly, two-hour professional development sessions (Table IV.6).[27] The PD sessions complemented the interactions between mentors and beginning teachers as described in the seven ETS events and NTC's FAS. On average, the PD sessions drew 72 and 65 percent of the beginning teachers (ETS and NTC, respectively, as shown in Tables IV.7 and IV.8). 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. The study groups gave teachers an opportunity to discuss with their mentors how they were progressing in their practice, the 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.

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.[28]

End-of-Year Colloquium. The two- to three-hour colloquium 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 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).

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

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

[28] To limit the time burden on teachers, no professional development session was held in the month(s) when the observations were conducted. Programs encouraged mentors to accompany beginning teachers for the observations, but it was challenging for mentors to accomplish this in the scheduling context of also maintaining their regular weekly charge of traveling 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.

Table IV.6. Topics for Monthly Professional Development Sessions, by Program

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

Source: The ETS program derives its content from Enhancing Professional Practice: A Framework for Teaching (Association for Supervision and Curriculum Development 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.7. Teacher Attendance at ETS Induction Activities (Percentages)

		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 (5 Sessions)	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.

n.a. = not applicable

Table IV.8. Teacher Attendance at NTC Induction Activities (Percentages)

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 (6 Sessions)	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.

n.a. = not applicable

B. Prevailing Induction Services and Comprehensive Induction Service Contrast

The 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 study schools 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. 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 as reported in Chapter V. 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. After discussing these results in detail, we provide a summary of the findings at the end of the chapter.

Our data derive from the induction activities survey that we administered twice during the 2005–2006 school year—during the fall/winter and spring (see Chapter III)—to characterize the induction services received by the treatment and control groups. Both surveys posed the same questions. Our discussion focuses on the spring survey because the fall results may cover a period affected by program start-up, which was delayed in several

schools due to some principals' initial resistance to participation or delays in obtaining lists of new teachers. We report on data from the spring survey, and note the instances in which the fall service contrast differs significantly from the spring service contrast.[29] Results from the fall survey are reported in Appendix F, Tables F.7-F.13. Comparisons of the service contrasts from the fall and spring surveys in terms of teacher-reported areas of mentor support and professional development during the past three months are presented in Appendix F, Tables F.14 and F.15, respectively.

Findings for each provider were consistent with the overall findings, except as noted. We present separate results by provider in Appendix F. Tables F.16-F.22 present results for teachers in ETS districts while Tables F.23-F.29 present results for teachers in NTC districts.

1. General Professional Supports and Duties

Treatment teachers were significantly more likely than control teachers to report being provided with induction programs focused on general support and guidance (70 versus 50 percent), and significantly less likely than control teachers to report programs focused on orientation to school/district (19 versus 40 percent, Table IV.9 at the end of this chapter).[30] However, general support and guidance was the most frequently cited purpose of the program for both groups of teachers.[31] The availability of the treatment intervention did not significantly affect teachers' other assistance and responsibilities, suggesting that principals did not differentially change treatment teachers' support or duties in response to the study (see Table IV.9 at the end of this chapter).

2. Mentor Meetings and Activities

As intensive mentoring was the centerpiece of the two comprehensive induction programs, an analysis of mentoring services received by teachers is critical. Tables IV.10 and IV.11 at the end of this chapter present estimates of treatment-control differences in mentor assignments, meetings, and activities. Results in the tables show that treatment teachers reported spending significantly more time during the most recent full week of teaching meeting with mentors than did control teachers (95 versus 74 minutes, Table IV.11) and reported significantly more assistance in a wide range of content areas than did control teachers. We summarize the findings in more detail below.

Treatment teachers were significantly more likely than control teachers to report having a mentor (94 versus 83 percent), as Table IV.10 shows.[32] They were also significantly more

[29] We test for significant differences between fall and spring service contrasts using the sample of teachers who responded to both the fall and spring Induction Activities surveys.

[30] All differences discussed in the text are statistically significant at the 0.05 level unless stated otherwise.

[31] The survey asks teachers to report the primary purpose of the induction program by selecting one purpose from a list (e.g., general support/guidance, orientation to the school/district, promotion of standards-based teaching, or other).

[32] The service contrast in having a mentor is greater in the fall than in the spring by 3 percentage points, p-value 0.025.

likely than control teachers to report having an assigned mentor; 93 percent of treatment teachers versus 75 percent of control teachers reported that the mentors were purposely assigned to help them rather than individuals whom the teachers found on their own.[33] Treatment teachers were significantly more likely than control teachers to report having multiple mentors (29 versus 17 percent) and to report having more than one mentor assigned to them (24 versus 11 percent). To capture all the mentoring services received by teachers, we define the measures in Tables IV.10 through IV.13 as covering all of a teacher's mentors. (For instance, under "Mentor Positions" in Table IV.10, the row labeled "Full-time mentor" indicates the percentages of teachers reporting any full-time mentor). Treatment teachers were also significantly more likely than control teachers to report having a full-time mentor (74 versus 13 percent), and significantly less likely than control teachers to report having a mentor who was another teacher (30 versus 66 percent).[34] Although all treatment teachers were assigned a full-time ETS or NTC mentor (referred to in Table IV.10 as a "study mentor"), 89 percent reported working with this person.

Treatment teachers reported spending significantly more time with their mentors than did control teachers. They were significantly more likely than control teachers to meet with their mentors at regularly scheduled times (86 versus 54 percent) and significantly more likely than control teachers to meet during school hours (77 versus 38 percent) and before or after school hours (38 versus 31 percent), as shown in Table IV.11. While treatment teachers' scheduled meetings were not significantly more frequent than control teachers' scheduled meetings (1.4 times a week for treatment teachers versus 1.2 times a week for control teachers), they tended to last significantly longer—59 versus 38 minutes during the most recent full week of teaching. In addition, treatment and control teachers both met informally with mentors for about 36 minutes on average during the most recent full week of teaching. Taking usual scheduled time and informal time during the most recent full week of teaching together, treatment teachers spent an average of 95 minutes in mentor meetings compared to 74 minutes for control teachers, with the statistically significant 21-minute difference attributable entirely to differences in the duration of the usual scheduled meetings. For a typical school year of 36 weeks, the total hours of mentor contact time during the year is estimated to be 56.8 hours for treatment teachers compared to 44.4 hours for control teachers, a difference of 12.5 hours. Estimates of the treatment-control difference in time spent with mentors are shown separately by district in Appendix H, Figure H.4. The figure shows the impact estimates as vertical bars, with negative impact estimates lying below the horizontal axis and positive impact estimates lying above the horizontal axis.

In addition to reporting spending more time with their mentors overall, treatment teachers reported spending significantly more time than control teachers in specific types of mentoring activities during the most recent full week of teaching. These activities included teachers being observed by mentors (26 versus 11 minutes), teachers meeting one-on-one

[33] The survey asks if teachers have a mentor, then asks whether and by whom the mentor(s) were assigned. Mentors could have been assigned by a teacher's district or principal, by a teacher preparation program, or by the study.

[34] Not all treatment teachers may have been aware that their ETS or NTC mentor was a full-time mentor.

with mentors (34 versus 21 minutes) or meeting with mentors together with other first-year teachers (27 versus 7 minutes), and mentors modeling lessons (11 versus 7 minutes). The total time spent in the six types of activities that the survey asked about averaged 121 minutes per week for treatment teachers and 66 minutes per week for control teachers, a significant difference of 56 minutes per week. (For a typical school year of 36 weeks, the total hours spent in the six activities during the year is estimated to be 72.8 hours for treatment teachers versus 39.3 hours for control teachers, a difference of 33.4 hours).

Despite the treatment teachers' significant additional scheduled meeting time, treatment and control teachers voiced no significant differences in the perceived adequacy of the length of their scheduled meeting time with mentors (84 percent of treatment teachers versus 81 percent of control teachers felt that the amount of scheduled time was adequate).

The service contrast in terms of time spent meeting with mentors and the likelihood of receiving guidance from mentors in various areas did not change significantly between the fall and the spring, except in three areas where it decreased over this period. (See Table F.9 in Appendix F for the fall service contrast estimates). The extra time that treatment teachers spent being observed by mentors during the last full week of teaching relative to control teachers was greater by 7 minutes in the fall than in the spring (service contrast of 22 minutes in the fall versus 15 minutes in the spring, p-value 0.003), and the extra time spent by treatment teachers relative to control teachers in all six activities per week was greater by 17 minutes in the fall than in the spring (service contrast of 73 minutes in the fall versus 56 minutes in the spring, p-value 0.047). In addition, the service contrast in terms of the percentage of teachers reporting having received encouragement or moral support during the last full week of teaching was greater by 5 percentage points in the fall than in the spring (service contrast of 21 percentage points in the fall versus 16 percentage points in the spring, p-value 0.030).

Treatment teachers' mentoring time came from different sources than did control teachers' time, reflecting the different types of mentors with whom they worked. The additional mentoring time that treatment teachers received relative to control teachers was the net result of their spending 64 minutes per week with their study mentors and spending significantly less time than control teachers with non-study mentors (31 versus 74 minutes), as shown in Table IV.11. Thus, the treatment mentoring partially substituted for mentoring provided by schools, districts, and other sources. Decomposing mentoring time in another way, treatment teachers reported spending significantly more time than control teachers with mentors who are full-time mentors (60 versus 8 minutes in the most recent full week of teaching), and significantly less time than control teachers with mentors who are other teachers (29 versus 64 minutes in the most recent full week of teaching).

3. Content of Mentor Meetings

Treatment teachers were significantly more likely than control teachers to report receipt of a wide range of types of assistance during the last full week of teaching and during the three months prior to the spring survey, as discussed below.

Table IV.11 shows that, during the most recent full week of teaching, treatment teachers were significantly more likely than control teachers to report receiving mentors' assistance in each of the topic areas asked about by the survey by 15 to 26 percentage points, with significant impacts above 20 percentage points on discussing instructional goals (70 versus 44 percent), receiving suggestions to improve practice (74 versus 52 percent), and receiving guidance on assessing students (62 versus 40 percent). Among treatment teachers, the percentage reporting each type of assistance ranged from 85 percent receiving encouragement or moral support to 61 percent receiving guidance in teaching to meet state or district standards. Among control teachers, the percentage reporting each type of assistance ranged from 69 percent receiving encouragement or moral support to 40 percent receiving guidance on how to assess students.

Treatment teachers were also significantly more likely than control teachers to report receiving a moderate amount or a lot of guidance in a broad array of areas during the three months prior to the survey, with differences ranging from 7 to 36 percentage points (see Table IV.12 for the spring service contrast estimates). There were statistically significant differences between the treatment and control teachers for each of the 22 types of guidance the survey asked about, with differences of 25 percentage points or more in reflecting on instructional practice (68 versus 33 percent); managing classroom activities, transitions, and routines (65 versus 40 percent); reviewing and assessing student work (55 versus 30 percent); and using student assessments to inform your teaching (54 versus 29 percent).

Treatment-control differences in the reported receipt of mentoring support in various areas during the three months prior to the survey increased significantly during the year for 10 of the 22 topic areas. The service contrast in these 10 areas rose by 10 to 20 percentage points between the fall and the spring for reviewing and assessing student work (20 percent); understanding/teaching toward state or district standards (15 percent); using student assessments to inform teaching (14 percent); teaching children with varying levels of achievement/ability (13 percent); teaching reading/language arts (13 percent); planning lessons (12 percent); selecting or adapting curriculum materials (11 percent); teaching students of varying ethnic/racial and socioeconomic backgrounds (11 percent); using multiple instructional strategies/techniques to teach students (11 percent); and teaching mathematics (10 percent) (see Table F.10 in Appendix F for the fall service contrast estimates and Table F.14 for the differences between the fall and spring contrast estimates).

4. Professional Development

During the three months prior to the spring survey, treatment teachers were significantly more likely than control teachers to report engaging in certain PD activities the survey asked about, as shown in Table IV.13 at the end of this chapter. These activities were keeping written logs (40 versus 28 percent), working with study groups of new teachers (68 versus 27 percent) and study groups of new and experienced teachers (47 versus 37 percent), and observing others teaching both in their classrooms (70 versus 42 percent) and in the study teacher's classroom (47 versus 38 percent)—all important parts of the comprehensive programs. Compared to control teachers, treatment teachers were also significantly more frequently observed by mentors during the three months prior to the spring survey (3.4 versus 1.5 times) and given feedback on teaching both as part of a formal evaluation (1.7

versus 1.5 times) and not as part of a formal evaluation (2.5 versus 2.0 times) during this period.

The treatment-control service contrast was significantly smaller in the fall than in the spring for teachers' reports on the likelihood of observing others teaching in their classrooms (service contrast of 14 percentage points in the fall versus 28 percentage points in the spring, p-value 0.001). The treatment-control contrast was also smaller in the fall than in the spring for keeping a written log of reflections (service contrast of 7 percentage points in the fall versus 12 percentage points in the spring, p-value 0.033). In addition, the service contrast in the frequency of principal observations was significantly different in the fall than in the spring (service contrast of –0.3 times in the fall versus 0.1 times in the spring, p-value 0.013). (See Table F.11 in Appendix F for the fall service contrast estimates).

Treatment and control teachers did not differ significantly in the likelihood that they found PD activities useful (22 percent of treatment teachers and 23 percent of control teachers found them "very useful", as shown in Table IV.13). The separate results by provider show that, among ETS districts, significantly fewer treatment teachers than control teachers found PD to be useful (16 versus 24 percent found PD "very useful," Table F.20).

Treatment and control teachers did not differ significantly in their reported attendance or time spent in PD, except in certain areas. Of the 17 PD topics that the survey asked about, a significantly larger percentage of treatment teachers compared to control teachers reported having attended PD in three areas: lesson planning (38 versus 26 percent), analyzing student work (56 versus 42 percent), and differentiated instruction (55 versus 46 percent). There were no statistically significant differences found between the treatment and control groups for the remaining 14 variables asked about within the survey (see Table IV.14 for the spring service contrast estimates for PD topic sessions attended by teachers during the past three months).

Treatment teachers reported spending significantly more time than control teachers in four PD areas: parent and community relations (23 versus 15 minutes), lesson planning (36 versus 26 minutes), analyzing student work (58 versus 41 minutes), and assigning grades/record keeping (17 versus 10 minutes). Treatment teachers reported spending significantly less time than control teachers in one area: preparing students for standardized testing (43 versus 53 minutes). There were no statistically significant differences found between the treatment and control groups for the remaining 12 PD topic areas (see Table IV.15 for the spring service contrast estimates for time spent on PD topics during the past three months).

One notable change in the service contrast over the course of the year was a significant increase in the treatment-control differences regarding PD related to analyzing student work. The service contrast in the likelihood of offering PD on this topic rose from –3 percentage points during the fall to 13 percentage points during the spring, and the service contrast in attendance at PD sessions covering this topic rose from –2 percentage points during the fall to 15 percentage points during the spring. The service contrast in time spent in PD on analyzing student work increased significantly from –0.1 minutes in the fall to 17 minutes in the spring. Other significant changes in the service contrast in PD receipt between the fall

and the spring are shown in Table F.15 (with fall service contrast estimates shown in Tables F.12 and F.13).

5. Summary of Service Contrast

Overall, results show statistically significant differences in the amount, types, and content of induction support received by treatment and control teachers. This service contrast was similar in the fall and the spring, with some exceptions. Below we summarize the key elements of the contrast.

- Treatment teachers were significantly more likely than control teachers to report having a mentor (94 versus 83 percent) and to report having multiple mentors (29 versus 17 percent).

- Treatment teachers were significantly more likely than control teachers to report having a mentor assigned to them (93 versus 75 percent) and to report having a full-time mentor (74 versus 13 percent).

- Treatment teachers reported spending significantly more total time meeting formally and informally with their mentors than control teachers did. Treatment teachers reported spending an average of 95 minutes during the most recent full week of teaching in mentor meetings compared to 74 minutes for control teachers, with the 21-minute difference attributable entirely to differences in the duration of scheduled meetings. For a typical school year of 36 weeks, the difference in total hours of mentor contact time is estimated to be 12.5 hours over the course of the school year.

- Treatment teachers were significantly more likely than control teachers to report having received mentors' assistance in a variety of topic areas (by 15 to 26 percentage points during the last full week of teaching, and by 7 to 36 percentage points during the three months prior to the spring survey). These areas include discussing instructional goals (70 versus 44 percent), suggestions to improve practice (74 versus 52 percent), and guidance on assessing students (62 versus 40 percent) during the last full week of teaching; as well as reflecting on instructional practice (68 versus 33 percent); managing classroom activities, transitions, and routines (65 versus 40 percent); reviewing and assessing student work (55 versus 30 percent); and using student assessments to inform your teaching (54 versus 29 percent) during the three months prior to the spring survey.

- Treatment teachers were significantly more likely than control teachers to report having participated in certain professional development activities during the three months before the spring survey, including keeping written logs (40 versus 28 percent), working with study groups of new teachers (68 versus 27 percent) and new and experienced teachers (47 versus 37 percent), and observing others

teaching in their classrooms (70 versus 42 percent) and in the study teachers' own classrooms (47 versus 38 percent).

- Treatment teachers reported having been observed significantly more frequently by mentors (3.4 versus 1.5 times) and having been more frequently given feedback on their teaching as part of a formal evaluation (1.7 versus 1.5 times) and not as part of a formal evaluation (2.5 versus 2.0 times) than control teachers during the three months prior to the spring survey.

- Treatment teachers were significantly more likely than control teachers to report attending PD in lesson planning (38 versus 26 percent), analyzing student work (56 versus 42 percent), and differentiated instruction (55 versus 46 percent). Treatment teachers reported spending more time than control teachers in PD in parent and community relations (23 versus 15 minutes), lesson planning (36 versus 26 minutes), analyzing student work (58 versus 41 minutes), and assigning grades/record keeping (17 versus 10 minutes). Compared to control teachers, treatment teachers reported spending significantly less time in PD in preparing students for standardized tests (43 versus 53 minutes).

- The treatment-control service contrast was significantly greater in the fall than in the spring in terms of teachers' reports of having had any mentor (14 versus 11 percentage points), receiving moral support from mentors (21 versus 16 percentage points), time spent being observed by mentors during the last full week of teaching (22 versus 15 minutes), and total time in six mentor activities during the last full week of teaching (73 versus 56 minutes).

Table IV.9. Teacher Reports on Professional Support and Duties (Spring) (Percentages)

	Treatment	Control	Difference	P-value
Services Offered				
Induction Program Provided by School or District	92.2	90.8	1.3	0.515
Primary Purpose of Program[a]				
General support/guidance	69.6	49.8	19.9*	0.000
Orientation to school/district	19.4	40.1	-20.7*	0.000
Standards-based teaching	8.9	8.9	0.0	0.990
Other	1.8	0.8	1.0	0.269
BT[b] Has a Mentor	93.8	82.7	11.1*	0.000
BT Has an Assigned Mentor	92.5	75.2	17.3*	0.000
Professional Development Activities Offered in Past 3 Months	97.8	99.0	-1.2	0.170
Assistance Received During Past 3 Months				
Reduced Teaching Schedule	7.5	6.3	1.1	0.512
Common Planning Time with Teachers at Grade Level	74.1	74.0	0.1	0.980
Received Teacher's Aide for Assistance	34.6	35.6	-1.0	0.767
Regular Communication with Administrators on Teaching Practice	57.7	63.1	-5.4	0.105
Duties Required During Past 3 Months				
Extracurricular Assignments	41.6	42.0	-0.4	0.914
Administrative Duties Including Lunchroom, Hall, or Recess Duties (but Not Staff Meetings)	44.7	43.7	1.0	0.794
Moved Between Classrooms to Teach	10.4	12.8	-2.4	0.230
Traveled to More Than One School to Teach	2.3	2.7	-0.4	0.642
Unweighted Sample Size (Teachers)	**468**	**417**		

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Difference in the distributions is statistically significant using a chi-squared test (p= 0.000).

[b]BT = beginning teacher.

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

Table IV.10. Impacts on Teacher-Reported Mentor Profiles (Spring) (Percentage)

Mentoring Characteristic	Treatment	Control	Difference	P-value
Number of Mentors				
Any Mentor (One or More)	93.8	82.7	11.1*	0.000
Multiple Mentors (More Than One)	29.0	16.5	12.5*	0.000
Number of Mentors[a]				
None	6.2	17.4	-11.1*	0.000
One	64.9	66.2	-1.3	0.734
Two	25.0	13.5	11.5*	0.000
Three	2.6	1.1	1.5	0.155
Four	0.4	0.2	0.3	0.483
Five	0.8	1.7	-0.9	0.241
Mentor Assignment				
Any Mentor Assigned	92.5	75.2	17.3*	0.000
Number of Mentors Assigned[a]				
No mentor assigned	7.5	24.8	-17.3*	0.000
One mentor assigned	68.3	63.8	4.5	0.234
Two mentors assigned	24.2	11.4	12.8*	0.000
BT[b] Reports Assigned Study Mentor	89.3	n.a.	n.a.	n.a.
Mentor Positions				
Full-Time Mentor	73.9	13.1	60.8*	0.000
Teacher	30.4	65.8	-35.4*	0.000
Administrator, School, or District	8.1	6.7	1.4	0.450
Staff External to District	4.0	2.5	1.5	0.220
No Mentor	6.3	17.5	-11.1*	0.000
Position of Mentor If Have Only One[a]				
Full-time mentor	74.4	14.8	59.6*	0.000
Teacher	16.9	78.0	-61.1*	0.000
Administrator	5.1	5.6	-0.5	0.805
Staff external to district	3.6	1.6	2.0	0.111
Combination of Mentor Positions If Have Two[a]				
Teacher and full-time mentor	64.9	19.0	45.8*	0.000
Both teachers	5.1	51.1	-46.0*	0.000
Teacher and administrator	7.0	16.7	-9.7	0.157
Teacher and staff external to district	0.0	9.8	-9.2*	0.026
Full-time mentor and administrator	12.6	1.8	10.8*	0.026
Full-time mentor and staff external to district	3.1	0.0	3.1	0.108
Other combination	8.7	3.5	5.2	0.232
Unweighted Sample Size (Teachers)	**468**	**417**		

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

n.a.= not applicable.

[a]Difference in the distributions is statistically significant using a chi-squared test (p = 0.000).

[b]BT = beginning teacher.

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

Table IV.11. Impacts on Teacher-Reported Mentor Services Received in the Most Recent Full Week of Teaching (Spring)

Mentor Service	Treatment	Control	Difference	Effect Size[b]	P-value
Teacher Has a Usual Meeting Time with a Mentor (Percentage)					
During school hours	76.9	38.1	38.8*	--	0.000
Before or after school hours	38.3	30.9	7.4*	--	0.024
On weekends	0.6	0.0	0.6	--	0.163
Varies	2.4	3.0	-0.7	--	0.558
Any usual meeting time	85.5	53.5	32.0*	--	0.000
"Usual" Meetings with Mentors					
Frequency (number of meetings)	1.4	1.2	0.2	0.09	0.187
Average duration (minutes)	23.4	11.1	12.3*	0.67	0.000
Total time (minutes)	59.0	38.0	21.0*	0.27	0.000
Informal Meetings with Mentors					
Total time (minutes)	35.8	36.1	-0.4	-0.01	0.905
Total Usual and Informal Time with Mentors (Minutes)	94.7	74.0	20.8*	0.20	0.007
Teacher Has Usual Meetings and Feels There Is Adequate Time to Meet with a Mentor (Percentage)	84.3	81.3	3.0	--	0.370
Meeting Time with Mentors in the Following Positions (Minutes)					
Study mentor	63.6	n.a.	n.a.	n.a.	n.a.
Non-study mentor	31.1	74.0	-42.8*	-0.46	0.000
Meeting Time with Mentors in the Following Positions (Minutes)					
Full-time mentor	60.3	7.6	52.7*	0.82	0.000
Teacher	28.8	63.6	-34.7*	-0.37	0.000
Administrator	3.9	2.5	1.4	0.08	0.276
Staff external to district	1.7	1.0	0.7	0.06	0.332
Mentor Time in the Following Activities (Minutes)					
Observing BT[a] teaching	26.0	11.1	14.9*	0.48	0.000
Meeting with BT one-on-one	33.7	21.0	12.8*	0.41	0.000
Meeting with BT and other first year teachers	27.1	7.0	20.1*	0.57	0.000
Meeting with BT and other teachers	16.2	14.8	1.4	0.05	0.500
Modeling a lesson	10.8	6.5	4.2*	0.19	0.004
Co-teaching a lesson	7.3	5.3	2.0	0.10	0.125
All six activities (all mentors)	121.3	65.5	55.7*	0.49	0.000
All six activities (study mentor only)	99.2	n.a	n.a	n.a	n.a
Types of Assistance a Mentor Provided (Percentage)					
Suggestions to improve practice	74.3	52.3	22.1*	--	0.000
Encouragement or moral support	84.7	69.0	15.7*	--	0.000
Opportunity to raise issues/ discuss concerns	82.5	63.9	18.6*	--	0.000
Help with administrative/ logical issues	67.6	51.8	15.8*	--	0.000
Help teaching to meet state or district standards	60.6	44.3	16.3*	--	0.000
Help identifying teaching challenges and solutions	71.8	52.2	19.7*	--	0.000
Discussed instructional goals and ways to achieve them	69.8	44.1	25.7*	--	0.000
Guidance on how to assess students	61.9	39.9	22.1*	--	0.000
Shared lesson plans, assignments, or other instructional activities	63.4	48.5	14.9*	--	0.000
Acted on something BT requested[c]	68.2	49.8	18.4*	--	0.000
Unweighted Sample Size (Teachers)	**468**	**417**			

Table IV.11 *(continued)*

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]BT = beginning teacher.

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

[c]Total sample size is 696. 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 IV.12. Impacts on Teacher-Reported Areas of Mentor Support During Past 3 Months (Spring) (Percentages)

Area		"Moderate Amount" or "A Lot" of Guidance			
		Treatment	Control	Difference	P-value
Areas of Guidance					
1.	Reflecting on your instructional practices	68.1	32.6	35.5*	0.000
2.	Managing classroom activities, transitions, and routines	64.6	39.9	24.7*	0.000
3.	Managing student discipline and behavior	62.2	42.2	20.0*	0.000
4.	Using multiple instructional strategies/techniques to teach students	61.0	37.8	23.2*	0.000
5.	Teaching children with varying levels of achievement/ability	58.2	35.8	22.3*	0.000
6.	Motivating students	56.7	36.2	20.5*	0.000
7.	Understanding/teaching toward state or district standards	56.5	33.8	22.7*	0.000
8.	Teaching reading/language arts	55.6	34.2	21.4*	0.000
9.	Reviewing and assessing student work	55.3	29.5	25.9*	0.000
10.	Understanding this school's culture, policies, and practices	54.3	44.8	9.4*	0.004
11.	Selecting or adapting curriculum materials	53.9	33.8	20.2*	0.000
12.	Using student assessments to inform your teaching	53.9	29.1	24.7*	0.000
13.	Planning lessons	52.4	32.9	19.5*	0.000
14.	Completing paperwork	50.8	37.3	13.4*	0.000
15.	Accessing district and community resources	48.8	29.1	19.7*	0.000
16.	Teaching students of varying ethnic/racial and socioeconomic backgrounds	46.1	30.0	16.2*	0.000
17.	Teaching mathematics	45.6	31.8	13.9*	0.000
18.	Teaching students with special needs	41.6	24.0	17.6*	0.000
19.	Working with other teachers to plan instruction	40.0	33.3	6.7*	0.038
20.	Working with other school staff, such as principal, counselors, disability specialist	39.3	32.7	6.5*	0.041
21.	Communicating with parents	38.0	30.6	7.4*	0.017
22.	Teaching English language learners	31.2	20.5	10.7*	0.002
Unweighted Sample Size (Teachers)		**468**	**417**		

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the 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 IV.13. Impacts on Teacher-Reported Professional Development During the Past 3 Months (Spring)

Aspect of Professional Development	Treatment	Control	Difference	Effect Size[b]	P-value
When Professional Development Took Place[a] (Percentages)					
Before or after school	57.1	52.0	5.1	--	0.128
During "regular" teaching hours	27.7	32.2	-4.5	--	0.129
In the evening or on Saturday	8.6	9.8	-1.2	--	0.612
During summer or PD days	5.9	5.0	0.9	--	0.596
Other	0.4	1.0	-0.6	--	0.300
Did not attend any PD activities	0.4	0.0	0.4	--	0.147
Activities Completed (Percentages)					
Kept a written log	40.0	27.9	12.1*	--	0.000
Kept a portfolio and analysis of student work	77.9	74.3	3.7	--	0.203
Worked with a study group of new teachers	68.2	27.2	41.0*	--	0.000
Worked with a study group of new and experienced teachers	47.1	37.4	9.7*	--	0.003
Observed others teaching in their classrooms	70.2	42.3	28.0*	--	0.000
Observed others teaching your class	46.5	37.8	8.7*	--	0.015
Met with principal to discuss teaching	67.6	69.4	-1.8	--	0.600
Met with a literacy or mathematics coach or other curricular specialist	68.6	66.4	2.2	--	0.524
Met with a resource specialist to discuss needs of particular students	60.4	62.6	-2.2	--	0.526
Frequency of Selected Activities (Number of Times During Past 3 Months)					
Teaching was observed by mentor	3.4	1.5	1.8*	0.77	0.000
Teaching was observed by principal	2.1	2.0	0.1	0.04	0.626
Given feedback on your teaching, not as part of formal evaluation	2.5	2.0	0.6*	0.28	0.000
Given feedback on your teaching, as part of formal evaluation	1.7	1.5	0.2*	0.15	0.028
Given feedback on your lesson plans	1.6	1.6	0.0	0.00	0.976
Professional Development Was "Very Useful" (Percentage)	21.9	23.2	-1.3	--	0.659
Unweighted Sample Size (Teachers)	**468**	**417**			

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

PD = professional development.

[a]Difference in the distributions is not statistically significant using a chi-squared test (p= 0.500).

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

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

Table IV.14. Impacts on Teacher-Reported Areas of Professional Development During the Past 3 Months (Spring)

Area of Professional Development		Attended PD (Percentages)			
		Treatment	Control	Difference	P-value
Areas Offered					
1.	Human resource policies/procedures	19.0	20.6	-1.7	0.527
2.	Parent and community relations	26.7	23.5	3.2	0.283
3.	School policies on student disciplinary procedures	35.0	40.2	-5.3	0.106
4.	Instructional techniques/strategies	75.0	73.4	1.6	0.597
5.	Understanding the composition of students in your class	25.8	21.4	4.4	0.139
6.	Content area knowledge (language arts, mathematics, science)	63.8	64.4	-0.6	0.851
7.	Lesson planning	37.6	25.9	11.7*	0.000
8.	Analyzing student work/assessment	56.3	41.8	14.5*	0.000
9.	Student motivation/engagement	35.8	31.0	4.8	0.140
10.	Differentiated instruction	54.9	45.5	9.4*	0.009
11.	Using computers to support instruction	30.0	33.0	-3.0	0.350
12.	Classroom management techniques	43.0	37.2	5.8	0.116
13.	Accessing school, district, or community resources	19.3	17.4	1.9	0.490
14.	Administrative paperwork	14.5	16.3	-1.8	0.497
15.	Handling non-classroom duties and responsibilities (e.g., supervision of lunch room, back to school night)	12.9	11.4	1.5	0.519
16.	Assigning grades/record keeping	22.8	19.6	3.3	0.261
17.	Preparing students for standardized testing	46.3	51.7	-5.4	0.066
Unweighted Sample Size (Teachers)		**468**	**417**		

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

PD = professional development.

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

Table IV.15. Impacts on Time Spent in Teacher-Reported Areas of Professional Development During the Past 3 Months (Spring)

Area of PD		Time Spent (Minutes)				
		Treatment	Control	Difference	Effect Size	P-value
Areas Offered						
1.	Human resource policies/procedures	15.8	15.5	0.3	0.01	0.919
2.	Parent and community relations	22.8	14.7	8.1*	0.19	0.005
3.	School policies on student disciplinary procedures	24.4	28.0	-3.6	-0.08	0.264
4.	Instructional techniques/strategies	85.8	85.6	0.2	0.00	0.974
5.	Understanding the composition of students in your class	24.1	18.6	5.4	0.11	0.099
6.	Content area knowledge (language arts, mathematics, science)	78.8	83.0	-4.2	-0.05	0.403
7.	Lesson planning	35.6	26.4	9.2*	0.16	0.014
8.	Analyzing student work/assessment	58.2	41.0	17.3*	0.27	0.000
9.	Student motivation/engagement	31.9	28.1	3.8	0.07	0.315
10.	Differentiated instruction	54.0	45.0	8.9	0.14	0.065
11.	Using computers to support instruction	23.9	29.5	-5.5	-0.11	0.114
12.	Classroom management techniques	44.1	36.9	7.2	0.11	0.124
13.	Accessing school, district, or community resources	12.7	10.8	1.9	0.06	0.440
14.	Administrative paperwork	10.2	9.1	1.1	0.04	0.597
15.	Handling non-classroom duties and responsibilities (e.g., supervision of lunch room, back to school night)	7.7	5.9	1.8	0.07	0.272
16.	Assigning grades/record keeping	17.1	9.8	7.3*	0.22	0.002
17.	Preparing students for standardized testing	43.2	53.1	-9.9*	-0.15	0.011
Unweighted Sample Size (Teachers)		**468**	**417**			

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

PD = professional development.

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

CHAPTER V

IMPACT FINDINGS

The main goal of this study is to estimate the impact of comprehensive teacher induction on teacher and student outcomes. This chapter presents the impact estimates for each of the major outcomes discussed at the beginning of the report. Readers may refer to Appendix C 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, there was some complexity in the study design and outcome measurement that required decisions on the part of the researcher that could potentially 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 that outcome's "benchmark analysis." We conducted a series of sensitivity analyses to demonstrate the robustness of the findings using alternative samples or specifications of covariates for each outcome.

A. NO IMPACTS ON TEACHER ATTITUDES

The impact of teacher induction on teacher attitudes was not one of the study's central research questions listed in Chapter I, 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 changing teacher practices, productivity, and retention. The induction activities surveys allowed us to examine whether comprehensive teacher induction made teachers feel more satisfied with or better prepared for their jobs as teachers. Results based on teachers' self-reports indicated that this was not the case. As shown below, there were no statistically significant positive impacts of treatment on teacher satisfaction or preparedness.

1. Methods

Using items from the induction activities surveys, we measured teachers' feelings of preparedness in 13 areas of teaching (such as preparedness to work with students with special challenges) and satisfaction in 19 areas of teaching (such as satisfaction with their workload). The surveys asked teachers to respond along a four-point scale (ranging from "not at all prepared" to "very well prepared" and from "very dissatisfied" to "very

satisfied"). We conducted a factor analysis to explore how items could be sensibly grouped together. The factor analyses suggested that teacher preparedness and teacher satisfaction each consisted of three categories: preparedness to (1) instruct, (2) work with students, and (3) work with others; and satisfaction with (1) school, (2) class, and (3) career. (Factor loadings for the teacher preparedness and teacher satisfaction items are shown in Tables G.1 and G.2 in Appendix G). We used these three domains of preparedness and three domains of satisfaction to summarize the data.

We found that the results did not vary according to estimation method or the set of control variables we used. We present our benchmark estimates based on a simple regression model that has district and grade fixed effects and no other covariates.

2. Impact Estimates

Overall, teachers from the treatment and control groups reported feelings of preparedness and satisfaction that differed by 0.1 or less on a 4-point scale. Out of 12 differences we examined, 3 were statistically significant and favored the control group (Table V.1). Compared to treatment teachers, control teachers reported higher levels of preparedness to instruct during the fall (2.9 versus 2.8), higher levels of preparedness to work with other staff during the spring (3.0 versus 2.9) and higher levels of satisfaction with teaching as a career during the spring (3.0 versus 2.9). The largest impacts were about one-tenth of the distance between two categories (e.g., between "somewhat" and "well" prepared).

The isolated impacts that appear statistically significant can be attributed to chance when we correct for the probability of finding false positives among a large number of hypothesis tests. By conducting several hypothesis tests (for different outcomes and time points), we increase the potential for a Type I error (concluding that there is an impact when the null hypothesis of no impact is true). When we applied a method developed by Benjamini and Hochberg (1995) for reducing Type I error, or more specifically, the rate of false discoveries, none of the negative impacts remained statistically significant.

3. Sensitivity Analysis

One concern with the above 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 difference between category 1 and 2 may be larger than the difference between 2 and 3. We recoded teacher preparedness into two categories: (1) "not at all prepared" or "somewhat prepared" or (2) "well prepared" or "very well prepared." We also 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 as a dichotomous variable. The results show statistically significant differences in 6 of the 13 areas of preparedness, with all of the significant differences favoring the control group. The significant differences in percentages between treatment and control group teachers' feelings of preparedness ranged from 6 to 10 percent. There were no statistically significant differences with regard to teachers' reports of

satisfaction (see Appendix Tables G.3 and G.4). All six items under preparedness—assessing students, being an effective teacher, addressing the needs of a diversity of learners, selecting and adapting instructional materials, planning effective lessons, and working with the principal or other instructional leaders—remained significant after applying the Benjamini-Hochberg correction within each area of preparedness.

Table V.1. Impacts on Teacher Attitudes (Scores on a 4-Point Scale)

	Fall					Spring				
Outcome	Treatment	Control	Difference	Effect Size	P-value	Treatment	Control	Difference	Effect Size	P-value
Feel Prepared at This Point in the School Year to:										
Instruct	2.8	2.9	-0.1*	-0.17	0.013	2.9	3.0	-0.1	-0.14	0.056
Work with students	2.7	2.7	0.0	-0.03	0.705	2.8	2.9	0.0	-0.02	0.729
Work with other staff	2.9	2.9	0.0	-0.06	0.380	2.9	3.0	-0.1*	-0.16	0.021
Feel Satisfied with:										
School	3.1	3.1	0.0	-0.05	0.501	3.0	3.0	0.0	-0.02	0.738
Class	3.1	3.0	0.0	0.04	0.544	3.0	3.0	0.0	-0.02	0.779
Teaching career	3.0	3.1	-0.1	-0.12	0.077	2.9	3.0	-0.1*	-0.16	0.025
Unweighted Sample Size (Teachers)	**472**	**426**	**898**			**468**	**417**	**885**		

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

Note: Data pertain to teachers in all districts participating in the study. Data are regression-adjusted using ordinary least squares to account for differences in baseline characteristics. Preparation scale: (1) not at all prepared, (2) somewhat prepared, (3) well prepared, or (4) very well prepared; Satisfaction scale: (1) very dissatisfied, (2) somewhat dissatisfied, (3) somewhat satisfied, or (4) very satisfied. Sample sizes vary due to item nonresponse.

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

None of the differences is significantly different from zero after applying a Benjamini-Hochberg correction.

B. NO IMPACTS ON TEACHER PRACTICES

The conceptual framework presented in Chapter I suggests that for teacher induction to improve student achievement, it must first change the way teachers teach. To test for changes in teacher practices, we sent trained observers into treatment and control classrooms to administer the Vermont Classroom Observation Tool (VCOT), described in

Chapter III and Appendix D. The VCOT measures three domains of teaching: lesson implementation, lesson content, and classroom culture. It captures the degree to which the lesson observed reflects evidence of what are believed to be effective practices. We observed literacy lessons (or reading/language arts) in over 600 classrooms in spring of the study year (2006). Teachers who were teaching special populations, were teaching subjects other than reading/English language arts, were no longer teaching, or had prior teaching experience were not observed (by design) and therefore not included in this component of the impact analysis. See Chapter III for a more detailed discussion of the sample exclusions.

1. Methods

We estimated impacts on classroom practices using the regression methods described in Appendix C. As discussed in Chapter III, observers scored teachers on a five-point scale in each of the three domains based on a set of 16 items believed to be indicators of effective practice. The three domains comprise five, four, and seven indicators, respectively. The full set of 16 indicators is shown in Appendix G (Tables G.5 and G.6). Covariates in the model include teacher demographic characteristics (age, sex), teacher's educational and professional background (teacher preparation type, certification status, highest degree attained, months of teaching experience), teaching assignment (grade level), school characteristics (school-lunch program eligibility rate and racial/ethnic composition), and district and grade fixed effects.

To summarize the information from the classroom observations across all 16 indicators, we produced three scores corresponding to the three domains captured by the observation protocol (into which the items had already been grouped). Our benchmark estimates used the average score of the indicators within each domain, and thus assumes that the intervals between each category are equal. For example, the difference between "no evidence" and "limited evidence" is the same as the difference between "moderate evidence" and "consistent evidence." It also assigns equal weight to the indicators within each domain. In other words, we assume that a score of 3 on two indicators of classroom culture—for example, "Classroom routines are clear and consistent" and "Behavior is respectful and appropriate"—is equivalent to a score of 4 on one of those indicators and 2 on the other. Histograms for treatment and control teachers' performance in each of the three domains are included in Appendix H (Figures H.1 – H.3). These histograms illustrate the pattern of variation (or distribution) of the classroom practices data.

2. Impact Estimates

We observed no statistically significant impact of the comprehensive induction programs on classroom practices (Table V.2). After controlling for important teacher and school characteristics, we observed no statistically significant differences between treatment and control teachers' performance on implementation of literacy lesson, content of literacy lesson, or classroom culture. We express the impact on each domain of classroom practice as the difference in scores on the five-point scale. An impact of 0.5 points, for example, would suggest that the intervention moves the average teacher from being able to demonstrate "moderate" evidence of a particular practice in that domain half of the distance to being able to demonstrate "consistent" evidence of that practice. (The observed estimates of the impacts were smaller than the 0.5 points of this example).

Table V.2. Impacts on Classroom Practices (Average Score on a 5-Point Scale)

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

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

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

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

3. Sensitivity Analysis

We re-estimated the impacts by using a variety of assumptions about item scoring and estimation and found that the results did not change substantially.

The results were not sensitive to how we grouped the individual items into constructs. We performed a factor analysis of the 16 classroom observation items to explore the degree to which the theoretical groupings were empirically justified. In finding the groupings justified, we maintained the three-construct scoring method (implementation of literacy lesson, content of literacy lesson, and classroom culture) described above. Though the factor analyses were consistent with the theoretical groupings, they did suggest that the implementation and content items could be grouped together, forming one construct rather than two. (Factor loadings for the 16 class observation items are shown in Table G.5 in Appendix G). When we substituted a single construct that included all implementation and content items in place of two constructs, there were no statistically significant impacts. (Impact estimates for each of the 16 class observation items are shown in Table G.6 Appendix G).

The results were not sensitive to the choice of summary score. In addition to scoring individual items under each domain, classroom observers reported a summary score for each of the three domains. They based the summary score on a five-point scale that could differ from our constructed domain scores in two ways. First, they reported the score as an integer such that they had to round off to the nearest whole number and thus could have recorded numbers that differ from the average score. Second and more significantly, observers could exercise their discretion in assigning an overall domain score. Thus, if indicator scores were 3, 3, 3, 4, and 4, for example, for the five indicators, respectively, then an observer, in reporting the overall domain score, could round up to 4 instead of down to 3 based on a

judgment that the last two domains are more important than the first three. Observers could also justify an overall score of 4 if the item scores of 3 were actually rounded down from, say, 3.4 and the item scores of 4 had been rounded down from 4.4. (The average of 3.4, 3.4, 3.4, 4.4, and 4.4 is 3.8, which rounds up to 4).

The two types of summary scores were not identical. Given that they each have advantages and disadvantages, we had to choose one arbitrarily to include for the benchmark estimates presented above. When we substituted the observer summary scores for the computed average scores, we reached the same conclusions: no statistically significant impact of treatment (see Table G.7).

The findings did not change when we collapsed the scale or divided the sample into two subgroups. As part of our sensitivity analyses, we estimated the model separately for each classroom observation item after recoding each score from a five-point scale into a binary variable: (1) no, limited, or moderate evidence or (2) consistent or extensive evidence of good practice. This dichotomous coding scheme allowed us to compare the percentage of treatment and control teachers who demonstrated "consistent" or "extensive" evidence of good practice in the classroom. The results, however, support the same conclusions of no impact. We also estimated the model separately for district-defined subgroups, such as school district and program provider (ETS or NTC, see Appendix Tables G.8 and G.9) and found that the impact estimates were not significantly different from zero. Separate impact estimates by each of the 17 individual districts are shown in Appendix Figure H.5. The vertical bars in that figure represent the impact estimate, with positive impacts lying above the horizontal axis and negative impact estimates lying below the horizontal axis.

C. No Positive Impacts on Student Test Scores in the First Year

One of the central research questions for this study asks, "What is the impact of comprehensive teacher induction on teachers' contributions to their students' achievement?" To answer this, we examined the adjusted average achievement growth by using student test scores linked from one year to the next, covering the year that a class is taught by a teacher in the study. Though district-administered test scores may 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 found the grade-specific estimates of impacts on achievement to be negative and statistically significant for the lower grades, but the average impacts across all grades were not significantly different from zero, suggesting that across all grades the program had no impact on student test scores in its first year. The central finding of no positive impact was robust to 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 15 districts, most of which administered different tests under different conditions and followed different recordkeeping practices. Although 17 districts participated in the study, we included 15 districts in the benchmark estimates of student achievement

effects and excluded the remaining 2. One of the districts in the study failed to provide usable data on student test scores for a sufficient number of teachers in tested grades. A second district was unable to match teachers in the study with students' 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. At specific grade levels, only a few districts contributed to the analysis; testing is not common in lower grades, and many elementary schools in the study do not include grade 6.

The covariates include the normalized student pre-test score, student characteristics (student gender, race/ethnicity, special education status, English-language learner status, free/reduced-price lunch status, and whether the student was over age for grade), teacher personal characteristics (age, gender, race/ethnicity, whether 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, whether teacher holds a degree in an education-related field, whether teacher is a first-year teacher, whether teacher was hired after school year began, whether teacher attended competitive college, whether teacher held a non-teaching job for five or more years), and district-by-grade fixed effects. Appendix C describes in more detail the aggregation method, treatment of missing data, regression model, and estimation strategies we used.

2. Impact Estimates

The impacts on math and reading scores were not significantly different from zero. (See Table V.3 for reading and V.4 for math scores). The overall impacts may have masked a trend by grade. The grade-specific estimates for lower grades were negative and significant. For reading, grade 2 results were negative and significant, with an effect size of -22.3 percent of a standard deviation.[35] For grades 3 to 5, grades that included larger samples of districts and schools, the estimates were not significantly different from zero. Applying a Benjamini-Hochberg correction to the four reading grade-level outcomes, none of the results remain statistically significant.

[35] The study was designed to be able to detect an overall impact of 10 percent of a standard deviation in test scores.

Table V.3. Impacts on Reading Test Scores

Grade	Adjusted Mean Test Scores		Difference	Effect Size	P-value	Unweighted Sample Sizes		
	Treatment	Control				Students	Teachers	Districts
2	-0.12	0.10	-0.22*	-0.22	0.034	543	42	4
3	-0.06	0.07	-0.13	-0.13	0.119	1,113	75	8
4	0.02	-0.02	0.04	0.04	0.421	1,679	108	14
5	0.01	-0.01	0.01	0.01	0.843	1,516	81	11
6	--	--	--	--		48	4	1
All Grades	**0.01**	**-0.01**	**0.01**	**0.01**	**0.735**	**4,899**	**283**	**15**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. Treatment and control group sample sizes are shown in Appendix Table G.22.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

None of the differences is significantly different from zero after applying a Benjamini-Hochberg correction.

The estimates for math scores were negative and statistically significant for grades 2 and 3. Impacts for grades 4 and 5 were not statistically significant. The average impact across all grades, which was based on a larger sample and therefore was more precise, was not significantly different from zero. 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. The evidence suggests that the intervention depressed math scores for the lower grades but overall had no impact. Applying a Benjamini-Hochberg correction to the four math grade-level outcomes, the negative and significant results remain statistically significant.

Table V.4. Impacts on Math Test Scores

Grade	Adjusted Mean Test Scores		Difference	Effect Size	P-value	Unweighted Sample Sizes		
	Treatment	Control				Students	Teachers	Districts
2	-0.20	0.18	-0.38*†	-0.38	0.000	472	35	3
3	-0.11	0.15	-0.26*†	-0.26	0.002	837	65	6
4	0.01	-0.01	0.03	0.03	0.617	1,545	99	13
5	-0.02	0.02	-0.04	-0.04	0.549	1,510	81	11
6	--	--	--	--		48	4	1
All Grades	**-0.03**	**0.03**	**-0.05**	**-0.05**	**0.184**	**4,412**	**261**	**14**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. Treatment and control group sample sizes are shown in Appendix Table G.23.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

†Significantly different from zero at the 0.05 level, two-tailed test, after applying a Benjamini-Hochberg correction.

3. Sensitivity Analysis

We confirmed that the overall impact (across all grades) was not statistically significant when we re-estimated the impacts using different samples or different sets of covariates. The results from these alternative estimations are shown in Appendix G. We also re-estimated the test score impacts by including one problematic district that we had excluded from the benchmark analysis and found it did not alter the findings. The district was problematic because it was not possible to reliably link students to teachers in the data. The district in question had provided test scores for all students in the relevant grades for each of its schools in the study, rather than for students in study classrooms only, so we tried various methods to isolate the students of participating teachers in that district. As one approach we assigned each student in the district a weight equal to the number of study teachers in his or her grade and school divided by the estimated number of classrooms in that grade-school combination. When we included the district in question, the estimated impacts on reading and math scores for all grades were consistent with the benchmark results presented above.

The benchmark impact estimates were not sensitive to our requirement that the student have both a pre-test and a post-test. Some students in our sample were missing test score data for the study year (post-test) or the year before the study (pre-test). 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 post-test and pre-test. To test whether the main analysis results were biased by excluding students with missing data, we estimated impacts without including the pre-test as a covariate. Excluding the pre-test also extends the analysis to first grade. The results, shown in Appendix G (Tables G.10 and G.11), indicate that the overall impact is not significantly different from zero for either subject. Sixth grade reading and second, third, and sixth grade math are negative and significant, both before and after applying the Benjamini-Hochberg correction.

An alternate means by which to enlarge the sample while retaining the pre-test covariates was to include Dynamic Indicators of Basic Early Literacy Skills (DIBELS) test scores in the analysis. As explained in Appendix C, test scores based on DIBELS could be problematic so they were dropped from the main analysis. Including DIBELS has the potential to affect reading scores in grades 2 and 3 and the overall impact on reading. Appendix Table G.12 shows that the overall impact (across all grades) is not significant when DIBELS test scores are included. Table G.12 also shows that effect sizes are similar to those from the benchmark model, but instead of grade 2 being statistically significant, the impact on grade 3 scores is negative and significant. No grade levels are statistically significant after applying a Benjamini-Hochberg correction.

When we instead used the same sample as the benchmark estimates but added control variables for the teacher's SAT/ACT score, the results, shown in Appendix Tables G.13 and G.14, still indicate no overall test score impact in either subject. For 46.7 percent of the teachers in this sample, the SAT/ACT scores were imputed due to missing data. None of the individual grade results are statistically significant for reading. Impact estimates for grades 2 and 3 in math are negative and significant. The math results remain significant after applying the Benjamini-Hochberg correction.

Finally, Tables G.15 to G.18 show reading and math results for districts using the ETS program and those using the NTC program. The covariates are identical to the benchmark model. For ETS, the impact estimate for grade 2 in math is negative and statistically significant, although the estimate is no longer significant after applying a Benjamini-Hochberg correction. None of the other grade-level results or the overall impact in either subject is statistically significant. All of the estimated effect sizes except grades 2 and 5 in math are below 10 percent of a standard deviation. For NTC, the effect sizes range from negative and significant in grade 2 for both subjects and grade 3 for math to positive and significant for grade 4 in math and grade 5 in reading. All statistically significant grade-specific impact estimates remain so after applying a Benjamini-Hochberg correction, but the findings for these grades are based on small numbers of districts. The overall impact for NTC districts is not statistically significant. Figures H.6 and H.7 in Appendix H show estimates of impacts on reading and math scores, respectively, separately by district.

D. NO POSITIVE IMPACTS ON TEACHER RETENTION AFTER ONE YEAR

An often-cited benefit 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 beginning their teaching career (e.g. Kapadia et al. 2007). To address

the question of turnover, we examined the effect of comprehensive induction programs on the retention of new teachers.

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. While staff turnover can be disruptive and costly, some turnover is inevitable in teaching and in most professions. A critical question is whether turnover raises quality (by weeding out the weakest teachers) or lowers it (by discouraging the strongest ones). The random assignment design allows us to test the effects of comprehensive teacher induction directly on the composition of the teaching force by comparing the characteristics of treatment stayers to control stayers. Under random assignment, the treatment and control teachers are equivalent, on average, prior to the intervention. As teachers leave both groups, the average quality or qualifications of the teachers who remain in the districts (or in teaching) begins to change, perhaps differentially between the two groups. By measuring teacher outcomes (observed classroom practice and estimated contribution to test score performance) at the end of year 1, we can see who returns to the district for a second year in terms of teacher performance measures and teacher qualifications, such as advanced degrees and certification, and capture the impact of those attrition-induced changes.

1. Methods

Teachers' mobility status can be defined in a variety of ways, but the most common way calls for creating 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 in 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.[36] Teachers' mobility status can vary over time; unless otherwise stated, we report mobility as of fall 2006, which indicates whether the teacher returned to the district for a second year after one year of intervention. Future reports will consider changes over time in teacher mobility status based on two additional years of follow-up surveys.

The impact estimates are derived from a logistic regression model that mimics the model used for other outcomes above, except that the outcome variable is binary. The model and covariates used are described in Appendix C. As part of the sensitivity tests, we

[36]Another definition of stayers and movers that may be policy relevant from society's perspective is mobility in terms of high-need schools. In future reports that include more follow-up data and larger numbers of movers we will code the characteristics of the schools to which movers transfer and thereby be able to categorize teachers as high-need school stayers or movers.

estimated the model with other assumptions as well, 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 classroom practices and student achievement by using the same methods as above but included in the analysis only the district stayers. 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 or showed more evidence of effective classroom practice 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 therefore that 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

Before discussing impacts, we first describe the levels of teacher mobility. We found that, after one year, 75 percent of study teachers returned to the same schools, with another 11 percent changing schools but remaining in the same district. An additional 9 percent stayed in teaching but changed districts or left the public sector. The remaining teachers, fewer than 6 percent, left teaching altogether. The regression-adjusted district retention rate was 86 percent and the total retention rate in teaching (including movers) was 95 percent (see Table V.5).

Table V.5. Impacts on Teacher Retention Rates (Percentages)

Outcome	All Teachers	Treatment	Control	Difference	P-value
Retained in the same school	74.9	74.5	75.3	-0.9	0.766
Retained in the same district	85.8	85.9	85.7	0.2	0.915
Retained in the teaching profession	94.6	94.7	94.5	0.2	0.897
Unweighted Sample Size (Teachers)	**882**	**457**	**425**		
Unweighted Sample Size (Schools)	**390**	**199**	**191**		

Source: MPR Mobility Survey administered in 2006-2007 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 found no impacts of treatment on this pattern of teacher mobility after one year. The control group's teacher mobility pattern was nearly identical to that of the treatment group. Table V.5 shows the result of the three hypothesis tests specifically focused on retention in the school, retention in the district, and retention in the profession as binary

outcomes. For each of the outcomes, there was no statistically significant impact. Table V.6 shows the percentages of treatment and control teachers who stayed in the same school, moved within the profession, and left the teaching profession. The difference in mobility patterns between the two groups was not statistically significant.

Table V.6. Impacts on Teacher Mobility, by Destination (Percentages)

Outcome	Treatment	Control	Difference
Stayers			
Stayed at original school	75.0	74.6	0.3
Movers			
Moved, same district	11.2	10.6	0.6
Moved, different district	6.3	7.4	-1.1
Moved, private, parochial, or other school	2.4	1.4	1.1
Leavers			
Left, to stay at home	0.8	1.3	-0.5
Left, in school or new job	3.9	4.2	-0.3
Left, other	0.4	0.5	-0.1
Unweighted Sample Size (Teachers)	**470**	**433**	**903**
Unweighted Sample Size (Schools)	**201**	**193**	**394**

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

Note: Data are weighted to account for the study design. Difference in the distributions is not statistically significant using a design-based F-test (p= 0.890).

We also examined movers' and leavers' self-reported reasons for leaving their schools and found no statistically significant impacts of treatment. The reasons for moving out of one's original school are listed in Table V.7 in terms of both reasons that were cited by movers and the top-cited reason by each mover. The reasons for leaving are not presented because there were too few cases to draw meaningful inferences, but the raw differences did not suggest a strong treatment effect. 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. We will repeat these analyses in the coming years when we collect additional follow-up data, at which point we expect there to be more teacher mobility to explain.

The reasons for moving provide some insight into the problem that teacher induction was meant to address. Dissatisfaction with the administrative support was the most commonly cited single reason for treatment group movers (21 percent) and involuntary transfer was most commonly cited by control group teachers (21 percent), although there were a variety of reasons given by teachers in both groups. We will know more about the teacher mobility question when we have collected the two additional rounds of data that will ultimately extend into the teachers' fourth year.

Table V.7. Impacts on Reasons for Moving out of the School (Percentages)

	"Very" or "Extremely" Important Reason				Single Most Important Reason		
Reason	Treatment	Control	Difference	P-Value	Treat-ment	Control	Difference
Moved out of area	16.5	18.1	-1.6	0.775	6.2	7.3	-1.1
Moved, spouse/partner's job	7.8	7.3	0.6	0.885	0.0	3.7	-3.7
Salary or benefits	16.7	11.5	5.1	0.313	3.2	2.6	0.6
Job security	22.2	14.0	8.2	0.174	0.0	0.0	0.0
Workplace conditions (e.g. facilities, classroom resources, school safety, parent and community support)	53.3	40.7	12.7	0.122	6.3	10.1	-3.7
Opportunities for desirable teaching assignment	51.5	40.8	10.7	0.154	7.4	10.5	-3.1
Dissatisfied with administrative support	53.3	43.3	10.1	0.205	20.5	8.9	11.6
Principal's leadership	51.0	41.7	9.3	0.222	8.6	11.1	-2.5
Changes in responsibilities	13.1	15.4	-2.3	0.684	0.0	0.0	0.0
Challenges of implementing new reform measures	17.1	6.0	11.1*	0.012	1.1	0.0	1.1
Difficulty with colleagues	13.5	6.7	6.8	0.111	0.0	0.0	0.0
Autonomy over the classroom	21.1	13.3	7.8	0.214	1.2	2.0	-0.8
Lesson planning time	18.6	11.6	6.9	0.249	0.0	0.0	0.0
Professional development opportunities	9.9	14.6	-4.7	0.398	0.0	1.2	-1.2
Involuntary transfer	28.0	37.0	-9.0	0.207	18.1	21.0	-3.0
Not asked to return	11.5	16.7	-5.1	0.329	2.5	4.1	-1.5
Unweighted Sample Size (Teachers)	**91**	**81**	**172**		**82**	**75**	**157**

Source: MPR Mobility Survey administered in fall/winter 2006-2007

Note: Data are weighted to account for the study design. Difference in the distributions of "single most important reason" responses is not statistically significant using a design-based F-test (p= 0.446).

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

None of the differences is significantly different from zero after applying a Benjamini-Hochberg correction.

We found that the treatment had no statistically significant positive impact on the district stayers in terms of classroom practices and student achievement after one year. Table V.8 presents the impacts on classroom practices and student achievement outcomes for those who returned to teach in the same district for the 2006–2007 school year. The impact estimates for stayers showed no evidence of a statistically significant treatment effect in terms of classroom practices and a significant negative impact on achievement. The intervention retained teachers who produced student achievement gains that were 8 percent of a standard deviation lower than the gains produced by teachers retained in the control group.

Table V.9 shows the background characteristics of teachers by mobility status. Across a wide variety of characteristics we found no differences between the treatment group stayers 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.

3. Sensitivity Analysis

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. We examined the impacts for the two types of districts separately and found no significant impact within ETS districts and no significant impact within NTC districts (results shown in Tables G.19 and G.20).

Finally, we considered nonresponse to the mobility survey. Though the overall response rate to the teacher mobility survey was high (89 percent), the response rates for treatment and control groups differed (92 and 86 percent, respectively). If nonrespondents differed from respondents in characteristics related to outcomes, then the differential nonresponse could bias the impact estimates.

We re-estimated the impacts under alternative assumptions about nonrespondents (Appendix G, Table G.21) and found no impacts of treatment except under the most extreme and implausible assumptions. For example, the conclusions did not change when we used an enhanced weight that incorporated information from the teacher background survey or when we used no weights.[37] Nor did they change when we assumed that all nonrespondents were stayers or all were leavers. The only exceptions were our 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 gives an

[37] 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 the 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, and whether they majored in an education-related field, and amount of student-teaching experience.

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. The impact estimates are presented separately by district in Appendix H, Figure H.8

Table V.8. Impacts on Classroom Practices and Student Achievement, District Stayers Only

Outcome	Treatment	Control	Difference	Effect Size	P-value
Classroom Practices (Average Score on a 5-point Scale)					
Implementation of literacy lesson	2.7	2.7	0.0	0.02	0.814
Content of literacy lesson	2.4	2.4	0.0	-0.05	0.586
Classroom culture	3.1	3.1	0.0	0.05	0.613
Unweighted Sample Size (Teachers)	**281**	**236**	**517**		
Unweighted Sample Size (Schools)	**153**	**141**	**294**		
Student Achievement (Effect Size)					
Reading scores (all grades)	0.00	0.01	-0.01	-0.01	0.785
Unweighted Sample Size (Students)	**2,100**	**1,750**	**3,850**		
Unweighted Sample Size (Teachers)	**122**	**99**	**221**		
Unweighted Sample Size (Schools)	**94**	**73**	**167**		
Math scores (all grades)	-0.04	0.04	-0.08*	-0.08	0.037
Unweighted Sample Size (Students)	**1,874**	**1,647**	**3,521**		
Unweighted Sample Size (Teachers)	**113**	**95**	**208**		
Unweighted Sample Size (Schools)	**87**	**70**	**157**		

Source: MPR classroom observations conducted in spring 2006; MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 and Mobility Survey administered in 2006-2007 to all study teachers.

Note: Classroom practice means are weighted and regression-adjusted using ordinary least squares to account for differences in baseline characteristics and the study design. Student achievement means are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools.

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

Table V.9. Characteristics of District Stayers, Movers, and Leavers by Treatment Status (Percentages Except Where Noted)

Teacher Characteristic	Treatment			Control			Difference		
	Stayers	Movers	Leavers	Stayers	Movers	Leavers	Stayers	Movers	Leavers
College Entrance Exam Scores (SAT combined score or equivalent)	1,000	1,009	1,016	1,009	997	1,059	-9	12	-44
Attended Highly Selective College	29.6	27.6	42.4	27.9	44.8	32.7	1.7	-17.2	9.7
Major or Minor in Education	73.5	61.8	70.3	74.0	81.8	51.9	-0.5	-20.0	18.4
Student Teaching Experience (Weeks)	14.5	14.5	12.3	13.9	13.6	11.7	0.6	0.8	0.6
Highest Degree Is Master's or Doctorate	20.7	20.5	22.7	21.6	30.5	23.2	-0.9	-10.0	-0.6
Entered the Profession Through Traditional Four-Year Program	64.1	61.7	35.7	60.3	62.3	37.6	3.8	-0.6	-1.8
Certified (Regular or Probationary)	92.1	97.1	95.7	94.2	94.3	87.0	-2.1	2.8	8.7
Career Changer	14.7	10.4	21.1	13.4	15.9	25.8	1.2	-5.5	-4.7
Unweighted Sample Size (Teachers)	**394**	**40**	**23**	**361**	**38**	**26**			
Unweighted Sample Size (Schools)	**188**	**34**	**21**	**180**	**36**	**25**			

Source: MPR calculations using data from the College Board and ACT, Inc.; MPR Teacher Background Survey administered in 2005-2006, Mobility Survey administered in fall/winter 2006-2007, and 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 (190/59/13 treatment stayers/movers/leavers and 183/44/11 control stayers/movers/leavers) and schools (111/36/7 treatment and 106/24/6 control). See Table V.6 for a definition of stayers, movers, and leavers.

None of the differences between treatment stayers and control stayers, between treatment movers and control movers, or between treatment leavers 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 VI

CORRELATIONAL ANALYSES

We have shown that the treatment and control groups were equivalent on baseline characteristics (Chapter II) and then were exposed to different levels of beginning teacher support (Chapter IV). We also showed, however, that the comprehensive induction services did not translate into positive impacts as we had hypothesized in our conceptual framework in Figure I.1. There were no statistically significant positive impacts on teacher attitudes, classroom practices, student achievement, or teacher retention after one year of the intervention (Chapter V). Given the prevalence of supports reported by control teachers, this lack of positive impacts led us to explore the relationship between induction and professional development services and outcomes, regardless of whether a teacher was in the control group or treatment group. We report on correlational (nonexperimental) analyses in which we examined how the variation in the type and intensity of teacher supports is related to outcome measures. For example, even though the average difference of 21 minutes of weekly mentor meeting time between the two treatment groups (as reported in Chapter IV) did not improve outcomes, it is possible that greater levels of support may be related to improvements in classroom practice, student test scores, or teacher retention.

We interpret the results with caution because the analyses are correlational and not causal. A high level of services for a particular teacher may result from the principal's decision to help struggling teachers who would likely have poor outcomes anyway. Alternately, a high level might be obtained if an assertive, motivated teacher, who would have had positive outcomes anyway, spends extra time with a mentor by taking the initiative to seek the extra help. In both of these examples, the estimate of the effect of induction services on 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.

1. Methods

We analyzed a set of key measures of the induction services received by both treatment and control teachers. The primary dimensions on which teacher induction programs can vary are the types of services teachers receive, the purpose of the induction program (e.g.

general support or instructional practice), and the duration and intensity of involvement (Ingersoll and Kralik 2004).[38] For the types of services received, we included strategies discussed in the teacher induction literature as important program components (Portner 2005) and that ETS and NTC emphasized in their comprehensive induction programs (see Chapter IV). For measures of program purpose, we selected items that captured a focus on instructional practice, given that this was the intended objective of the comprehensive induction programs. The duration and intensity measures capture the amount of time engaged in core mentoring activities and associated professional development.

Four variables measure the type of services the beginning teacher received:

- Whether the beginning teacher was assigned a mentor (yes/no)
- Whether the beginning teacher met with a literacy or math coach in the prior three months (yes/no)
- Whether the beginning teacher worked with a study group (with new or both new and experienced teachers) during the prior three months (yes/no)
- Whether the beginning teacher observed others teaching during the prior three months (yes/no)

Four measures capture the instructional purpose of the support:

- Whether the mentor gave the beginning teacher suggestions to improve his/her practices during the most recent full week of teaching (yes/no)
- Whether the beginning teacher received a "moderate amount" or "a lot" of guidance in math content during the prior three months (yes/no)
- Whether the beginning teacher received a "moderate amount" or "a lot" of guidance in literacy content during the prior three months (yes/no)
- Frequency with which beginning teacher received feedback on teaching, whether or not as part of a formal evaluation, during the prior three months (number of times in three-month period)

For program duration and intensity, two measures capture the amount of time spent on aspects of mentoring:

- Time the beginning teacher spent in mentoring sessions (both scheduled and informal) during a typical week (hours per week)

[38] 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.

- Time the mentor spent observing the beginning teacher teaching during the most recent full week of teaching (hours per week)

The final two measures capture time spent on professional development activities other than direct mentoring:

- Time spent on instructional techniques and strategies as part of professional development activities (for example, in-service workshops, study groups, seminars, and continuing education courses) during the prior three months (hours per three-month period)

- Time spent on content area knowledge (language arts, math, science) as part of professional development activities (for example, in-service workshops, study groups, seminars, and continuing education courses) during the prior three months (hours per three-month period)

The analyses mimic the experimental analyses discussed in Chapter V, but instead of assignment to treatment status, which was randomly determined, the key explanatory variables are the measures of the types of services received and the intensity of services listed above. For each analysis (classroom practices, student achievement, teacher retention), we control for the same baseline characteristics that we did in the corresponding experimental analysis. See Appendix C for details. We conducted the analyses twice, once using measures of induction services reported by teachers in the fall and once using their spring reports. If more induction services and more intense services are associated with better teacher and student outcomes, our measures of the level of services provided should be positively related to each outcome.

2. Nonexperimental Results

We find varying levels of evidence of positive relationships between mentoring and outcomes across the three domains. Table VI.1 shows the results for three classroom practice outcomes (lesson content, lesson implementation, and culture of classroom), Table VI.2 for two student achievement measures (math and reading), and Table VI.3 for two teacher mobility outcomes (remaining in the same school district and remaining in the teaching profession). Of the 12 measures of induction support and intensity, there are 11 measures used for each classroom practice outcome (receiving guidance in math content is not included), 11 measures used for student achievement (receiving guidance in math content is not included for reading scores and vice versa), and all 12 measures for teacher mobility. Combining the three classroom practice outcomes for 11 measures across 2 time periods (fall and spring), 2 of 66 estimated relationships are positive and significant and 5 are negative and significant. Combining results for the two student achievement outcomes for 11 measures across two time periods, 8 of 44 relationships are positive and significant and one is negative and significant. For teacher mobility, combining results for two outcomes for 12 measures across two time periods, 19 of 48 are positive and significant and none is negative and significant. We discuss each domain separately, with additional information on how to interpret the magnitude of the estimates, which relationships remain significant after

applying a Benjamini-Hochberg correction for multiple comparisons, and the magnitude of these associations.

a. Classroom Practices

Fifty-nine of the 66 relationships between teacher induction measures and classroom practices measured in the first year were not statistically significant. Table VI.1 shows the relationship between individual teacher support variables and classroom practice scores. Each estimate (coefficient) represents the predicted difference in classroom practice scores (on a five-point scale) that is associated with a one-unit change in the induction variable. Before applying a Benjamini-Hochberg correction, 2 of the 66 estimates is significantly positive and 5 are significantly negative. After applying a Benjamini-Hochberg correction to the 22 coefficients obtained for any of these three classroom practice outcomes, all results are statistically insignificant.

b. Student Achievement

The associations between teacher induction supports and student achievement are shown in Table VI.2. Each estimate in Table VI.2 is stated in terms of a standard unit of test scores. Because test scores have been standardized to have mean zero and standard deviation one, the magnitude of each estimate can be interpreted as an effect size. For example, according to fall measures, students of beginning teachers who were assigned a mentor score lower on the math test by 9 percent of a standard deviation or 0.09 of a standard deviation, all else equal. Just as with the benchmark model in Chapter V, each regression includes a pre-test control variable for each student. Results are shown separately for math and reading.

Eight of 44 associations between induction services and test scores are positive and significant and one is negative and significant before applying a Benjamini-Hochberg correction. After applying this correction to the 22 math estimates, there are two significant estimates, both positive: the frequency with which the beginning teacher received feedback (fall) and whether or not the beginning teacher met with a literacy or math coach (fall). For reading, after applying an analogous correction, there is a significant positive relationship between feedback (measured in the fall) and reading. All else equal, the students of teachers who met with a subject coach in the fall score higher on math tests by 0.14 of a standard deviation. The students of teachers who received feedback on teaching score higher on both math and reading tests by 0.02 of a standard deviation per instance that the teacher received feedback during three months in the fall.

Table VI.1.Association Between Beginning Teacher (BT) Support and Classroom Practices

	Fall						Spring					
	Lesson Content		Lesson Implementation		Culture of Classroom		Lesson Content		Lesson Implementation		Culture of Classroom	
Induction Measure	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
BT was assigned a mentor	-0.07	0.486	-0.04	0.697	-0.08	0.531	-0.02	0.814	-0.04	0.715	-0.09	0.487
BT met with a literacy or math coach	0.07	0.409	0.06	0.467	-0.07	0.388	0.09	0.282	0.12	0.135	0.15	0.082
BT worked with a study group	-0.01	0.884	0.02	0.751	0.00	0.987	0.17*	0.006	0.17*	0.014	0.14	0.051
BT observed others teaching	-0.13	0.071	-0.17*	0.037	-0.18*	0.023	-0.06	0.370	-0.04	0.601	-0.02	0.762
Mentor gave the BT suggestions in last week	-0.07	0.379	-0.10	0.210	-0.06	0.466	-0.09	0.168	-0.10	0.162	-0.09	0.292
BT received guidance in literacy content	-0.09	0.167	-0.05	0.462	-0.09	0.217	-0.05	0.416	-0.03	0.614	-0.03	0.643
BT received feedback on teaching (number of times)	0.00	0.704	0.01	0.637	0.02	0.169	-0.02	0.083	-0.01	0.587	0.00	0.721
Time BT spent in mentoring sessions (hours per week)	-0.02	0.363	-0.03	0.122	-0.02	0.282	-0.02	0.380	-0.04*	0.039	-0.03	0.100
Time mentor spent observing the teacher (hours per week)	-0.05	0.405	-0.03	0.703	-0.02	0.794	-0.07	0.318	-0.11	0.084	-0.05	0.505
Time spent on instructional techniques (hours per three months)	0.04	0.160	0.03	0.269	0.05	0.075	0.05	0.091	0.03	0.239	0.03	0.305

Table VI.1 *(continued)*

	Fall						Spring					
	Lesson Content		Lesson Implementation		Culture of Classroom		Lesson Content		Lesson Implementation		Culture of Classroom	
Induction Measure	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
Time spent on content area knowledge (hours per three months)	0.02	0.545	0.02	0.482	-0.03	0.237	-0.03	0.229	-0.06*	0.032	-0.09*	0.004
Unweighted Sample Size (Teachers)	573		573		573		585		585		585	

Source: MPR classroom observations conducted in spring 2006, MPR Teacher Background Survey administered in 2005-2006 and MPR First and Second Induction Activities Surveys administered in fall/winter 2005-2006 and spring 2006 to all study teachers.

Note: Each coefficient shown in the table is a coefficient estimate from a separate regression. The explanatory variable is listed in the row label. The dependent variable is listed in the column label. Sample sizes vary due to item nonresponse.

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

None of the differences is significantly different from zero after applying a Benjamini-Hochberg correction.

Table VI.2. Association Between Beginning Teacher (BT) Support and Test Scores

	Fall				Spring			
	Math Score		Reading Score		Math Score		Reading Score	
Induction Measure	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
BT was assigned a mentor	-0.09	0.108	0.01	0.810	-0.03	0.609	0.04	0.494
BT met with a literacy or math coach	0.14*†	0.002	0.10*	0.021	0.09*	0.027	0.05	0.175
BT worked with a study group	0.00	0.938	0.02	0.522	0.03	0.417	0.00	0.927
BT observed others teaching	-0.06	0.109	-0.08	0.065	-0.01	0.793	0.02	0.658
Mentor gave the BT suggestions in last week	-0.09*	0.024	0.00	0.974	0.01	0.883	0.00	0.990
BT received guidance in math content	0.08	0.060	n.e.	n.e.	0.07*	0.047	n.e.	n.e.
BT received guidance in literacy content	n.e.	n.e.	0.08	0.052	n.e.	n.e.	0.05	0.127
BT received feedback on teaching (number of times)	0.02*†	0.004	0.02*†	0.001	0.02*	0.012	0.01*	0.014
Time BT spent in mentoring sessions (hours per week)	-0.01	0.598	0.01	0.315	0.01	0.254	0.00	0.766
Time mentor spent observing the teacher (hours per week)	-0.05	0.228	-0.01	0.745	0.00	0.954	-0.01	0.767
Time spent on instructional techniques (hours per three months)	0.00	0.779	0.00	0.854	0.01	0.397	0.00	0.776
Time spent on content area knowledge (hours per three months)	0.01	0.478	0.00	0.994	0.02	0.123	-0.02	0.263
Unweighted Sample Size (Districts)	**14**		**15**		**14**		**15**	
Unweighted Sample Size (Schools)	**169**		**186**		**169**		**186**	
Unweighted Sample Size (Teachers)	**234**		**265**		**233**		**262**	
Unweighted Sample Size (Students)	**3,939**		**4,489**		**3,974**		**4,486**	

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 and First and Second Induction Activities Surveys administered in fall/winter 2005-2006 and spring 2006 to all study teachers.

Note: Each coefficient shown in the table is a coefficient estimate from a separate regression. The explanatory variable is listed in the row label. The dependent variable is listed in the column label. Sample sizes vary due to item nonresponse.

n.e. = not estimated

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

†Significantly different from zero at the 0.05 level, two-tailed test, after applying a Benjamini-Hochberg correction.

Table VI.3. Association Between Beginning Teacher (BT) Support and Teacher Mobility

	Fall				Spring			
	Remains in District		Remains in Teaching		Remains in District		Remains in Teaching	
Induction Measure	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
BT was assigned a mentor	0.01	0.821	0.00	0.755	0.09*	0.017	0.04*†	0.006
BT met with a literacy or math coach	0.01	0.728	0.01	0.439	-0.01	0.768	0.01	0.211
BT worked with a study group	0.01	0.564	0.01	0.610	0.05*	0.015	0.01	0.515
BT observed others teaching	-0.05	0.063	0.00	0.918	0.02	0.491	0.01	0.259
Mentor gave the BT suggestions in last week	0.00	0.995	0.00	0.938	0.04	0.091	0.02*	0.035
BT received guidance in math content	0.01	0.611	0.03*	0.023	0.06*†	0.004	0.02*†	0.010
BT received guidance in literacy content	0.01	0.737	0.03*	0.018	0.03	0.162	0.03*†	0.000
BT received feedback on teaching (number of times)	0.00	0.688	0.00	0.604	0.01*	0.013	0.00	0.061
Time BT spent in mentoring sessions (hours per week)	0.00	0.757	0.01	0.304	0.02*	0.024	0.01*	0.046
Time mentor spent observing the teacher (hours per week)	0.00	0.986	0.01	0.495	0.07*	0.020	0.02*	0.045
Time spent on instructional techniques (hours per three months)	0.03*†	0.002	0.01*†	0.009	0.02*†	0.005	0.01	0.077
Time spent on content area knowledge (hours per three months)	0.02	0.079	0.02*†	0.000	0.02	0.051	0.01*	0.038
Unweighted Sample Size (Teachers)	**668**		**606**		**663**		**601**	

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

Note: Each coefficient shown in the table is the marginal effect in a logit model calculated at the mean of the independent variables from a separate regression. The explanatory variable is listed in the row label. The dependent variable is listed in the column label. Sample sizes vary due to item nonresponse.

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

†Significantly different from zero at the 0.05 level, two-tailed test, after applying a Benjamini-Hochberg correction.

c. Teacher Mobility

Table VI.3 presents results using an indicator variable for whether or not the teacher continued teaching in the same school district the following year (columns 1 and 3) or remained in the teaching profession (columns 2 and 4). Similar to the experimental analysis, the model is a logit model. The results presented are marginal effects calculated from a logit model with the independent variables set at their means for the full sample. Thus each estimate shows how receiving a given support increases or decreases the likelihood that a teacher remains in the school district or teaching profession, all else equal. The estimates are measured in percentage point changes. For example, based on the fall measures a teacher who is assigned a mentor is one percent more likely than a teacher not assigned a mentor to remain teaching in the district the following year.

Before applying a Benjamini-Hochberg correction for multiple comparisons, 19 of 48 associations between induction services and teacher mobility are positive and significant and none is negative and significant. More associations are positive and significant using spring measures than fall measures: 14 of 24 using spring measures compared to 5 of 24 using fall measures. Eight of 24 measures are positive and significant for remaining in the district and 11 of 24 for remaining in teaching. After applying the Benjamini-Hochberg correction to the 24 variables within each outcome, 3 associations are positive and significant for remaining in the district, and 5 associations are positive and significant remaining in teaching.

Considering only associations that remain significant after applying the Benjamini-Hochberg correction, the measures that are significantly related to remaining in the district are receiving guidance in math content (spring) and the professional development time spent on instructional techniques and strategies (spring and fall). For remaining in teaching, the measures are being assigned a mentor (spring), receiving guidance in math content (spring), receiving guidance in literacy content (spring), professional development time spent on instructional techniques and strategies (fall), and professional development time spent on content area knowledge (fall). The magnitude of the relationships varies. Although 94 percent of teachers continued to the second year, being assigned a mentor in the spring is associated with a 4 percentage point increase in remaining in teaching. Receiving guidance in math or literacy content in the spring is associated with an increased likelihood of remaining in the district of 6 percentage points and of remaining in teaching of between 2 and 3 percentage points. Each hour per three-month interval spent on instructional techniques is associated with an increased likelihood of remaining in the district of between 2 and 3 percentage points (higher with the fall measure) and of one percentage point on the likelihood of remaining in teaching (measured in the fall). Each hour per three-month interval spent on content area knowledge (measured in the fall) is associated with a 2 percentage point increase of the likelihood of remaining in teaching

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). U.S. Department of Education. Washington, DC: 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, G., E. Crowe, and B. 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.

Cawelti, Gordon (ed.). *Handbook of Research on Improving Student Achievement*. Third edition. Arlington, VA: Educational Research Service, 2004.

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

Darling–Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, vol. 8 , no. 1, <http://epaa.asu.edu>.

Decker, Paul, John Deke, Amy Johnson, Daniel Mayer, John Mullens, and Peter Schochet. "The Evaluation of Teacher Preparation Models: Design Report." Princeton, NJ: Mathematica Policy Research, Inc., October 2005.

Dorans, N.J., Lyu, C.F., Pommerich, M, and Houston, W.M. "Concordance Between ACT Assessment and Recentered SAT I Sum Scores." College and University, vol. 73, 1997, pp. 24-31.

Dwyer, Carol A. *Development of the Knowledge Base for the Praxis III: Classroom Performance Assessments Assessment Criteria.* Princeton, NJ: Educational Testing Service, 1994.

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

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

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, R. H. and R. 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." Statistical analysis report. Washington, DC: 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. 1967. "The Behavior of Maximum Likelihood Estimation under Nonstandard Conditions." Pp. 221-33 in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1, edited by L. M. LeCam and J. Neyman. Berkeley, CA: University of California Press.

Ingersoll, Richard M. "Is There Really a Teacher Shortage?" Research report co-sponsored by the Center for the Study of Teaching and Policy and the Consortium for Policy Research in Education. 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, R. 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, D., Johnson, S. M., Kardos, S. M., Liu, E., & Peske, H. G. "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.B., and Newmann, F.M. "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: 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." Daniel J. Evans School of Public Affairs, University of Wisconsin, Madison Working Paper 13, 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, issue 23, 2003.

National Commission on Teaching and America's Future. *The High Cost of Teacher Turnover. Policy Brief*. Washington, DC: National Commission on Teaching and America's Future, 2007.

National Institute of Child Health and Human Development. *Report of the National Reading Panel. Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction—Reports of the Subgroups.* NIH Publication No. 00-4754. Washington, DC: U.S. Government Printing Office, 2000.

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.

Saginor, Nicole, and Phil Hyjek. "Observing Standards-Based Classrooms: The Vermont Classroom Observation Tool (VCOT)." Montpelier, VT: Vermont Institutes, March 2005.

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.

Tourkin, Steven C., Toni Warner, Randall Parmer, Cornette Cole, Betty Jackson, Andrew Zuckerberg, Shawna Cox, and Andrew Soderborg. *Documentation for the 2003-04 Schools*

and Staffing Survey (NCES 2007-337). U.S. Department of Education. Washington, DC: National Center for Education Statistics, 2007.

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: 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: National Commission on Teaching and America's Future, 2002.

APPENDIX A

NATIONAL DATA ON TEACHER INDUCTION

We sought evidence about the degree to which the districts included in the study are typical of large urban school districts with respect to the prevailing level of teacher induction supports. The study would be less generalizable if the districts included in the study had levels of prevailing teacher induction that were unusually high compared to other districts in the United States. Although we screened districts to exclude those that already had comprehensive teacher induction, it is possible that the screening was imperfect. Therefore, we analyzed data from the Schools and Staffing Survey, or SASS (Tourkin et al. 2007) to provide some context for interpreting the current study.

The SASS analysis draws on the 2003-2004 school year, two years prior to the study. The data are used to characterize the prevailing conditions in the study's control group and provide teacher-level reports of induction support received in a set of comparison districts drawn from a national sample. The comparison districts included all study districts plus any other districts in the SASS that had at least 570 teachers in elementary schools and 50 percent of students eligible for free or reduced-price meals under the federal School Lunch Program. In other words, the comparison districts compose the universe of school districts that could have been selected for inclusion in the study under the criteria we specify in Chapter II.

Table A.1. Induction Support Reported by First-Year K-6 Teachers in 2003-2004 by Type of District (Percentages)

Level of Induction Support[a]	All Districts	Comparison Districts[b]	Study Districts	Difference Between Study Districts and All Districts[c]	Difference Between Study Districts and Comparison Districts[d]
Less than Basic Induction	37.2	50.9	70.1	32.9	19.2
Basic Induction	27.6	18.9	13.2	-14.4	-5.7
Basic Induction + Collaboration	35.2	30.1	16.8	-18.5	-13.3
Number of Teachers	**642**	**129**	**16**		
Number of Districts	**505**	**58**	**9**		

Source: MPR calculations using data from the 2003-2004 Schools and Staffing Survey.

Notes: Data are weighted using teacher-level weights provided by the SASS to account for survey design and nonresponse.

[a]Levels of induction support are defined in accordance with Smith and Ingersoll (2004) to reflect teacher responses to a series of yes/no questions about whether they received each of several types of support.

[b]Comparison districts included all study districts plus any other districts in the SASS that had at least 570 teachers in elementary schools and 50 percent of students eligible for free or reduced-price meals under the federal School Lunch Program.

[c]The p-value from the chi-squared test of independence is 0.000.

[d]The p-value from the chi-squared test of independence is 0.076.

To characterize how comprehensive the induction supports were, we divided teachers into three categories, "less than basic induction," "basic induction," and "basic induction + collaboration," definitions used by Smith and Ingersoll (2004), who investigated teacher induction using an earlier wave (1999-2000) of the SASS.[39] Each cell in the top portion of Table A.1 indicates the percentage of first-year K-6 teachers in the three levels of induction support. The results are presented for three sets of districts. The first column presents percentages for all districts participating in the SASS that had first-year K-6 teachers. The second column presents percentages for the comparison districts. The third column is for the study districts, all 17 of which participated in the 2003-04 SASS but only 9 of which had

[39] The levels of induction support are: (1) Less Than Basic Induction, defined as teachers who may have received some services but lacked two key elements—a mentor and supportive communication with school administrators or their department chair; (2) Basic Induction, defined as teachers who had a mentor and supportive communication but not all the elements in the next highest package; (3) Basic + Collaboration, defined as teachers who had a mentor in the same field, supportive communication, common planning time or regularly scheduled collaboration with other teachers in their subject area, and participation in a seminar for beginning teachers.

first-year teachers in the SASS sample. The final two columns subtract the percentages for the study districts from the national sample and the comparison districts, respectively.

We did not find evidence of a statistically significant difference between induction supports reported by teachers in the study districts and those reported by teachers in the comparison districts. A chi-squared test of independence fails to reject the null hypothesis that the two observed distributions of induction support levels are the same ($p=0.076$).

A chi-squared test comparing study districts to all districts in the SASS, however, does reject the null hypothesis. Seventy percent of teachers in study districts had "less than basic support" compared to 37 percent in the national sample and 17 percent had "basic induction + collaboration" compared to 35 percent of teachers in districts nationally. The significant difference suggests that the study was successful in identifying the sub-population of school districts that offer fewer induction supports than districts nationally.

APPENDIX 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, 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, 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, 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 (B.1):

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

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.

APPENDIX C

IMPACT ESTIMATION METHODS

To implement the regression approach, we used a two-level model where level 1 corresponds to teachers and level 2 to schools.[40] Treatment effects are estimated in the level 2 model, where the sample size is dictated by the number of schools, not teachers. The basic form of the model is given in Equations (C.1) and (C.2), which express teacher-level and school-level analyses, respectively:

$$(C.1) \quad Y_{ij} = c_j + \beta' X_{ij} + e_{ij}$$

$$(C.2) \quad c_j = \mu + \delta T_j + \gamma' Z_j + u_j,$$

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 must estimate the variance of the school effects u_j.

By substituting Equation (C.2) into Equation (C.1), we can express the unified model as Equation (C.3):

$$(C.3) \quad Y_{ij} = \mu + \delta T_j + \beta' X_{ij} + \gamma' Z_j + [u_j + e_{ij}].$$

In this formulation, the coefficient for the treatment group indicator represents the impact of the receipt of comprehensive induction services and is the main parameter of

[40] For the test score analysis, level 1 represents 6,666 students.

interest. The standard error of this impact estimate accounts for the design effects attributable to the clustering of teachers within schools, which occurs because teachers within schools tend to have similar outcomes.

Specification of the Outcome Variable. The model in Equation (C.1) has a generic outcome *Y*. In place of *Y*, we can substitute any outcome, such as quality of classroom lesson content. Some outcomes will be binary or categorical. For example, teacher mobility may be expressed as an indicator for whether the sample member returned for a second year of teaching, or it may be expressed as a variable with separate categories for remaining in, moving within, or leaving the profession. In the case of categorical outcome variables, we use bivariate or multinomial logistic regression to estimate the parameters of Equation (C.1).

Specification of the Treatment Variable. The simplest specification of the treatment variable (*T*) is to include a single indicator for whether the teacher's assigned school at the point of random assignment was in the treatment or control group. For the sensitivity analysis, we estimated separate treatment impacts for each district. We then examined the distribution of district-level impacts to determine whether there is evidence for a common treatment effect across districts or whether outliers or trends might otherwise be masked by reporting only the average impact across all districts.

Specification of the Explanatory Variables. A teacher background questionnaire, discussed in Chapter III, provides a long list of potential explanatory variables for inclusion in the model (the *X* vector), including demographic and household characteristics, information on teachers' education and professional background, and teaching assignment. In addition, we have access to school-level variables from the National Center for Education Statistics' Common Core of Data (CCD).[41]

The benchmark analyses included the following variables as covariates. The analysis of teacher attitudes (Table V.1) had district and grade fixed effects and no other covariates. The analysis of classroom practices (Table V.2) included teacher demographic characteristics (age, sex), teacher's educational and professional background (teacher preparation type, certification status, highest degree attained, months of teaching experience), teaching assignment (grade level), school characteristics (school-lunch program eligibility rate and racial/ethnic composition), and district and grade fixed effects.

The student achievement analyses (Tables V.3 and V.4) had normalized student pre-test score, student characteristics (student gender, race/ethnicity, special education status, English-language learner status, free/reduced-price lunch status, and whether the student was over age for grade), teacher personal characteristics (age, gender, race/ethnicity, whether 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, whether teacher holds a degree in an education-related field, whether

[41] CCD data are reported with a lag; therefore, the school-level information describes schools in 2004–2005, one year before the study year.

teacher is a first-year teacher, whether teacher was hired after school year began, whether teacher attended competitive college, whether teacher held a non-teaching job for five or more years), and district-by-grade fixed effects. We collected teacher SAT scores (or ACT scores converted to an SAT equivalent) for 53.3 percent of the teachers in the analysis sample. Due to the large amount of missing data, we did not use these scores as a control variable in the baseline analysis but do so in the sensitivity analysis.

Finally, the teacher retention analysis (Table V.5) included teacher personal characteristics (age, gender, race/ethnicity, whether teacher race/ethnicity matches that of a majority of students, marital status, whether the teacher has children), teacher professional characteristics (months of relevant teaching experience, certification status, whether teacher holds a degree in an education-related field, whether teacher was hired after school year began, whether teacher attended competitive college, whether teacher held a non-teaching job for five or more years, whether the teacher taught a single grade level), teacher neighborhood characteristics (commuting distance, whether the teacher is a homeowner, whether the teacher lives in the school district, and whether the 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), and district and grade fixed effects.

Estimation of the Variance Components. Equation (C.3) can be thought of as a mixed model or a hierarchical model. It is "mixed" because it contains fixed effects (represented by a, δ, and β) as well as random effects (represented by e and u). It is hierarchical because it embeds a school-level model (indexed by j) within a classroom-level model (indexed by i). Several techniques are available for estimating such a model, including ordinary least squares (OLS) with robust standard errors (see Huber 1967; White 1980); Generalized Least Squares (GLS) estimates of a random effects model; maximum likelihood; and restricted maximum likelihood. We estimated the standard errors of the model by using each of these methods, but the findings did not change. Therefore, we report findings in this report based on the robust standard errors that adjust for clustering of students and teachers within schools.

Nonexperimental Analysis. Chapter VI presents findings from nonexperimental analyses that are very similar in structure to the experimental analyses. Those analyses are based on Equation (C.3), except that we replace the treatment status indicator with a variable describing the level or intensity of teacher induction services reported by the teacher. The result is Equation (C.4):

$$(C.4) \quad Y_{ij} = \mu + \theta' Q_{ij} + \beta' X_{ij} + \gamma' Z_j + [u_j + e_{ij}]$$

where Qij, representing a measure of the type or intensity of induction services, replaces T, the indicator variable for assignment to the treatment group in Equation (C.3). The θ coefficient captures the relationship between an induction intensity measure and the outcome Y. We estimated the relationships between measures of types of services or their intensity and each of the three main outcomes of interest—classroom practices, student

achievement, and teacher mobility—by substituting measures of the outcomes for Y_{ij}. The same vector of X variables used in the experimental section is used here. The regressions are unweighted.[42]

We conducted the analysis twice, once using measures of induction reported by teachers in the fall and once using their spring reports. If more induction services and more intense services are associated with better teacher and student outcomes, our measures of the level of services provided should be positively related to each outcome.

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 to one of 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.

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. As an example, we consider the case where we wish to compare the gains for a grade 4 teacher named Ms. Smith in Seattle with those of a grade 5 teacher named Mr. Cone in Cleveland.[43] We assume that if Ms. Smith's students scored at the average level for all Seattle third graders in the pre-test year and 10 percent of a standard deviation above the grade 4 average at the end of the post-test year on a Washington State math assessment, then that would be considered equivalent to Mr. Cone's class in Cleveland performing at one-half of a standard deviation above the mean at the end of grade 4 on Ohio's state math assessment and then 0.6 of a standard deviation at the end of grade 5 (0.1 – 0.0 = 0.6 – 0.5). Both sets of students moved up one-tenth of a standard deviation relative to their local reference groups on their own state's assessment.

It is also possible to aggregate by subject matter. We kept two broad subject areas distinct—math and reading (which includes "English" and "language arts")—and present

[42] We also analyzed Equation (C.4) using a vector of induction variables for Q_{ij} instead of a single variable and obtained similar results to those reported in Chapter VI. For the purpose of easy exposition, results were presented in which induction variables were analyzed one at a time.

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

the findings separately for those two subjects. Within math or reading, some districts provided subtest scores or scores from several tests. In such cases, we combined the test or subtest scores by using an equal-weighted average of the normalized scores on the subtests. Whenever a student had not taken one of the tests, we did not include that test score in the average score to be used in the analysis. In the benchmark analysis, we dropped reading z-scores based on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) tests, a commonly used set of orally administered early reading assessments for young students (Good and Kaminski 2002). Three districts used the DIBELS test for grades K-2. We also dropped two lower grades in a district in which the z-scores were based on a combination of DIBELS-like tests. DIBELS tests are problematic because they were scored subjectively, raised concerns about missing student data, and were sometimes administered in the middle of a school year rather than at the beginning or end of the year. As part of the sensitivity analyses, we estimated models that included DIBELS tests. We excluded other subjects from the main impact analysis, such as foreign languages, social studies, or science, which are not available in enough districts to yield meaningful findings.

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, could be missing a test score because of repeated absence, or could not have been enrolled during the test period. These problems can result in a missing pre-test or post-test 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, the missing cases 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 2.0 percentage points for reading and 3.2 percentage points for math, we assumed that the data were missing at random.

An important consideration when interpreting findings based on district-administered tests is that the findings apply only to teachers in the tested grades and subjects. Because we relied on the pre-test from the year before program implementation, we also excluded the youngest grade at which testing begins. For example, in districts that test only in grades 3 through 8 and operate K through 5 elementary schools (the most common case), we were able to estimate impacts on achievement for grades 4 and 5. As part of the sensitivity analyses, we examined differences in post-test scores only and thus were able to consider more grades and include more students in the analysis.

APPENDIX D

CLASSROOM OBSERVATION METHODS

This Appendix describes the instrument and procedure used to observe teachers' classroom practices in the study.

THE VERMONT CLASSROOM OBSERVATION TOOL (VCOT)

The observers were trained to use the Vermont Classroom Observation Tool (VCOT) to assess instruction practices. The VCOT is a proprietary tool for classroom observations developed by the Vermont Institutes (see Saginor and Hyjek 2005). Researchers who first worked with Science and Math Program Improvement (SAMPI), a research group at Western Michigan University, developed the VCOT over several years. SAMPI had developed an instrument to measure the quality of standards-based, investigative science and mathematics instruction based on research conducted by Horizon Research, Inc.

In developing the VCOT, the Vermont Institutes staff used the SAMPI Observation Tool as a starting point and carefully reviewed Charlotte Danielson's Framework for Teaching (1996), on which the widely used Praxis III observational assessment (and ETS induction program) is based (Dwyer, 1994). In parallel with the Praxis III content, the VCOT developers included examples of evidence for each indicator, added systematic and ongoing formative and summative assessment of student learning as a major indicator, and simplified and shortened the tool. The VCOT underwent further refinement through its use in the field by a group of trained teacher-leaders who observed classrooms. In 2004, several of those involved in the original design of the VCOT adapted it for use in the observation of literacy lessons. The standards and practices included in the National Council of Teachers of English (NCTE) Standards and the National Reading Panel (NICHHD 2000) also helped inform development of the literacy version of the VCOT.

The VCOT describes teaching practices in four areas:

1. Planning and Organization of a Lesson
2. Implementation of a Lesson
3. Content of a Lesson
4. Classroom Culture

In this study we attempted to measure all but the first construct, lesson planning and organization. The procedure for assessing lesson planning and organization is more suited for individual teacher feedback than for research and requires measurement of activities before the start of a lesson and a separate teacher interview of varying length and content.

IMPLEMENTING THE VCOT

Staff from the Vermont Institutes trained the classroom observers. Much of the training relied on videotaped classes but also included practice observations conducted in pairs in "live" school settings. During the practice observations, observers scored independently and then debriefed to reach consensus on any individual items for which the discrepancy exceeded a single point. In addition to practice observations, observers participated in training for a total of nine days over the course of three training sessions.

After observing and scoring a videotaped class, observers were deemed "certified" to conduct the observations based on a comparison of their 16-item scores to the observations of a "gold standard" panel. The "gold standard" panel consisted of the tool's developer and two trained observers who demonstrated a clear understanding of the items measured in the tool and showed high rates of agreement in scoring. Trainees had two opportunities to come within 0.75 points of the "gold standard" average score for the three constructs (implementation, content, and culture) during a test observation. Trainees who did not meet the standard were not allowed to conduct observations. To address the possibility that observers' scoring would start to drift in one direction or another after conducting some observations, we asked the tool developer to observe a classroom with each observer in the field at least once to verify scoring. As mentioned in Chapter III, observers were always blind to teachers' treatment status and therefore did not know if they were observing someone who had received the comprehensive induction support.

INTERPRETING VCOT SCORES

The estimated impacts on classroom practices described in Chapter V can be better understood by relating the VCOT scores to student achievement. Consistent with the nonexperimental analyses presented in Chapter VI, we conducted correlational analyses to explore whether there is a relationship between student achievement gains and VCOT scores, regardless of whether a teacher was assigned to the control group or treatment group. After fielding the VCOT for this study and comparing the results with student achievement gains, we found a statistically significant positive relationship, with a 1-scale-point increase in

VCOT scores being associated with a 4-to 6-point increase in students' test score gains as measured in percentage of standard deviation units. We interpret the results of the correlational analyses with caution, however, because the analyses are correlational and not necessarily causal.

APPENDIX E

SUPPLEMENTAL TABLES FOR CHAPTER II

Table E.1. School Characteristics by Treatment Status in ETS Districts (Percentages)

School Characteristic	All Schools	Treatment	Control	Difference	P-value
Percent Eligible for School Lunch Program					0.879
<50%	6.3	7.7	4.9	2.8	
50–75%	27.7	27.0	28.4	-1.4	
75–100%	53.2	53.0	53.3	-0.4	
Unknown	12.8	12.3	13.3	-1.0	
Race/Ethnicity					0.803
Majority African American	24.7	24.5	25.0	-0.5	
Majority Hispanic	21.8	24.6	19.0	5.6	
Majority white	28.6	27.9	29.3	-1.4	
Majority other	0.0	0.0	0.0	0.0	
Other/mixed	24.9	23.1	26.7	-3.6	
Grade Configuration					0.226
Pre-K or K–5	85.4	85.6	85.2	0.4	
Pre-K or K–6	1.0	0.0	2.0	-2.0	
Pre-K or K–8	9.6	8.3	10.8	-2.5	
Other	4.1	6.1	2.1	4.1	
Number of Sample Teachers					0.518
1	32.0	32.0	32.1	-0.1	
2	23.7	21.7	25.7	-4.0	
3	18.9	22.3	15.4	6.9	
4	10.1	12.2	7.9	4.3	
5	9.2	6.6	11.8	-5.2	
More than 5	6.1	5.2	7.0	-1.8	
Unweighted Sample Size (Schools)	**203**	**100**	**103**		

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

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

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

Table E.2. School Characteristics by Treatment Status in NTC Districts (Percentages)

School Characteristic	All Schools	Treatment	Control	Difference	P-value
Percent Eligible for School Lunch Program					0.215
<50%	9.6	10.7	8.4	2.3	
50–75%	13.8	8.9	18.8	-9.9	
75–100%	74.8	78.5	71.1	7.4	
Unknown	1.8	1.9	1.7	0.2	
Race/Ethnicity					0.551
Majority African American	62.7	62.5	63.0	-0.5	
Majority Hispanic	21.6	24.3	18.8	5.5	
Majority white	5.6	5.0	6.2	-1.2	
Majority other	0.5	1.0	0.0	1.0	
Other/mixed	9.6	7.3	12.0	-4.8	
Grade Configuration					0.066
Pre-K or K–5	57.2	60.1	54.6	5.5	
Pre-K or K–6	4.7	0.8	8.3	-7.5	
Pre-K or K–8	31.2	30.7	31.6	-0.9	
Other	6.9	8.4	5.5	3.0	
Number of Sample Teachers					0.159
1	43.3	39.0	47.6	-8.5	
2	24.2	28.9	19.4	9.5	
3	17.4	19.2	15.6	3.6	
4	7.4	6.9	7.9	-1.0	
5	3.2	4.3	2.1	2.2	
More than 5	4.6	1.7	7.5	-5.8	
Unweighted Sample Size (Schools)	**215**	**110**	**104**		

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

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

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

APPENDIX F

SUPPLEMENTAL TABLES FOR CHAPTER IV

Table F.1. Mentor Characteristics—ETS

Characteristic	Percent	
Race/ethnicity		
White, non-Hispanic	70.0	
Black or African American, non-Hispanic	20.0	
Hispanic	a.	
Asian or Pacific Islander	0.0	
Other/multiple	0.0	
Gender (percent female)	100.0	
Education: Highest Degree Attained		
Bachelor's degree	18.2	
Master's degree	81.8	
Working Toward Advanced Degree or Additional Credits	35.0	
Certification		
Not certified	0.0	
Certified in one area	45.5	
Certified in multiple areas	54.6	
Area of Certification		
General elementary education	90.9	
Bilingual education	a.	
Special education	27.3	
Special subject area(s)	18.2	
Other area	31.8	
Certified Through National Board of Professional Teaching Standards (NBPTS)	0.0	
Working Toward Additional Certification	30.0	
Working Toward Certification Through NBPTS	a.	
Teaching Experience		
Taught within one year of hire as a mentor	81.8	
Have not taught for at least one year	18.2	
Ever worked in non-teaching position(s) within education	36.4	
	Average	Range (Min, Max)
Age in 2005 (Years)	45.0	(30, 61)
Teaching Experience (Years)	19.2	(5, 35)
Experience in Non-teaching Position(s) Within Education (Years)	1.1	(0, 5)
Caseload (Number of Beginning Teachers)	12.0	(9, 14)
Unweighted Sample Size (Mentors)	**22**	

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

Note: a. Values suppressed to protect respondent confidentiality.

Table F.2. Mentor Experience and Training—ETS

Characteristic	Percent	
Any Mentoring Experience	68.2	
Years of Mentoring Experience		
0	33.3	
1	a.	
2 or more	52.4	
Types of Teachers Mentored (If Have Mentoring Experience)		
Beginning teachers	46.7	
Veteran teachers	0.0	
Both beginning and veteran teachers	53.3	
Any Previous Mentoring Training (If Have Mentoring Experience)	41.4	
Areas of Mentoring Training (If Received Mentor Training)		
Classroom management	81.8	
Giving effective feedback	100.0	
Mentor roles	72.7	
Coaching strategies	75.0	
Lesson planning	81.8	
Classroom observations	72.7	
Helping adult learners set goals	45.5	
Analyzing student work	41.7	
Leading study groups	45.5	
Coaching in literacy/language	27.3	
Coaching in math	25.0	
	Average	Range (Min, Max)
Mentoring Experience (Years)	7.0	(1, 30)
Unweighted Sample Size (Mentors)	**22**	

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

Note: a. Values suppressed to protect respondent confidentiality.

Table F.3. Activities in Previous Mentor Experience—ETS

Activities in Previous Mentoring	Frequency of Activity (Percentages)			
	Never	A Few Times a Year or on Request	Monthly	Weekly or Bi-monthly
Helped teachers with strategies for effective instruction	31.8	27.3	13.6	27.3
Helped teachers plan lessons	31.8	31.8	a.	31.8
Helped teachers with classroom management	36.4	27.3	0.0	36.4
Observed teachers and provided feedback	36.4	40.9	a.	18.2
Helped teachers set goals to improve practice	40.9	31.8	a.	22.7
Provide opportunities for teachers to observe others	45.5	36.4	0.0	18.2
Reviewed teacher portfolios	68.2	31.8	0.0	0.0
Led study groups on teaching	68.2	18.2	a.	a.
Unweighted Sample Size (Mentors)	**22**			

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

Note: a. Values suppressed to protect respondent confidentiality.

Table F.4. Mentor Characteristics—NTC

Characteristic	Percent	
Race/ethnicity		
White, non-Hispanic	33.3	
Black or African American, non-Hispanic	42.9	
Hispanic	23.8	
Asian or Pacific Islander	0.0	
Other/multiple	0.0	
Gender (percent female)	90.9	
Education: Highest Degree Attained		
Bachelor's degree	a.	
Master's degree	90.9	
Working Toward Advanced Degree or Additional Credits	30.0	
Certification		
Not certified	0.0	
Certified in one area	45.5	
Certified in multiple areas	54.6	
Area of Certification		
General elementary education	90.9	
Bilingual education	a.	
Special education	a.	
Special subject area(s)	36.4	
Other area	31.8	
Certified Through National Board of Professional Teaching Standards (NBPTS)	27.3	
Working Toward Additional Certification	a.	
Working Toward Certification Through NBPTS	a.	
Teaching Experience		
Taught within one year of hire as a mentor	81.8	
Have not taught for at least one year	a.	
Ever worked in non-teaching position(s) within education	54.6	
	Average	Range (Min, Max)
Age in 2005 (Years)	41.3	(28, 57)
Teaching Experience (Years)	46.6	(5, 32)
Experience in Non-teaching Position(s) Within Education (Years)	1.7	(0, 6.8)
Caseload (Number of Beginning Teachers)	11.4	(8, 14)
Unweighted Sample Size (Mentors)	**22**	

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

Note: a. Values suppressed to protect respondent confidentiality.

Table F.5. Mentor Experience and Training—NTC

Characteristic	Percent	
Any Mentoring Experience	86.4	
Years of Mentoring Experience		
0	a.	
1	a.	
2 or more	72.7	
Types of Teachers Mentored (If Have Mentoring Experience)		
Beginning teachers	31.6	
Veteran teachers	a.	
Both beginning and veteran teachers	63.2	
Any Previous Mentoring Training (If Have Mentoring Experience)	68.4	
Areas of Mentoring Training (If Received Mentor Training)		
Classroom management	a.	
Giving effective feedback	76.9	
Mentor roles	100.0	
Coaching strategies	84.6	
Lesson planning	76.9	
Classroom observations	58.3	
Helping adult learners set goals	58.3	
Analyzing student work	58.3	
Leading study groups	33.3	
Coaching in literacy/language	46.2	
Coaching in math	16.7	
	Average	Range (Min, Max)
Mentoring Experience (Years)	5.6	(1, 20)
Unweighted Sample Size (Mentors)	**22**	

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

Note: a. Values suppressed to protect respondent confidentiality.

Table F.6. Activities in Previous Mentor Experience—NTC

Activities in Previous Mentoring	Frequency of Activity (Percentages)			
	Never	A Few Times a Year or on Request	Monthly	Weekly or Bi-monthly
Helped teachers with strategies for effective instruction	a.	a.	18.2	59.1
Helped teachers plan lessons	a.	a.	22.7	50.0
Helped teachers with classroom management	a.	19.1	a.	57.1
Observed teachers and provided feedback	18.2	22.7	a.	50.0
Helped teachers set goals to improve practice	22.7	36.4	27.3	13.6
Provide opportunities for teachers to observe others	27.3	36.4	a.	22.7
Reviewed teacher portfolios	38.1	33.3	23.8	a.
Led study groups on teaching	42.9	a.	a.	38.1
Unweighted Sample Size (Mentors)	**22**			

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

Note: a. Values suppressed to protect respondent confidentiality.

Table F.7. Teacher Reports on Professional Support and Duties (Fall/Winter)

	Treatment	Control	Difference	P-value
Services Offered				
Induction program provided by school or district	91.1	94.6	-3.5*	0.049
Primary purpose of program[a]				
General support/guidance	66.9	53.1	13.8*	0.000
Orientation to school/district	20.2	35.3	-15.1*	0.000
Standards-based teaching	10.8	9.4	1.4	0.530
Other	1.5	2.2	-0.7	0.532
BT has a mentor[b]	95.2	81.2	14.0*	0.000
BT has an assigned mentor	91.9	73.8	18.1*	0.000
Professional development activities offered in past 3 months	98.7	99.3	-0.5	0.417
Assistance Received During Past 3 Months				
BT was compensated for attendance at professional development activities	32.5	26.5	6.0	0.070
Reduced teaching schedule	6.6	7.5	-0.9	0.611
Common planning time with teachers at grade level	71.8	74.0	-2.2	0.469
Received teacher's aide for assistance	35.5	39.1	-3.6	0.275
Regular communication with administrators on teaching practice	66.3	70.3	-4.0	0.205
Duties Required During Past 3 Months				
Extracurricular assignments	38.2	42.9	-4.7	0.164
Administrative duties including lunchroom, hall or recess duties (but not staff meetings)	41.6	38.9	2.7	0.439
Moved between classrooms to teach	9.5	11.2	-1.8	0.353
Traveled to more than one school to teach	1.8	3.0	-1.3	0.154
Unweighted Sample Size (Teachers)	472	426		

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Difference in the distributions is statistically significant using a chi-squared test (p-value 0.000).

[b]BT = beginning teacher.

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

Table F.8. Impacts on Teacher-Reported Mentor Profiles (Fall/Winter)

	Treatment	Control	Difference	P-value
Number of Mentors				
Any Mentor (One or More)	95.2	81.2	14.0*	0.000
Multiple Mentors (More Than One)	30.8	18.2	12.6*	0.000
Number of Mentors[a]				
None	4.8	18.9	-14.0*	0.000
One	64.3	62.9	1.4	0.706
Two	25.7	12.5	13.2*	0.000
Three	3.0	1.4	1.6	0.157
Four	1.4	2.7	-1.3	0.193
Five	0.7	1.6	-0.9	0.146
Mentor Assignment				
Any Mentor Assigned	91.9	73.8	18.1*	0.000
Number of Mentors Assigned[a]				
No mentor assigned	8.1	26.1	-18.1*	0.000
One mentor assigned	67.6	64.0	3.6	0.338
Two mentors assigned	24.3	9.9	14.4*	0.000
BT Reports Assigned Study Mentor[b]	85.9	n.a.	n.a.	n.a.
Mentor Positions				
Full-Time Mentor	73.2	11.2	62.1*	0.000
Teacher	30.7	62.9	-32.3*	0.000
Administrator, School, or District	8.3	7.2	1.1	0.578
Staff External to District	3.9	4.6	-0.7	0.621
No Mentor	4.9	19.0	-14.1*	0.000
Position of Mentor If Have Only One[a]				
Full-time mentor	77.6	10.0	67.6*	0.000
Teacher	16.8	79.0	-62.2*	0.000
Administrator	4.3	6.6	-2.2	0.310
Staff external to district	1.3	4.5	-3.1*	0.015
Combination of Mentor Positions If Have Two[a]				
Teacher and full-time mentor	50.8	18.4	32.5*	0.000
Both teachers	9.9	45.9	-36.0*	0.000
Teacher and administrator	4.4	14.1	-9.7	0.055
Teacher and staff external to district	1.8	4.2	-2.4	0.478
Full-time mentor and administrator	11.6	0.0	11.6*	0.001
Full-time mentor and staff external to district	3.6	4.6	-0.9	0.749
Other combination	17.9	12.9	4.9	0.329
Unweighted Sample Size (Teachers)	**472**	**426**		

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a] Difference in the distributions is statistically significant using a chi-squared test (p-value 0.000).

[b] BT = beginning teacher.

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

n.a. = not applicable

Table F.9. Impacts on Teacher-Reported Mentor Services Received in the Most Recent Full Week of Teaching (Fall/Winter)

Mentor Service	Treatment	Control	Difference	Effect Size[b]	P-value
Teacher Has a Usual Meeting Time with Mentor (%)					
During school hours	74.8	37.0	37.8*	--	0.000
Before or after school hours	41.2	31.7	9.5*	--	0.006
On weekends	0.5	0.2	0.2	--	0.600
Varies	1.5	4.3	-2.7*	--	0.019
Any usual meeting time	84.1	54.5	29.6*	--	0.000
"Usual" Meetings with Mentors					
Frequency (number of meetings)	1.5	1.3	0.2	0.13	0.063
Average duration (minutes)	23.8	10.6	13.2*	0.73	0.000
Total time (minutes)	67.5	37.8	29.8*	0.40	0.000
Informal Meetings with Mentors					
Total time (Minutes)	37.1	35.3	1.8	0.04	0.553
Total Usual & Informal Time with Mentors (Minutes)	104.3	72.6	31.7*	0.32	0.000
Teacher Has Usual Meetings and Feels There is Adequate Time to Meet with a Mentor (Percent)	84.6	83.7	0.9	--	0.749
Meeting Time with Mentors in the Following Positions (Minutes)					
Study mentor	66.8	n.a.	n.a.	n.a.	n.a.
Non-study mentor	37.5	72.6	-35.1*	-0.40	0.000
Meeting Time with Mentors in the Following Positions (Minutes)					
Full-time mentor	67.3	5.2	62.2*	0.91	0.000
Teacher	30.7	64.0	-33.3*	-0.40	0.000
Administrator	5.1	2.2	2.9*	0.16	0.027
Staff external to district	3.2	1.7	1.5	0.06	0.437
Mentor Time in the Following Activities (Minutes)					
Observing beginning teacher (BT) teaching	35.2	13.3	21.9*	0.65	0.000
Meeting with BT one-on-one[a]	38.0	22.9	15.1*	0.47	0.000
Meeting with BT and other first year teachers	33.0	10.2	22.8*	0.59	0.000
Meeting with BT and other teachers	21.2	15.6	5.6*	0.16	0.028
Modeling a lesson	12.3	7.5	4.8*	0.21	0.001
Co-teaching a lesson	9.1	6.4	2.7	0.11	0.084
All six activities (all mentors)	148.7	75.9	72.8*	0.59	0.000
All six activities (study mentor only)	72.0	n.a.	n.a.	n.a.	n.a.
Types of Assistance a Mentor Provided (%)					
Suggestions to improve practice	79.1	57.2	21.9*	--	0.000
Encouragement or moral support	89.3	68.8	20.5*	--	0.000
Opportunity to raise issues/discuss concerns	87.6	66.6	21.0*	--	0.000
Help with administrative/ logical issues	69.9	56.0	13.9*	--	0.000
Help teaching to meet state or district standards	64.1	47.1	17.0*	--	0.000
Help identifying teaching challenges/solutions	82.3	56.0	26.3*	--	0.000
Discussed instructional goals and ways to achieve them	73.8	48.2	25.6*	--	0.000
Guidance on how to assess students	61.3	45.7	15.6*	--	0.000
Shared lesson plans, assignments, or other instructional activities	62.5	50.8	11.8*	--	0.001
Acted on something BT requested[c]	74.4	50.4	24.0*	--	0.000
Unweighted Sample Size (Teachers)	**472**	**426**			

Table F.9 *(continued)*

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]BT = beginning teacher.

n.a. = not applicable.

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

[c]Total sample size is 711. 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.

Table F.10. Impacts on Teacher-Reported Areas of Mentor Support During Past 3 Months (Fall/Winter) (Percentages)

Area		"Moderate Amount" or "A Lot" of Guidance			
		Treatment	Control	Difference	P-value
Areas of Guidance					
1.	Reflecting on your instructional practices	68.0	36.5	31.5*	0.000
2.	Managing student discipline and behavior	67.1	49.8	17.3*	0.000
3.	Managing classroom activities, transitions, and routines	66.2	49.0	17.2*	0.000
4.	Using multiple instructional strategies/techniques to teach students	55.9	43.4	12.6*	0.000
5.	Understanding this school's culture, policies, and practices	55.0	50.8	4.3	0.223
6.	Motivating students	54.0	39.9	14.1*	0.000
7.	Teaching children with varying levels of achievement/ability	52.9	42.4	10.5*	0.002
8.	Teaching reading/language arts	52.6	43.3	9.3*	0.004
9.	Completing paperwork	52.4	44.2	8.2*	0.013
10.	Understanding/teaching toward state or district standards	49.5	41.5	8.0*	0.018
11.	Selecting or adapting curriculum materials	48.1	38.5	9.6*	0.003
12.	Accessing district and community resources	46.1	33.2	12.9*	0.000
13.	Planning lessons	44.3	36.5	7.9*	0.017
14.	Reviewing and assessing student work	42.0	34.5	7.4*	0.024
15.	Using student assessments to inform your teaching	41.3	30.9	10.4*	0.001
16.	Teaching students of varying ethnic/racial and socioeconomic backgrounds	40.4	34.1	6.3	0.067
17.	Communicating with parents	39.0	37.0	2.0	0.545
18.	Teaching students with special needs	38.7	25.1	13.6*	0.000
19.	Teaching mathematics	38.4	34.7	3.7	0.275
20.	Working with other school staff, such as principal, counselors, disability specialist	38.0	34.5	3.5	0.275
21.	Working with other teachers to plan instruction	36.4	33.4	3.1	0.328
22.	Teaching English language learners	28.9	23.2	5.7	0.089
Unweighted Sample Size (Teachers)		**472**	**426**		

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the 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 F.11. Impacts on Teacher-Reported Professional Development During the Past 3 Months (Fall/Winter)

	Treatment	Control	Difference	Effect Size[b]	P-value
When Professional Development Activities Took Place[a] (Percentages)					
Before or after school	51.9	48.6	3.3	--	0.302
During "regular" teaching hours	27.8	28.1	-0.3	--	0.914
In the evening or Saturday	13.1	12.1	1.0	--	0.664
During summer or PD days	6.5	10.0	-3.5	--	0.067
Other	0.6	1.2	-0.6	--	0.395
Did not attend any PD activities	0.2	0.0	0.2	--	0.318
Activities Completed (Percentages)					
Kept a written log	39.8	32.9	6.8*	--	0.038
Kept a portfolio and analysis of student work	76.3	78.0	-1.7	--	0.543
Worked with a study group of new teachers	66.2	29.8	36.3*	--	0.000
Worked with a study group of new and experienced teachers	48.4	42.0	6.4	--	0.051
Observed others teaching in their classrooms	59.9	46.2	13.7*	--	0.000
Observed others teaching your class	49.0	49.0	0.0	--	0.993
Met with principal to discuss teaching	71.3	71.8	-0.5	--	0.877
Met with a literacy or mathematics coach or other curricular specialist	72.9	76.9	-4.0	--	0.194
Met with a resource specialist to discuss needs of particular students	66.6	70.1	-3.4	--	0.276
Frequency of Selected Activities (Number of Times During Past 3 Months)					
Teaching was observed by mentor	3.7	1.7	2.0*	0.80	0.000
Teaching was observed by principal	2.2	2.5	-0.3*	-0.16	0.032
Given feedback on your teaching, not as part of formal evaluation	3.0	2.4	0.6*	0.26	0.000
Given feedback on your teaching, as part of formal evaluation	1.7	1.4	0.2*	0.15	0.035
Given feedback on your lesson plans	1.8	1.8	-0.1	-0.03	0.691
Professional Development Activities Were "Very Useful" (Percent)	27.8	23.8	4.0	--	0.209
Unweighted Sample Size (Teachers)	**472**	**426**			

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

PD = professional development.

n.a. = not applicable.

[a]Difference in the distributions is not statistically significant using a chi-squared test (p= 0.454).

[b]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.

Table F.12 Impacts on Teacher-Reported Areas of Professional Development (PD) During the Last 3 Months (Fall/Winter)

Area of PD	PD Was Offered (Percentages)				Attended PD (Percentages)			
	Treatment	Control	Difference	P-value	Treatment	Control	Difference	P-value
Areas Offered								
1. Human resource policies/procedures	42.8	52.0	-9.3*	0.006	36.5	42.8	-6.4	0.053
2. Parent and community relations	43.8	41.6	2.2	0.500	35.9	29.6	6.3*	0.050
3. School policies on student disciplinary procedures	49.8	59.3	-9.6*	0.003	45.2	53.0	-7.8*	0.020
4. Instructional techniques/strategies	82.4	87.5	-5.1*	0.045	76.3	80.8	-4.5	0.128
5. Understanding the composition of students in your class	32.7	31.8	0.9	0.787	28.0	24.7	3.3	0.276
6. Content area knowledge (language arts, mathematics, science)	70.2	80.4	-10.2*	0.000	62.2	71.9	-9.8*	0.001
7. Lesson planning	37.9	39.9	-2.0	0.546	33.5	34.3	-0.8	0.812
8. Analyzing student work/assessment	50.0	52.6	-2.6	0.441	45.0	46.8	-1.9	0.585
9. Student motivation/engagement	47.3	43.9	3.3	0.333	41.8	37.0	4.8	0.164
10. Differentiated instruction	61.9	55.4	6.5	0.071	54.0	48.0	6.0	0.105
11. Using computers to support instruction	40.9	50.1	-9.2*	0.007	30.4	35.4	-5.0	0.114
12. Classroom management techniques	65.9	62.7	3.2	0.352	56.7	51.5	5.2	0.138
13. Accessing school, district, or community resources	20.3	28.6	-8.4*	0.005	17.1	22.8	-5.6*	0.037
14. Administrative paperwork	17.8	25.6	-7.8*	0.008	15.4	22.9	-7.5*	0.006
15. Handling non-classroom duties and responsibilities (e.g., supervision of lunch room, back to school night)	15.0	18.5	-3.5	0.165	13.9	17.1	-3.1	0.199
16. Assigning grades/record keeping	35.6	34.3	1.3	0.707	31.5	30.5	0.9	0.773
17. Preparing students for standardized testing	36.0	47.7	-11.8*	0.001	29.9	38.6	-8.8*	0.008
Unweighted Sample Size (Teachers)	**472**	**426**						

Table F.12 *(continued)*

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the 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 F.13 Impacts on Time Spent in Teacher-Reported Areas of Professional Development (PD) During the Last 3 Months (Fall/Winter)

Area of PD	Time Spent (Minutes)				
	Treatment	Control	Difference	Effect Size	P-value
Areas Offered					
1. Human resource policies/procedures	35.2	41.7	-6.5	-0.11	0.105
2. Parent and community relations	33.1	19.9	13.3*	0.27	0.000
3. School policies on student disciplinary procedures	31.9	39.3	-7.4*	-0.14	0.046
4. Instructional techniques/strategies	97.1	102.4	-5.3	-0.07	0.283
5. Understanding the composition of students in your class	25.9	22.0	4.0	0.08	0.254
6. Content area knowledge (language arts, mathematics, science)	83.0	92.7	-9.6	-0.12	0.061
7. Lesson planning	32.3	32.5	-0.2	0.00	0.958
8. Analyzing student work/assessment	47.2	47.3	-0.1	0.00	0.981
9. Student motivation/engagement	43.0	34.7	8.3*	0.14	0.049
10. Differentiated instruction	60.3	52.1	8.2	0.12	0.109
11. Using computers to support instruction	24.0	30.4	-6.3	-0.13	0.064
12. Classroom management techniques	63.5	58.0	5.5	0.08	0.256
13. Accessing school, district, or community resources	12.8	14.3	-1.5	-0.04	0.516
14. Administrative paperwork	10.5	14.7	-4.2	-0.13	0.070
15. Handling non-classroom duties and responsibilities (e.g., supervision of lunch room, back to school night)	8.1	9.9	-1.8	-0.06	0.367
16. Assigning grades/record keeping	19.1	19.9	-0.8	-0.02	0.781
17. Preparing students for standardized testing	28.0	39.3	-11.4*	-0.19	0.005
Unweighted Sample Size (Teachers)	**472**	**426**			

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

Note: Data pertain to teachers in all districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the 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 F.14. Differences in Impacts on Teacher-Reported Areas of Mentor Support During Past 3 Months (Percentages)

Area		"Moderate Amount" or "A Lot" of Guidance			
		Spring	Fall/Winter	Difference in Impacts	P-value
Areas of Guidance					
1.	Reflecting on your instructional practices	37.6	32.3	5.3	0.184
2.	Managing classroom activities, transitions, and routines	25.4	19.1	6.3	0.121
3.	Managing student discipline and behavior	20.5	17.7	2.8	0.532
4.	Using multiple instructional strategies/techniques to teach students	24.0	13.5	10.5*	0.015
5.	Teaching children with varying levels of achievement/ability	22.6	9.7	12.9*	0.001
6.	Motivating students	21.5	14.9	6.6	0.106
7.	Understanding/teaching toward state or district standards	23.4	8.7	14.7*	0.001
8.	Teaching reading/language arts	22.0	9.1	12.9*	0.001
9.	Reviewing and assessing student work	27.5	7.5	20.0*	0.000
10.	Understanding this school's culture, policies, and practices	10.2	5.6	4.6	0.219
11.	Selecting or adapting curriculum materials	21.2	10.1	11.1*	0.007
12.	Using student assessments to inform your teaching	25.3	11.4	13.9*	0.000
13.	Planning lessons	20.0	8.1	11.9*	0.001
14.	Completing paperwork	14.0	8.7	5.3	0.175
15.	Accessing district and community resources	20.1	14.5	5.6	0.220
16.	Teaching students of varying ethnic/racial and socioeconomic backgrounds	18.8	7.9	10.9*	0.008
17.	Teaching mathematics	14.4	4.4	10.0*	0.010
18.	Teaching students with special needs	18.1	14.2	3.9	0.586
19.	Working with other teachers to plan instruction	7.4	5.1	2.3	0.548
20.	Working with other school staff, such as principal, counselors, disability specialist	7.1	4.5	2.6	0.649
21.	Communicating with parents	7.1	1.7	5.4	0.132
22.	Teaching English language learners	12.1	7.0	5.1	0.197
Unweighted Sample Size (Teachers)		**846**	**846**		
Unweighted Sample Size (Treatment Teachers)		**453**	**453**		
Unweighted Sample Size (Control Teachers)		**393**	**393**		

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

Note: Data pertain to teachers who responded to both surveys in all districts participating in the study. Sample sizes vary due to item nonresponse.

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

Table F.15. Differences in Impacts on Teacher-Reported Areas of Professional Development (PD) During the Last 3 Months

Area of PD	PD Was Offered (Percentages)				Attended PD (Percentages)				Time Spent (Minutes)			
	Spring	Fall/ Winter	Difference in Impacts	P-value	Spring	Fall/ Winter	Difference in Impacts	P-value	Spring	Fall/ Winter	Difference in Impacts	P-value
Areas Offered												
1. Human resource policies/procedures	-2.0	-9.3	7.3*	0.039	-1.7	-6.4	4.7	0.075	0.3	-6.5	6.8*	0.016
2. Parent and community relations	3.1	2.2	0.9	0.469	3.2	6.3	-3.1	0.271	8.1	13.3	-5.2	0.150
3. School policies on student disciplinary procedures	-5.3	-9.6	4.3	0.232	-5.3	-7.8	2.5	0.504	-3.6	-7.4	3.8	0.239
4. Instructional techniques/strategies	-0.7	-5.1	4.4	0.089	1.6	-4.5	6.1	0.087	0.2	-5.3	5.5	0.292
5. Understanding the composition of students in your class	4.4	0.9	3.5	0.323	4.4	3.3	1.1	0.774	5.4	4.0	1.4	0.432
6. Content area knowledge (language arts, mathematics, science)	-3.2	-10.2	7.0	0.056	-0.6	-9.8	9.2*	0.025	-4.2	-9.6	5.4	0.366
7. Lesson planning	10.0	-2.0	12.0*	0.006	11.7	-0.8	12.5*	0.008	9.2	-0.2	9.4	0.117
8. Analyzing student work/assessment	12.8	-2.6	15.4*	0.001	14.5	-1.9	16.4*	0.000	17.3	-0.1	17.4*	0.003
9. Student motivation/engagement	2.0	3.3	-1.3	0.958	4.8	4.8	0.0	0.999	3.8	8.3	-4.5	0.379
10. Differentiated instruction	11.2	6.5	4.7	0.542	9.4	6.0	3.4	0.511	8.9	8.2	0.7	0.845
11. Using computers to support instruction	-4.7	-9.2	4.5	0.727	-3.0	-5.0	2.0	0.711	-5.5	-6.3	0.8	0.798
12. Classroom management techniques	2.5	3.2	-0.7	0.939	5.8	5.2	0.6	0.786	7.2	5.5	1.7	0.636
13. Accessing school, district, or community resources	-0.8	-8.4	7.6*	0.039	1.9	-5.6	7.5*	0.040	1.9	-1.5	3.4	0.284
14. Administrative paperwork	-2.9	-7.8	4.9	0.090	-1.8	-7.5	5.7*	0.042	1.1	-4.2	5.3*	0.009

Area of PD	PD Was Offered (Percentages)				Attended PD (Percentages)				Time Spent (Minutes)			
	Spring	Fall/ Winter	Difference in Impacts	P-value	Spring	Fall/ Winter	Difference in Impacts	P-value	Spring	Fall/ Winter	Difference in Impacts	P-value
Areas Offered												
15. Handling non-classroom duties and responsibilities (e.g., supervision of lunch room, back to school night)	0.2	-3.5	3.7	0.456	1.5	-3.1	4.6	0.243	1.8	-1.8	3.6	0.210
16. Assigning grades/record keeping	4.5	1.3	3.2	0.316	3.3	0.9	2.4	0.449	7.3	-0.8	8.1*	0.009
17. Preparing students for standardized testing	-8.2	-11.8	3.6	0.233	-5.4	-8.8	3.4	0.341	-9.9	-11.4	1.5	0.881
Unweighted Sample Size (Teachers)	**885**	**898**			**885**	**898**			**885**	**898**		
Unweighted Sample Size (Treatment Teachers)	**468**	**472**			**468**	**472**			**468**	**472**		
Unweighted Sample Size (Control Teachers)	**417**	**426**			**417**	**426**			**417**	**426**		

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

Note: Data pertain to teachers in all districts participating in the study. Sample sizes vary due to item nonresponse. P-values are computed from tests comparing changes in outcomes over time between treatment and control teachers, using the sample of teachers who responded to both surveys. This is equivalent to testing for differences between spring and fall impacts for this sample.

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

Table F.16. Teacher Reports on Professional Support and Duties (ETS)

	Treatment	Control	Difference	P-value
Services Offered				
Induction program provided by school or district	89.5	90.6	-1.1	0.685
Primary purpose of program[a]				
General support/guidance	66.3	52.4	13.9*	0.005
Orientation to school/district	21.8	43.5	-21.7*	0.000
Standards-based teaching	8.1	4.1	4.0	0.135
Other	3.4	0.0	3.4*	0.006
BT has a mentor[b]	91.8	80.4	11.5*	0.000
BT has an assigned mentor	90.6	72.9	17.7*	0.000
Professional development activities offered in past 3 months	97.6	99.2	-1.6	0.152
Assistance Received During Past 3 Months				
Assistance Received During Past 3 Months				
BT was compensated for attendance at professional development activities	31.0	14.0	17.0*	0.000
Reduced teaching schedule	5.2	2.8	2.4	0.165
Common planning time with teachers at grade level	72.7	75.1	-2.4	0.600
Received teacher's aide for assistance	33.2	38.1	-5.0	0.329
Regular communication with administrators on teaching practice	56.5	65.2	-8.7	0.058
Duties Required During Past 3 Months				
Extracurricular assignments	39.6	39.7	-0.1	0.985
Administrative duties including lunchroom, hall or recess duties (but not staff meetings)	46.7	42.7	4.0	0.448
Moved between classrooms to teach	8.8	10.6	-1.8	0.465
Traveled to multiple schools to teach	1.5	1.4	0.1	0.880
Unweighted Sample Size (Teachers)	**239**	**226**		

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

Note: Data pertain to teachers in ETS districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Difference in the distributions is statistically significant using a chi-squared test (p= 0.000).

[b]BT = beginning teacher.

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

Table F.17. Impacts on Teacher-Reported Mentor Profiles (ETS)

	Treatment	Control	Difference	P-value
Number of Mentors				
Any Mentor (One or More)	91.8	80.4	11.5*	0.000
Multiple Mentors (More Than One)	33.7	10.6	23.1*	0.000
Number of Mentors[a]				
None	8.2	19.6	-11.5*	0.000
One	58.4	69.8	-11.4*	0.023
Two	28.3	8.3	20.0*	0.000
Three	3.7	0.9	2.8	0.113
Four	0.4	0.4	0.0	0.939
Five	1.0	1.0	0.0	0.983
Mentor Assignment				
Any Mentor Assigned	90.6	72.9	17.7*	0.000
Number of Mentors Assigned[a]				
No mentor assigned	9.4	27.1	-17.7*	0.000
One mentor assigned	62.2	66.7	-4.5	0.373
Two mentors assigned	28.4	6.2	22.1*	0.000
BT Reports Assigned Study Mentor[b]	87.8	n.a.	n.a.	n.a.
Mentor Positions				
Full-Time Mentor	70.0	9.5	60.5*	0.000
Teacher	35.2	64.8	-29.6*	0.000
Administrator, School, or District	7.5	7.2	0.3	0.918
Staff External to District	4.7	2.5	2.1	0.243
No Mentor	8.4	19.8	-11.4*	0.000
Position of Mentor If Have Only One[a]				
Full-time mentor	68.7	10.5	58.1*	0.000
Teacher	23.7	80.8	-57.1*	0.000
Administrator	3.9	6.3	-2.5	0.438
Staff external to district	3.8	2.4	1.4	0.451
Combination of Mentor Positions If Have Two[a]				
Teacher and full-time mentor	65.8	23.9	41.9*	0.042
Both teachers	5.1	38.0	-32.8*	0.016
Teacher and administrator	-1.60	30.5	-30.5*	0.013
Teacher and staff external to district	0.4	9.1	-8.6	0.172
Full-time mentor and administrator	13.0	0.0	13.0	0.136
Full-time mentor and staff external to district	7.6	0.0	7.6	0.203
Other combination	9.7	0.0	9.7	0.120
Unweighted Sample Size (Teachers)	**239**	**226**		

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

Note: Data pertain to teachers in ETS districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Difference in the distributions is statistically significant using a chi-squared test (p= 0.000).

[b]BT = beginning teacher.

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

Table F.18. Impacts on Teacher-Reported Mentor Services Received in the Most Recent Full Week of Teaching (ETS)

Mentor Service	Treatment	Control	Difference	Effect Size[b]	P-value
Teacher Has a Usual Meeting Time with a Mentor (%)					
During school hours	74.4	36.8	37.6*	--	0.000
Before or after school hours	42.0	31.7	10.3*	--	0.043
On weekends	0.7	0.0	0.7	--	0.311
Varies	1.7	2.3	-0.6	--	0.690
Any usual meeting time	83.1	51.5	31.7*	--	0.000
"Usual" Meetings with Mentors					
Frequency (number of meetings)	1.5	1.1	0.4*	0.23	0.016
Average duration (minutes)	19.6	9.1	10.5*	0.71	0.000
Total time (minutes)	53.1	29.5	23.5*	0.41	0.000
Informal Meetings with Mentors					
Total time (Minutes)	36.2	33.8	2.3	0.06	0.518
Total Usual and Informal Time with Mentors (Minutes)	89.2	63.3	25.9*	0.30	0.002
Teacher Has Usual Meetings and Feels There is Adequate Time to Meet with a Mentor (Percent)	87.4	82.6	4.8	--	0.265
Meeting Time with Mentors in the Following Positions (Minutes)					
Study mentor	48.9	n.a.	n.a.	n.a.	n.a.
Non-study mentor	40.3	63.4	-23.1*	-0.30	0.003
Meeting Time with Mentors in the Following Positions (Minutes)					
Full-time mentor	45.6	4.7	40.9*	0.83	0.000
Teacher	38.8	55.8	-17.0*	-0.21	0.017
Administrator	3.0	2.8	0.2	0.01	0.897
Staff external to district	1.1	0.9	0.2	0.02	0.760
Mentor Time in the Following Activities (Minutes)					
Observing beginning teacher (BT) teaching	22.6	8.1	14.5*	0.51	0.000
Meeting with BT one-on-one[a]	29.2	20.7	8.5*	0.29	0.001
Meeting with BT and other first year teachers	24.8	6.0	18.9*	0.59	0.000
Meeting with BT and other teachers	16.6	13.6	2.9	0.10	0.297
Modeling a lesson	7.8	5.1	2.7	0.15	0.132
Co-teaching a lesson	5.1	3.3	1.8	0.12	0.210
All six activities (all mentors)	106.1	56.7	49.3*	0.51	0.000
All six activities (study mentor only)	79.1	n.a.	n.a.	n.a.	n.a.
Types of Assistance a Mentor Provided (%)					
Suggestions to improve practice	66.3	49.7	16.6*	--	0.000
Encouragement or moral support	81.4	66.4	15.0*	--	0.000
Opportunity to raise issues/ discuss concerns	80.6	61.5	19.0*	--	0.000
Help with administrative/ logical issues	63.3	45.9	17.4*	--	0.000
Help teaching to meet state or district standards	55.0	40.4	14.6*	--	0.001
Help identifying teaching challenges and solutions	64.9	48.9	16.0*	--	0.001
Discussed instructional goals and ways to achieve them	61.0	39.9	21.2*	--	0.000
Guidance on how to assess students	56.9	35.2	21.7*	--	0.000
Shared lesson plans, assignments, or other instructional activities	57.8	44.8	13.0*	--	0.004
Acted on something BT requested[b]	61.2	47.5	13.7*	--	0.008
Unweighted Sample Size (Teachers)	**239**	**226**			

Table F.18 *(continued)*

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

Note: Data pertain to teachers in ETS districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

n.a. = not applicable.

[a]BT = beginning teacher.

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

[c]Total sample size is 355. The question did not apply to teachers who did not make a request to their mentors.

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

Table F.19. Impacts on Teacher-Reported Areas of Mentor Support During Past 3 Months (ETS) (Percentages)

Area		"Moderate Amount" or "A Lot" of Guidance			
		Treatment	Control	Difference	P-value
Areas of Guidance					
1.	Reflecting on your instructional practices	63.8	29.2	34.7*	0.000
2.	Managing classroom activities, transitions, and routines	56.0	36.3	19.6*	0.000
3.	Managing student discipline and behavior	56.5	40.1	16.4*	0.000
4.	Using multiple instructional strategies/techniques to teach students	51.6	30.0	21.7*	0.000
5.	Teaching children with varying levels of achievement/ability	48.9	31.3	17.7*	0.000
6.	Motivating students	49.4	32.1	17.3*	0.000
7.	Understanding/teaching toward state or district standards	50.6	29.9	20.7*	0.000
8.	Teaching reading/language arts	49.0	30.4	8.5*	0.000
9.	Reviewing and assessing student work	51.3	26.5	24.7*	0.000
10.	Understanding this school's culture, policies, and practices	53.2	40.5	2.7*	0.006
11.	Selecting or adapting curriculum materials	45.1	30.0	5.1*	0.000
12.	Using student assessments to inform your teaching	50.0	25.3	24.7*	0.000
13.	Planning lessons	43.6	31.2	12.4*	0.008
14.	Completing paperwork	49.8	35.1	14.8*	0.001
15.	Accessing district and community resources	43.9	29.4	14.5*	0.002
16.	Teaching students of varying ethnic/racial and socioeconomic backgrounds	39.7	27.6	12.0*	0.012
17.	Teaching mathematics	43.9	28.2	5.7*	0.000
18.	Teaching students with special needs	36.4	21.8	14.6*	0.003
19.	Working with other teachers to plan instruction	41.4	29.9	11.6*	0.009
20.	Working with other school staff, such as principal, counselors, disability specialist	37.7	27.3	0.4*	0.016
21.	Communicating with parents	38.7	29.6	9.1*	0.036
22.	Teaching English language learners	24.1	18.6	5.5	0.220
Unweighted Sample Size (Teachers)		**239**	**226**		

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

Note: Data pertain to teachers in ETS districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the 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 F.20. Impacts on Teacher-Reported Professional Development During the Past 3 Months (ETS)

	Treatment	Control	Difference	Effect Size[b]	P-value
When Professional Development Took Place[a] (Percentages)					
Before or after school	54.0	52.6	1.4	--	0.765
During "regular" teaching hours	31.9	35.6	-3.7	--	0.380
In the evening or Saturday	5.4	7.2	-1.8	--	0.477
During summer or PD days	7.4	4.6	2.8	--	0.276
Other	0.9	0.0	0.9	--	0.143
Did not attend any PD activities	0.3	0.0	0.3	--	0.302
Activities Completed (Percentages)					
Kept a written log	35.9	25.7	10.2*	--	0.015
Kept a portfolio and analysis of student work	78.1	70.2	7.9	--	0.075
Worked with a study group of new teachers	79.0	27.2	51.8*	--	0.000
Worked with a study group of new and experienced teachers	48.0	38.3	9.7*	--	0.048
Observed others teaching in their classrooms	55.0	43.4	11.7*	--	0.017
Observed others teaching your class	44.0	38.3	5.7	--	0.227
Met with principal to discuss teaching	70.3	74.3	-4.0	--	0.364
Met with a literacy or mathematics coach or other curricular specialist	71.5	66.3	5.3	--	0.264
Met with a resource specialist to discuss needs of particular students	61.7	64.9	-3.2	--	0.497
Frequency of Selected Activities (Number of times during past 3 months)					
Teaching was observed by mentor	3.1	1.3	1.9*	0.82	0.000
Teaching was observed by principal	2.2	2.2	0.0	0.00	0.992
Given feedback on your teaching, not as part of formal evaluation	2.4	2.0	0.4	0.18	0.073
Given feedback on your teaching, as part of formal evaluation	1.7	1.6	0.1	0.05	0.589
Given feedback on your lesson plans	1.5	1.6	-0.2	-0.08	0.384
Professional Development Activities Were "Very Useful" (Percent)	15.5	23.5	-8.1*	--	0.038
Unweighted Sample Size (Teachers)	**239**	**226**			

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

Note: Data pertain to teachers in ETS districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

PD = professional development.

n.a. = not applicable.

[a]Difference in the distributions is not statistically significant using a chi-squared test (p= 0.451).

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

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

Table F.21 Impacts on Teacher-Reported Areas of Professional Development During the Past 3 Months (ETS)

Area of Professional Development		Attended PD (Percentages)			
		Treatment	Control	Difference	P-value
Areas Offered					
1.	Human resource policies/procedures	19.7	21.3	-1.6	0.657
2.	Parent and community relations	30.2	25.3	5.0	0.268
3.	School policies on student disciplinary procedures	36.8	39.0	-2.2	0.634
4.	Instructional techniques/strategies	71.4	74.3	-2.8	0.497
5.	Understanding the composition of students in your class	26.3	23.1	3.2	0.441
6.	Content area knowledge (language arts, mathematics, science)	60.6	64.3	-3.8	0.401
7.	Lesson planning	31.1	24.7	6.4	0.149
8.	Analyzing student work/assessment	59.0	40.1	18.8*	0.000
9.	Student motivation/engagement	35.8	37.6	-1.9	0.665
10.	Differentiated instruction	56.5	42.9	13.6*	0.005
11.	Using computers to support instruction	27.2	30.3	-3.1	0.469
12.	Classroom management techniques	43.5	41.0	2.4	0.649
13.	Accessing school, district, or community resources	18.4	14.5	3.9	0.271
14.	Administrative paperwork	16.3	14.9	1.4	0.702
15.	Handling non-classroom duties and responsibilities (e.g., supervision of lunch room, back to school night)	10.2	10.4	-0.2	0.948
16.	Assigning grades/record keeping	22.9	18.1	4.8	0.220
17.	Preparing students for standardized testing	42.5	49.5	-7.0	0.096
Unweighted Sample Size (Teachers)		**239**	**226**		

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

Note: Data pertain to teachers in ETS districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

PD = professional development.

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

Table F.22 Impacts on Time Spent in Teacher-Reported Areas of Professional Development During the Past 3 Months (ETS)

Area of PD	Time Spent (Minutes)				
	Treatment	Control	Difference	Effect Size	P-value
Areas Offered					
1. Human resource policies/procedures	17.3	16.1	1.2	0.03	0.799
2. Parent and community relations	28.2	15.4	12.8*	0.28	0.002
3. School policies on student disciplinary procedures	28.8	27.9	0.9	0.02	0.846
4. Instructional techniques/strategies	83.2	86.8	-3.6	-0.05	0.629
5. Understanding the composition of students in your class	26.3	19.8	6.4	0.13	0.179
6. Content area knowledge (language arts, mathematics, science)	77.5	81.9	-4.4	-0.06	0.531
7. Lesson planning	28.8	26.1	2.6	0.05	0.626
8. Analyzing student work/assessment	64.2	39.2	25.0*	0.38	0.000
9. Student motivation/engagement	31.6	33.4	-1.8	-0.03	0.747
10. Differentiated instruction	58.8	40.3	18.5*	0.28	0.005
11. Using computers to support instruction	21.1	28.9	-7.7	-0.16	0.077
12. Classroom management techniques	45.6	38.9	6.7	0.11	0.321
13. Accessing school, district, or community resources	11.6	7.6	4.1	0.14	0.130
14. Administrative paperwork	11.8	8.3	3.5	0.12	0.209
15. Handling non-classroom duties and responsibilities (e.g., supervision of lunch room, back to school night)	7.1	5.6	1.5	0.07	0.473
16. Assigning grades/record keeping	19.5	8.8	10.7*	0.29	0.002
17. Preparing students for standardized testing	39.3	48.9	-9.6	-0.15	0.068
Unweighted Sample Size (Teachers)	**239**	**226**			

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

Note: Data pertain to teachers in ETS districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

PD = professional development.

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

Table F.23. Teacher Reports on Professional Support and Duties (NTC)

	Treatment	Control	Difference	P-value
Services Offered				
Induction program provided by school or district	94.7	91.1	3.6	0.170
Primary purpose of program[a]				
General support/guidance	73.4	46.8	26.7*	0.000
Orientation to school/district	16.2	36.2	-20.0*	0.000
Standards-based teaching	10.1	14.5	-4.4	0.233
Other	0.3	1.8	-1.5	0.238
BT has a mentor[b]	95.2	85.3	9.9*	0.002
BT has an assigned mentor	93.5	77.8	15.7*	0.000
Professional development activities offered in past 3 months	98.0	98.8	-0.8	0.552
Assistance Received During Past 3 Months				
BT was compensated for attendance at professional development activities	18.4	17.4	1.0	0.801
Reduced teaching schedule	9.4	10.4	-1.0	0.734
Common planning time with teachers at grade level	75.6	72.7	2.9	0.534
Received teacher's aide for assistance	36.4	32.7	3.6	0.426
Regular communication with administrators on teaching practice	58.3	60.7	-2.4	0.626
Duties Required During Past 3 Months				
Extracurricular assignments	44.0	44.6	-0.5	0.915
Administrative duties including lunchroom, hall or recess duties (but not staff meetings)	42.8	44.9	-2.1	0.724
Moved between classrooms to teach	11.9	15.3	-3.4	0.289
Traveled to more than one school to teach	2.8	4.3	-1.5	0.379
Unweighted Sample Size (Teachers)	**229**	**191**		

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

Note: Data pertain to teachers in NTC districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Difference in the distributions is statistically significant using a chi-squared test (p= 0.000).

[b]BT = beginning teacher.

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

Table F.24. Impacts on Teacher-Reported Mentor Profiles (NTC)

	Treatment	Control	Difference	P-value
Number of Mentors				
Any Mentor (One or More)	95.2	85.3	9.9*	0.002
Multiple Mentors (More Than One)	24.6	23.2	1.3	0.773
Number of Mentors[a]				
None	4.8	14.7	-9.9*	0.002
One	70.6	62.0	8.6	0.097
Two	22.3	19.5	2.8	0.526
Three	1.2	1.3	-0.1	0.926
Four	0.5	0.0	0.5	0.317
Five	0.6	2.5	-1.9	0.141
Mentor Assignment				
Any Mentor Assigned	93.5	77.8	15.7*	0.000
Number of Mentors Assigned[a]				
No mentor assigned	6.5	22.2	-15.7*	0.000
One mentor assigned	73.2	60.6	12.6*	0.018
Two mentors assigned	20.4	17.2	3.2	0.455
BT Reports Assigned Study Mentor[b]	91.3	n.a.	n.a.	n.a.
Mentor Positions				
Full-Time Mentor	78.8	17.3	61.5*	0.000
Teacher	24.1	66.9	-42.8*	0.000
Administrator, School, or District	8.7	6.2	2.5	0.396
Staff External to District	3.8	2.5	1.4	0.436
No Mentor	4.8	14.8	-10.0*	0.002
Position of Mentor If Have Only One[a]				
Full-time mentor	82.3	20.2	62.1*	0.000
Teacher	8.6	74.5	-65.9*	0.000
Administrator	0.9	4.7	1.2	0.667
Staff external to district	0.1	0.6	-0.5	0.336
Combination of Mentor Positions If Have Two[a]				
Teacher and full-time mentor	63.8	16.5	47.3*	0.000
Both teachers	0	58.7	-58.7*	0.000
Teacher and administrator	11.6	10.2	1.4	0.879
Teacher and staff external to district	0.3	7.3	-7.0	0.119
Full-time mentor and administrator	11.9	2.8	9.1	0.126
Full-time mentor and staff external to district	3.3	0.0	3.3	0.196
Other combination	9.9	5.3	4.6	0.426
Unweighted Sample Size (Teachers)	**229**	**191**		

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

Note: Data pertain to teachers in NTC districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

[a]Difference in the distributions is statistically significant using a chi-squared test (p= 0.000).

[b]BT = beginning teacher.

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

Table F.25. Impacts on Teacher-Reported Mentor Services Received in the Most Recent Full Week of Teaching (NTC)

Mentor Service	Treatment	Control	Difference	Effect Size[b]	P-value
Teacher Has a Usual Meeting Time with a Mentor (%)					
During school hours	79.7	39.6	40.1*	--	0.000
Before or after school hours	33.2	30.0	3.3	--	0.415
On weekends	0.5	0.0	0.5	--	0.316
Varies	3.2	4.0	-0.8	--	0.655
Any usual meeting time	88.1	55.8	32.3*	--	0.000
"Usual" Meetings with Mentors					
Frequency (number of meetings)	1.3	1.3	-0.1	-0.05	0.634
Average duration (minutes)	28.0	13.4	14.6*	0.70	0.000
Total time (minutes)	68.6	47.7	20.9*	0.22	0.048
Informal Meetings with Mentors					
Total time (Minutes)	35.2	38.8	-3.5	-0.08	0.452
Total Usual & Informal Time with Mentors (Minutes)	103.8	86.1	17.7	0.15	0.183
Teacher Has Usual Meetings and Feels There is Adequate Time to Meet with a Mentor (Percent)	81.6	79.9	1.7	--	0.739
Meeting Time with Mentors in the Following Positions (Minutes)					
Study mentor	79.5	n.a.	n.a.	n.a.	n.a.
Non-study mentor	24.3	86.1	-61.8*	-0.57	0.000
Meeting Time with Mentors in the Following Positions (Minutes)					
Full-time mentor	77.6	10.9	66.7*	0.88	0.000
Teacher	20.2	72.3	-52.1*	-0.50	0.000
Administrator	4.5	2.2	2.2	0.14	0.207
Staff external to district	2.5	1.0	1.4	0.09	0.279
Mentor Time in the Following Activities (Minutes)					
Observing beginning teacher (BT) teaching	30.6	14.5	16.1*	0.49	0.000
Meeting with BT one-on-one[a]	38.7	21.3	17.4*	0.53	0.000
Meeting with BT and other first year teachers	30.1	8.1	22.0*	0.57	0.000
Meeting with BT and other teachers	16.5	16.1	0.3	0.01	0.918
Modeling a lesson	14.5	8.2	6.3*	0.25	0.010
Co-teaching a lesson	9.8	7.7	2.2	0.09	0.339
All six activities (all mentors)	140.7	75.6	65.1*	0.50	0.000
All six activities (study mentor only)	119.9	n.a.	n.a.	n.a.	n.a.
Types of Assistance a Mentor Provided (%)					
Suggestions to improve practice	82.3	55.2	27.1*	--	0.000
Encouragement or moral support	88.0	71.9	16.1*	--	0.000
Opportunity to raise issues/discuss concerns	84.1	66.6	17.6*	--	0.000
Help with administrative/ logical issues	71.3	58.5	12.8*	--	0.009
Help teaching to meet state or district standards	65.6	48.7	16.9*	--	0.001
Help identifying teaching challenges and solutions	79.1	55.9	23.2*	--	0.000
Discussed instructional goals/ways to achieve them	78.5	48.9	29.6*	--	0.000
Guidance on how to assess students	66.9	45.2	21.7*	--	0.000
Shared lesson plans, assignments, or other instructional activities	68.6	52.7	15.9*	--	0.002
Acted on something BT requested[c]	75.1	52.5	22.6*	--	0.000
Unweighted Sample Size (Teachers)	**229**	**191**			

Table F.25 *(continued)*

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

Note: Data pertain to teachers in NTC districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

n.a. = not applicable.

[a]BT = beginning teacher.

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

[c]Total sample size is 341. The question did not apply to teachers who did not make a request to their mentors.

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

Table F.26. Impacts on Teacher-Reported Areas of Mentor Support During Past 3 Months (NTC) (Percentages)

Area		"Moderate Amount" or "A Lot" of Guidance			
		Treatment	Control	Difference	P-value
Areas of Guidance					
1.	Reflecting on your instructional practices	73.4	36.5	36.9*	0.000
2.	Managing classroom activities, transitions, and routines	73.8	44.0	29.8*	0.000
3.	Managing student discipline and behavior	68.4	44.5	23.9*	0.000
4.	Using multiple instructional strategies/techniques to teach students	71.2	46.7	24.5*	0.000
5.	Teaching children with varying levels of achievement/ability	68.1	41.0	27.1*	0.000
6.	Motivating students	64.1	40.7	23.3*	0.000
7.	Understanding/teaching toward state or district standards	62.5	38.4	24.1*	0.000
8.	Teaching reading/language arts	63.2	38.5	24.7*	0.000
9.	Reviewing and assessing student work	60.1	32.8	27.3*	0.000
10.	Understanding this school's culture, policies, and practices	55.2	49.9	5.4	0.259
11.	Selecting or adapting curriculum materials	63.8	38.0	25.8*	0.000
12.	Using student assessments to inform your teaching	57.9	33.5	24.4*	0.000
13.	Planning lessons	62.6	34.8	27.7*	0.000
14.	Completing paperwork	51.5	39.9	11.5*	0.020
15.	Accessing district and community resources	53.3	28.8	24.5*	0.000
16.	Teaching students of varying ethnic/racial and socioeconomic backgrounds	53.5	32.8	20.6*	0.000
17.	Teaching mathematics	47.1	35.8	11.3*	0.024
18.	Teaching students with special needs	48.1	26.7	21.4*	0.000
19.	Working with other teachers to plan instruction	38.3	37.2	1.2	0.813
20.	Working with other school staff, such as principal, counselors, disability specialist	41.7	39.0	2.8	0.571
21.	Communicating with parents	37.2	31.7	5.4	0.224
22.	Teaching English language learners	38.3	22.6	15.7*	0.003
Unweighted Sample Size (Teachers)		**229**	**191**		

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

Note: Data pertain to teachers in NTC districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Unweighted Sample Sizes vary due to item nonresponse.

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

Table F.27. Impacts on Teacher-Reported Professional Development During the Past 3 Months (NTC)

	Treatment	Control	Difference	Effect Size	P-value
Timing of Professional Development Activities[a] (Percentages)					
Before or after school	59.3	51.3	8.0	0.16	0.091
During "regular" teaching hours	23.7	28.3	-4.5	-0.10	0.287
In the evening or Saturday	12.2	12.8	-0.6	-0.02	0.893
During summer or PD days	4.4	5.4	-1.1	-0.05	0.624
Other	-0.1	2.2	-2.3*	-0.23	0.021
Did not attend any PD activities	0.4	0.0	0.4	0.08	0.313
Activities Completed (Percentages)					
Kept a written log	43.3	30.4	12.9*	0.27	0.009
Kept a portfolio and analysis of student work	78.9	79.0	-0.1	0.00	0.981
Worked with a study group of new teachers	55.8	27.2	28.5*	0.57	0.000
Worked with a study group of new and experienced teachers	46.4	36.4	10.0*	0.20	0.031
Observed others teaching in their classrooms	87.5	41.0	46.5*	0.98	0.000
Observed others teaching your class	47.0	37.3	9.7	0.20	0.068
Met with principal to discuss teaching	64.1	63.8	0.4	0.01	0.948
Met with a literacy or mathematics coach or other curricular specialist	64.8	66.6	-1.8	-0.04	0.719
Met with a resource specialist to discuss needs of particular students	58.5	59.9	-1.4	-0.03	0.778
Frequency of Selected Activities (Number of times during past 3 months)					
Teaching was observed by mentor	3.7	1.9	1.9*	0.76	0.000
Teaching was observed by principal	2.0	1.8	0.2	0.09	0.394
Given feedback on your teaching, not as part of formal evaluation	2.7	1.9	0.8*	0.40	0.000
Given feedback on your teaching, as part of formal evaluation	1.7	1.3	0.4*	0.26	0.009
Given feedback on your lesson plans	1.7	1.5	0.2	0.08	0.447
Professional Development Activities Were "Very Useful" (Percent)	27.6	22.9	4.8	0.11	0.276
Unweighted Sample Size (Teachers)	**229**	**191**			

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

Note: Data pertain to teachers in NTC districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

PD = professional development.

[a]Difference in the distributions is not statistically significant using a chi-squared test (p= 0.222).

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

Table F.28 Impacts on Teacher-Reported Areas of Professional Development During the Past 3 Months (NTC)

Area of Professional Development		Attended PD (Percentages)			
		Treatment	Control	Difference	P-value
Areas Offered					
1.	Human resource policies/procedures	17.5	19.9	-2.3	0.522
2.	Parent and community relations	23.0	21.5	1.5	0.701
3.	School policies on student disciplinary procedures	32.5	41.7	-9.2	0.055
4.	Instructional techniques/strategies	78.4	72.3	6.1	0.175
5.	Understanding the composition of students in your class	25.1	19.4	5.7	0.198
6.	Content area knowledge (language arts, mathematics, science)	67.8	64.4	3.4	0.461
7.	Lesson planning	44.7	27.2	17.4*	0.000
8.	Analyzing student work/assessment	53.1	43.7	9.4	0.060
9.	Student motivation/engagement	35.3	23.4	11.8*	0.014
10.	Differentiated instruction	53.8	48.5	5.3	0.319
11.	Using computers to support instruction	33.0	36.1	-3.2	0.537
12.	Classroom management techniques	42.8	32.8	10.1*	0.048
13.	Accessing school, district, or community resources	20.5	20.6	-0.2	0.969
14.	Administrative paperwork	12.1	17.8	-5.7	0.147
15.	Handling non-classroom duties and responsibilities (e.g., supervision of lunch room, back to school night)	15.4	12.6	2.8	0.406
16.	Assigning grades/record keeping	22.3	21.3	1.0	0.811
17.	Preparing students for standardized testing	50.4	54.2	-3.8	0.360
Unweighted Sample Size (Teachers)		**229**	**191**		

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

Note: Data pertain to teachers in NTC districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

PD = professional development.

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

Table F.29 Impacts on Time Spent in Teacher-Reported Areas of Professional Development During the Past 3 Months (NTC)

Area of PD		Time Spent (Minutes)				
		Treatment	Control	Difference	Effect Size	P-value
Areas Offered						
1.	Human resource policies/procedures	13.1	14.9	-1.8	-0.05	0.621
2.	Parent and community relations	17.2	14.0	3.2	0.08	0.420
3.	School policies on student disciplinary procedures	19.0	28.0	-9.0*	-0.22	0.045
4.	Instructional techniques/strategies	88.6	84.3	4.3	0.06	0.536
5.	Understanding the composition of students in your class	21.3	17.3	4.0	0.09	0.398
6.	Content area knowledge (language arts, mathematics, science)	81.1	84.3	-3.2	-0.04	0.661
7.	Lesson planning	43.3	26.7	16.5*	0.28	0.001
8.	Analyzing student work/assessment	51.1	43.0	8.1	0.13	0.185
9.	Student motivation/engagement	31.5	22.0	9.4	0.18	0.073
10.	Differentiated instruction	48.9	50.4	-1.5	-0.02	0.828
11.	Using computers to support instruction	26.9	30.1	-3.2	-0.06	0.561
12.	Classroom management techniques	42.7	34.6	8.1	0.13	0.212
13.	Accessing school, district, or community resources	14.6	14.4	0.2	0.01	0.959
14.	Administrative paperwork	8.6	10.0	-1.4	-0.05	0.680
15.	Handling non-classroom duties and responsibilities (e.g., supervision of lunch room, back to school night)	8.2	6.4	1.8	0.07	0.450
16.	Assigning grades/record keeping	14.0	10.9	3.2	0.10	0.359
17.	Preparing students for standardized testing	46.6	57.8	-11.3	-0.17	0.058
Unweighted Sample Size (Teachers)		**229**	**191**			

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

Note: Data pertain to teachers in NTC districts participating in the study. Data are weighted and regression-adjusted using ordinary least squares to account for differences in districts, teacher grade assignments, the study design, and the clustering of teachers within schools. Sample sizes vary due to item nonresponse.

PD = professional development.

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

APPENDIX G

SUPPLEMENTAL TABLES FOR CHAPTER V

Table G.1. Teacher Preparedness Constructs: Factor Loadings

Variable	Factor Loading 1	2	3
Prepared to instruct (Cronbach's alpha = .89)			
Managing classroom activities, transitions, and routines	.677	.397	.045
Using a variety of instructional methods	.747	.182	.225
Assessing your students	.621	.211	.399
Selecting and adapting curriculum and instructional materials	.690	.154	.345
Planning effective lessons	.644	.148	.497
Being an effective teacher	.693	.340	.298
Addressing the needs of a diversity of learners	.621	.337	.292
Prepared to work with students (Cronbach's alpha = .77)			
Handling a range of classroom behavior or discipline situations	.573	.599	.001
Motivating students	.448	.604	.133
Working effectively with parents	.077	.725	.447
Working with students who have special behavioral, emotional, developmental, or physical challenges	.264	.691	.226
Prepared to work with other school staff (Cronbach's alpha = .82)			
Working with other teachers to plan instruction	.268	.166	.809
Working with the principal or other instructional leaders	.282	.287	.779

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

Notes: Data pertain to teachers in all districts participating in the study. The following items were not included in factor analyses or subscales: Teaching reading/language arts, Teaching mathematics, Working with English language learners. The extraction method was principal components analysis and the rotation method was varimax with Kaiser normalization.

Table G.2. Teacher Satisfaction Constructs: Factor Loadings

Variable	Factor Loading 1	2	3
Satisfaction with School (Cronbach's alpha = .88)			
Support from administration for beginning teachers	.757	.330	.043
Availability of resources and materials/equipment for your classroom	.576	.264	.153
Input into school policies and practices	.665	.296	.202
Opportunities for professional development	.473	.250	.338
Principals' leadership and vision	.765	.281	.015
Professional caliber of colleagues	.709	.046	.251
Supportive atmosphere among faculty/collaboration with colleagues	.728	.075	.191
School facilities such as the building or grounds	.557	.215	.141
School policies	.631	.449	.183
Satisfaction with Class (Cronbach's alpha = .80)			
Autonomy or control over own classroom	.397	.551	.038
Student motivation to learn	.194	.736	.194
Student discipline and behavior	.167	.795	.177
Parental involvement in the school	.210	.498	.336
Grade assignment	.239	.558	-.021
Students assigned	.156	.734	.143
Satisfaction with Teaching Career (Cronbach's alpha = .72)			
Salary and benefits	.035	.008	.851
Professional prestige	.425	.271	.623
Intellectual challenge	.414	.346	.460
Workload	.313	.386	.475

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

Notes: Data pertain to teachers in all districts participating in the study. The following item was not included in factor analyses or subscales: emphasis on standardized test scores. The extraction method was principal components analysis and the rotation method was varimax with Kaiser normalization.

Table G.3. Impacts on Teacher Feelings of Preparedness (Percentages Feeling "Well Prepared" or "Very Well Prepared")

Area of Preparedness	Treatment	Control	Difference	Effect Size	P-value
Prepared to Instruct					
Managing classroom activities, transitions, and routines	69.7	73.6	-4.0	-0.09	0.195
Using a variety of instructional methods	62.5	67.2	-4.7	-0.10	0.123
Assessing your students	57.9	68.0	-10.1*†	-0.21	0.001
Selecting and adapting instructional materials	53.6	61.0	-7.3*†	-0.15	0.018
Planning effective lessons	72.6	78.9	-6.3*†	-0.15	0.028
Being an effective teacher	69.3	76.7	-7.4*†	-0.17	0.011
Addressing the needs of a diversity of learners	58.9	67.3	-8.4*†	-0.17	0.012
Prepared to Work with Students					
Handling a range of classroom behavior or discipline situations	64.4	66.3	-1.9	-0.04	0.563
Motivating students	73.2	75.1	-1.9	-0.04	0.491
Working effectively with parents	61.6	62.2	-0.6	-0.01	0.838
Working with students with special challenges	38.1	41.5	-3.3	-0.07	0.303
Prepared to Work with Other School Staff					
Working with other teachers to plan instruction	72.3	75.5	-3.2	-0.07	0.272
Working with the principal or other instructional leaders	64.1	71.6	-7.5*†	-0.16	0.011
Unweighted Sample Size (Teachers)	**471**	**426**	**897**		

Source: MPR First Induction Activities Survey administered in fall/winter 2005-2006 to all study teachers

Note: Data pertain to teachers in all study districts. Sample sizes vary due to item nonresponse.

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

†Significantly different from zero at the 0.05 level, two-tailed test, after applying a Benjamini-Hochberg correction.

Table G.4. Impacts on Teacher Satisfaction (Percent "Somewhat Satisfied" or "Very Satisfied")

Area of Satisfaction	Treatment	Control	Difference	Effect Size	P-value
Satisfaction with School					
Administration support for beginning teachers	75.6	75.9	-0.3	-0.01	0.932
Availability of resources and materials/equipment for your classroom	67.3	68.0	-0.7	-0.01	0.844
Input into school policies and practices	67.6	71.6	-4.0	-0.09	0.199
Opportunities for professional development	85.5	83.8	1.7	0.05	0.504
Principals' leadership and vision	80.6	78.2	2.4	0.06	0.434
Professional caliber of colleagues	81.7	86.1	-4.5	-0.12	0.079
Supportive atmosphere among faculty/collaboration with colleagues	83.3	81.9	1.5	0.04	0.611
School facilities such as the building or grounds	76.6	75.0	1.6	0.04	0.609
School policies	81.2	79.7	1.5	0.04	0.576
Satisfaction with Class					
Autonomy or control over own classroom	86.5	86.7	-0.2	0.00	0.939
Student motivation to learn	75.2	72.8	2.4	0.05	0.427
Student discipline and behavior	66.8	62.3	4.5	0.09	0.161
Parental involvement in the school	46.2	46.2	0.0	0.00	0.994
Grade assignment	89.3	87.4	1.8	0.06	0.361
Students assigned	83.5	84.4	-0.9	-0.02	0.721
Satisfaction with Teaching Career					
Salary and benefits	76.3	78.1	-1.8	-0.04	0.491
Professional prestige	81.5	82.5	-1.0	-0.03	0.698
Intellectual challenge	87.9	90.0	-2.1	-0.07	0.345
Workload	55.6	59.9	-4.3	-0.09	0.179
Unweighted Sample Size (Teachers)	**471**	**426**	**897**		

Source: MPR First Induction Activities Survey administered in fall/winter 2005-2006 to all study teachers

Note: Data pertain to teachers in all study districts. Sample sizes vary due to item nonresponse.

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

Table G.5. VCOT Classroom Practices Constructs: Factor Loadings

	Factor Loading	
Variable	1	2
Literacy Implementation (Cronbach's alpha = .89)		
Best Practices	.808	.364
Institutional Choices	.719	.509
Student Choices	.805	.241
Pace	.595	.581
Literacy Content (Cronbach's alpha = .80)		
Understanding content and close reading	.756	.321
Assessment	.473	.275
Skill Development	.784	.332
Connections between reading and writing	.771	.138
Literacy Classroom Culture (Cronbach's alpha = .93)		
Maximizes learning opportunities	.315	.868
Routines clear and consistent	.256	.817
Respectful behavior, safe atmosphere	.278	.867
Literacy Valued	.644	.439
Teacher works collaboratively with students	.536	.652
Students work collaboratively with students	.458	.654
Equal access to teacher and resources	.285	.776

Source: MPR classroom observations conducted in spring 2006.

Note: The extraction method was principal components analysis and the rotation method was varimax with Kaiser normalization.

Table G.6. Impacts on Classroom Practices (Percentages with Consistent or Extensive Evidence of Practice)

Classroom Observation Item	Treatment	Control	Difference	P-value
Implementation of Literacy Lesson				
Best practices	23.4	27.2	-3.8	0.306
Institutional choices	28.8	30.7	-1.8	0.614
Student choices	18.2	18.4	-0.2	0.952
Pace	24.2	26.3	-2.1	0.559
Student-student interaction	16.8	15.5	1.3	0.682
Content of Literacy Lesson				
Understanding content and close reading	23.5	25.4	-1.9	0.593
Assessment	7.2	7.4	-0.2	0.935
Skill development	17.9	17.8	0.1	0.983
Connections between reading and writing	15.9	17.0	-1.1	0.737
Classroom Culture				
Maximizes learning opportunities	44.4	46.4	-2.0	0.619
Routines clear and consistent	46.1	49.4	-3.3	0.434
Behavior respectable, atmosphere safe	45.3	44.0	1.2	0.756
Literacy valued	28.1	31.1	-3.0	0.429
Teacher works collaboratively with students	39.5	37.2	2.2	0.594
Students work collaboratively with other students	25.0	23.8	1.2	0.735
Equal access to teacher and resources	41.3	46.0	-4.6	0.291
Unweighted Sample Size (Teachers)	**342**	**289**		

Source: MPR classroom observations conducted in spring 2006.

Notes: Data are weighted and regression-adjusted to account for differences in baseline characteristics and the study design.

None of the differences is statistically different from zero at the 0.05 level, two-tailed test.

Table G.7. Impacts on Classroom Practices (Observer Summary Scores)

Outcome	Treatment	Control	Difference	Effect Size	P-value
Implementation of literacy lesson	2.7	2.7	0.0	-0.01	0.942
Content of literacy lesson	2.5	2.5	0.0	-0.01	0.859
Classroom culture	3.1	3.0	0.0	0.02	0.804
Unweighted Sample Size (Teachers)	**342**	**289**			

Source: MPR classroom observations conducted spring 2006.

Notes: Data are regression-adjusted to account for differences in baseline characteristics. Scoring scale: (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence.

None of the differences is statistically different from zero at the 0.05 level, two-tailed test.

Table G.8. Impacts on Classroom Practices for ETS Districts

Outcome	Treatment	Control	Difference	Effect Size	P-value
Implementation of Literacy Lesson	2.6	2.6	0.0	0.03	0.770
Content of Literacy Lesson	2.3	2.3	0.1	0.06	0.573
Classroom Culture	3.1	3.0	0.1	0.11	0.341
Unweighted Sample Size (Teachers)	**185**	**165**			

Source: MPR classroom observations conducted in spring 2006.

Note: Data are regression-adjusted to account for differences in baseline characteristics. Scoring scale: (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence.

None of the differences is statistically different from zero at the 0.05 level, two-tailed test.

Table G.9. Impacts on Classroom Practices for NTC Districts

Outcome	Treatment	Control	Difference	Effect Size	P-value
Implementation of Literacy Lesson	2.6	2.7	-0.1	-0.07	0.551
Content of Literacy Lesson	2.4	2.5	-0.1	-0.18	0.106
Classroom Culture	3.0	3.1	-0.1	-0.10	0.419
Unweighted Sample Size (Teachers)	**157**	**124**			

Source: MPR classroom observations in spring 2006.

Note: Data are regression-adjusted to account for differences in baseline characteristics. Scoring scale: (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence.

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

Table G.10. Impacts of Reading Test Scores – No Pretests

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
1	-0.01	0.01	-0.02	-0.02	0.827	643	46	4
2	-0.04	0.05	-0.09	-0.09	0.283	1,070	58	5
3	-0.06	0.08	-0.14	-0.14	0.163	1,845	108	12
4	0.01	-0.01	0.02	0.02	0.774	1,971	109	14
5	0.02	-0.02	0.04	0.04	0.599	2,127	101	13
6	-0.45	0.43	-0.88*†	-0.88	0.000	55	4	1
All Grades	**-0.02**	**0.02**	**-0.04**	**-0.04**	**0.362**	**7,711**	**389**	**15**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. Pre-test scores are excluded from the baseline characteristics. Treatment and control group sample sizes are shown in Table G24.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

†Significantly different from zero at the 0.05 level, two-tailed test, after applying a Benjamini-Hochberg correction.

Table G.11. Impacts of Math Test Scores – No Pretests

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
1	0.08	-0.10	0.17	0.17	0.174	534	30	2
2	-0.15	0.18	-0.32*†	-0.32	0.001	971	52	4
3	-0.10	0.12	-0.22*†	-0.22	0.023	1,784	106	11
4	0.01	-0.01	0.02	0.02	0.807	1,989	110	14
5	0.03	-0.02	0.05	0.05	0.440	2,112	101	13
6	-0.24	0.24	-0.48*†	-0.48	0.000	55	4	1
All Grades	**-0.02**	**0.03**	**-0.05**	**-0.05**	**0.293**	**7,445**	**366**	**15**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. Pre-test scores are excluded from the baseline characteristics. Treatment and control group sample sizes are shown in Table G.25.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

†Significantly different from zero at the 0.05 level, two-tailed test, after applying a Benjamini-Hochberg correction.

Table G.12. Impacts of Reading Test Scores with DIBELS scores included

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
1	--	--	--	--		95	6	1
2	-0.10	0.08	-0.18	-0.18	0.067	580	44	5
3	-0.08	0.10	-0.18*	-0.18	0.033	1,155	78	9
4	0.02	-0.02	0.04	0.04	0.421	1,679	108	14
5	0.01	-0.01	0.01	0.01	0.843	1,516	81	11
6	--	--	--	--		48	4	1
All Grades	**0.00**	**0.01**	**-0.01**	**-0.01**	**0.771**	**5,073**	**294**	**15**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. Treatment and control group sample sizes are shown in Table G.26.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

None of the differences is significantly different from zero after applying a Benjamini-Hochberg correction.

Table G.13. Impacts on Reading Test Scores with SAT/ACT Scores Included in Regression Model

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
Reading Scores								
2	-0.10	0.08	-0.18	-0.18	0.058	543	42	4
3	-0.07	0.09	-0.15	-0.15	0.069	1,113	75	8
4	0.01	-0.01	0.03	0.03	0.613	1,679	108	14
5	0.00	0.00	0.00	0.00	0.932	1,516	81	11
6	--	--	--	--		48	4	1
All Grades	**0.00**	**0.00**	**0.00**	**0.00**	**0.931**	**4,899**	**283**	**15**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR calculations using data from the College Board and ACT, Inc.; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. Teacher ACT/SAT scores have been included as baseline characteristics. Treatment and control group sample sizes are shown in Table G.27.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

Table G.14. Impacts on Math Test Scores with SAT/ACT Scores Included in Regression Model

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
Math Scores								
2	-0.13	0.12	-0.26*†	-0.26	0.010	472	35	3
3	-0.12	0.15	-0.28*†	-0.28	0.001	837	65	6
4	0.00	0.00	0.00	0.00	0.928	1,545	99	13
5	-0.03	0.03	-0.06	-0.06	0.447	1,510	81	11
6	--	--	--	--		48	4	1
All Grades	**-0.03**	**0.03**	**-0.06**	**-0.06**	**0.179**	**4,412**	**261**	**14**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR calculations using data from the College Board and ACT, Inc.; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. Teacher ACT/SAT scores have been included as baseline characteristics. DIBELS tests are not included. Treatment and control group unweighted sample sizes are shown in Table G.28.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

†Significantly different from zero at the 0.05 level, two-tailed test, after applying a Benjamini-Hochberg correction.

Table G.15. Impacts on Reading Test Scores – ETS Districts

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
2	-0.01	0.01	-0.02	-0.02	0.675	443	32	3
3	0.02	-0.02	0.04	0.04	0.775	369	26	3
4	0.03	-0.03	0.06	0.06	0.320	980	59	7
5	0.04	-0.02	0.06	0.06	0.391	742	38	6
6	--	--	--	--		48	4	1
All Grades	**0.00**	**0.00**	**0.00**	**0.00**	**0.986**	**2,582**	**150**	**8**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. Teacher ACT/SAT scores have not been included as baseline characteristics. DIBELS tests are not included. Treatment and control group sample sizes are shown in Table G.29.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

Table G.16. Impacts on Math Test Scores – ETS Districts

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
2	-0.11	0.13	-0.23*	-0.23	0.048	372	25	2
3	0.03	-0.04	0.08	0.08	0.406	314	22	2
4	-0.01	0.01	-0.02	-0.02	0.806	838	51	6
5	0.08	-0.05	0.14	0.14	0.287	744	38	6
6	--	--	--	--		48	4	1
All Grades	**-0.01**	**0.01**	**-0.03**	**-0.03**	**0.585**	**2,316**	**133**	**7**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. Teacher ACT/SAT scores have not been included as baseline characteristics. DIBELS tests are not included. Treatment and control group sample sizes are shown in Table G.30.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

None of the differences is significantly different from zero after applying a Benjamini-Hochberg correction.

Table G.17. Impacts on Reading Test Scores – NTC Districts

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
2	-1.01	0.32	-1.33*†	-1.33	0.000	100	10	1
3	-0.03	0.04	-0.07	-0.07	0.586	744	49	5
4	0.02	-0.02	0.04	0.04	0.504	699	49	7
5	0.15	-0.17	0.32*†	0.32	0.005	774	43	5
6	--	--	--	--		0	0	0
All Grades	**0.00**	**0.00**	**0.00**	**0.00**	**0.954**	**2,317**	**133**	**7**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. Teacher ACT/SAT scores have not been included as baseline characteristics. DIBELS tests are not included. Treatment and control group sample sizes are shown in Table G.31.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

†Significantly different from zero at the 0.05 level, two-tailed test, after applying a Benjamini-Hochberg correction.

Table G.18. Impacts on Math Test Scores – NTC Districts

	Adjusted Mean Test Scores					Unweighted Sample Sizes		
Grade	Treatment	Control	Difference	Effect Size	P-value	Students	Teachers	Districts
2	-0.58	0.18	-0.76*†	-0.76	0.000	100	10	1
3	-0.13	0.16	-0.29*†	-0.29	0.002	523	43	4
4	0.09	-0.11	0.21*†	0.21	0.005	707	48	7
5	0.01	-0.01	0.02	0.02	0.869	766	43	5
6	--	--	--	--		0	0	0
All Grades	**-0.05**	**0.05**	**-0.10**	**-0.10**	**0.121**	**2,096**	**128**	**7**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Notes: Data are regression-adjusted using a least squares model to account for baseline characteristics and clustering of students within schools. Teacher ACT/SAT scores have not been included as baseline characteristics. DIBELS tests are not included. Treatment and control group sample sizes are shown in Table G.32.

-- Impacts cannot be estimated due to treatment status being confounded with covariates.

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

†Significantly different from zero at the 0.05 level, two-tailed test, after applying a Benjamini-Hochberg correction.

Table G.19. Impacts on Teacher Retention Rates for ETS Districts (Percentages)

Outcome	Total	Treatment	Control	Difference	P-value
Retained in the same school	70.4	69.4	71.4	-1.9	0.783
Retained in the same district	81.9	81.2	82.5	-1.4	0.682
Retained in the teaching profession	93.7	93.8	93.6	0.2	0.937
Unweighted Sample Size (Teachers)	**469**	**233**	**236**		
Unweighted Sample Size (Schools)	**191**	**94**	**97**		

Source: MPR Mobility Survey administered in 2006-2007 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.

Table G.20. Impacts on Teacher Retention Rates for NTC Districts (Percentages)

Outcome	Total	Treatment	Control	Difference	P-value
Retained in the same school	78.5	77.9	79.2	-1.3	0.790
Retained in the same district	89.2	90.6	87.6	3.0	0.378
Retained in the teaching profession	93.0	93.2	92.7	0.5	0.858
Unweighted Sample Size (Teachers)	**413**	**224**	**189**		
Unweighted Sample Size (Schools)	**199**	**105**	**94**		

Source: MPR Mobility Survey administered in 2006-2007 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.

Table G.21. Mobility Impacts Under Alternative Assumptions

Outcome and Assumption	Treatment Group Mean	Control Group Mean	Difference (Estimated Impact)
Retention in the District			
Respondents			
Benchmark weights (benchmark estimates)	85.6	85.7	-0.1
No weights	85.6	85.3	0.3
Enhanced weights	85.3	85.8	-0.6
Respondents and Nonrespondents			
Assume 100% of treatment nonrespondents are movers, 0% of controls	80.1	86.6	-6.5*
Assume 0% of nonrespondents are movers	86.7	86.5	0.2
Assume 25% of nonrespondents are movers	84.7	84.2	0.5
Assume 50% of nonrespondents are movers	82.6	80.7	1.9
Assume 100% of nonrespondents are movers	80.2	77.7	2.5
Assume 0% of treatment nonrespondents are movers, 100% of controls	86.7	77.6	9.1*
Retention in the Teaching Profession			
Respondents			
Benchmark weights (benchmark estimates)	93.8	93.8	-0.1
No weights	93.7	93.7	0.1
Enhanced weights	93.7	93.9	-0.2
Respondents and Nonrespondents			
Assume 100% of treatment nonrespondents are leavers, 0% of controls	88.2	94.7	-6.5*
Assume 0% nonrespondents are leavers	94.2	94.1	0.1
Assume 25% of nonrespondents are leavers	92.2	91.1	1.1
Assume 50% of nonrespondents are leavers	90.1	87.6	2.5
Assume 100% of nonrespondents are leavers	88.4	85.7	2.7
Assume 0% of treatment nonrespondents are leavers, 100% of controls	94.5	84.8	9.7*
Unweighted Sample Size (Teachers)			
Respondents	**463**	**432**	
Respondents and Nonrespondents	**506**	**503**	

Source: MPR Mobility Survey administered in 2006-2007 to all study teachers.

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

Table G.22. Treatment and Control Sample Sizes for Impacts on Reading Test Scores (Benchmark Model)

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
2	243	18	18	4	300	24	19	4
3	629	43	35	8	484	32	25	8
4	919	56	51	14	760	52	50	14
5	707	38	33	11	809	43	38	11
6	24	3	3	1	24	1	1	1
All Grades	**2,522**	**147**	**102**	**15**	**2,377**	**136**	**90**	**15**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts.

Table G.23. Treatment and Control Sample Sizes for Impacts on Math Test Scores (Benchmark Model)

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
2	226	16	16	3	246	19	17	3
3	469	37	29	6	368	28	21	6
4	805	50	47	13	740	49	47	13
5	699	38	33	11	811	43	38	11
6	24	3	3	1	24	1	1	1
All Grades	**2,223**	**133**	**94**	**14**	**2,189**	**128**	**84**	**14**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts.

Table G.24. Treatment and Control Sample Sizes for Impacts on Reading Test Scores – No Pretests

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
1	364	26	21	4	279	20	18	4
2	557	29	25	5	513	29	23	5
3	1,019	58	50	12	826	50	43	12
4	1,067	56	51	14	904	53	50	14
5	968	46	40	13	1,159	55	50	13
6	27	3	3	1	28	1	1	1
All Grades	**4,002**	**203**	**122**	**15**	**3,709**	**186**	**114**	**15**

Source: MPR analysis of data provided by participating school districts covering 2004-2005 and 2005-2006.

Table G.25. Treatment and Control Sample Sizes for Impacts on Math Test Scores – No Pretests

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
1	292	16	12	2	242	14	12	2
2	532	27	23	4	439	25	21	4
3	992	57	49	11	792	49	42	11
4	1,079	56	51	14	910	54	51	14
5	952	46	40	13	1,160	55	50	13
6	27	3	3	1	28	1	1	1
All Grades	**3,874**	**190**	**119**	**15**	**3,571**	**176**	**109**	**15**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts.

Table G.26. Treatment and Control Sample Sizes for Impacts on Reading Test Scores with DIBELS scores included

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
1	33	2	2	1	62	4	2	1
2	264	19	19	5	316	25	20	5
3	656	45	37	9	499	33	26	9
4	919	56	51	14	760	52	50	14
5	707	38	33	11	809	43	38	11
6	24	3	3	1	24	1	1	1
All Grades	**2,603**	**152**	**105**	**15**	**2,470**	**142**	**92**	**15**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts.

Table G.27. Treatment and Control Sample Sizes for Impacts on Reading Test Scores with SAT/ACT Scores Included in Regression Model

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Reading Scores								
2	243	18	18	4	300	24	19	4
3	629	43	35	8	484	32	25	8
4	919	56	51	14	760	52	50	14
5	707	38	33	11	809	43	38	11
6	24	3	3	1	24	1	1	1
All Grades	**2,522**	**147**	**102**	**15**	**2,377**	**136**	**90**	**15**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts.

Table G.28. Treatment and Control Sample Sizes for Impacts on Math Test Scores with SAT/ACT Scores Included in Regression Model

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
Math Scores								
2	226	16	16	3	246	19	17	3
3	469	37	29	6	368	28	21	6
4	805	50	47	13	740	49	47	13
5	699	38	33	11	811	43	38	11
6	24	3	3	1	24	1	1	1
All Grades	**2,223**	**133**	**94**	**14**	**2,189**	**128**	**84**	**14**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts.

Table G.29. Treatment and Control Sample Sizes for Impacts on Reading Test Scores – ETS Districts

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
2	219	15	15	3	224	17	12	3
3	219	15	13	3	150	11	8	3
4	531	32	27	7	449	27	26	7
5	295	17	16	6	447	21	20	6
6	24	3	3	1	24	1	1	1
All Grades	**1,288**	**78**	**55**	**8**	**1,294**	**72**	**49**	**8**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts.

Table G.30. Treatment and Control Sample Sizes for Impacts on Math Test Scores –ETS Districts

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
2	202	13	13	2	170	12	10	2
3	179	12	10	2	135	10	7	2
4	411	26	23	6	427	25	24	6
5	295	17	16	6	449	21	20	6
6	24	3	3	1	24	1	1	1
All Grades	**1,111**	**67**	**49**	**7**	**1,205**	**66**	**45**	**7**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts.

Table G.31. Treatment and Control Sample Sizes for Impacts on Reading Test Scores – NTC Districts

	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
2	24	3	3	1	76	7	7	1
3	410	28	22	5	334	21	17	5
4	388	24	24	7	311	25	24	7
5	412	21	17	5	362	22	18	5
6	0	0	0	0	0	0	0	0
All Grades	**1,234**	**69**	**47**	**7**	**1,083**	**64**	**41**	**7**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts.

Table G.32. Treatment and Control Sample Sizes for Impacts on Math Test Scores – NTC Districts

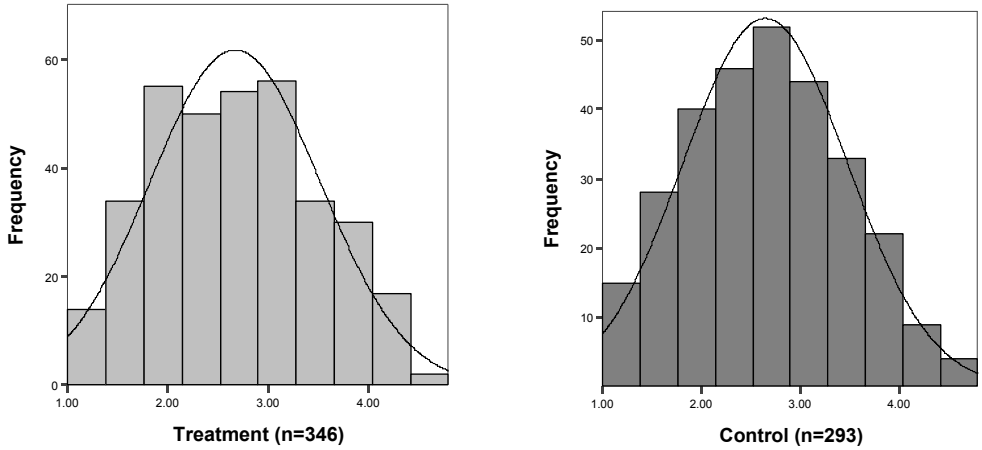
	Unweighted Sample Sizes: Treatment Group				Unweighted Sample Sizes: Control Group			
Grade	Students	Teachers	Schools	Districts	Students	Teachers	Schools	Districts
2	24	3	3	1	76	7	7	1
3	290	25	19	4	233	18	14	4
4	394	24	24	7	313	24	23	7
5	404	21	17	5	362	22	18	5
6	0	0	0	0	0	0	0	0
All Grades	**1,112**	**66**	**45**	**7**	**984**	**62**	**39**	**7**

Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts.

APPENDIX H

SUPPLEMENTAL FIGURES

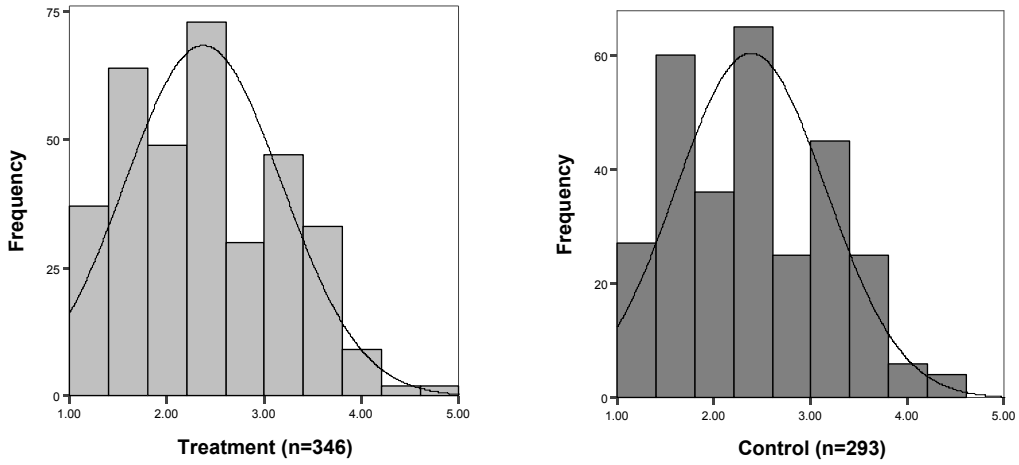
Figure H.1. Distribution of Classroom Observation Scores: Literacy Implementation



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

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

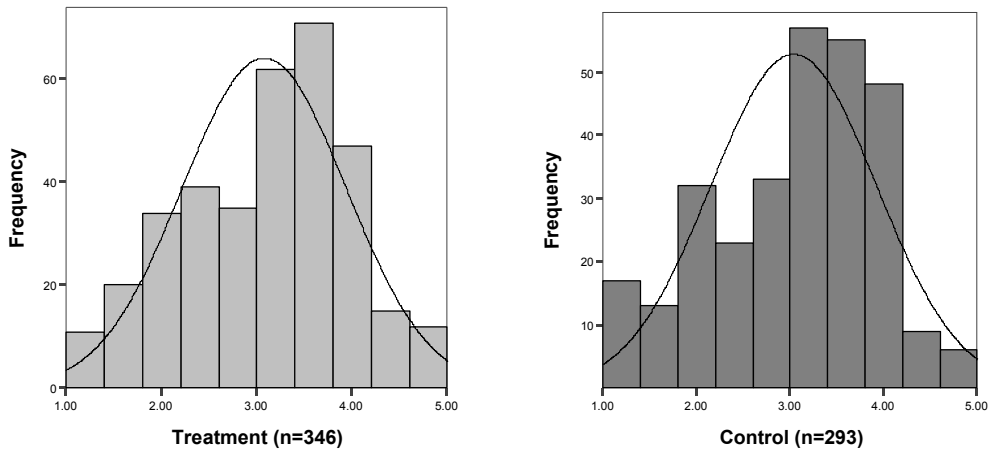
Figure H.2. Distribution of Classroom Observation Scores: Literacy Content



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

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

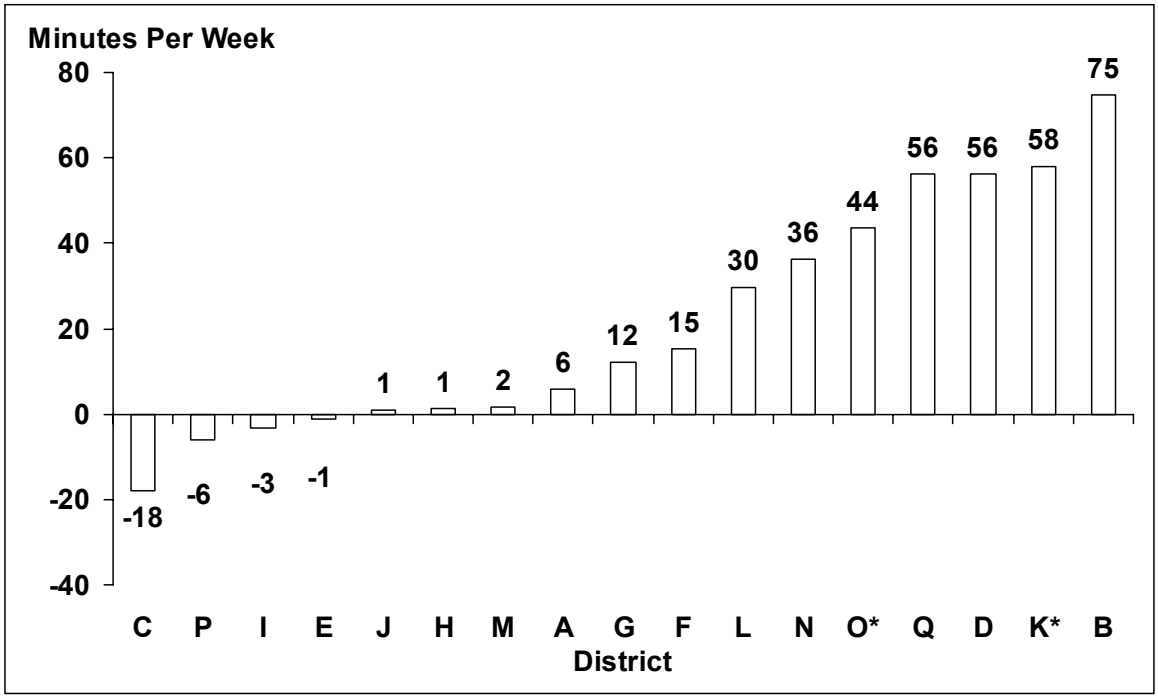
Figure H.3. Distribution of Classroom Observation Scores: Literacy Culture



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

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

Figure H.4. Impacts on Total Minutes Spent in Mentoring Per Week by District

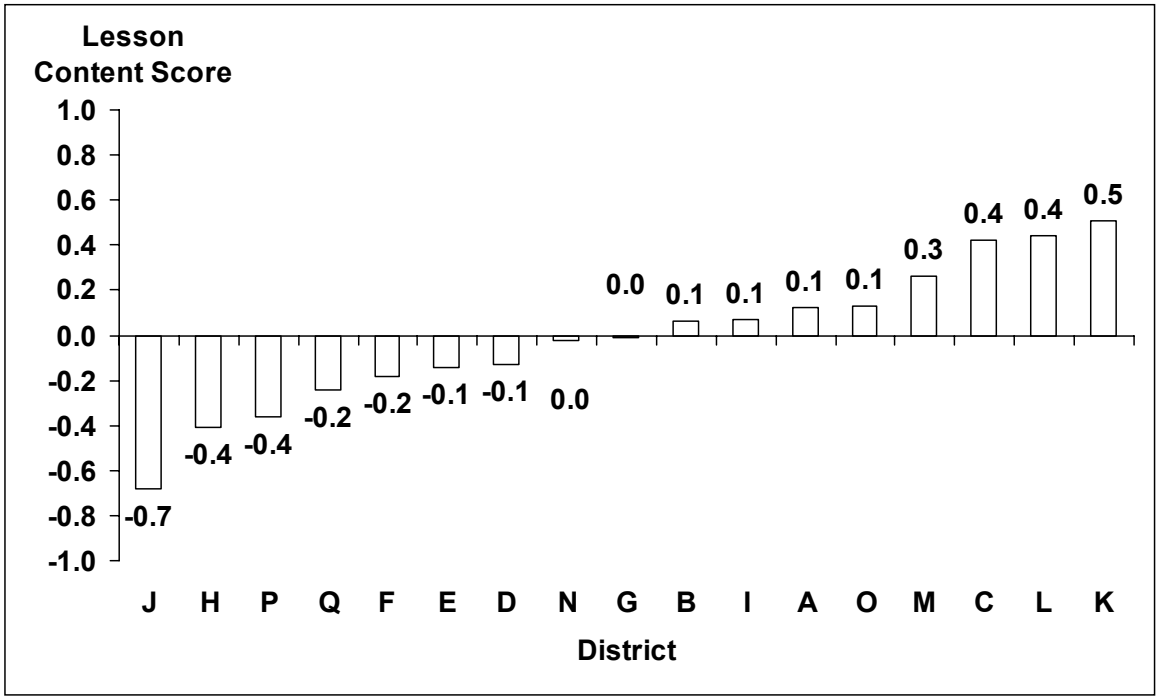


Source: MPR First Induction Activities Survey administered in fall/winter 2005-2006 to all study teachers.

Note: 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 "P" are arbitrary.

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

Figure H.5. Impacts on Classroom Practices by District



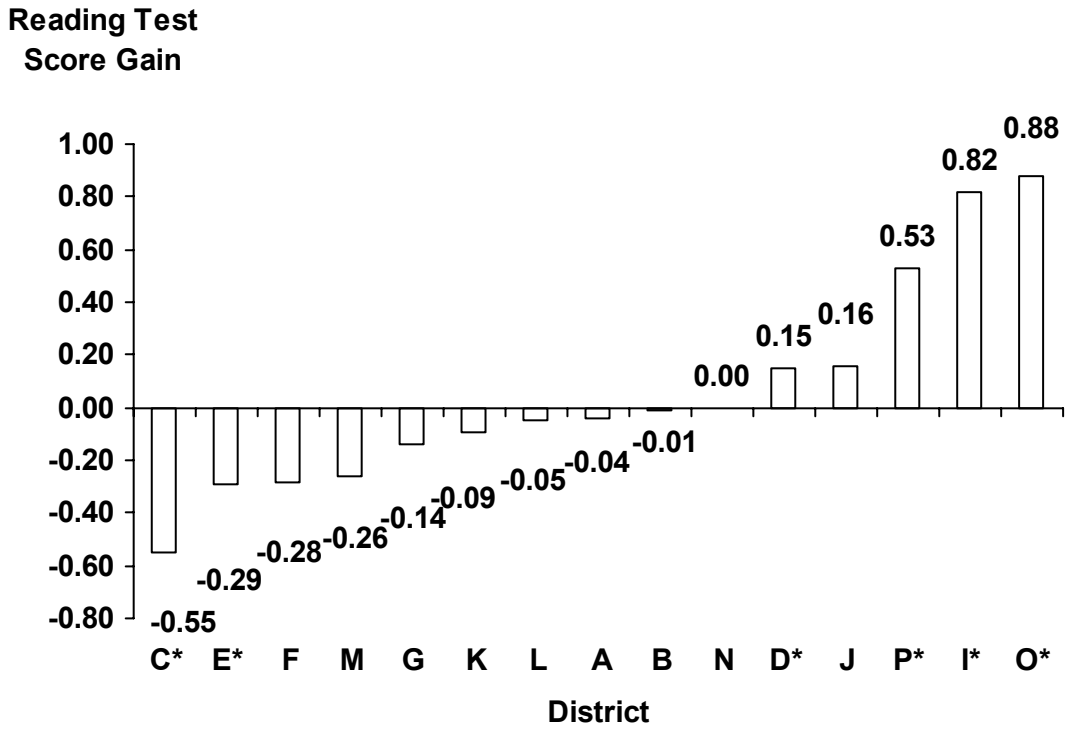
Source: MPR classroom observations conducted in spring 2006; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Note: 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 "P" are arbitrary.

Scoring scale: (1) no evidence, (2) limited evidence, (3) moderate evidence, (4) consistent evidence, or (5) extensive evidence of effective teaching practice.

District-specific impacts are not statistically significant for any district.

Figure H.6. Impacts on Reading Test Scores by District



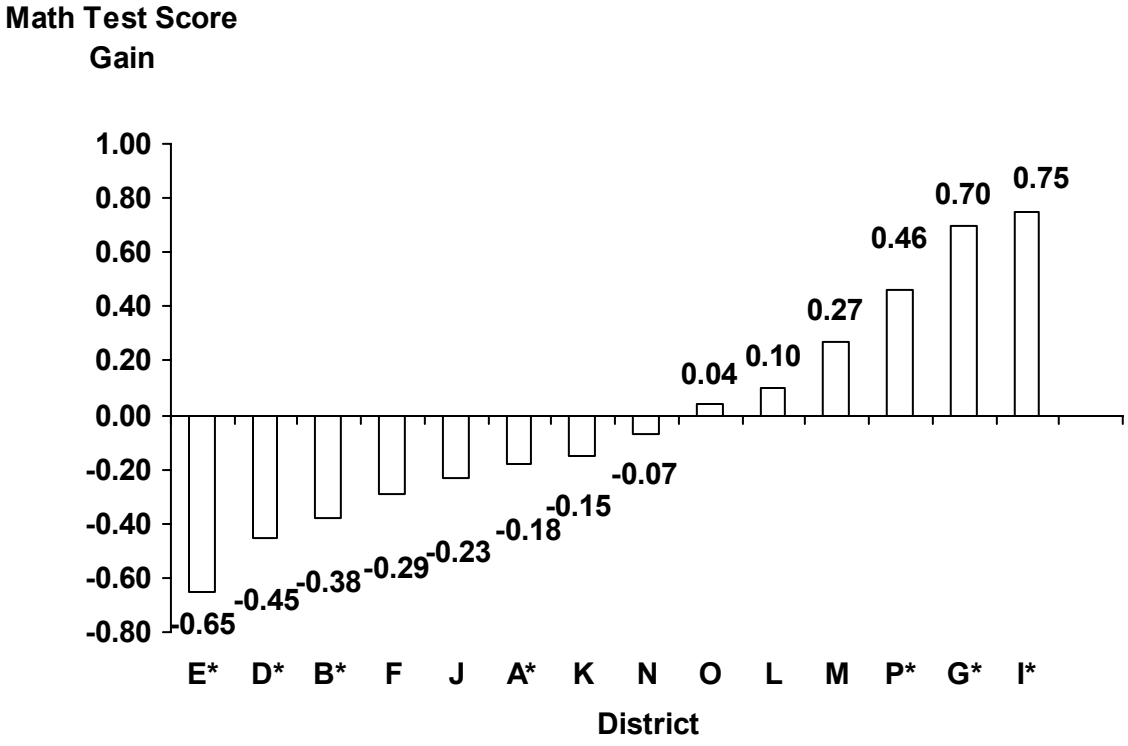
Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Note: 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 "P" 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 correction is applied for multiple comparisons).

Figure H.7. Impacts on Math Test by District



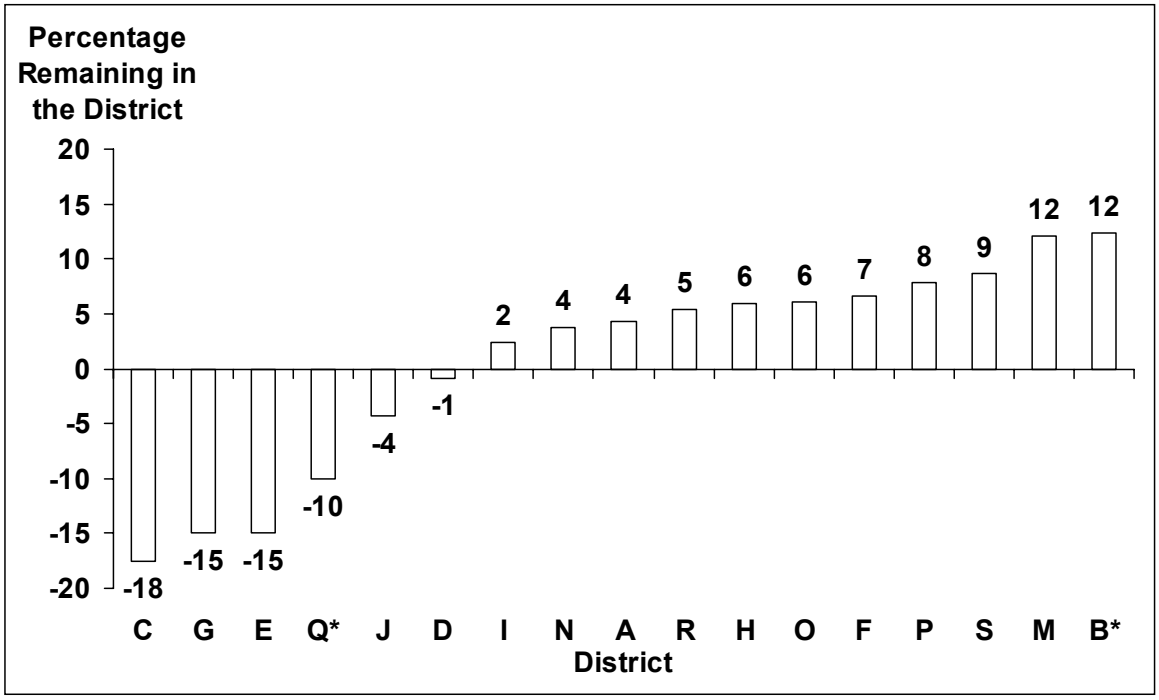
Source: MPR analysis of data from 2004-2005 and 2005-2006 provided by participating school districts; MPR Teacher Background Survey administered in 2005-2006 to all study teachers.

Note: 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 "P" 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 correction is applied for multiple comparisons).

Figure H.8. Impacts on Teacher Retention, by District



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

Note: 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 "P" are arbitrary.

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