NCEE 2010-4028

U.S. DEPARTMENT OF EDUCATION

Impacts of Comprehensive Teacher Induction

Final Results from a Randomized Controlled Study

Executive Summary

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

Impacts of Comprehensive Teacher Induction

Final Results from a Randomized Controlled Study

Executive Summary

June 2010

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

Melanie Ali
Project Officer
Institute of Education Sciences

NCEE 2010-4028
U.S. DEPARTMENT OF EDUCATION

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

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

June 2010

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

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

To order copies of this report,

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

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

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

ACKNOWLEDGMENTS

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

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

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

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

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST[1]

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

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

EXECUTIVE SUMMARY

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

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

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

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

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

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

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

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

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

Key findings:

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

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

- **The extra induction support for treatment teachers did not translate into impacts on classroom practices in the first year.** We observed teachers giving a literacy lesson in the spring of their first year and found no impacts on teachers' implementation of the literacy lesson, content of the literacy lesson, or classroom culture.
- **For teachers who received one year of comprehensive induction, there was no impact on student achievement.** In each of the first three years of teachers' careers, students of treatment teachers receiving one year of comprehensive induction support performed no better on average than students of the corresponding control teachers.
- **For teachers who received two years of comprehensive induction, there was no impact on student achievement in the first two years. In the third year, there was a positive and statistically significant impact on student achievement.**
  - In the third year, in districts and grades in which students' test scores from the current and prior year are available, students of treatment teachers outperformed students of the corresponding control teachers on average. These impacts are equivalent to effect sizes of 0.11 in reading and 0.20 in math, which is enough to move the average student from the 50th percentile up 4 percentile points in reading and 8 percentile points in math.
  - These results are based on the subset of data for which students' test scores from the current and prior year are available. If the analyses are conducted without requiring test scores from the prior year, we do not find an impact on math or reading scores. This alternative approach nearly doubles the available sample of study teachers but the lack of data on students' prior achievement results in a less precise estimate. This means that we are less likely to detect a true impact if it exists, despite the larger sample size.
- **Neither exposure to one year nor exposure to two years of comprehensive induction had a positive impact on retention or other teacher workforce outcomes:**
  - Treatment teachers did not report being more satisfied or feeling more prepared to teach than control teachers at any of the six time points over the four school years in which we collected data.
  - There was no impact on teacher retention over the first four years of the teachers' careers. This was true of retention in the original school, the original school district, and the teaching profession.
  - We found no evidence that comprehensive induction improved the composition of the teacher workforce through selective retention. There were no statistically significant positive differences between treatment teachers retained in the district and control teachers retained in the district in teacher characteristics such as college entrance exam scores, college selectivity, and advanced degrees; nor in performance measures such as first year classroom observation scores or third year student test scores.

Study Design

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

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

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

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

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

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

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

Data

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

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

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

Induction Support for Beginning Teachers

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

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

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

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

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

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

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

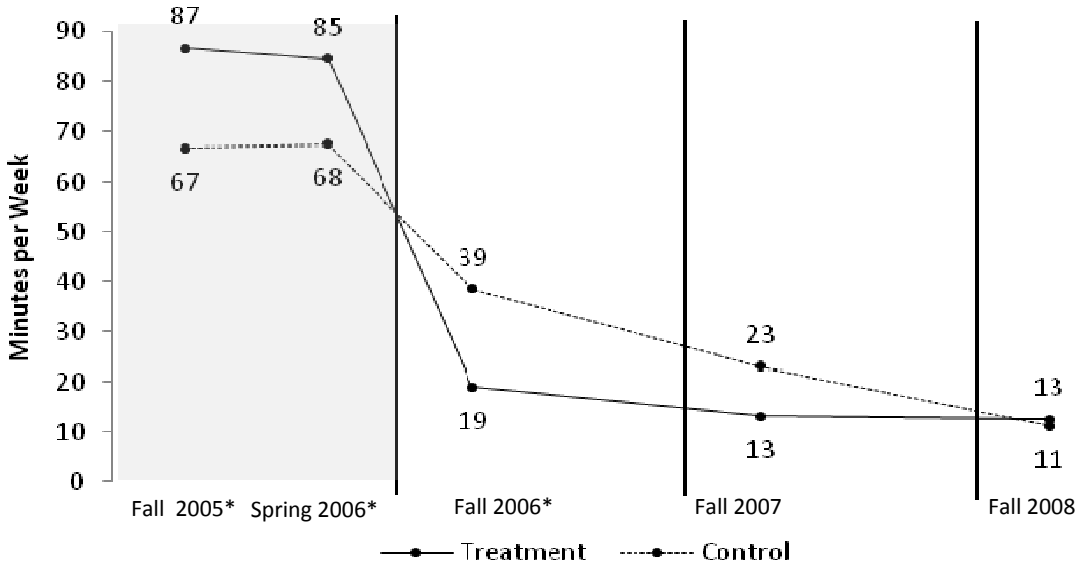
We found that assignment to the treatment group changed the pattern of induction services reported by beginning teachers. Figures ES.1 and ES.2 summarize the timing and intensity of the program's core support—mentoring—measured in minutes per week spent with mentors. Figure ES.1 presents data for one-year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

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

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

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

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

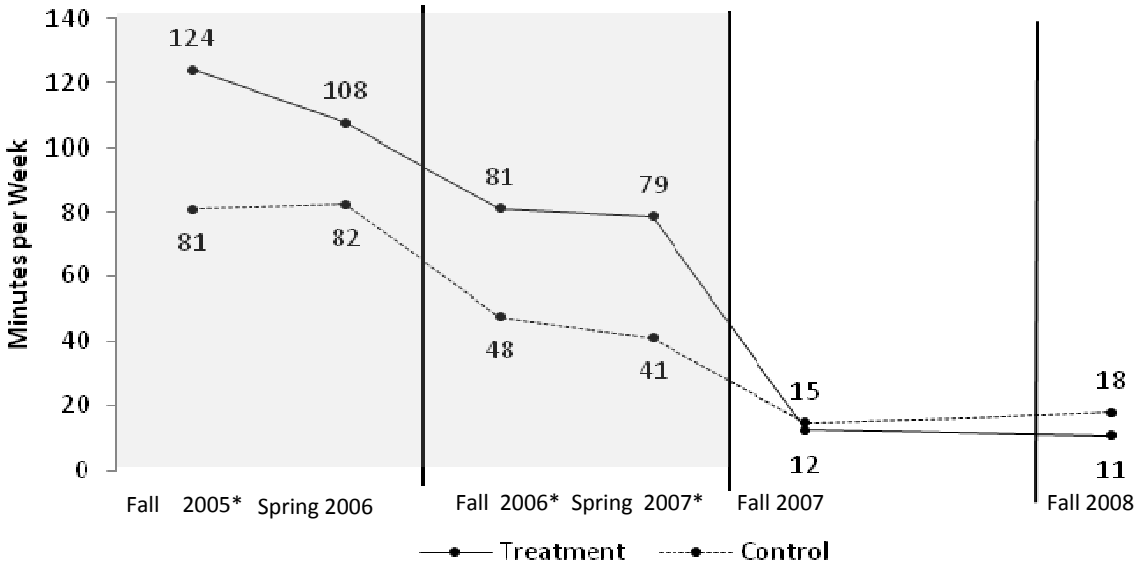


Source: Mathematica First, Second, Third, Fifth, and Sixth Induction Activities Surveys administered in fall 2005, spring 2006, fall 2006, fall 2007, and fall 2008 to all study teachers.

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

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

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



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

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

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

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

Impacts on the Classroom

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

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

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

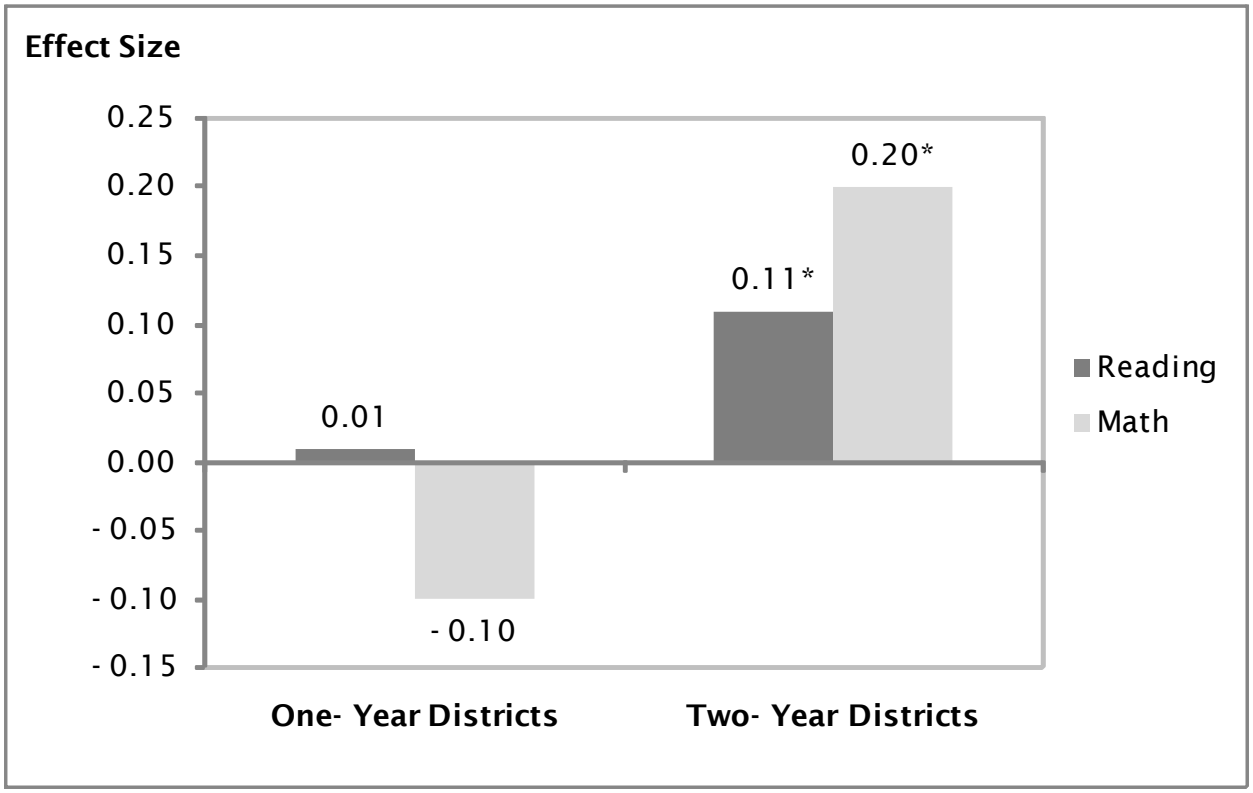
Student Achievement. To measure impacts on student achievement we compared the test scores for students of treatment teachers at the end of the year (posttest) to those of control teachers, accounting for any preexisting differences in prior achievement (pretest) and background characteristics of students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

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

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

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

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



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

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

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

Impacts on the Teaching Workforce

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

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

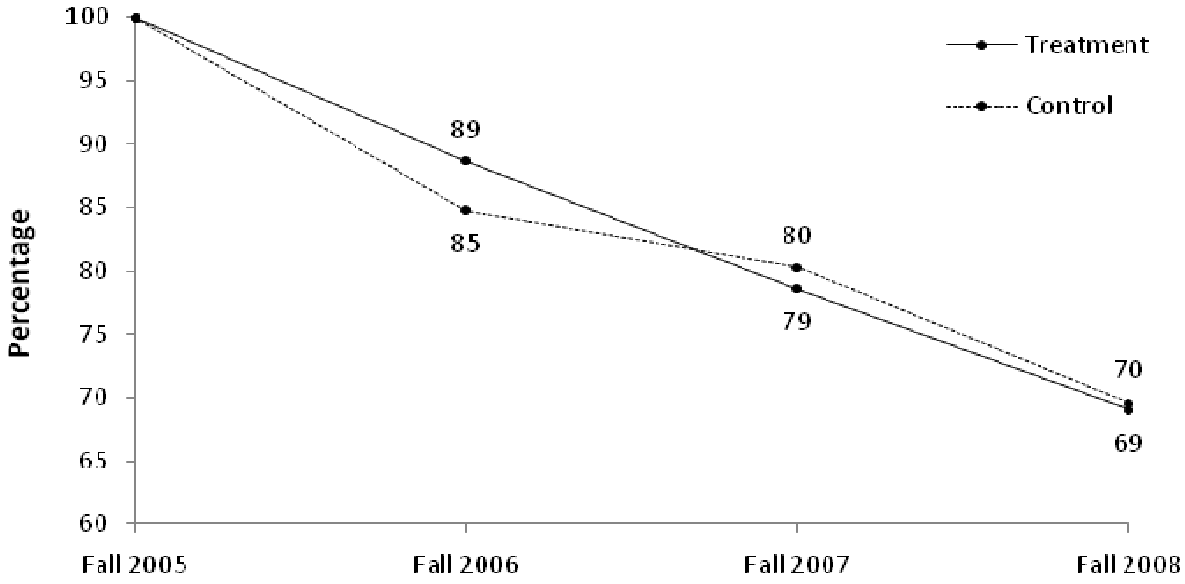
Comprehensive induction did not make teachers feel more satisfied or prepared. Teachers from the treatment and control groups reported feelings of satisfaction and preparedness that differed by 0.1 or less on the four-point scales at all points in time at which we measured these attitudes. These results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

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

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

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

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

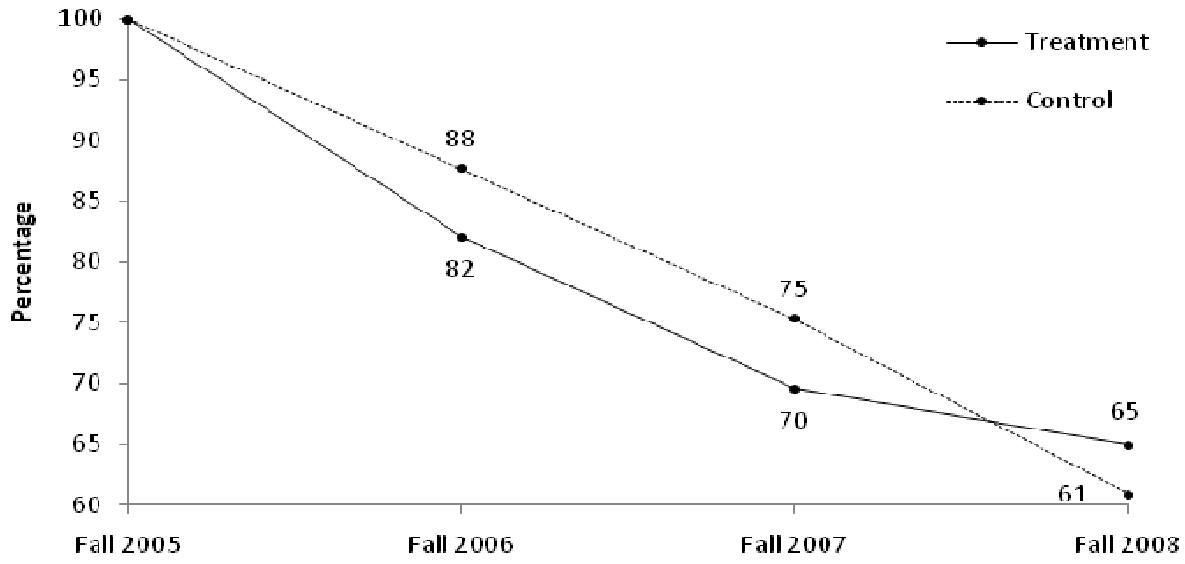


Source: Mathematica First, Second, and Third Teacher Mobility Surveys administered in fall 2006, fall 2007, and fall 2008 to all study teachers.

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

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

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



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

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

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

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

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

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

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

Association Between Levels of Induction Support and Outcomes

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

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

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

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

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

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

*Significantly different from zero at the 0.05 level.

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

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

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

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

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

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

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

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

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

REFERENCES

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

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.

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

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

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

Ingersoll, Richard M., and Thomas M. Smith. "Do Teacher Induction and Mentoring Matter?" *NASSP Bulletin*, vol. 88, no. 638, 2004, pp. 28-40.

Ingersoll, Richard, and Thomas M. Smith. "The Wrong Solution to the Teacher Shortage." *Educational Leadership*, vol. 60, 2003, pp. 30-33.

Interstate New Teacher Assessment and Support Consortium. *Model Standards for Beginning Teacher Licensing, Assessment and Development: A Resource for State Dialogue*. Washington, DC: Council of Chief State School Officers, 1992.

Isenberg, Eric, Steven Glazerman, Martha Bleeker, Amy Johnson, Julieta Lugo-Gil, Mary Grider, Sarah Dolfin, and Edward Britton. *Impacts of Comprehensive Teacher Induction: Results from the Second Year of a Randomized Controlled Study* (NCEE 2009-4072). Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2009.

Kauffman, David, Susan Moore Johnson, Susan M. Kardos, Edward Liu, and Heather G. Peske. "'Lost at Sea': New Teachers' Experiences with Curriculum and Assessment." *Teachers College Record,* vol. 104, no. 2, 2002, pp. 273-300.

Kelly, Linda Molner. "Why Induction Matters." *Journal of Teacher Education*, vol. 55, no. 5, winter 2004, pp. 438-448.

King, M. Bruce, and Fred M. Newmann. "Will Teacher Learning Advance School Goals?" *Phi Delta Kappan,* vol. 81, no. 8, 2000, pp. 576-580.

Serpell, Zewelanji, and Leslie 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.


DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA