NCEE 2009-4035

U.S. DEPARTMENT OF EDUCATION

Impacts of Comprehensive Teacher Induction

Results From the First Year of a Randomized Controlled Study

Executive Summary

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

Executive Summary

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 2009-4035
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 teacher 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

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.

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.