

Does Mentoring Reduce Turnover and Improve Skills of New Employees? Evidence from Teachers in New York City

Jonah Rockoff
Columbia Business School

February, 2008

Mentoring has become an extremely popular policy for improving the retention and performance of new teachers, but we know little about its effects on teacher and student outcomes. I study the impact of mentoring in New York City, which adopted a nationally recognized mentoring program in 2004. I use detailed program data to examine the relationship between teacher and student outcomes and measures of mentoring quality, such as hours of mentoring received and the characteristics of mentors. Although assignment of teachers to mentors was non-random, I use instrumental variables and school fixed effects to address potential sources of bias. I find strong relationships between measures of mentoring quality and teachers' claims regarding the impact of mentors on their success in the classroom, but weaker evidence of effects on teacher absences, retention, and student achievement. The most consistent finding is that retention within a particular school is higher when a mentor has previous experience working in that school, suggesting that an important part of mentoring may be the provision of school specific knowledge. I also find evidence that student achievement in both reading and math were higher among teachers that received more hours of mentoring, supporting the notion that time spent working with a mentor does improve teaching skills.

¹ jonah.rockoff@columbia.edu I thank Leigh Linden, Randy Reback, Doug Staiger, Miguel Urquiola, and seminar participants at the University of Michigan, the Federal Reserve Bank of New York, the New York City Department of Education, and the New Teacher Center at UC Santa Cruz for helpful comments and suggestions. I am also grateful to Dara Barlin, Fred King, Diana Levengood, Nitzan Pelman, Santo Shuvi, and the Regional Directors of the mentoring program who helped me put together the data for this project and answered many questions regarding the program. This work was supported by Atlantic Philanthropic, the Ford Foundation, and the Smith Richardson Foundation. Karla Diaz-Hadzisadikovic provided outstanding research assistance.

1. Introduction

A longstanding theoretical and empirical literature in labor economics focuses on human capital and its relation to work experience and on-the-job training (e.g., Becker (1964), Brown (1989), and Topel (1991)). In education, researchers have documented an important relationship between productivity and work experience: teachers with experience are more effective at raising student achievement than new teachers (see Rockoff (2004), Rivkin et al. (2005), Harris and Sass (2006), or Clotfelter et al. (2007)). Thus, policies that reduce turnover or provide teachers with training that accelerates the return to experience are likely to provide a significant benefit to students in terms of improved educational outcomes.

Both academics and policymakers have been concerned with reducing overall rates of attrition among new teachers and, particularly, addressing high rates of departure from schools serving disadvantaged student populations.¹ Recently, much attention has been paid to how programs that give teachers pecuniary incentives can reduce turnover and increase productivity (e.g., Lavy (2002), Clotfelter et al. (2006), Figlio and Kenny (2006)). These studies generally indicate that teachers do respond to incentives, though other work suggests their career choices are not strongly influenced by variation in pay (e.g., Stinebrickner et al. (2008)). Nevertheless, the use of monetary incentives for public school teachers in the U.S. is quite uncommon and has been so for decades (Ballou (2001)). In the most recent School and Staffing Survey (2003-04), just 8.5 percent of districts reported using monetary incentives to reward excellent teaching.

In contrast to the scarce use of pecuniary incentives, school districts spend considerable resources on services designed to help their new employees develop professional skills and

¹ There is a much larger literature on teachers' labor supply decisions which I do not attempt to summarize here (e.g., Murnane (1981), Murnane and Olsen (1989, 1990), Dolton and van der Klaauw (1995, 1999), Brewer (1996), Stinebrickner (1998, 2001, 2002), Hanushek et al. (2004)). Whether overall rates of attrition are higher in teaching than in similar occupations is a topic of debate among economists (see Harris and Adams (2007)).

overcome difficulties they face at the start of their careers. Mentoring has become one of the most common services of this type.² In the 2003-04 Schools and Staffing Survey, nearly 70 percent of recently hired teachers reported that they received help from a mentor in their first year of teaching, up from roughly 25 percent in 1990. Moreover, a majority of states now require mentoring programs for new teachers (Education Week (2003, 2006)).

Despite the popularity of mentoring, little is known about its impact on employee turnover and skill acquisition. Nearly all published and unpublished evaluations of mentoring programs have used research methodologies that fall short of providing credible estimates of the causal impacts of mentoring (see reviews by Serpell (2000), Ingersoll and Kralik (2004), Lopez et al. (2004), and Strong (2005)).³ Over one million new teachers received mentoring between 1993 through 2003, but we know little about the magnitude of the benefits they have received or how the impact of mentoring varied across different types of programs.

I study the impact of mentoring on new teachers working for the New York City Department of Education (hereafter “the DOE”). In 2004, following the passage of a state law requiring mentoring for new teachers, the DOE adopted a well known mentoring program used in many school districts throughout the country. New York City is the largest school district in the nation with an extremely diverse group of students, many of them from disadvantaged backgrounds. Thus, the impact of mentoring in New York is an important policy outcome in and of itself. Of greater interest, however, is an understanding of what determines the effects of

² “Mentoring” programs are often described as “new teacher induction” and may involve services other than mentoring, (e.g., a workload reduction). I use the term mentoring for simplicity, because it is the most common component of induction programs, and because it best describes the program I analyze here.

³ In addition to the work covered by these reviews, more recent studies, such as Fletcher et al. (2005), Strong (2006), and Villar and Strong (2007), use small samples of teachers and/or cross sectional identification strategies, again limiting the conclusions one can draw from their results. Smith and Ingersoll (2004) and Smith (2007) also find some evidence that mentoring increased retention among new teachers in the 1999 Schools and Staffing Survey. However, these studies face a potential bias from omitted variables that is difficult to resolve without better data (e.g., schools with greater turnover may be more likely to adopt mentoring programs).

mentoring on employee outcomes. The large scale of the DOE program and the availability of detailed data make this an ideal setting to address this more general issue.

Over 500 mentors worked in the DOE program, each working in multiple schools with multiple teachers. This variation allows me to ask whether certain mentors, or mentors with particular characteristics, systematically lead teachers to better outcomes. This line of inquiry is similar to work on whether teachers with certain characteristics are more effective in improving student outcomes (e.g., Harris and Sass (2006), Aaronson et al. (2007), Clotfelter et al. (2007)). In addition, I investigate whether the similarity of mentor and teacher characteristics has implications for the success of mentoring, much in the way similarity has been shown to influence the impact of teachers on students (Dee (2004, 2005)).

To measure the overall impact of the DOE program, I employ a difference-in-differences methodology, taking advantage of the fact that teachers hired with prior experience were not targeted by the program and were much less likely to be assigned a mentor. However, the variation in program implementation is quite rudimentary in that I can only compare outcomes for true “rookie” teachers in pre- and post- program years with yearly outcomes for experienced hires. I do find evidence of a positive effect of mentoring on whether a teacher completes the school year, but no evidence of effects on other dimensions of teacher retention, teacher absences, or student achievement.

The bulk of my analysis focuses on how outcomes were related to variation in measures of mentoring quality, including the hours of mentoring received by teachers, the characteristics of their mentors, interactions between mentor and teacher characteristics, and teachers’ evaluations of mentor performance. The DOE program was not designed with the purpose of evaluation, and teachers were not randomly assigned to mentors. Fortunately, because mentors

typically worked with teachers in a number of different schools, there exist plausible instrumental variables for several key indicators of mentoring quality.⁴ In addition, since most schools employed multiple mentors, I can use school fixed effects to remove bias due to correlation between the characteristics of mentors and the schools to which they were assigned.

I find that a number of indicators of mentoring quality have significant power to predict teachers' claims regarding the impact of mentors on their success in the classroom. However, only a small subset of these measures of mentoring quality had significant impacts on teacher retention or student achievement. On the issue of retention, there is particularly strong evidence that having a mentor who previously worked in the same school as a mentor or teacher has an important impact on whether a teacher decides to remain in the school the following year. This suggests that an important part of mentoring may be the provision of school specific knowledge. I also find evidence that student achievement in both reading and math improved for teachers who were given additional hours of mentoring. Notably, I find very little evidence that teacher or student outcomes are improved when a mentor matches a teacher's subject area, despite the fact that this type of matching is often stressed by state laws and supporters of mentoring programs.⁵

Last, but not least, I find evidence indicating that new teachers who received other types of support (i.e., common planning time, professional development) were more likely to remain teaching in the district and also to return to the same school. This finding is consistent with work by Smith and Ingersoll (2004) using the 1999-2000 Schools and Staffing Survey. As I explain

⁴ For example, I instrument for the number of hours of mentoring a teacher received using the hours logged by his/her mentor with teachers in other schools. This helps to avoid several sources of bias, e.g., mentors cannot meet with teachers who are absent, mentors allocate more time to teachers who are struggling, etc. Similarly, I instrument for a teacher's perception of his/her mentor's performance using evaluations of the mentor by teachers in other schools. This reduces a concern that teachers who are unsuccessful in the classroom lay blame on their mentors and rate their performance as poor.

⁵ Of the 28 states that required or funded mentoring in 2000, 16 also required or strongly recommended that mentors be matched by subject area (Education Week (2000)).

below, data on other types of support come from a survey which purposefully asked new teachers the same questions as those used in the Schools and Staffing Survey.

2. Mentoring of New Teachers in New York City

In February of 2004, New York State established a requirement that all teachers with less than a year of teaching experience receive a “mentored experience.” The law supplied regulations on how districts could fulfill this requirement, but provided no funding to school districts for this purpose. The DOE mentoring program began in the following school year. Although the timing of the program was arguably motivated by the change in state law, it is highly likely that the district could have fulfilled their legal obligations with a much cheaper and less intensive “mentored experience.” Instead, the DOE created a significant initiative with an annual budget of about \$40 million.⁶

The program was designed in partnership with the New Teacher Center (NTC), based at the University of California at Santa Cruz, which created what is arguably the most widely recognized mentoring program in the nation.⁷ The goals of the mentoring program were to reduce turnover among new teachers and increase the achievement of students assigned to these teachers. These goals were made explicit in statements by the head of the NTC and New York City’s mayor, and were directly in line with the state law which required the program:

“This program will help to retain new teachers and accelerate their development.”
- *Ellen Moir, Executive Director of NTC, August 23, 2004*

⁶ The program was discontinued after the school year 2006-07, due to a dramatic change in the organizational structure of the DOE which gave school principals the power to make many financial decisions, including the amount of money to spend on mentoring. Thus, although new teachers in the DOE still receive “mentored experiences,” there is no longer a centralized, uniform, mentoring program nor any centralized data collection on mentoring. Program data for the school year 2006-2007 are as yet unavailable, and I therefore I analyze outcomes from the first two years of the program.

⁷ NTC has a 20 year history of work on developing “a systematic, mentor-based teacher induction model.” Although many districts working with NTC are located in California, NTC has initiatives in 23 other states, Washington DC, and Puerto Rico. More information on NTC can be found at www.newteachercenter.org.

“This new mentoring program ... will help produce and retain even more outstanding teachers for New York City schools.”
- *Mayor Michael R. Bloomberg, August 23, 2004*

“The purpose of the mentoring program shall be...increasing retention of teachers in the public schools, and to increase the skills of new teachers in order to improve student achievement in accordance with the State learning standards.”
- *Regulations of the New York Commissioner of Education, Section 100.2(dd)*

Thus, evaluation of the program based on teacher retention and student achievement outcomes is quite appropriate.

The DOE mentoring program possessed almost all of the features of the NTC model. All newly hired teachers with less than one year of teaching experience were assigned a mentor who was expected to meet with them on a weekly basis. Mentors were given training prior to the start of the school year and were provided with a detailed program to help improve teachers’ instructional skills. The only notable departures from the recommended NTC program were that NTC preferred a two year program and a ratio of teachers to mentors lower than that set by the DOE (17 to 1). Both recommendations were ruled out due to budgetary considerations.

Mentors worked within one of 10 geographic regions of the DOE (see Figure 1) or an 11th “region” devoted to special education programs. Each region had a director who hired mentors and monitored them throughout the year. Mentors did not teach or work as school administrators in addition to mentoring, and almost all of them (92 percent) worked full-time. They were generally selected from current DOE employees but their backgrounds varied considerably, as I discuss below. According to the DOE, there were over 1,600 applicants for the roughly 300 mentoring positions available in the program’s first year. Mentors met regularly during the year with other mentors in their region and with their regional director.

While the target ratio of the DOE program was 17 teachers for each mentor, there was considerable variation around this target in practice. The top panel of Figure 2 shows the

distribution of mentor caseloads, separately for mentors who worked full-time versus part-time. Because some teachers entered and left during the school year, I present these distributions separately for start-of-year and end-of-year caseloads.⁸ The average full time mentor caseload at the start of the school year was 13.8. By the end of the school year, the average had grown to 15.8. There was also substantial variation in caseloads across mentors. The difference between the 90th and 10th percentiles of full-time mentor caseloads was 9 teachers at the start of the school year and 6 teachers at the end of the year. As might be expected, part-time mentors had substantially lower caseloads than full-time mentors.

The bottom panel of Figure 2 plots average end-of-year caseloads against mentors' start-of-year caseloads. There was a clear tendency for caseloads to increase for mentors who began with a relatively light load. For example, full-time mentors who began the school year with only 8 teachers ended with 14.2 teachers on average. In contrast, mentors who began with 17 teachers ended the year with 17.1 on average. This tendency drives the decrease in the 90-10 gap mentioned above.

3. Data

The main data used in my analysis come from the mentoring program. These data provide me with teacher and mentor characteristics (e.g., demographics, school, subject area, etc.), teacher-mentor assignment, and the hours of mentoring the teacher received (as logged by the mentor).⁹ The data also include information on mentors' years employed at the DOE, years since the mentor taught in a classroom, whether the mentor was an administrator, and whether

⁸ Specifically, to measure caseloads, I calculate the number of teachers assigned to the mentor who were hired by September (start of year) and the number who appear in payroll data for the month of May (end of year).

⁹ Because the program targeted teachers with less than a year of experience, some teachers mentored each year were actually hired during the previous spring. In addition, teachers that received zero or few hours of mentoring are often marked as ineligible (i.e., they were assigned a mentor but were discovered to have prior teaching experience). For simplicity, I consider a teacher as mentored only if he/she was mentored during the year he/she was hired and was provided with at least 5 hours of mentoring. This sample restriction does not qualitatively affect my results.

he/she worked full-time. In addition, for mentors that did not return in the second year of the program, I have an indicator for whether the mentor left voluntarily or was asked not to return.

Administrative data from the DOE payroll system provide me with information on all full-time teachers in the DOE in September, November, and May of the school years 2000-2001 through 2005-2006. In particular, the data provide information on each teacher's gender and ethnicity, certification, subject area (based on license), recruitment through special programs (e.g., Teach for America), prior teaching experience (as proxied by their position on a salary schedule), number of absences, and whether they have left the DOE or switched schools. Almost all mentors (95 percent) also appear in the DOE payroll data, providing me with their years of teaching experience (again, proxied by salary) and the school(s) in which they recently taught.

I also use data from a survey of teachers administered by the DOE program in the spring of 2006 that focused on evaluating mentor activity and performance. The response rate—among teachers still working when the survey was administered—was about 75 percent. I use two questions from this survey to create an index of mentor quality: (1) “Of the success you’ve had as a beginning teacher, what proportion would you attribute to help from your mentor?” and (2) “Overall, how helpful has your mentor been in developing your confidence, knowledge, and skills in teaching?” Teachers replied to each question on a five point scale.¹⁰ For simplicity, I normalize the distribution of each question and then average them to create a “mentor evaluation” for each teacher. Although only 70 percent of teachers in the program during the school year 2005-2006 answered this question, the sizeable number of teachers per mentor

¹⁰ These two questions were placed in different sections of the survey and the scales were worded differently. The possible replies to first question (“Of the success...”) were: “A great deal”; “Quite a bit”; “Some”; “Hardly at all”; “Not at all”. The possible replies to first question (“Overall; how helpful...”) were: “Extremely Helpful”; “Helpful”; “Neutral”; “Somewhat Helpful”; “Not Helpful at All”. Of the 4767 respondents, 4563 submitted answers to these questions. The correlation between responses was 0.75, and the results presented below are similar whether I use one or both questions to measure mentoring quality.

means that 90 percent of the mentors working during 2005-2006 had at least five teachers evaluate them and the median number of evaluations per mentor was 12.

The survey also provides me with several other variables I use in my analysis. The survey asked teachers about the subject area and grade level in which they worked and whether their mentor's area of expertise matched their subject area and grade level. It also included several questions replicated from the 1999-2000 SASS and analyzed by Smith and Ingersoll (2004). Teachers were asked whether they had common planning time with other teachers, whether they attended professional development workshops and conferences, and whether they had a reduced work load. As noted above, Smith and Ingersoll (2004) find that teachers who had common planning time with other teachers were significantly less likely to leave teaching after their first year. I therefore include these variables in my analysis below.

I measure student achievement using data on standardized test scores in math and reading for students in grades four through eight. These data follow students over time and provide links to their reading and math teachers. They also include information on student demographics, receipt of free and reduced price lunch, and status for special education and English Language Learner services. A more detailed description of the data can be found in Kane et al. (2006). Because of the restricted subject areas and grades for which achievement data are available, I can only estimate the impact of mentoring on student achievement using a subset of teachers.

3.1 Descriptive Statistics

Figure 3 plots trends in outcomes for new teachers with and without prior experience over the school years 2000-2001 to 2005-2006. The top panel shows the fraction that completed their first year of teaching, who returned to the DOE in their second year, and who returned to teach in the same school, as well as the total number of absences for teachers who worked over

the entire school year. The bottom panel shows the average reading and math test scores of students in the fourth through eighth grades who were assigned to these newly hired teachers. Because the scaled scores of these tests differ across grades and years, they have been normalized by grade and year to have a mean of zero and a standard deviation of one. The top panel shows that at the end of this period retention rates and absences improved for new, inexperienced hires, relative to new, experienced teachers, but these relative shifts appear to begin before the start of the mentoring program. With regard to achievement, the trends bounce around quite a bit and there is not much evidence of a significant change in the raw differences in achievement between students assigned to teachers with and without prior experience.

Importantly, the trends shown in Figure 3 do not take into account any changes in the composition of new teachers. Table 1 presents summary statistics on all teachers hired in the school years 2000-2001 to 2005-2006, broken out by period hired and whether the teacher had prior experience. There was a major compositional shift in the characteristics of newly hired teachers over this period. For instance, the fraction of newly hired teachers that had prior experience nearly doubled from 17 percent to 30 percent, the fraction of inexperienced hires without certification fell dramatically from 60 to 6 percent, and the fraction of experienced hires recruited internationally shrunk from 21 to 8 percent. Thus, it is difficult to conclude much about the impact of mentoring without accounting for the composition of new hires. Additional graphical analysis of teacher and student outcomes, after conditioning on their observable characteristics, is presented in Section 5.

Summary statistics on mentoring related variables for mentored teachers are shown in Table 2. In the school years 2004-2005 and 2005-2006 the mean hours of mentoring provided were 32 and 39, respectively, with standard deviations of 17 and 14, respectively. The average

caseload and the number of schools assigned to the teachers' mentors were similar across program years, with a mean of 16 and a standard deviation of 3. Thus, it seems there may have been a significant amount of variation in mentoring services received by teachers.¹¹ One can see some evidence of this variation by looking at teachers' perceptions of their mentors' performances. The average responses to the two survey questions on mentor performance (on a scale from one to five) were 4.04 for "Overall Rating" and 3.48 for "Amount of Success Attributed to Mentor," with standard deviations for both ratings slightly greater than 1.

Using the hours of mentoring a teacher received or his/her evaluation of the mentor's performance as measures of mentoring quality creates potential endogeneity problems. For example, mentors may have allocated more hours to teachers who are struggling, and teachers who perform poorly for other reasons may have blamed their mentor. In an attempt to avoid these types of biases, I construct teacher-specific instrumental variables corresponding to the average value (of hours or evaluation) among teachers in *other* schools who were assigned the *same* mentor. These variables isolate plausibly exogenous variation across mentors in the quality and quantity of mentoring services they provided. The means and standard deviations for these variables are also shown in Table 2.

Summary statistics for mentors, broken out by year, are shown in Table 3. The number of mentors rose from 348 to 425 between 2004-2005 and 2005-2006, but their demographic characteristics and prior experience were roughly the same. This stability is partly due to the fact that nearly three quarters of the mentors from 2004-2005 remained with the program. About 60 percent of mentors were white, and 80 percent were female. On average, mentors were almost

¹¹ Mentors are responsible for logging the hours they spend with each teacher, and logged hours are thus susceptible to misreporting. According to DOE program officials, mentors were less apt to keep up to date logs of their activities in the first year of the program, and this seems to be the most likely explanation for the greater average hours reported in 2005-2006. I find an increase in hours among mentors who worked in both years, particularly among those who reported fewer hours in 2004-2005.

50 years old and had taught for 13 years, but there was considerable variation in both age (standard deviation 9 years) and teaching experience (standard deviation 6 years). Notably, roughly a quarter of the mentors worked in a school in which they used to teach.

4. Program Assignment

Officially, the DOE program was supposed to provide mentoring to teachers with less than one year of teaching experience. In practice, there were a number of issues that complicated this process. Some teachers who in the data appear eligible for mentoring did not receive it, and some teachers who appear ineligible for mentoring did receive it. As I show below, observable characteristics of teachers can explain only part of the variation in who received mentoring.

In order to assign teachers to mentors, the administrators of the DOE program put together lists of eligible teachers based on hire date and reported experience and distributed them to regional program directors before the start of the school year. For several reasons, these lists were highly imperfect. First, data for a significant fraction of new teachers are unavailable before the start of the school year. About 15 percent of teachers were hired in the middle of the year, i.e., between October and June, and about 10 percent of teachers whose records indicate they were hired between July and September do not show up in the September payroll data, indicating a delay in processing their information. Second, identifying teachers with no prior experience at the start of the school year was difficult. New teachers submit a form in order to be placed above the first step of the salary scale, and many of them do not submit this form by the start of the year. In the 2004-2005 and 2005-2006 school years, about 30 percent of newly hired teachers that eventually received a salary increase due to prior experience actually started out on the first rung of scale at the beginning of the year. Moreover, in school years prior to

2004-2005, most new teachers started the year at the bottom of the salary scale regardless of prior experience because of delays in processing these forms.

As a result, upon visiting schools, mentors often discovered newly hired teachers that needed to be included in the program or discovered that some of their assigned teachers had previous teaching experience. In my conversations with administrators from the program, it seemed that mentors and regional directors were unsure of the exact criteria for eligibility, and decisions about which teachers would be added or removed from mentors' assignments seem to have been made case-by-case in an *ad hoc* manner.

The top panel of Figure 4 shows the fraction of newly hired teachers in school years 2004-2005 and 2005-2006 that received mentoring, by prior teaching experience and whether the teacher was hired "on-time" (i.e., between July and September) or mid-year (October to June). The bottom panel of Figure 4 shows the absolute number of teachers in each of the categories used to generate the top panel. Although assignment into the mentoring program was clearly related to whether a teacher had prior experience and whether or not the teacher was hired mid-year, these relationships were by no means strict. Of the teachers hired on-time, over 15 percent of teachers whose salary indicated they had no prior experience did not receive mentoring, and about 35 percent of teachers whose salary indicated 3 or more years of experience did receive mentoring. The probability of receiving mentoring was lower among teachers hired mid-year, particularly for inexperienced teachers in the first year of the program.

The potential endogeneity of mentoring and the loose relationship between eligibility and receipt mean that one cannot fairly judge the impact of mentoring based on a comparison of teachers who received mentoring to those who did not. Instead, I pursue a difference-in-differences strategy that compares outcomes for new teachers with and without prior experience

and uses the period prior to the school year 2004-2005 to control for preexisting differences in unobservable differences between these groups.

4.1 Assignment of Teachers to Mentors

In order to consistently estimate the relationship between teachers' outcomes and mentors' characteristics or measures of mentoring quality (e.g., hours of mentoring), it is also important to take into consideration the process by which teachers were assigned to particular mentors. A key issue is whether assignment was related to teachers' likelihood of successful outcomes in ways that cannot be controlled for by observable mentor and teacher characteristics.

The only explicit, official goal of the assignment process was to match teachers with mentors who had expertise in their subject areas (e.g., mathematics, special education, science, etc.) and grade levels (e.g., Elementary, High School, etc.). This is a common feature of mentoring programs: 8 of the 30 states with mandated mentoring programs require subject matching, and 73 percent of mentored teachers in the 2003-2004 SASS data claimed that their mentor matched their subject area. In the DOE program, 47.3 percent of teachers are matched with a mentor in their subject according to administrative data. However, in survey responses (available only for a subset of teachers in the school year 2005-2006), 77 percent of new teachers claimed that their mentor's area of expertise matched their subject. The discrepancy between administrative data and teachers' opinions regarding subject matching turns out to be important for my analysis, and I discuss this issue in more detail below.

In my conversations with regional directors, they invariably stated that their first concern in making assignments was to match teachers to mentors in their subject areas. A secondary concern was to minimize the time that mentors spent traveling between teachers. Travel time was reduced by assigning multiple teachers in the same school or in nearby schools to a single

mentor. For example, an ideal assignment pattern would have been to match four teachers in the same subject and school with a single mentor; he/she could then visit the school once per week, meet with each teacher, and avoid all between-teacher travel time. However, such scenarios were typically not possible: only 30 percent of teachers were hired into a school at the same time as three or more colleagues in the same subject.

I examine systematic matching of mentors with teachers by comparing the rates with which teachers and mentors match on observable characteristics with match rates based on simulations where mentors are assigned randomly.¹² I find strong evidence that regional directors placed considerable weight on matching based on subject area, but little evidence to suggest they used characteristics such as race or gender.

I also asked regional directors whether they made assignments based on specific teacher characteristics such as personality, style, or other characteristics that would be unobservable in my data. They stated that it was impossible to do this sort of matching due to the lack of information on teachers and the large number of matches to be made. This reduces some concerns regarding endogeneity, e.g., teachers who were in greater need of mentoring were assigned to mentors with more impressive qualifications.¹³ However, regional directors did say that if they had information about schools' curriculum or educational style that they would assign mentors to schools where they thought the mentor might match well. Thus, there may be some unobservable factors at the school level that caused teachers and mentors of a certain type to locate there. By looking at variation in mentor and teacher characteristics within schools, I hope to avoid any spurious correlation between teacher outcomes and these unobservable factors. I

¹² For details on this simulation analysis, please see the appendix.

¹³ Concerns regarding the endogenous assignment of teachers to mentors of higher or lower quality are further assuaged by my estimation results, which indicate that teacher outcomes were not significantly different with mentors who were asked not to return in the second year of the program, relative to mentors who continued into the second year. Approximately 10 percent of mentors were asked not to return.

also directly examine variables that may indicate that a mentor matched well with a school (e.g., they previously worked there).

Last, but not least, one might be worried about whether assignments were changed after regional directors had more information regarding the mentor-teacher relationship. Regional directors told me that these were made only in very rare situations (i.e., two used the phrases “less than 5” and “maximum of 3” for the number of changes they had ever made, and directors matched about 500 teachers on average per year). Though I do not have data specifically on re-assignments, I have data from the 2005-2006 year that list the teacher-mentor interactions logged by all mentors over the year. From this I can discern that, in that year, 93 percent of teachers interacted with only one mentor, and 97 percent received more than 80 percent of their hours of mentoring from one mentor (see Table 2). Thus, the actual amount of reshuffling may have been slightly higher than stated by the program directors, but it was still quite low.

5. Difference-in-Differences Estimates of the Mentoring Program’s Impact

The difference-in-differences approach to estimating the impact of mentoring on teacher and student outcomes is motivated by the fact that the program targeted only new teachers with less than one year of experience, who constituted roughly 70 percent of the teachers hired during the two program years. The basic difference-in-differences methodology can be implemented using the specification shown by Equation 1:

$$Y_i = \delta D_i^{post} + \tau D_i^{no\ prior} + \lambda D_i^{post} * D_i^{no\ prior} + \beta X_i + \sum_s \alpha_s D_i^s + \xi_i \quad (1)$$

Y_i is the outcome of interest, D_i^{post} indicates whether teacher i was hired after the start of the mentoring program, $D_i^{no\ prior}$ indicates whether a teacher had no prior teaching experience, X_i are

teacher characteristics, and α_s are school fixed effects.¹⁴ The interaction of D_i^{post} and $D_i^{no\ prior}$ is an instrumental variable for whether a teacher was mentored, and the estimate of the parameter λ measures the impact of the higher probability of receiving mentoring for inexperienced teachers targeted by program. This is often referred to in the program evaluation literature as the “Intent to Treat” (ITT) effect, and does not capture the impact of actually receiving mentoring. In order to measure the impact of mentoring, one can scale the estimate of λ by dividing the increased probability that a teacher with no prior experience received mentoring, which was approximately 40 percent. This is typically referred to as a “Treatment on Treated” (TOT) estimate in the program evaluation literature.

While the estimate of λ from Equation 1 should be consistent, the OLS standard error is likely to be severely biased. The instrumental variable ($D_i^{post} * D_i^{no\ prior}$) does not vary among teachers hired in the same year who have similar prior experience, and these teachers may have correlated outcomes due to other unobservable factors they have in common. This is a well known problem in difference-in-difference estimators (Bertrand et al. (2004)) and worsens when important explanatory variable moves in a simple way, e.g., changing for one of two groups at a single period in time. I address this using a two step procedure shown by Equations 2 and 3.¹⁵

$$Y_i = \sum_g \alpha_g D_i^g + \beta X_i + \sum_s \alpha_s D_i^s + \xi_i \quad (2)$$

$$\hat{\alpha}_g = \delta D_g^{post} + \tau D_g^{no\ prior} + \lambda D_g^{post} * D_g^{no\ prior} + \varepsilon_g \quad (3)$$

First, the dependent variable is regressed on a set of fixed effects that group teachers based on year hired and possession of prior experience, plus a set of control variables for teacher

¹⁴ If the outcome is student achievement, one can simply recast D^{post} as an indicator for whether student i 's teacher was hired after the start of the mentoring program, $D^{no\ prior}$ as whether the student's teacher had no prior teaching experience, and X as both student and teacher characteristics.

¹⁵ I thank John DiNardo for suggesting this approach. For details on other methods for dealing with this issue, see Conley and Taber (2005), Cameron et al. (2007), and Donald and Lang (2007).

characteristics and school fixed effects.¹⁶ The coefficients for the group fixed effects are then used as dependent variables in a second-stage, differences-in-differences specification, and the standard errors of the group fixed effect estimates can be used as regression weights. The second stage regression will have just 14 observations, and should provide conservative standard errors for statistical inference.

The identifying assumption in the differences-in-differences specification is that there were no other differential changes over time in the unobservable characteristics of new teachers with and without prior experience (or the students to which they were assigned) that affected the outcomes of interest.¹⁷ This assumption not directly testable, but I investigate its plausibility by examining the group level coefficients graphically (Figure 5). In a number of cases, outcomes improved more for new teachers without prior experience than for those with prior experience, particularly in the final years of my sample. However, as with the raw trends shown in Figure 3, the timing of these improvements is not purely coincident with the start of the program. Moreover, the trends over time in the coefficients for teachers with and without prior experience do not always move together in the pre-program years, particularly for retention within the same school. Thus, to be careful, I include a linear time trend interacted with an indicator variable for having no prior experience as additional controls in my second stage regressions.

The results of my analysis are shown in Table 4. In Column 1, the dependent variable is an indicator for receiving mentoring on the interaction of indicators for program years and being

¹⁶ Controls for teacher characteristics include dummy variables for a teacher's subject area, level, ethnicity, gender, prior experience, age, age squared, whether they were hired after the start of the school year, and type of certification or alternative recruitment program. Some teachers have missing values for some of these control variables (e.g., a handful of teachers are missing ethnicity information). Instead of dropping these observations, I include dummy variables for missing data and set missing values to a constant.

¹⁷ Another potential concern would be if teachers strategically underreported experience (and forfeited an increase in salary) in order to gain access to mentoring services. I believe this was very unlikely for two reasons. First, while teachers submitted for salary increases near the start of the school year, they typically did not meet their mentor until several weeks after the year began. Second, a program official whom I asked about this issue felt that teachers were usually unaware of the program's existence until they were contacted by their mentors.

hired with no prior experience. Not surprisingly, there is a highly significant coefficient of roughly 0.4 on the $D_i^{post} * D_i^{no\ prior}$ interaction, reflecting the expected increase of about 40 percent in the probability that an inexperienced hire received mentoring. In Columns 2 through 5, I show the estimates for teacher absences and a series of indicators for retention: completing the school year, returning the following year, and returning to the same school. While the regressions for absences and student achievement do not show a statistically significant effect of the program, there is a marginally significant and positive effect of 4.5 percent on whether teachers complete their first year (p-value 0.07). This is a sizeable effect, and supports the idea that mentoring can assist struggling teachers work through a difficult first year.

However, there are two reasons to be cautious about this result. First, the estimates for returning the following year and returning to the same school are similarly positive, but less precisely estimated, suggesting that some of the teachers who were induced to complete their first year may still have left the DOE or transferred at the close of the school year. Second, the graphical analysis shown in Figure 5 suggests that this estimate is driven in part by a decrease in completion rates for teachers with prior experience that occurred when the mentoring program was created. If this adverse change for experienced teachers is not the appropriate counterfactual for their counterparts without experience, then my estimates may be biased towards finding a positive effect of the program.

While these estimates provide limited evidence that the mentoring had significant benefits for teachers and students, it is important to recognize that my estimation strategy is only informative about the impact of mentoring on the marginal teachers who received it because they did not have prior teaching experience. If there is important heterogeneity in the impact of mentoring on teachers, and teachers who benefited the most from mentoring were most likely to

receive it, then the estimates from my specification may be considerably lower than the true, average effect of mentoring.¹⁸ The difference between the local and average effects of mentoring may be particularly relevant in this case, given that roughly 40 percent of the “control” teachers, i.e., new teachers with prior experience received mentoring.

6. Mentor Evaluations and the Impact of Mentoring Quality

As mentioned above, the bulk of my analysis focuses on how measures of mentoring quality affected teachers who received mentoring. Equation 4 shows the basic specification I use to estimate the impact of mentoring quality on teachers.

$$Y_{it} = \beta X_{it} + \lambda M_{it} + \phi S_i + \sum_r \alpha_s D_{it}^s + \sum_t \delta_t D_{it}^t + \varepsilon_{it} \quad (4)$$

Y_{it} is the outcome of interest for teacher i in year t . X_{it} and M_{it} are teacher and mentor characteristics, respectively, and S_i represents the three survey response variables. α_s and δ_t represent school and year fixed effects. Standard errors are adjusted for clustering within mentor.

There are a number of mentor characteristics that one might reasonably think would predict the quality of services that mentors provide to teachers. First, I look at mentor experience, both in the mentoring program itself, in the classroom, and in the DOE in general. Research on teaching and other occupations finds that work experience is an important determinant of productivity. As mentioned above, mentors varied considerably in their level of teaching and DOE experience. For example, the 10th and 90th percentiles of teaching experience for mentors were 21 and 6 years. Variation in program experience is more limited (i.e., some

¹⁸ To illustrate, suppose that there are two types of teachers, A and B, and only A’s receive a benefit to mentoring. Further, suppose that all A’s are mentored, but only type B’s without prior experience are mentored. The difference-in-differences estimates would show no benefits to mentoring, and these estimates would still be unbiased (i.e., the type B teachers without prior experience truly did receive no benefits from mentoring). However, it would be incorrect to conclude that mentoring was not beneficial to any of the teachers that received it.

mentors are in their second year on the job in the school year 2005-2006). However, I can also measure whether a mentor is working in a school for the second time, as distinct from general experience in the program. Likewise, I can examine whether teachers who did not return to work as mentors in the second year of the program were less effective than those which remained, and whether this depended on whether the mentor left voluntarily or was asked not to return.

Mentors also vary in the time since they last taught. If mentors who had been out of the classroom for a long time had lost skills or found it difficult to relate to new teachers, this may have stood in the way of their providing high quality mentoring. On the other hand, mentors may have gained valuable experience outside the classroom that helps them in their new position, e.g., working in administration, teacher coaching, or professional development.

The caseload assigned to the mentor may also affect the quality of mentoring services received by teachers. Mentors with larger caseloads and/or caseloads spread across many schools may have less time to work with each teacher, and may therefore deliver lower quality mentoring. It is clear that caseload must have a negative effect on mentor service quality at some point (e.g., a mentor could not reasonably work with 100 teachers spread over 50 different schools), but whether the variation that existed within the program affected teacher outcomes is an empirical question. Indeed, some increases in caseload may be beneficial to the extent that mentoring skills involve learning-by-doing or if there are positive spillovers from mentoring multiple teachers (e.g., incorporating successful practices across teachers, avoiding repetition of unsuccessful practices, etc.). I construct caseload measures on a teacher specific basis to reflect the average number of teachers assigned to mentors and the average number of schools the mentor was assigned during the months that teachers worked.¹⁹

¹⁹ For example, suppose that Mrs. A worked during the first 5 months of the year, Mrs. B and Mrs. C worked the second five months, Mrs. D worked all ten months, and all four were assigned the same mentor. The teacher

I test several hypotheses with regard to mentor-teacher matching. First and foremost, I test the hypothesis that teacher outcomes improve when the teacher's subject area falls within the mentor's area of expertise. Given that primary role of the mentor is to improve the instructional skills of the teacher, and the weight given to subject-based matching in the DOE program, this is an important aspect of the analysis. It is also quite plausible that mentoring relationships are more successful when mentors and teachers are more "alike." I therefore test whether teacher outcomes are related to mentor-teacher demographic matches on gender, ethnicity, and age. Unfortunately, I do not have other measures of similarity, such as personality, that might plausibly affect the mentor-teacher working relationship.

One might also believe that mentors who previously taught in the school in which they are mentoring may be more effective because they are familiar with the student populations or the way these schools function. I therefore examine whether teachers' outcomes are related to whether their mentor had previously worked in the same school, which occurs for 6 percent of teachers. I group this variable with the mentor-teacher match variables. I also include a control for whether a mentor was assigned to any school in which he/she used to teach. This helps me to distinguish any impact of selection (i.e., mentors who work in their previous school may be positively selected) from an effect due to school specific skills.

All of the above variables are measures of inputs, rather than direct indicators of mentor quality. Evaluations of mentors from the DOE survey provide a direct indicator of the quality of mentoring provided to teachers. Since it may be problematic to use a teacher's own evaluation to predict his/her outcomes (due to reverse causality), I also calculate a teacher-specific mentor

caseload variable for A, B, C, and D would be 2, 3, 3, and 2.5, respectively. This reflects the number of average number of teachers their mentor was assigned during the months they were working. Then suppose that A worked in school 1 and B, C and D worked in school 2. The schools caseload variable for A, B, C, and D would be 2, 1, 1, and 1.5, respectively. This reflects the number of average number of schools their mentor was assigned during the months they were working.

quality variable by averaging the index across all other teachers working with the same mentor but in a different school.²⁰ This provides an arguably exogenous measure of the quality of services a mentor provided. Note that, although no survey was administered in the school year 2004-2005, 75 percent of mentors in that year were evaluated in 2005-2006. If this measure of quality reflects persistent differences in mentoring services provided, then it is reasonable to use the 2005-2006 survey evaluations to proxy for mentor quality in the prior year.

Last, but not least, I examine survey responses to the three replicated SASS questions for comparison with Smith and Ingersoll (2004). These are dummy variables for whether teachers had common planning time with other teachers, whether they attended professional development workshops and conferences, and whether they had a reduced work load.²¹

6.1 Evaluations of Mentor Performance by Teachers

Table 5 presents results on the impact of mentor characteristics on a teacher's evaluation of the mentor's performance. For ease of interpretation, the dependent variable has been normalized to have mean zero and standard deviation of one. The sample is limited to the roughly 4,000 new teachers in the school year 2005-2006 that responded to the survey.²²

The results in Column 1 of Table 5 show a number of significant relationships between teachers' evaluations of mentors and the predictive variables. Teachers who received more hours

²⁰ By excluding the evaluations of other teachers working in the same school, I avoid correlations between this measure of mentor quality and any school level unobservable factors that would create a spurious positive relationship between mentor evaluations and teacher outcomes. In practice, there is a high correlation between measures created when excluding all teachers in the same school or excluding only a teacher's own rating ($\rho=.96$).

²¹ Of the teachers responding to these questions, 62, 92, and 29 percent, respectively, said that they had common planning time, attended professional development, and had a reduced work load. There is considerable variation in the responses to these questions among teachers working in the same school, allowing me to include these as explanatory variables in regressions that include school fixed effects.

²² For simplicity, I do not report coefficients on other controls for mentor and teacher characteristics, but a few of these coefficients may be of interest to the reader. Mentor evaluations were significantly lower among alternatively certified teachers (e.g., Teach for America, Teaching Fellows) relative to traditionally certified teachers, and lower among female teachers relative to male teachers. Evaluations were significantly higher among Science and Foreign Language teachers relative to elementary ("Common Branch") teachers, and higher among Black, Hispanic, and Asian teachers relative to White teachers.

of mentoring gave better evaluations of their mentors. On average, evaluations rose by 0.19 standard deviations when hours of mentoring increased by 10 (Column 1). Evaluations were not significantly related to the number of schools or teachers in the mentor's caseload. However, the sign on coefficient for the number of schools is negative, while the coefficient on the number of teachers is positive. In later specifications, these effects are statistically significant, and indicate that (conditional on other covariates) adding a teacher in a new school may lower the quality of services a mentor can provide, while adding a teacher in a school where the mentor already works may increase the quality of mentoring he/she provides. Of course, an alternative explanation is that the number of teachers in a mentors caseload is simply correlated with mentoring quality (e.g., because "better" mentors get bigger caseloads).

A mentor who was the same gender as a teacher received higher evaluations of 0.13 standard deviations on average. This was not true of mentors who previously taught in the school in which they were working. More interestingly, teachers did not give higher evaluations to mentors who matched their subject area or grade level, based on administrative records. This is surprising, given the significant degree to which program administrators attempted to match teachers based on these criteria.

Last, but not least, I find that none of the experience related measures (years since last classroom teaching experience, total teaching experience, total DOE experience, program experience, experience in the school) improved mentor evaluations. The lack of any impact of program experience is quite interesting. Individuals that had spent a year as mentors, perhaps even in the same school where they were currently working, were not rated any better than those who were doing it for the first time. This stands much in contrast to research on teaching

experience and student achievement, which finds that gains to the first year of experience are typically large in their impact on student achievement.

One issue with this finding is that matches based on administrative data may not have aligned with reality. Column 2 shows results from a regression where subject and level matches are based on teachers' survey responses instead of administrative data. Here, in stark contrast to the earlier finding, subject and grade level matches were very strongly related to evaluations: evaluations rose by 0.51 and 0.30 standard deviations on average with, respectively, a subject or grade level match. Thus, teachers' perceptions regarding whether their mentor had expertise relevant to their area was very important in determining their overall evaluation.

To investigate what is driving this divergence of results between self-reported and administrative match, I first examine whether the administrative data on a teacher's subject and level match with the teacher's own belief as to his/her subject and level. I find that they match up very well; Appendix Figure A1 shows that the vast majority of all survey responses agree with the administrative data.²³ Thus, the divergence must come from teachers' beliefs regarding the subject and level expertise of their mentors. I examine this by asking whether the administrative mentor-teacher match indicator has predictive power for teacher's beliefs regarding mentor-teacher match. In other words, are the administrative data good instruments for survey responses to the match questions? Appendix Figure A2 shows the fraction of teachers who claim they match their mentor on subject (level), grouped by whether the administrative data indicate a match. For most subjects, an administrative match raises the probability of a survey reported match by about 40 percentage points. However, in the general elementary

²³ The only subjects where agreement was not close to 100 percent were Common Branches and Early Childhood. This likely stems from some confusion among lower elementary school teachers regarding their official subject area. If I recode the data so that (administrative) Common Branch teachers are allowed to agree with a survey response of Early Childhood, and vice versa, then agreement rates are close to 100 percent.

subjects (i.e., “Common Branch” and “Early Childhood”), teachers with administrative matches were only slightly more likely to state that their mentor matched their subject.

With regard to level, the administrative match data have very little power to explain the survey responses. However, it is important to note the distinction between the survey question on a teacher’s school level (e.g., Elementary) and whether their mentor’s area of expertise matched their “grade level.” It is theoretically possible that the administrative data on school level could match their responses to the former question well but poorly match the latter. It may be that teachers’ responses to these questions are simply a general assessment of their mentor’s overall teaching expertise, and having a mentor who taught at a similar level is a poor proxy for this assessment. In any case, it is clear that the administrative data on level matches may be a poor proxy for the mentoring quality as it relates to grade level specific issues.

One potentially important source of endogeneity that could bias the results in Columns 1 and 2 is if mentors spent more hours with teachers who needed help and these teachers then claimed the mentor was more helpful. In Column 3, I instrument for hours of mentoring using the average hours of mentoring received by teachers in other schools assigned to the same mentor. I do not report results from the first stage due to space constraints, but this instrument has significant power to predict hours of mentoring for each teacher, with a coefficient of over 0.6 and a t-statistic above 10. The coefficient on mentoring hours in this specification is now 0.1 standard deviations, and is marginally significant. This provides support for the notion that additional hours of mentoring raises teachers’ evaluations of the quality of service, but that mentors do endogenously allocate some hours among teachers in their caseload.

Column 4 of Table 5 shows results when the average evaluation of mentors by other teachers is included as a covariate. Perhaps not surprisingly, evaluations by a given teacher

exhibit a strong positive relation with other teachers' evaluations of the same mentor. Teachers' evaluations rise by 0.2 standard deviations, on average, when other teachers' evaluations rise by one standard deviation. Notably, the inclusion of the evaluation variable further reduces the estimated coefficient on mentoring hours, which is now statistically insignificant. One potential explanation is that one of the main elements of mentor quality upon which teachers agree is the amount of mentoring hours the mentors provide.

In sum, several key indicators of mentoring quality are significantly related to teachers' opinions regarding how helpful their mentors have been to them. At the same time, a host of mentor characteristics that reflect previous work experience (in the DOE, in the classroom, full vs. part time, etc.) are not related to teachers' evaluations of mentor performance.²⁴

6.2 Mentoring Quality and Teacher Retention

The main goal of mentoring programs is to reduce turnover among new teachers. Among the teachers who received mentoring in the DOE program, 97 percent remained teaching until the end of the school year, 90 percent returned to teach in the DOE and 80 percent returned to the same school.²⁵ From the point of view of the DOE administrators, retention within the DOE was of paramount importance. However, from the point of view of a principal, retention in the school is likely of equal or greater concern. Furthermore, research on the returns to experience among teachers has not ruled out the possibility that some of this return is school specific. I therefore analyze all three retention outcomes, and, in addition, I examine teacher absences.

²⁴ In addition, these results do not appear to be heterogeneous across the teacher population. I find similar results when running these regressions separately by gender, ethnicity, type of certification, subject area, or whether the school is above or below the city median in the fraction of its students receiving free lunch.

²⁵ One shortcoming in my analysis, relative to other studies, is that I cannot observe teachers who leave the DOE but continue to teach in other school districts. Boyd et al. (2007) find that 30 percent of teachers who leave the DOE after their first year end up teaching elsewhere in New York State. I also cannot observe teachers who remain within the DOE in a non-teaching capacity. However, this kind of movement (e.g., to becoming an administrator) is highly unlikely for new teachers.

Table 6 presents results for teacher absences. I only observe the total number of absences for the year, and therefore restrict my attention to teachers who were working by September and completed the school year. The average days absent was 6.3 for these teachers. Column 1 displays results that include the full sample and do not use information from the teacher survey. Hours of mentoring are significantly negatively correlated with absences; a 10 hour increase in mentoring is associated with a 0.2 day decrease in absences. One difficulty in interpreting this result is that teachers who are often absent will miss appointments with their mentors, and thus causality may run the other way. In Column 2, hours of mentoring are instrumented using teachers in different schools with the same mentor. While the estimated coefficient is still negative, it is smaller in magnitude and too imprecise to reject the hypothesis that additional mentoring hours had no impact on absences.

There are several other significant findings. First, having a mentor who previously taught in the same school is associated with a decrease in absences of roughly 0.6 days.²⁶ However, absences are higher among teachers assigned to mentors whose caseload includes their previous schools. This indicates that it is something particular about the school match, rather than a general marker of mentoring quality, that is driving the result. In addition, teachers who are close in age to their mentor are less frequently absent (0.4 days), and teachers whose mentor has a relatively large caseload of teachers are absent 0.1 days less on average (conditional on hours of mentoring provided, the number of schools, etc.).

An additional result is noteworthy and will remain consistent throughout all of the remaining tables. Teacher absences (and other outcomes) were not significantly worse with mentors who were asked not to return in the second year of the program, relative to mentors who

²⁶ Note, one must combine the coefficient on school match (-.966) with the coefficient on whether a mentor was assigned to a school in which they previously taught (.369).

continued into the second year. Thus, directors of the program did not base their personnel decisions on aspects of mentor characteristics or actions that are unobservable in the data and significantly related to teacher outcomes conditional on the observable information. This suggests that it is unlikely that directors were able to assign the “better” mentors to caseloads with particular unobservable characteristics, addressing a potential concern on endogeneity.

In the remainder of Table 6, I present results that restrict the sample to survey respondents and include controls for mentor evaluations and three questions regarding other support for new teachers. A teacher’s own evaluation of the mentor is not significantly related to absences, but teachers who claim to have common planning time with others are less absent on average by 0.7 days (Column 3). Because the coefficient on a teacher’s own evaluation may be biased (e.g., mentors have difficulty helping teachers who are frequently absent), I use other teachers’ evaluations as an instrument, taking my cue from the results in Table 5 Column 4. This, however, causes the coefficient to grow in magnitude and become extremely imprecise.²⁷

All of the tables of results for retention outcomes follow the same sequence as for absences. The only exception is the results on whether a teacher completes the school year, shown in Table 7; the survey was given out in the spring of 2005-2006, and nearly every teacher who responded also completed the year. In Table 7, hours of mentoring appear positively related to completion in OLS but not in IV specifications. Here the reverse causality seems quite strong; mechanically, teachers who leave early get far fewer hours of mentoring. There is some evidence that having a mentor with a larger caseload or that previously taught in the same school reduce the probability that a teacher completes the school year. This is surprising, since both variables were associated with fewer absences among teachers who worked a full year.

²⁷ The instrument, however, is statistically significant, with an F-statistic of over 50.

Table 8 focuses on whether a teacher returned to teach in the DOE the following year. For ease of interpretation, I condition the sample of having completed the school year. Unconditional estimates are shown in Appendix Table A1. Here the impact of mentoring hours on retention is positive and marginally significant in OLS, and the IV coefficient is the same magnitude but less precisely estimated. Here again there is a significant negative relationship between the number of teachers in a caseload and retention. Notably, there is a positive and marginally significant relationship between retention to the next school year and having a mentor who matches (administratively) on subject area, and a negative significant relationship with matches based on gender. Moreover, the relationship between retention and subject match based on survey responses is statistically insignificant. These results suggest that the variables closely related to teacher evaluations of mentors may be different than those related to teacher retention.

Parallel to the result shown on absences, teachers who claim to have common planning time with other teachers are significantly more likely to return the following school year, while a teacher's own evaluation of his/her mentor's performance is unrelated to this decision (Table 8 Columns 3 and 4). This pattern is seen again when examining whether a teacher returns to the same school, conditional on returning to the DOE (Table 9 Columns 3 and 4). In addition, teachers who respond that they did not receive professional development are much less likely to return to the same school.²⁸

Several additional results from Table 9 are worth noting.²⁹ First, the mentor's previous experience in a school seems to play a significant role in a teacher's decision to return the following year. Both having taught and having previously worked as a mentor in a school have large positive effects on teachers' propensity to return. Again, these do not seem to be driven by

²⁸ While this result lends some support to the provision of professional development, is worth noting that 93 percent of teachers responding to the survey claimed that they received it.

²⁹ Unconditional estimates are shown in Appendix Table A2.

general quality of these mentors—the main effects of assignment to previous school and program experience are negative. In addition, if I include an indicator in the regression for whether the mentor is returning to mentor in any school for a second time, its coefficient is insignificant and the results are unaffected.

One possible interpretation of this finding is that mentors with experience in a school are able to provide school-specific information and guidance that is very useful to teachers. Another possible interpretation however is that mentors with experience in a school tend to spend much of their energy and attention on teachers in that school and to neglect the other teachers in their caseload. While this interpretation is less upbeat, it still would imply that mentor behavior can have a significant impact on teacher retention decisions, at least when it comes to choosing to return to a school.

Another notable finding in Table 9 is that the caseload variables now have statistically significant effects in the directions that one would have expected given their effects on survey evaluations and absences. More schools in a caseload are associated with a reduction in the propensity to return, while more teachers are associated with an increase in this propensity.

6.3 Mentoring Quality and Teacher's Impacts on Student Achievement

The final outcome of interest I examine is the impact of teachers on student achievement in math and reading. As mentioned above, I make use of matched student-teacher panel data on student test scores in grades 4 to 8, and the number of mentors in the sample is therefore considerably smaller. My analysis here is similar the analysis for retention, though the unit of observation is a student instead of a teacher. In addition to the teacher and mentor covariates, the regressions include year and grade fixed effects, controls for students' prior test scores, demographics and other factors (e.g. special education status), average student characteristics at

the classroom level, and class size. These factors are correlated with student achievement but, arguably, should not be attributed to teachers. For a fuller explanation of this data and methodology, see Kane et al. (2006).

In my analysis of absences and retention, I controlled for school-specific factors by including school fixed effects in the regression specifications. Using this specification for my analysis of student achievement is problematic in that there are far fewer instances in which more than one mentor works with teachers in the same school once the sample is limited to teachers who match with student test scores. In the full sample of mentored teachers, only 24 percent of schools have one mentor, while over 70 percent of schools have just one mentor in the student achievement sample. When I restrict the student achievement sample to the school year 2005-2006, 88 percent of schools have just one mentor. For this reason, I include school level covariates (i.e., school level means of student covariates) in lieu of school fixed effects. In the interest of space, I do not present estimates that include the actual evaluation of his/her mentor (corresponding to Column 3 in the earlier tables).

Table 10 presents results using both the full sample of teachers and those with survey data. The most interesting finding is that hours of mentoring received has a positive, statistically significant effect on student achievement when I instrument for hours (Columns 2, 3, 5, and 6). Notably, the estimated effect is considerably smaller (and insignificant for reading achievement) when actual hours are included (Columns 1 and 4 Table 5), consistent with the notion that mentors provide additional hours of service to teachers who need more help. The magnitude of these effects are substantial, with an additional ten hours of mentoring expected to raise student achievement by 0.05 standard deviations in math (0.10 in the survey sample) and 0.04 standard

deviations in reading (0.06 in the survey sample). If truly causal, these effects would lend considerable support for the notion that mentoring has an impact on student achievement.

The estimated effects of mentors' caseload are somewhat unexpected, with a positive effect for the number of schools and a negative effect for the number of teachers. These effects are statistically significant in some, but not all, specifications. In contrast to the results for retention, teachers receipt of additional support (e.g., common planning time) are not significantly related to student achievement. With regard to mentor characteristics, having a mentor with more DOE experience was negatively related to student achievement in both subjects, while having a mentor with one year of program experience is associated with a significant improvement in math achievement. In the survey sample, having a mentor who matched based on subject (according to the teacher) had a marginally significant, positive relationship with math achievement, while having a mentor close in age to the teacher was associated with a decrease in achievement.

7. Conclusion

In the analysis above, I find strong relationships between various measures of mentoring quality and teachers' evaluations of the impact of mentors on their success in the classroom. I find more limited evidence that these measures of mentoring quality had large impacts on teacher absences, teacher retention, and student achievement. My most consistent findings are that teachers whose mentor had prior experience working in their school were more likely to return to teaching in their schools, and that other types of support for new teachers (e.g., common planning time) help increase retention, consistent with Smith and Ingersoll (2004). I also do find that student achievement in both math and reading was higher with teachers who received more hours of mentoring. Notably, this positive association is obscured in OLS estimation, likely

because mentors spent more time with teachers who needed more help. It is also worth noting that I do not find much evidence that having a mentor whose area of subject matter expertise matches a teacher's subject is associated with better outcomes.

Some limitations of this study are worth noting. First, I use observational data to examine variation in measures of mentoring quality among teachers who received mentoring. While my data are quite detailed, and I have tried to be careful in my use of econometrics to estimate causal effects, the best way to address whether mentoring affected teachers would be to evaluate outcomes from an intervention where teachers were randomly assigned various types of mentoring, including no mentoring. Second, perhaps the most important outcome that mentoring is supposed to address is retention, but I can only observe teachers within New York City. Boyd et al. (2007) find that 30 percent of teachers who leave the New York City after their first year end up teaching elsewhere in New York State. From the city's point of view, I am examining the most important outcome, but from the point of the view of state—which did not pay for these services but required mentoring—it would be interesting to know which teachers went to work in other districts in New York.

Mentoring programs for new teachers have become a widespread policy tool for school districts across the country and are now required by many state governments. This study begins to fill the wide gap in our understanding of how these services affect career outcomes. The large sample of teachers and mentors in New York City and the detailed data I analyze allow me to estimate how mentors influence outcomes for new teachers while paying careful attention to the attribution of causal effects. While my results lend some support to the use of mentoring as a tool for increasing retention and skills of new teachers, the substantial resources that have been devoted to this policy provide strong motivation for additional research.

Bibliography

Aaronson, D., Barrow, L. & Sander, W. (2007) "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics*, 25(1): 95-135.

Ballou, D. (2001) "Pay for Performance in Public and Private Schools," *Economics of Education Review* 20(1): 51-61

Becker, G. S. (1964) Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. New York: Columbia University Press, 1964.

Bertrand, M., Duflo, E. & Mullainathan, S. (2004) "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119(1): 249-275.

Boyd, D., Grossman, P., Lankford, H., Loeb, S. & Wyckoff, J. (2007) "Teacher Attrition, Teacher Effectiveness and Student Achievement" Unpublished Manuscript.

Brewer, D.J. (1996) "Career Paths and Quit Decisions: Evidence from Teaching," *Journal of Labor Economics* 14(2): 313-339.

Brown, J. N. (1989) "Why Do Wages Increase with Tenure? On-the-Job Training and Life-Cycle Wage Growth Observed within Firms," *American Economic Review*, December 1989: 971-991.

Cameron, C., Gelbach, J.B., Miller, D.L. (2007) "Bootstrap-Based Improvements for Inference with Clustered Errors" Florida State University Law and Economics Research Paper Series, Paper No. 07/002

Clotfelter, C. T., Glennie, E., Ladd, H. F., & Vigdor, J. L. (2006) "Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina," Working Paper, Sanford Institute of Public Policy, Duke University

Clotfelter, C. T., Ladd, H. F. & Vigdor, J. L. (2007) "How and Why do Teacher Credentials Matter for Student Achievement?" NBER Working Paper 12828.

Conley, T. & Taber, C. (2005) "Inference with "Difference In Differences with a Small Number of Policy Changes" NBER Technical Working Paper 312.

Dee, T. (2004) "Teachers, Race and Student Achievement in Randomized Experiment." *The Review of Economics and Statistics*. 86(1): 195-210.

Dee, T. (2005) "A Teacher Like Me: Does Race, Ethnicity or Gender Matter?" *American Economic Review*. 95(2): 158-165.

- Dolton, P. & van der Klaauw, W. (1995) "Leaving Teaching in the UK: A Duration Analysis," *The Economic Journal*, 105(429): 431-444.
- Dolton, P. & van der Klaauw, W. (1999) "The Turnover of Teachers: A Competing Risks Explanation," *Review of Economics and Statistics* 81(3): 543-550.
- Donald, S.G. & Lang, K. (2007) "Inference with Difference-In-Differences and Other Panel Data," *Review of Economics and Statistics* 89(2): 221–233.
- Education Week (2000) "Quality Counts 2000: Who Should Teach?" *Education Week* 19 (18), <http://www.edweek.org/media/ew/qc/archives/QC00full.pdf>.
- Education Week (2003) "Quality Counts 2003: If I Can't Learn from You," *Education Week* 22 (17), <http://counts.edweek.org/sreports/qc03/index.cfm>.
- Education Week (2006) "Quality Counts at 10: A Decade of Standards-Based Education" *Education Week* 25 (17), <http://www.edweek.org/ew/toc/2006/01/05/index.html>.
- Figlio, D.N. & Kenny, L. (2006) "Individual Teacher Incentives and Student Performance," NBER Working Paper 12627.
- Fletcher, S., Strong, M. & Villar, A. (2005) *An Investigation of the Effects of Variations in Mentor-Based Induction on the Performance of Students in California*. Santa Cruz, CA: New Teacher Center.
- Hanushek, E.A., Kain, J.F., & Rivkin, S.G. (2004), "Why Public Schools Lose Teachers," *Journal of Human Resources*, 39 (2): 326-354.
- Harris, D.N. & Sass, T. (2006) "The Effects of Teacher Training on Teacher Value Added," Working Paper, Department of Economics, Florida State University.
- Harris, D.N. & Adams, S.J. (2007) "Understanding the Level and Causes of Teacher Turnover: A Comparison with Other Professions," *Economics of Education Review* 26(3): 325–337
- Ingersoll, R.M. & Kralik, J.M. (2004) "The Impact of Mentoring on Teacher Retention: What the Research Says," ECS Research Review.
- Kane, T. J., Rockoff, J. E. & Staiger, D. (2006) "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City," NBER Working Paper 12155.
- Lavy, V. (2002) "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement," *Journal of Political Economy* 110(6): 1286-1317.

Lopez, A., Lash, A., Schaffner, M., Shields, P., and Wagner, M. (2004) "Review of Research on the Impact of Beginning Teacher Induction on Teacher Quality and Retention" Menlo Park, CA: SRI International.

Murnane, R.J., (1981) "Teacher Mobility Revisited," *Journal of Human Resources*, 16(1): 3-19.

Murnane, R.J., Olsen, R.J. (1989) "The Effect of Salaries and Opportunity Costs on Duration in Teaching: Evidence from Michigan," *Review of Economics and Statistics*, 71(2): 347-352.

Murnane, R.J., Olsen, R.J. (1990) "The Effect of Salaries and Opportunity Costs on Length of Stay in Teaching: Evidence from North Carolina," *Journal of Human Resources* 25(1): 106-124.

Rivkin, S.G., Hanushek, E. A. & Kain, J.F. (2005) "Teachers, Schools, and Academic Achievement." *Econometrica* 73 (2): 417-458.

Rockoff, J. E. (2004) "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review*. 94(2): 247-252.

Serpell, Z. (2000) *Beginning Teacher Induction: A Review of the Literature*. Washington, DC: American Association of Colleges for Teacher Education.

Smith, T. M. & Ingersoll, R. M. (2004) "What are the Effects of Induction and Mentoring on Beginning Teacher Turnover?" *American Educational Research Journal*, 41(3), pp.681-714

Smith, T.M. (2007) "How Do State-Level Induction and Standards-Based Reform Policies Affect Induction Experiences and Turnover among New Teachers?" *American Journal of Education* 113(2): 273-309.

Stinebrickner, T.R. (1998) "An Empirical Investigation of Teacher Attrition," *Economics of Education Review* 17(2): 127-136.

Stinebrickner, T.R. (2001) "A Dynamic Model of Teacher Labor Supply," *Journal of Labor Economics* 19(1): 196-230.

Stinebrickner, T.R. (2002) "An Analysis of Occupational Change and Departure from the Labor Force: Evidence of the Reasons that Teachers Leave," *Journal of Human Resources* 37(1): 192-216.

Stinebrickner, T.R., Scafidi, B. & Sjoquist, D.L. (2008) "Do Teachers Really Leave for Higher Paying Jobs in Alternative Occupations?" Unpublished Manuscript.

Strong, M. (2005) *Induction, Mentoring and Teacher Retention: A Summary of the Research*. Santa Cruz, CA: ATE Commission on Mentoring and Teacher Induction and The New Teacher Center, University of California.

Strong M. (2006) Does New Teacher Support Affect Student Achievement: A Look at the Research. *Research Brief#06-01*, Santa Cruz, CA: New Teacher Center, University of California, Santa Cruz.

Topel, R. H. (1991) "Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority," *Journal of Political Economy* 99(1): 145-176.

Understanding New York City's Groundbreaking Induction Initiative. (2006) *NTC Policy Paper*. Santa Cruz, CA: New Teacher Center, University of California, Santa Cruz.

Villar, A. & Strong, M. (2007) "Is Mentoring Worth the Money? A Benefit-Cost Analysis and Five-year Rate of Return of a Comprehensive Mentoring Program for Beginning Teachers." Unpublished Manuscript.

Appendix: Analysis of Teacher-Mentor Assignment

I use Monte Carlo simulations to examine whether the assignment of teachers to mentors was systematically related to observable characteristics of these individuals. In particular, I ask whether the rates with which mentors and teachers have matching characteristics (e.g., same subject area, same gender) are greater than one would expect if assignments were random. Random assignment is based on the actual distribution of mentor characteristics within region and year, since this is the level of aggregation at which actual assignment was done. (For example, if 50 percent of mentors in Region 1 during the school year 2004-2005 were female, then I assign female gender to mentors in that region and year with 50 percent probability.) After random assignment, I calculate the percentage of teachers that match their mentors' characteristics, and I repeat the exercise 1000 times.

Figure A3 displays the results of these simulations as well as the actual rate of matching for subject area, race, and gender. The solid lines in the figures show the actual match rate in the data, and the dotted lines show the 95th percentile match rate from the random assignment simulations. Note that match rates vary systematically across groups of teachers because of group size. For example, there are relatively few Asian mentors or Asian teachers, so the likelihood that an Asian teacher has a mentor of the same ethnicity will be small unless ethnicity plays a very important role in assignment.

For simplicity, Figure A3 shows the results for the six most common teacher subjects: “Common Branch” (i.e., the general term in the DOE for elementary teachers), English, English as a Second Language (ESL), Math, Science, and Special Education. Match rates in subject area are considerably higher than one would expect from random assignment. For example, the average random assignment match rate for Science teachers is less than 10 percent, the 95th percentile is less than 15 percent, but the actual match rate was over 50 percent. This strongly supports the notion that mentors were matched based on their subject areas of expertise.

Match rates on ethnicity and gender were also higher than one would expect from random assignment within region and year, though to a much less dramatic extent than for subject area. For Black and Hispanic teachers, the 95th percentile simulation match rates are 34 and 12 percent, respectively, while the actual match rates were 37 and 16 percent. For males and females, the 95th percentile simulations were 24 and 81 percent, respectively, and the actual match rates were 25 and 83 percent. Though actual match rates on demographics exceed expected rates under random assignment, it is not clear that regional directors assigned mentors based on ethnicity and gender, since demographic characteristics may have been correlated with other factors that affected mentor assignment, such as subject area or experience working with particular student populations. For example, male teachers and male mentors are both less likely to work in Common Branch (general elementary) classes than females.

To investigate this issue further, I repeat the Monte Carlo simulations at the school level instead of the region-year level. In other words, I take assignments of mentors to schools as given, and then ask whether actual assignment of teachers to mentors within the school resemble random assignment. Because of the large number of schools (over 1000), I examine schools rather than school-year cells, and drop all schools with only one mentor, which, if anything, will bias the results towards finding differences between simulated and actual match rates. Results of these simulations are shown in Figure A4. While the actual match rates on subject are clearly higher than the within-school random simulations, nearly all of the actual match rates for demographics are under the 95th percentile simulation match rate. The sole exception is match rates for Hispanic teachers with Hispanic mentors, although the actual match rate (16 percent) is very similar to the average random match rate (13 percent). I conclude from this analysis that it is highly unlikely that regional mentors explicitly considered matching mentors and teachers on demographic characteristics.

Table 1: Summary Statistics on Newly Hired Teachers, 2000-2001 to 2005-2006

| <i>Year of Hire:</i> | Inexperienced Teachers | | | Experienced Teachers | | |
|-----------------------|------------------------|------------|-----------|----------------------|------------|-----------|
| | '01 & '02 | '03 & '04 | '05 & '06 | '01 & '02 | '03 & '04 | '05 & '06 |
| <i>N</i> | 12124 | 11013 | 9931 | 2556 | 3454 | 4189 |
| Received Mentoring | <i>n/a</i> | <i>n/a</i> | 80.64% | <i>n/a</i> | <i>n/a</i> | 40.37% |
| Absences (9 months) | 6.33 | 6.32 | 6.29 | 7.01 | 6.67 | 6.21 |
| Completed Year 1 | 89.6% | 91.8% | 94.1% | 98.6% | 97.6% | 95.2% |
| Returned in Year 2 | 78.8% | 83.8% | 88.1% | 91.2% | 89.6% | 86.8% |
| Same School in Year 2 | 65.2% | 71.9% | 84.6% | 77.7% | 74.2% | 79.1% |
| <i>Certification</i> | | | | | | |
| Traditional | 27.9% | 43.1% | 49.0% | 44.4% | 58.2% | 75.2% |
| Teaching Fellow | 10.6% | 34.0% | 34.6% | 2.7% | 9.4% | 8.7% |
| Teach for America | 1.7% | 4.4% | 7.8% | 0.3% | 0.3% | 0.1% |
| International Program | 1.3% | 2.3% | 0.9% | 20.9% | 21.6% | 8.0% |
| No Certification | 57.4% | 14.9% | 6.3% | 30.5% | 7.8% | 6.0% |
| Female | 70.5% | 71.6% | 73.0% | 77.1% | 71.5% | 76.0% |
| <i>Ethnicity/Race</i> | | | | | | |
| White | 55.3% | 67.2% | 70.0% | 54.3% | 62.1% | 66.9% |
| Black | 24.4% | 15.3% | 11.9% | 32.3% | 18.9% | 11.1% |
| Hispanic | 15.0% | 10.7% | 10.3% | 8.1% | 8.3% | 7.4% |
| Asian | 5.0% | 6.3% | 7.0% | 5.1% | 10.4% | 13.1% |
| Native American | 0.2% | 0.2% | 0.2% | 0.2% | 0.1% | 0.2% |
| Other/Missing | 0.3% | 0.3% | 0.7% | 0.0% | 0.3% | 1.4% |
| Has Value Added Data | 25.7% | 25.2% | 18.4% | 28.8% | 22.7% | 14.9% |

Note: Statistics based on payroll data covering all teachers in the DOE. See text for details.

Table 2: Summary Statistics, Mentored Teachers, by Program Year

| | 2004-2005 | | 2005-2006 | |
|---------------------------------------------|------------|------------|-----------|------|
| | 4774 | | 5626 | |
| | Mean | SD | Mean | SD |
| Number of Teachers | | | | |
| Hours of Mentoring Received | 32.0 | 16.7 | 39.1 | 14.1 |
| Hours Mentor Spent with Other Teachers | 32.2 | 13.1 | 40.9 | 10.2 |
| Teachers in Mentor's Caseload | 15.5 | 2.8 | 15.8 | 2.6 |
| Schools in Mentor's Caseload | 6.4 | 2.5 | 6.5 | 2.6 |
| Interacted with Only One Mentor | <i>n/a</i> | | 93% | |
| 80+% Hours of Mentoring from One Mentor | <i>n/a</i> | | 97% | |
| Evaluation of Mentor | | | | |
| Overall Rating of Mentor | <i>n/a</i> | <i>n/a</i> | 4.0 | 1.2 |
| Amount of Success Attributed to Mentor | <i>n/a</i> | <i>n/a</i> | 3.5 | 1.1 |
| Avg. Evaluation of Mentor by Other Teachers | | | | |
| Overall Rating of Mentor | <i>n/a</i> | <i>n/a</i> | 4.0 | 0.6 |
| Amount of Success Attributed to Mentor | <i>n/a</i> | <i>n/a</i> | 3.5 | 0.6 |
| Mentor Matches with Teacher Characteristic | | | | |
| Age (within 5 years) | 9.4% | | 10.1% | |
| Gender | 68.2% | | 68.4% | |
| Ethnicity | 49.1% | | 47.6% | |
| Subject Area | 46.4% | | 48.0% | |
| Grade Level | 37.8% | | 34.7% | |
| Subject Area (Survey Response) | <i>n/a</i> | | 77.2% | |
| Grade Level (Survey Response) | <i>n/a</i> | | 84.9% | |
| Received Reduced Workload (Survey) | <i>n/a</i> | | 28.9% | |
| Received Professional Development (Survey) | <i>n/a</i> | | 92.5% | |
| Received Common Planning Time (Survey) | <i>n/a</i> | | 61.8% | |

Notes: Evaluations of mentors by teachers were given a scale from 1-5; for more details see the text. The "average evaluation of mentor by other teachers," uses data from teachers in other school-year cells who were assigned the same mentor; the "average hours mentor spent with other teachers," uses data from teachers in other schools who were assigned the same mentor during the same year and worked an entire school year.

Table 3: Summary Statistics, Mentors, by Program Year

| | 2004-2005 | 2005-2006 |
|--------------------------------------------|-----------------|-----------------|
| Number of Mentors | 348 | 425 |
| Demographics | | |
| Female | 80.7% | 81.6% |
| White | 60.9% | 57.4% |
| Black | 26.4% | 26.8% |
| Hispanic | 8.6% | 9.4% |
| Asian | 3.7% | 3.3% |
| Native American | 0.3% | 0.2% |
| Missing Ethnicity Info | 0.0% | 2.8% |
| Age | 48.2 [9.27] | 47.1 [9.75] |
| Years of DOE Experience | 16.67 [9.39] | 15.79 [9.27] |
| Years of Teaching Experience | 13.61 [5.70] | 12.62 [5.76] |
| Years Since Last Teaching Experience | 3.68 [4.12] | 4.23 [3.99] |
| Working Part Time | 9.2% | 8.2% |
| Assigned to School Where Previously Taught | 25.0% | 20.0% |
| Returned in Following Year | 73.6% | <i>n/a</i> |
| Has Program Experience | <i>n/a</i> | 60.5% |
| Has Program Experience in School | <i>n/a</i> | 24.0% |

Note: Standard deviations of continuous variables are shown in brackets.

Table 4: Difference-in-Differences Estimates of the Impact of the Mentoring Program

| | Received Mentoring | Total Absences | Stayed Through End of Year | Stayed in NYC DOE | Stayed in Same School | Math Achievement | Reading Achievement |
|----------------------------------------|-----------------------|-------------------|-------------------------------|----------------------|--------------------------|---------------------|------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| No Prior Experience * Post-Program | 0.403 (0.071)** | 0.127 (0.513) | 0.045 (0.022)+ | 0.043 (0.031) | 0.036 (0.037) | -0.012 (0.034) | 0.020 (0.026) |
| No Prior Experience | -0.062 (0.038) | 0.017 (0.288) | -0.105 (0.012)** | -0.111 (0.018)** | -0.109 (0.021)** | -0.049 (0.019)* | -0.019 (0.014) |
| Post-Program | 0.401 (0.053)** | -0.019 (0.374) | -0.027 (0.016) | -0.014 (0.023) | 0.058 (0.027)+ | -0.037 (0.024) | 0.007 (0.018) |
| Year | 0.003 (0.011) | -0.012 (0.085) | -0.002 (0.003) | -0.010 (0.005)+ | -0.020 (0.006)* | -0.006 (0.005) | -0.010 (0.004)* |
| No Prior Experience * Year | -0.007 (0.015) | -0.100 (0.116) | 0.007 (0.005) | 0.014 (0.007)+ | 0.020 (0.008)* | 0.010 (0.008) | 0.001 (0.006) |
| Observations | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| Only "On-time" Hires Completing Year 1 | N | Y | N | N | N | Y | Y |
| Only Math/Reading Teachers Grades 4-8 | N | N | N | N | N | Y | Y |

Note: The dependent variables are coefficient estimates of interactions between year fixed effects and an indicator for having no prior experience, taken from teacher or student level regressions that include teacher or both teacher and student level control variables and school fixed effects.

Table 5: Predictors of Teacher Evaluations of Mentor Performance

| | (1) | (2) | (3) | (4) |
|------------------------------------------------------------------------------------|--------------------|--------------------|--------------------|--------------------|
| <i>Mentoring Quality Indicators</i> | | | | |
| Hours of Mentoring Received /10 <i>Instrumented in (3) and (4)</i> | 0.186 (0.021)** | 0.139 (0.019)** | 0.096 (0.050)+ | 0.038 (0.047) |
| Avg. # Schools in Mentor's Caseload <i>~N(0,1)</i> | -0.045 (0.037) | -0.068 (0.033)* | -0.071 (0.033)* | -0.070 (0.030)* |
| Avg. # Teachers in Mentor's Caseload <i>~N(0,1)</i> | 0.053 (0.041) | 0.053 (0.038) | 0.046 (0.038) | 0.040 (0.034) |
| Average Evaluation by Other Teachers <i>~N(0,1), Same Mentor, Other Schools</i> | | | | 0.200 (0.028)** |
| <i>Mentor and Teacher Matching</i> | | | | |
| Age | 0.024 (0.075) | 0.021 (0.069) | 0.017 (0.069) | 0.005 (0.069) |
| Gender | 0.131 (0.052)* | 0.089 (0.048)+ | 0.090 (0.048)+ | 0.086 (0.047)+ |
| Ethnicity | 0.049 (0.055) | 0.009 (0.052) | 0.015 (0.053) | 0.022 (0.052) |
| School <i>(i.e., Mentor Taught in School)</i> | -0.077 (0.156) | -0.035 (0.134) | -0.011 (0.140) | 0.080 (0.135) |
| Subject (<i>Administrative Data</i>) | 0.039 (0.053) | | | |
| Level (<i>Administrative Data</i>) | -0.000 (0.081) | | | |
| Subject (<i>Survey Response</i>) | | 0.511 (0.052)** | 0.522 (0.052)** | 0.528 (0.052)** |
| Level (<i>Survey Response</i>) | | 0.296 (0.066)** | 0.297 (0.066)** | 0.287 (0.066)** |
| Observations | 4009 | 4009 | 4009 | 4009 |
| R-squared | 0.48 | 0.56 | <i>n/a</i> | <i>n/a</i> |

Note: Dependent variable is the average response to two survey questions asking teachers to evaluate the performance of their mentor (see text for exact wording). Each response was normalized before they were averaged and then normalized again. Variables marked with $\sim N(0,1)$ below them have also been normalized. Regression also includes controls for a teacher's subject area, level, age, age squared, ethnicity, gender, prior experience, whether they were hired during the school year, month of hire if hired during the school year, and type of certification or alternative recruitment program. It also includes controls for mentors' subject area, level, age, ethnicity, and gender, and for school fixed effects. In columns (3) and (4), the hours of mentoring a teacher received is instrumented with the average hours received by teachers in other schools assigned to the same mentor. Standard errors (in parentheses) are clustered by mentor. + significant at 10%; * significant at 5%; ** significant at 1%

Table 5: Predictors of Teacher Evaluations of Mentor Performance (cont'd)

| | (1) | (2) | (3) | (4) |
|--------------------------------------------|-------------------|-------------------|-------------------|--------------------|
| <i>Mentor Characteristics:</i> | | | | |
| Assigned to School Where Previously Taught | -0.143 (0.093) | -0.121 (0.087) | -0.118 (0.086) | -0.148 (0.078)+ |
| Experience in New York City DOE | 0.004 (0.006) | 0.001 (0.005) | 0.001 (0.005) | 0.002 (0.004) |
| 1 Year of Program Experience in School | -0.077 (0.088) | -0.040 (0.079) | -0.040 (0.078) | -0.025 (0.078) |
| 1 Year of Program Experience | 0.054 (0.073) | 0.010 (0.062) | 0.012 (0.063) | 0.010 (0.053) |
| Years Since Taught in Classroom | -0.002 (0.009) | -0.002 (0.008) | -0.003 (0.008) | -0.000 (0.007) |
| Observations | 4009 | 4009 | 4009 | 4009 |
| R-squared | 0.48 | 0.56 | <i>n/a</i> | <i>n/a</i> |

Note: Dependent variable is the average response to two survey questions asking teachers to evaluate the performance of their mentor (see text for exact wording). Each response was normalized before they were averaged and then normalized again. Variables marked with $\sim N(0,1)$ below them have also been normalized. Regression also includes controls for a teacher's subject area, level, age, age squared, ethnicity, gender, prior experience, whether they were hired during the school year, month of hire if hired during the school year, and type of certification or alternative recruitment program. It also includes controls for mentors' subject area, level, age, ethnicity, and gender, and for school fixed effects. In columns (3) and (4), the hours of mentoring a teacher received is instrumented with the average hours received by teachers in other schools assigned to the same mentor. Standard errors (in parentheses) are clustered by mentor. + significant at 10%; * significant at 5%; ** significant at 1%

Table 6: Estimated Effects of Mentoring Quality on Teacher Absences

| | (1) | (2) | (3) | (4) |
|------------------------------------------------------------------------------------|---------------------|--------------------|---------------------|---------------------|
| Mentoring Quality Indicators | OLS | IV | IV | IV |
| Hours of Mentoring the Teacher Received / 10 <i>Instrumented in (2) (3) (4)</i> | -0.226 (0.047)** | -0.079 (0.093) | -0.106 (0.215) | -0.121 (0.207) |
| Avg. # Schools in Mentor's Caseload <i>~N(0,1)</i> | -0.059 (0.068) | -0.056 (0.069) | -0.000 (0.157) | 0.017 (0.166) |
| Avg. # Teachers in Mentor's Caseload <i>~N(0,1)</i> | -0.090 (0.050)+ | -0.092 (0.050)+ | -0.153 (0.166) | -0.172 (0.168) |
| Teacher's Own Evaluation of Mentor <i>~N(0,1), Instrumented in (4)</i> | | | 0.174 (0.132) | 0.455 (0.634) |
| SASS Support Questions | | | | |
| Reduced Work Load | | | -0.104 (0.215) | -0.189 (0.288) |
| Prof. Development, Workshops, Conferences | | | -0.013 (0.387) | -0.067 (0.405) |
| Common Planning w/Other Teachers | | | -0.717 (0.207)** | -0.721 (0.207)** |
| Mentor and Teacher Matching | | | | |
| Subject Area <i>(Admin. in (1) (2), Survey in (3) (4))</i> | 0.069 (0.145) | 0.041 (0.146) | -0.253 (0.251) | -0.400 (0.406) |
| Grade Level <i>(Admin. in (1) (2), Survey in (3) (4))</i> | -0.227 (0.232) | -0.234 (0.233) | -0.143 (0.232) | -0.240 (0.335) |
| School <i>(i.e., Mentor Previously Taught in School)</i> | -0.966 (0.394)* | -0.997 (0.397)* | -1.557 (0.732)* | -1.528 (0.742)* |
| Age <i>(Within 5 Years)</i> | -0.447 (0.223)* | -0.431 (0.222)+ | -0.242 (0.348) | -0.255 (0.352) |
| Gender | 0.018 (0.128) | 0.006 (0.128) | 0.065 (0.252) | 0.050 (0.257) |
| Ethnicity | 0.060 (0.159) | 0.045 (0.159) | -0.386 (0.251) | -0.393 (0.255) |
| Mentor Characteristics: | | | | |
| Assigned to School Where Previously Taught | 0.369 (0.187)* | 0.324 (0.188)+ | 0.127 (0.352) | 0.136 (0.351) |
| Experience in New York City DOE | -0.005 (0.011) | -0.006 (0.011) | 0.006 (0.025) | 0.006 (0.025) |
| 1 Year of Program Experience in School | -0.075 (0.222) | -0.069 (0.223) | 0.004 (0.335) | 0.014 (0.341) |
| 1 Year of Program Experience | 0.269 (0.200) | 0.268 (0.200) | 0.363 (0.285) | 0.362 (0.284) |
| Mentor Voluntarily Not Returning <i>(2004-2005 Only)</i> | 0.132 (0.248) | 0.165 (0.249) | | |
| Mentor Not Invited to Return <i>(2004-2005 Only)</i> | -0.244 (0.349) | -0.199 (0.355) | | |
| Years Since Taught in Classroom | -0.004 (0.016) | 0.002 (0.016) | | |
| Observations | 8399 | 8399 | 3,692 | 3,692 |
| R-squared | 0.32 | n/a | n/a | n/a |

Note: Variables marked with $\sim N(0,1)$ have been normalized. For details on additional controls see text. Standard errors (in parentheses) are clustered by mentor. Hours of mentoring received is instrumented with the average hours of mentoring received by teachers outside of the same school but with the same mentor among teachers who worked a full year. Survey evaluations are instrumented with the average evaluation given by teachers outside of the same school but with the same mentor.

Table 7: Estimated Effects of Mentoring Quality on Completing Year 1

| | (1) | (2) |
|---------------------------------------------------|-----------|------------|
| | OLS | IV |
| Mentoring Quality Indicators | | |
| Hours of Mentoring the Teacher Received / 10 | 0.039 | -0.003 |
| <i>Instrumented in (2)</i> | (0.003)** | (0.003) |
| Avg. # Schools in Mentor's Caseload | 0.003 | 0.000 |
| <i>~N(0,1)</i> | (0.002) | (0.003) |
| Avg. # Teachers in Mentor's Caseload | -0.006 | -0.008 |
| <i>~N(0,1)</i> | (0.002)** | (0.003)** |
| Mentor and Teacher Matching | | |
| Subject Area | -0.007 | 0.002 |
| <i>(Administrative)</i> | (0.005) | (0.005) |
| Grade Level | -0.005 | -0.004 |
| <i>(Administrative)</i> | (0.008) | (0.008) |
| School | -0.032 | -0.029 |
| <i>(i.e., Mentor Previously Taught in School)</i> | (0.015)* | (0.014)* |
| Age | 0.004 | 0.001 |
| <i>(Within 5 Years)</i> | (0.008) | (0.008) |
| Gender | -0.006 | -0.004 |
| | (0.005) | (0.006) |
| Ethnicity | -0.009 | -0.002 |
| | (0.005)+ | (0.005) |
| Mentor Characteristics: | | |
| Assigned to School Where Previously Taught | -0.010 | 0.001 |
| | (0.007) | (0.006) |
| Experience in New York City DOE | -0.001 | -0.001 |
| | (0.000) | (0.000)+ |
| 1 Year of Program Experience in School | 0.000 | 0.001 |
| | (0.007) | (0.007) |
| 1 Year of Program Experience | -0.006 | -0.006 |
| | (0.008) | (0.007) |
| Mentor Voluntarily Not Returning | 0.007 | -0.006 |
| <i>(2004-2005 Only)</i> | (0.010) | (0.008) |
| Mentor Not Invited to Return | 0.006 | -0.007 |
| <i>(2004-2005 Only)</i> | (0.011) | (0.009) |
| Years Since Taught in Classroom | 0.002 | 0.001 |
| | (0.001)** | (0.001) |
| Observations | 9924 | 9924 |
| R-squared | 0.24 | <i>n/a</i> |

Note: Variables marked with $\sim N(0,1)$ have been normalized. For details on additional controls see text. Standard errors (in parentheses) are clustered by mentor. Hours of mentoring received is instrumented with the average hours of mentoring received by teachers outside of the same school but with the same mentor among teachers who worked a full year.

Table 8: Estimated Effects of Mentoring Quality on Retention to Year 2

| | (1) | (2) | (3) | (4) |
|---------------------------------------------------|----------|------------|------------|------------|
| Mentoring Quality Indicators | OLS | IV | IV | IV |
| Hours of Mentoring the Teacher Received / 10 | 0.005 | 0.005 | 0.007 | 0.009 |
| <i>Instrumented in (2) (3) (4)</i> | (0.002)+ | (0.004) | (0.012) | (0.012) |
| Avg. # Schools in Mentor's Caseload | 0.003 | 0.005 | 0.011 | 0.010 |
| <i>~N(0,1)</i> | (0.004) | (0.004) | (0.010) | (0.010) |
| Avg. # Teachers in Mentor's Caseload | -0.006 | -0.006 | -0.013 | -0.013 |
| <i>~N(0,1)</i> | (0.003)* | (0.004) | (0.013) | (0.013) |
| Teacher's Own Evaluation of Mentor | | | -0.000 | -0.026 |
| <i>~N(0,1), Instrumented in (4)</i> | | | (0.008) | (0.039) |
| SASS Support Questions | | | | |
| Reduced Work Load | | | -0.005 | 0.003 |
| | | | (0.013) | (0.017) |
| Prof. Development, Workshops, Conferences | | | 0.013 | 0.018 |
| | | | (0.025) | (0.027) |
| Common Planning w/Other Teachers | | | 0.029 | 0.030 |
| | | | (0.013)* | (0.013)* |
| Mentor and Teacher Matching | | | | |
| Subject Area | 0.012 | 0.012 | -0.009 | 0.004 |
| <i>(Admin. in (1) (2), Survey in (3) (4))</i> | (0.007)+ | (0.007)+ | (0.014) | (0.025) |
| Grade Level | -0.003 | -0.003 | 0.007 | 0.016 |
| <i>(Admin. in (1) (2), Survey in (3) (4))</i> | (0.013) | (0.013) | (0.013) | (0.019) |
| School | 0.028 | 0.028 | 0.099 | 0.098 |
| <i>(i.e., Mentor Previously Taught in School)</i> | (0.018) | (0.018) | (0.051)+ | (0.051)+ |
| Age | 0.015 | 0.015 | -0.002 | -0.002 |
| <i>(Within 5 Years)</i> | (0.011) | (0.011) | (0.021) | (0.021) |
| Gender | -0.015 | -0.015 | -0.005 | -0.003 |
| | (0.008)* | (0.008)* | (0.014) | (0.015) |
| Ethnicity | 0.003 | 0.002 | -0.002 | -0.001 |
| | (0.008) | (0.008) | (0.016) | (0.017) |
| Mentor Characteristics: | | | | |
| Assigned to School Where Previously Taught | 0.004 | 0.004 | -0.008 | -0.011 |
| | (0.009) | (0.009) | (0.022) | (0.024) |
| Experience in New York City DOE | -0.001 | -0.001 | 0.001 | 0.001 |
| | (0.001)+ | (0.001) | (0.002) | (0.002) |
| 1 Year of Program Experience in School | -0.001 | -0.000 | -0.002 | -0.002 |
| | (0.011) | (0.011) | (0.018) | (0.018) |
| 1 Year of Program Experience | 0.007 | 0.006 | 0.011 | 0.012 |
| | (0.010) | (0.010) | (0.017) | (0.017) |
| Mentor Voluntarily Not Returning | -0.020 | -0.020 | | |
| <i>(2004-2005 Only)</i> | (0.015) | (0.015) | | |
| Mentor Not Invited to Return | -0.014 | -0.013 | | |
| <i>(2004-2005 Only)</i> | (0.017) | (0.017) | | |
| Years Since Taught in Classroom | 0.001 | 0.001 | | |
| | (0.001) | (0.001) | | |
| Observations | 9,637 | 9,637 | 3,924 | 3,924 |
| R-squared | 0.17 | <i>n/a</i> | <i>n/a</i> | <i>n/a</i> |

Note: Variables marked with $\sim N(0,1)$ have been normalized. For details on additional controls see text. Standard errors (in parentheses) are clustered by mentor. Hours of mentoring received is instrumented with the average hours of mentoring received by teachers outside of the same school but with the same mentor among teachers who worked a full year. Survey evaluations are instrumented with the average evaluation given by teachers outside of the same school but with the same mentor.

Table 9: Estimated Effects of Mentoring Quality on Returning to School in Year 2

| | (1) | (2) | (3) | (4) |
|---------------------------------------------------|-----------|------------|------------|------------|
| Mentoring Quality Indicators | OLS | IV | IV | IV |
| Hours of Mentoring the Teacher Received / 10 | -0.002 | -0.002 | -0.051 | -0.051 |
| <i>Instrumented in (2) (3) (4)</i> | (0.003) | (0.008) | (0.022)* | (0.023)* |
| Avg. # Schools in Mentor's Caseload | -0.047 | -0.049 | -0.055 | -0.055 |
| <i>~N(0,1)</i> | (0.006)** | (0.007)** | (0.019)** | (0.020)** |
| Avg. # Teachers in Mentor's Caseload | 0.017 | 0.006 | -0.011 | -0.013 |
| <i>~N(0,1)</i> | (0.005)** | (0.006) | (0.024) | (0.024) |
| Teacher's Own Evaluation of Mentor | | | 0.007 | 0.002 |
| <i>~N(0,1), Instrumented in (4)</i> | | | (0.009) | (0.086) |
| SASS Support Questions | | | | |
| Reduced Work Load | | | 0.005 | 0.006 |
| | | | (0.013) | (0.030) |
| Prof. Development, Workshops, Conferences | | | 0.048 | 0.049 |
| | | | (0.023)* | (0.027)+ |
| Common Planning w/Other Teachers | | | 0.022 | 0.022 |
| | | | (0.013)+ | (0.013)+ |
| Mentor and Teacher Matching | | | | |
| Subject Area | 0.005 | 0.005 | 0.019 | 0.022 |
| <i>(Admin. in (1) (2), Survey in (3) (4))</i> | (0.010) | (0.010) | (0.016) | (0.048) |
| Grade Level | 0.006 | 0.004 | 0.007 | 0.009 |
| <i>(Admin. in (1) (2), Survey in (3) (4))</i> | (0.017) | (0.017) | (0.014) | (0.032) |
| School | 0.108 | 0.103 | 0.075 | 0.075 |
| <i>(i.e., Mentor Previously Taught in School)</i> | (0.037)** | (0.037)** | (0.088) | (0.088) |
| Age | -0.002 | -0.004 | -0.010 | -0.011 |
| <i>(Within 5 Years)</i> | (0.012) | (0.013) | (0.024) | (0.024) |
| Gender | 0.008 | 0.008 | -0.013 | -0.013 |
| | (0.010) | (0.010) | (0.019) | (0.019) |
| Ethnicity | -0.005 | -0.005 | -0.018 | -0.018 |
| | (0.009) | (0.009) | (0.017) | (0.017) |
| Mentor Characteristics: | | | | |
| Assigned to School Where Previously Taught | -0.058 | -0.057 | -0.037 | -0.038 |
| | (0.017)** | (0.017)** | (0.033) | (0.036) |
| Experience in New York City DOE | -0.001 | -0.001 | 0.001 | 0.001 |
| | (0.001) | (0.001) | (0.002) | (0.003) |
| 1 Year of Program Experience in School | 0.128 | 0.127 | 0.217 | 0.217 |
| | (0.018)** | (0.018)** | (0.035)** | (0.035)** |
| 1 Year of Program Experience | -0.052 | -0.049 | -0.054 | -0.053 |
| | (0.018)** | (0.018)** | (0.033) | (0.034) |
| Mentor Voluntarily Not Returning | 0.011 | 0.009 | | |
| <i>(2004-2005 Only)</i> | (0.022) | (0.023) | | |
| Mentor Not Invited to Return | -0.031 | -0.036 | | |
| <i>(2004-2005 Only)</i> | (0.033) | (0.035) | | |
| Years Since Taught in Classroom | 0.003 | 0.003 | | |
| | (0.001)+ | (0.002)+ | | |
| Observations | 8,977 | 8,977 | 3,663 | 3,663 |
| R-squared | 0.30 | <i>n/a</i> | <i>n/a</i> | <i>n/a</i> |

Note: Variables marked with $\sim N(0,1)$ have been normalized. For details on additional controls see text. Standard errors (in parentheses) are clustered by mentor. Hours of mentoring received is instrumented with the average hours of mentoring received by teachers outside of the same school but with the same mentor among teachers who worked a full year. Survey evaluations are instrumented with the average evaluation given by teachers outside of the same school but with the same mentor.

Table 10: Estimated Effects of Mentoring Quality on Student Achievement

| | Math | | | Reading | | |
|----------------------------------------------------------------------------------------|--------------------|--------------------|---------------------|-------------------|--------------------|--------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Mentoring Quality Indicators | OLS | IV | IV | OLS | IV | IV |
| Hours of Mentoring the Teacher Received / 10 <i>Instrumented in (2) (3) (5) (6)</i> | 0.013 (0.004)** | 0.046 (0.015)** | 0.100 (0.021)** | 0.004 (0.003) | 0.037 (0.011)** | 0.061 (0.015)** |
| Avg. # Schools in Mentor's Caseload <i>~N(0,1)</i> | 0.017 (0.006)** | 0.045 (0.011)** | 0.018 (0.019) | 0.009 (0.005)+ | 0.008 (0.009) | 0.003 (0.012) |
| Avg. # Teachers in Mentor's Caseload <i>~N(0,1)</i> | -0.003 (0.005) | -0.016 (0.010) | 0.010 (0.018) | -0.004 (0.005) | -0.019 (0.009)* | -0.023 (0.014)+ |
| Teacher's Own Evaluation of Mentor <i>~N(0,1)</i> | | | -0.210 (0.139) | | | -0.083 (0.055) |
| SASS Support Questions | | | | | | |
| Reduced Work Load | | | 0.060 (0.054) | | | 0.039 (0.028) |
| Prof. Development, Workshops, Conferences | | | -0.014 (0.042) | | | 0.034 (0.042) |
| Common Planning w/Other Teachers | | | 0.038 (0.024) | | | 0.005 (0.016) |
| Mentor and Teacher Matching | | | | | | |
| Subject Area <i>(Admin. in (1) (2) (4) (5), Survey in (3) (6))</i> | -0.012 (0.017) | -0.030 (0.023) | 0.107 (0.062)+ | 0.001 (0.014) | 0.022 (0.021) | 0.044 (0.029) |
| Grade Level <i>(Admin. in (1) (2) (4) (5), Survey in (3) (6))</i> | -0.004 (0.018) | -0.000 (0.027) | 0.065 (0.081) | -0.010 (0.016) | 0.008 (0.023) | 0.032 (0.031) |
| School <i>(i.e., Mentor Previously Taught in School)</i> | -0.025 (0.024) | -0.067 (0.040)+ | 0.049 (0.094) | -0.012 (0.021) | -0.043 (0.039) | -0.015 (0.052) |
| Age <i>(Within 5 Years)</i> | 0.019 (0.019) | -0.007 (0.033) | -0.103 (0.037)** | 0.017 (0.017) | -0.019 (0.024) | -0.024 (0.030) |
| Gender | -0.003 (0.013) | -0.012 (0.021) | 0.056 (0.051) | -0.019 (0.017) | 0.020 (0.026) | 0.029 (0.034) |
| Ethnicity | 0.016 (0.012) | 0.041 (0.021)+ | 0.075 (0.039)+ | 0.004 (0.012) | 0.012 (0.020) | 0.049 (0.033) |
| Mentor Characteristics: | | | | | | |
| Assigned to School Where Previously Taught | 0.014 (0.015) | 0.037 (0.026) | 0.014 (0.033) | -0.011 (0.013) | 0.001 (0.018) | 0.018 (0.026) |
| Experience in New York City DOE | -0.002 (0.001)+ | -0.003 (0.002)+ | -0.004 (0.002)* | -0.001 (0.001) | -0.000 (0.001) | -0.004 (0.001)* |
| 1 Year of Program Experience in School | 0.029 (0.020) | 0.068 (0.025)** | 0.060 (0.028)* | 0.012 (0.018) | 0.003 (0.019) | -0.024 (0.024) |
| 1 Year of Program Experience | 0.011 (0.018) | 0.005 (0.020) | 0.038 (0.028) | -0.009 (0.018) | -0.007 (0.017) | -0.030 (0.017)+ |
| Mentor Voluntarily Not Returning <i>(2004-2005 Only)</i> | 0.042 (0.028) | | | 0.009 (0.018) | | |
| Mentor Not Invited to Return <i>(2004-2005 Only)</i> | 0.017 (0.023) | | | -0.011 (0.025) | | |
| Years Since Taught in Classroom | -0.001 (0.002) | -0.004 (0.003) | | 0.001 (0.002) | 0.003 (0.002)+ | -0.002 (0.004) |
| Observations | 42,519 | 42,519 | 17,940 | 39,809 | 39,809 | 39,809 |
| Number of Mentors | 313 | 313 | 214 | 320 | 320 | 215 |
| Number of Teachers | 1,281 | 1,281 | 507 | 1,314 | 1,314 | 523 |

Note: Variables marked with $\sim N(0,1)$ have been normalized. For details on additional controls see text. Standard errors (in parentheses) are clustered by mentor. Hours of mentoring received is instrumented with the average hours of mentoring received by teachers outside of the same school but with the same mentor among teachers who worked a full year. Survey evaluations are instrumented with the average evaluation given by teachers outside of the same school but with the same mentor.

Table A1: Estimated Effects of Mentoring Quality on Retention to Year 2 (Unconditional)

| | (1) | (2) | (3) | (4) |
|---------------------------------------------------|-----------|-----------|----------|----------|
| Mentoring Quality Indicators | OLS | IV | IV | IV |
| Hours of Mentoring the Teacher Received / 10 | 0.038 | 0.002 | 0.005 | 0.008 |
| <i>Instrumented in (2) (3) (4)</i> | (0.004)** | (0.005) | (0.012) | (0.012) |
| Avg. # Schools in Mentor's Caseload | 0.006 | 0.005 | 0.009 | 0.006 |
| <i>~N(0,1)</i> | (0.004) | (0.004) | (0.008) | (0.009) |
| Avg. # Teachers in Mentor's Caseload | -0.011 | -0.011 | -0.017 | -0.015 |
| <i>~N(0,1)</i> | (0.003)** | (0.003)** | (0.008)* | (0.008)+ |
| Teacher's Own Evaluation of Mentor | | | 0.001 | -0.031 |
| <i>~N(0,1), Instrumented in (4)</i> | | | (0.008) | (0.041) |
| SASS Support Questions | | | | |
| Reduced Work Load | | | -0.004 | 0.005 |
| | | | (0.013) | (0.018) |
| Prof. Development, Workshops, Conferences | | | 0.011 | 0.017 |
| | | | (0.025) | (0.027) |
| Common Planning w/Other Teachers | | | 0.031 | 0.032 |
| | | | (0.013)* | (0.013)* |
| Mentor and Teacher Matching | | | | |
| Subject Area | 0.006 | 0.014 | -0.009 | 0.007 |
| <i>(Admin. in (1) (2), Survey in (3) (4))</i> | (0.008) | (0.008)+ | (0.014) | (0.025) |
| Grade Level | -0.007 | -0.006 | 0.006 | 0.017 |
| <i>(Admin. in (1) (2), Survey in (3) (4))</i> | (0.014) | (0.014) | (0.013) | (0.019) |
| School | -0.004 | -0.000 | 0.096 | 0.096 |
| <i>(i.e., Mentor Previously Taught in School)</i> | (0.023) | (0.022) | (0.050)+ | (0.050)+ |
| Age | 0.018 | 0.016 | -0.001 | -0.001 |
| <i>(Within 5 Years)</i> | (0.013) | (0.013) | (0.021) | (0.021) |
| Gender | -0.019 | -0.018 | -0.003 | -0.001 |
| | (0.008)* | (0.008)* | (0.015) | (0.015) |
| Ethnicity | -0.003 | 0.002 | 0.001 | 0.002 |
| | (0.008) | (0.008) | (0.016) | (0.017) |
| Mentor Characteristics: | | | | |
| Assigned to School Where Previously Taught | -0.001 | 0.007 | -0.007 | -0.011 |
| | (0.010) | (0.010) | (0.022) | (0.024) |
| Experience in New York City DOE | -0.001 | -0.001 | 0.001 | 0.001 |
| | (0.001)+ | (0.001)+ | (0.002) | (0.002) |
| 1 Year of Program Experience in School | 0.001 | 0.002 | -0.005 | -0.006 |
| | (0.012) | (0.012) | (0.017) | (0.018) |
| 1 Year of Program Experience | 0.001 | -0.000 | 0.017 | 0.017 |
| | (0.012) | (0.011) | (0.016) | (0.016) |
| Mentor Voluntarily Not Returning | -0.016 | -0.027 | | |
| <i>(2004-2005 Only)</i> | (0.018) | (0.016)+ | | |
| Mentor Not Invited to Return | -0.009 | -0.020 | | |
| <i>(2004-2005 Only)</i> | (0.020) | (0.018) | | |
| Years Since Taught in Classroom | 0.002 | 0.001 | | |
| | (0.001)* | (0.001) | | |
| Observations | 9924 | 9924 | 3,932 | 3,932 |
| R-squared | 0.20 | n/a | n/a | n/a |

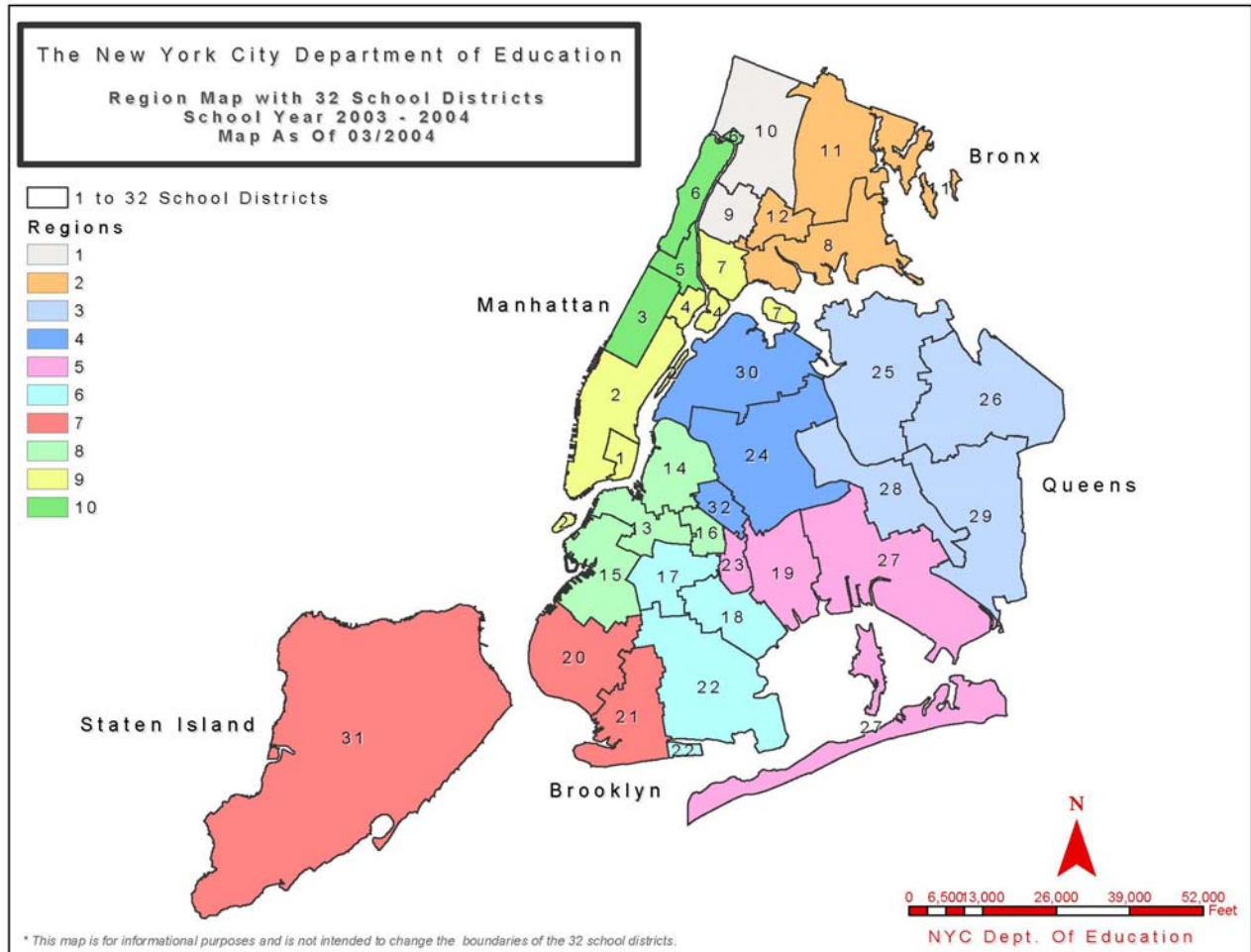
Note: Variables marked with $\sim N(0,1)$ have been normalized. For details on additional controls see text. Standard errors (in parentheses) are clustered by mentor. Hours of mentoring received is instrumented with the average hours of mentoring received by teachers outside of the same school but with the same mentor among teachers who worked a full year. Survey evaluations are instrumented with the average evaluation given by teachers outside of the same school but with the same mentor.

Table A2: Estimated Effects of Mentoring Quality on Returning to School in Year 2 (Unconditional)

| | (1) | (2) | (3) | (4) |
|---------------------------------------------------|-----------|------------|------------|------------|
| | OLS | IV | IV | IV |
| Mentoring Quality Indicators | | | | |
| Hours of Mentoring the Teacher Received / 10 | 0.032 | 0.000 | -0.029 | -0.027 |
| <i>Instrumented in (2) (3) (4)</i> | (0.004)** | (0.008) | (0.024) | (0.023) |
| Avg. # Schools in Mentor's Caseload | -0.037 | -0.038 | -0.040 | -0.042 |
| <i>~N(0,1)</i> | (0.007)** | (0.007)** | (0.018)* | (0.020)* |
| Avg. # Teachers in Mentor's Caseload | 0.007 | 0.007 | 0.013 | 0.015 |
| <i>~N(0,1)</i> | (0.005) | (0.005) | (0.018) | (0.018) |
| Teacher's Own Evaluation of Mentor | | | 0.003 | -0.022 |
| <i>~N(0,1), Instrumented in (4)</i> | | | (0.010) | (0.084) |
| SASS Support Questions | | | | |
| Reduced Work Load | | | -0.002 | 0.006 |
| | | | (0.016) | (0.029) |
| Prof. Development, Workshops, Conferences | | | 0.059 | 0.064 |
| | | | (0.030)+ | (0.034)+ |
| Common Planning w/Other Teachers | | | 0.051 | 0.052 |
| | | | (0.017)** | (0.017)** |
| Mentor and Teacher Matching | | | | |
| Subject Area | 0.011 | 0.018 | 0.009 | 0.022 |
| <i>(Admin. in (1) (2), Survey in (3) (4))</i> | (0.012) | (0.012) | (0.019) | (0.047) |
| Grade Level | 0.002 | 0.003 | 0.013 | 0.022 |
| <i>(Admin. in (1) (2), Survey in (3) (4))</i> | (0.019) | (0.019) | (0.018) | (0.034) |
| School | 0.096 | 0.099 | 0.166 | 0.166 |
| <i>(i.e., Mentor Previously Taught in School)</i> | (0.040)* | (0.039)* | (0.090)+ | (0.090)+ |
| Age | 0.010 | 0.008 | -0.007 | -0.007 |
| <i>(Within 5 Years)</i> | (0.017) | (0.017) | (0.031) | (0.031) |
| Gender | -0.009 | -0.008 | -0.008 | -0.006 |
| | (0.011) | (0.011) | (0.021) | (0.022) |
| Ethnicity | -0.008 | -0.003 | -0.021 | -0.021 |
| | (0.011) | (0.011) | (0.021) | (0.021) |
| Mentor Characteristics: | | | | |
| Assigned to School Where Previously Taught | -0.053 | -0.045 | -0.042 | -0.045 |
| | (0.018)** | (0.017)** | (0.036) | (0.038) |
| Experience in New York City DOE | -0.002 | -0.002 | 0.001 | 0.001 |
| | (0.001)+ | (0.001)+ | (0.003) | (0.003) |
| 1 Year of Program Experience in School | 0.115 | 0.116 | 0.198 | 0.197 |
| | (0.019)** | (0.019)** | (0.037)** | (0.037)** |
| 1 Year of Program Experience | -0.043 | -0.044 | -0.045 | -0.044 |
| | (0.020)* | (0.019)* | (0.034) | (0.035) |
| Mentor Voluntarily Not Returning | -0.007 | -0.016 | | |
| <i>(2004-2005 Only)</i> | (0.025) | (0.024) | | |
| Mentor Not Invited to Return | -0.026 | -0.036 | | |
| <i>(2004-2005 Only)</i> | (0.033) | (0.032) | | |
| Years Since Taught in Classroom | 0.004 | 0.003 | | |
| | (0.002)** | (0.002)+ | | |
| Observations | 9924 | 9924 | 3,932 | 3,932 |
| R-squared | 0.23 | <i>n/a</i> | <i>n/a</i> | <i>n/a</i> |

Note: Variables marked with $\sim N(0,1)$ have been normalized. For details on additional controls see text. Standard errors (in parentheses) are clustered by mentor. Hours of mentoring received is instrumented with the average hours of mentoring received by teachers outside of the same school but with the same mentor among teachers who worked a full year. Survey evaluations are instrumented with the average evaluation given by teachers outside of the same school but with the same mentor.

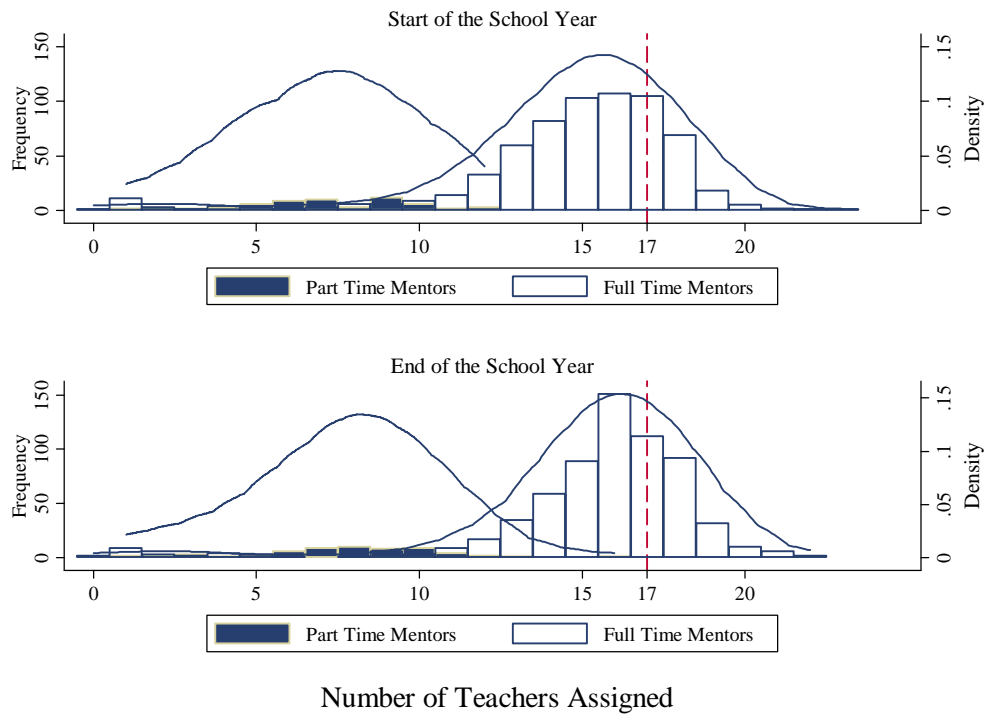
Figure 1: Regional Map of the New York City Department of Education



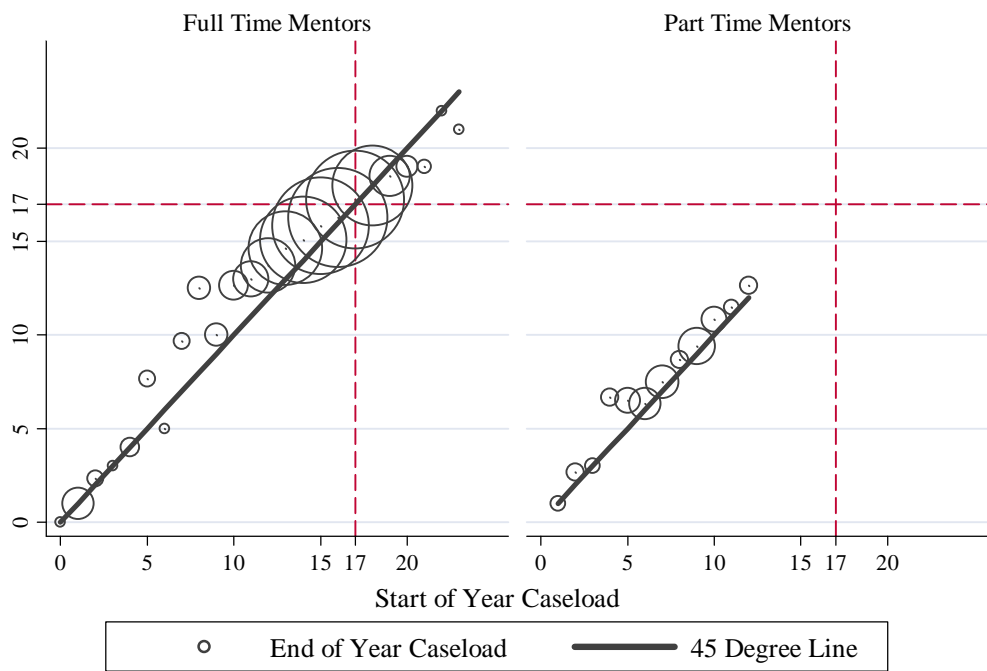
Note: Schools serving students with disabilities are treated as a separate, 11th, administrative region for the purpose of mentoring. Source: schools.nyc.gov/pdf/citywidemap.pdf

Figure 2: Mentor Caseloads at the Start and End of the School Year

Distributions by Time of Year and Full-Time Status

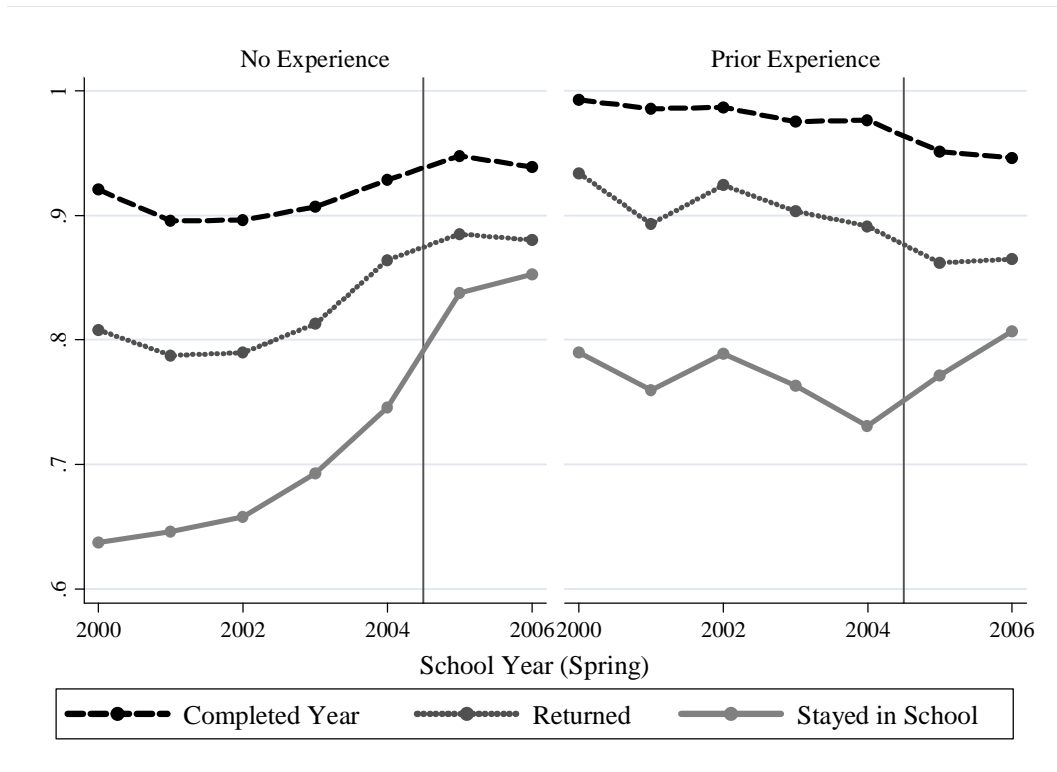


End of Year vs. Start of Year

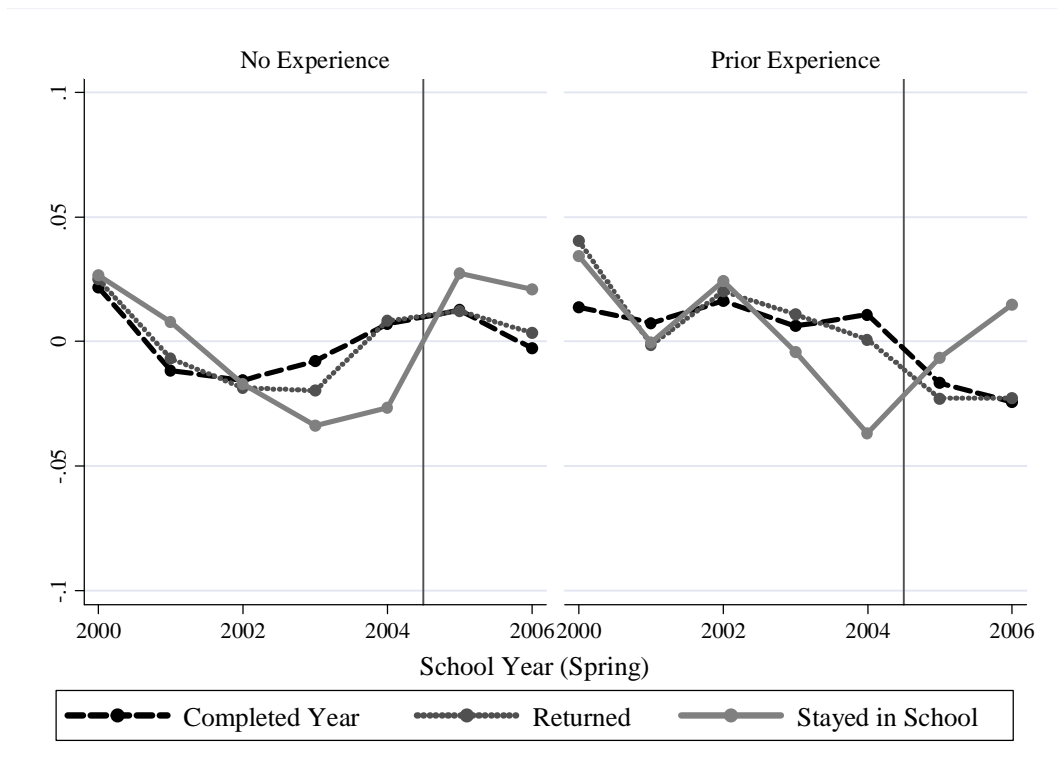


Note: Circle sizes weighted by number of mentors.

Figure 3: New Teacher Retention, by Prior Experience, 1999-2000 to 2005-2006
Raw Trends



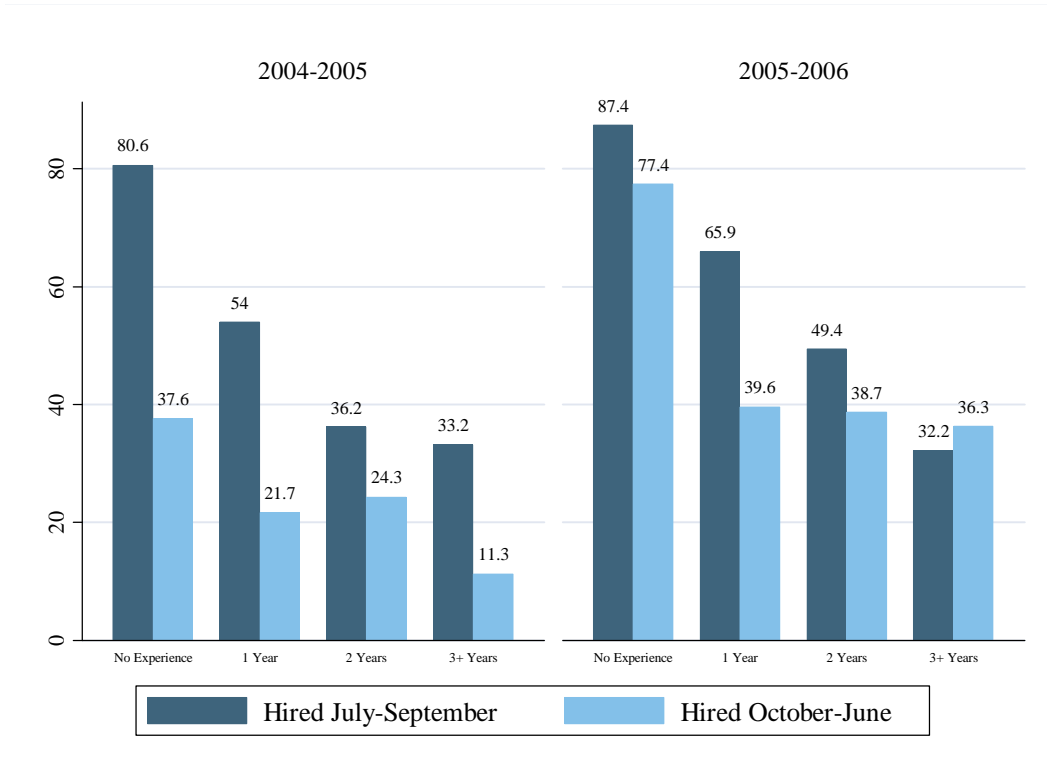
Residual Trends



Note: The top figure shows the portion of newly hired teachers who complete the school year (Completed Year), return the next year (Returned) and return to the same school (Stayed in School). The bottom figure shows average residuals from ordinary least squares regressions of these outcomes on controls for prior experience, ethnicity, gender, and type of certification, school fixed effects, a year trend, and an interaction of year with an indicator for no prior teaching experience.

Figure 4: Mentoring of New Teachers
by Prior Experience, Timing of Hire, and School Year

Percent of Teachers Receiving Mentoring



Number of Teachers Hired

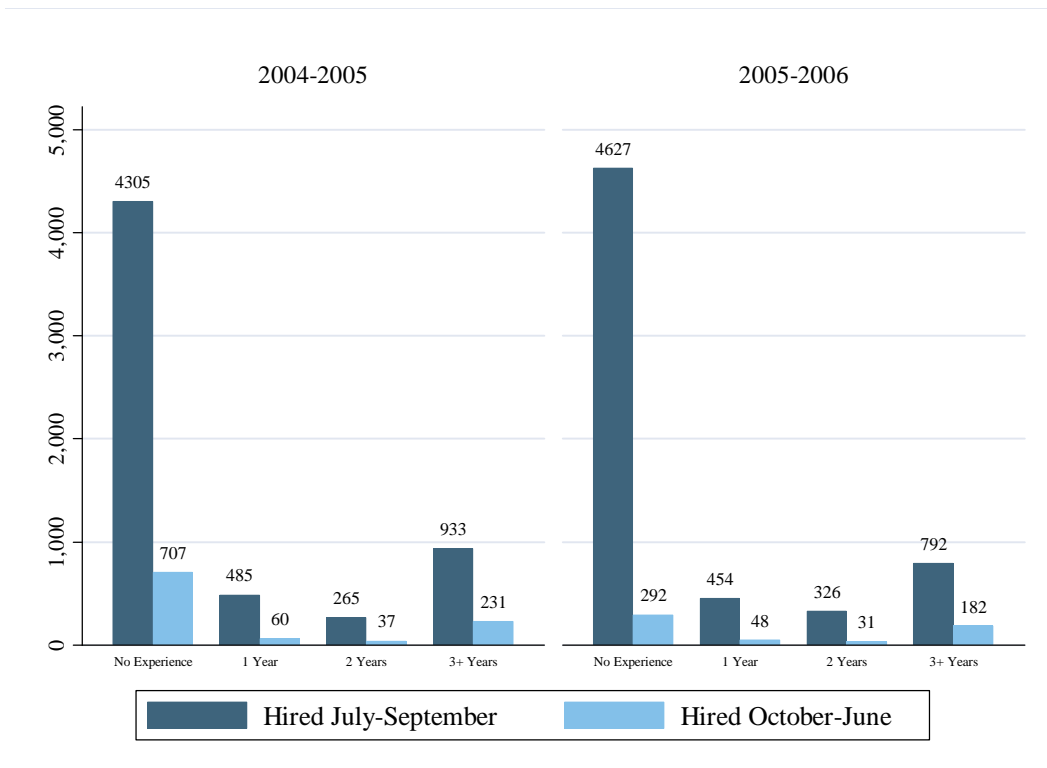
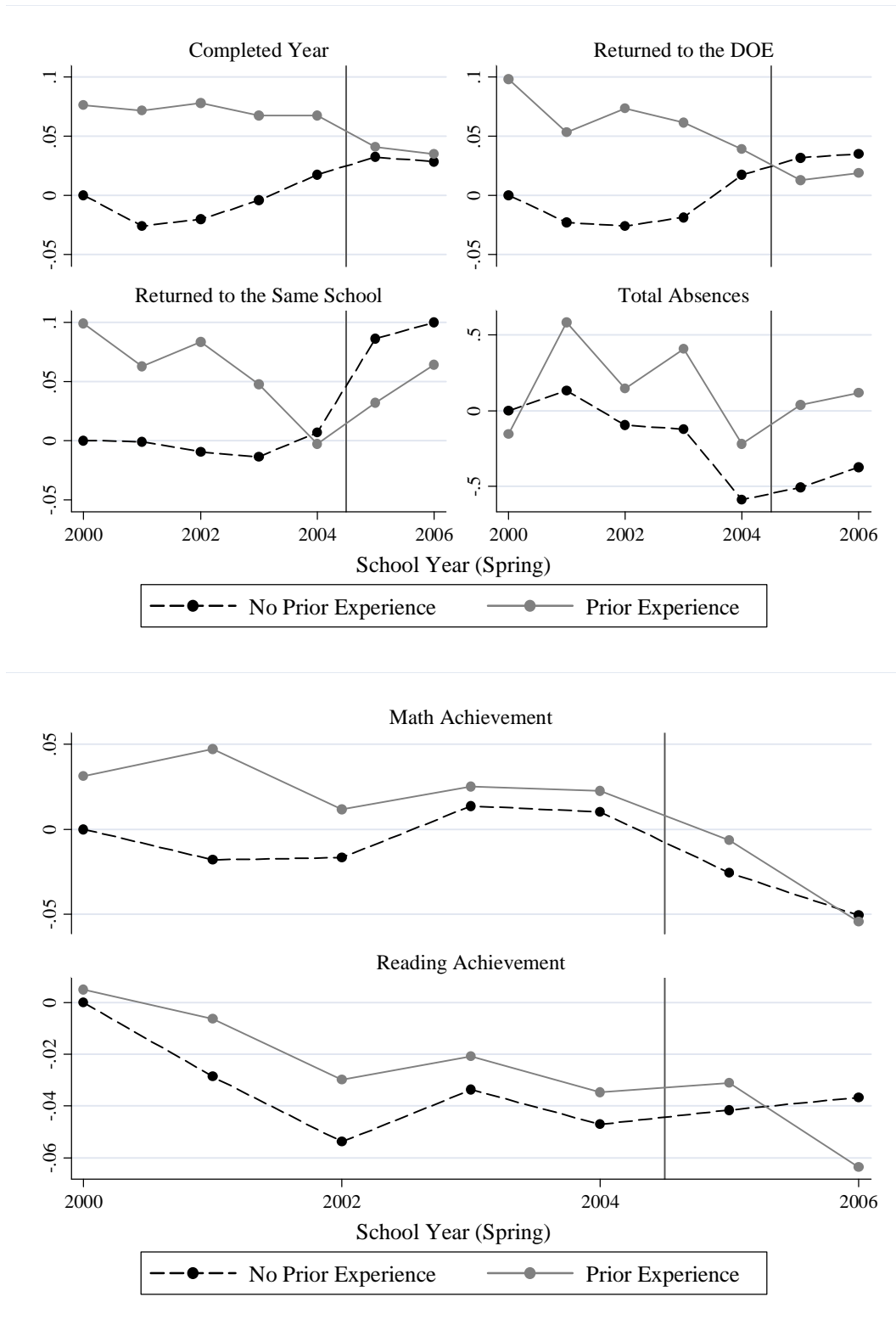


Figure 5: New Teacher Outcomes, Conditional on Observables
by Prior Experience, 1999-2000 to 2005-2006



Note: These figures show point estimates for interactions of year an indicator for prior experience taken from least squares regressions that include controls for teachers' prior experience, ethnicity, gender, age, age squared, and type of certification. Achievement regressions also include controls for students' prior achievement, gender, ethnicity, free lunch receipt, special education status, English Language Learner status, classroom means of these variables, class size, and school fixed effects.

Figure A1: Survey and Administrative Data Agreement

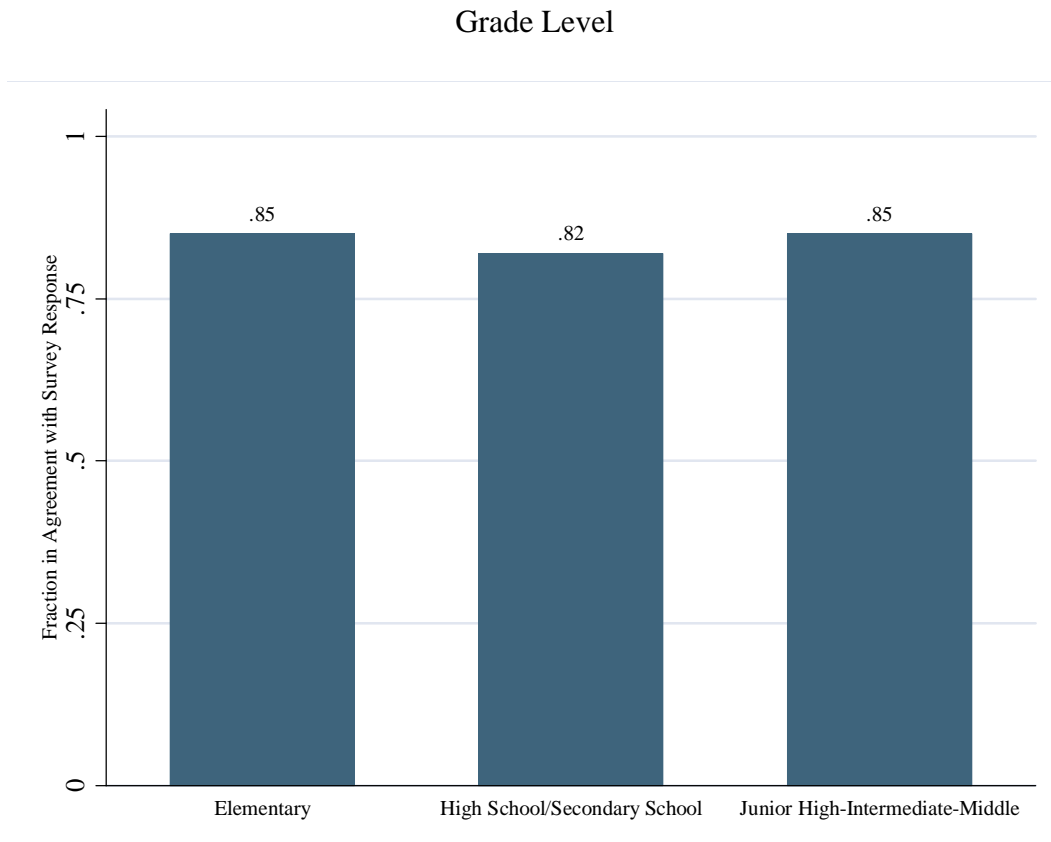
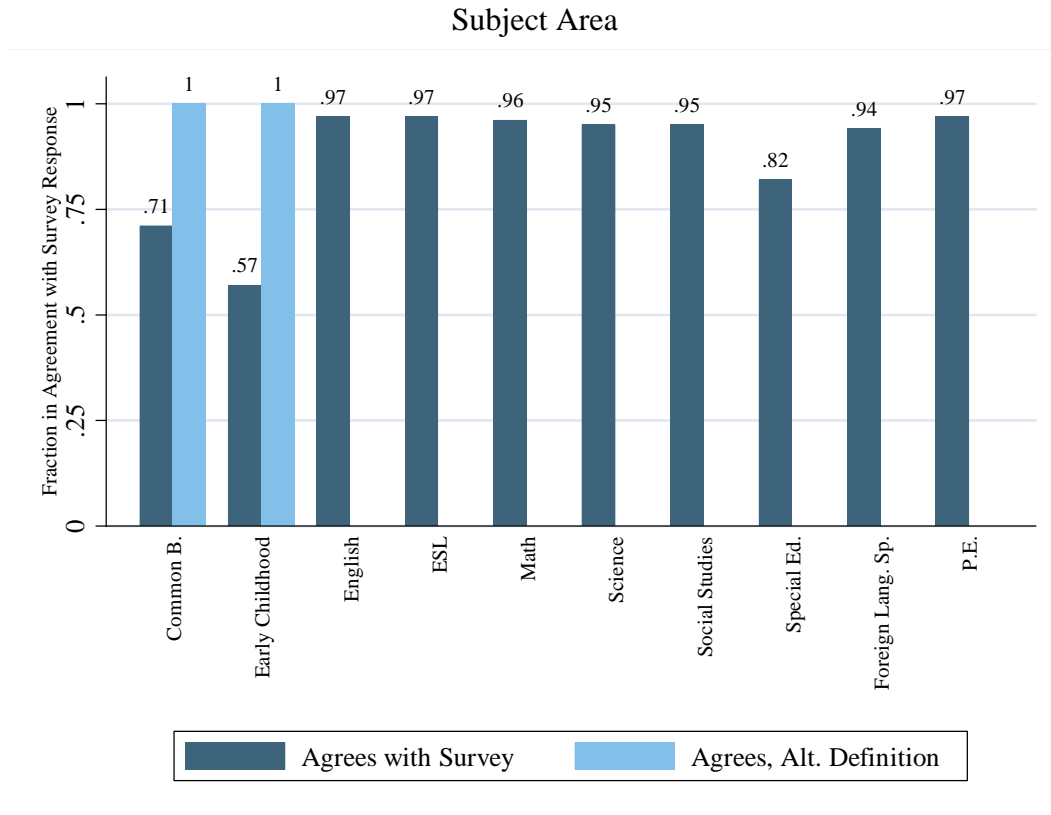


Figure A2: Power of Administrative Data to Predict Survey Response

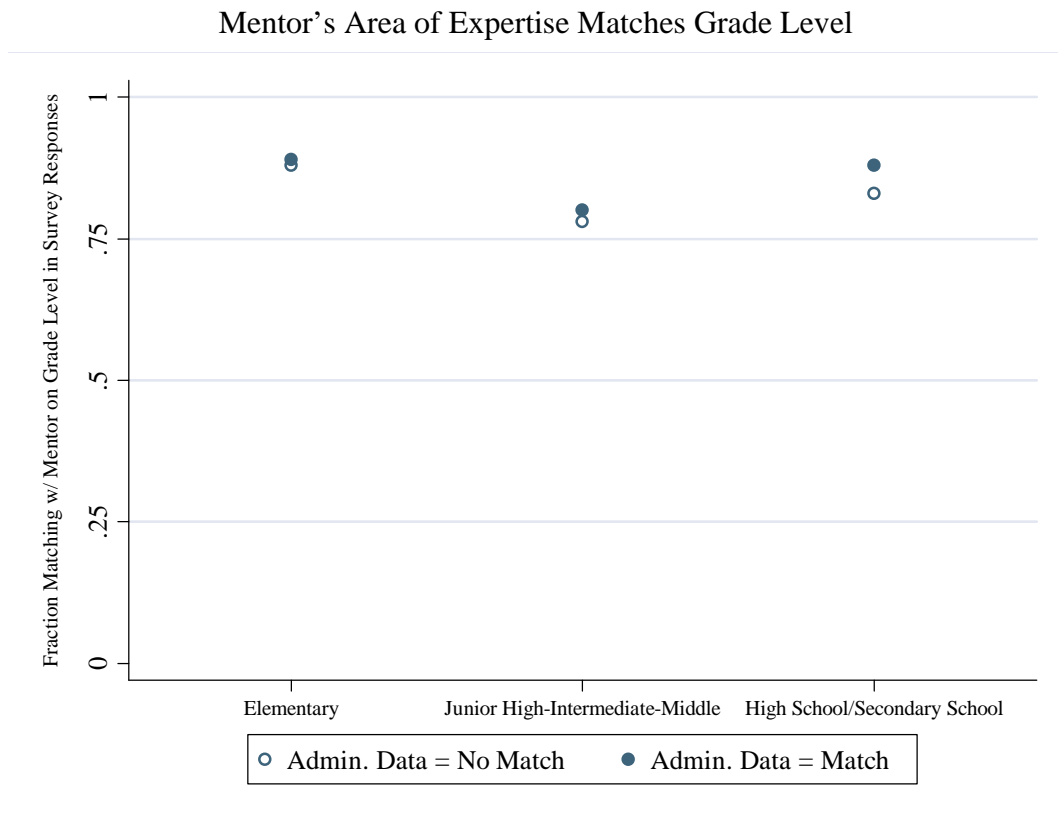
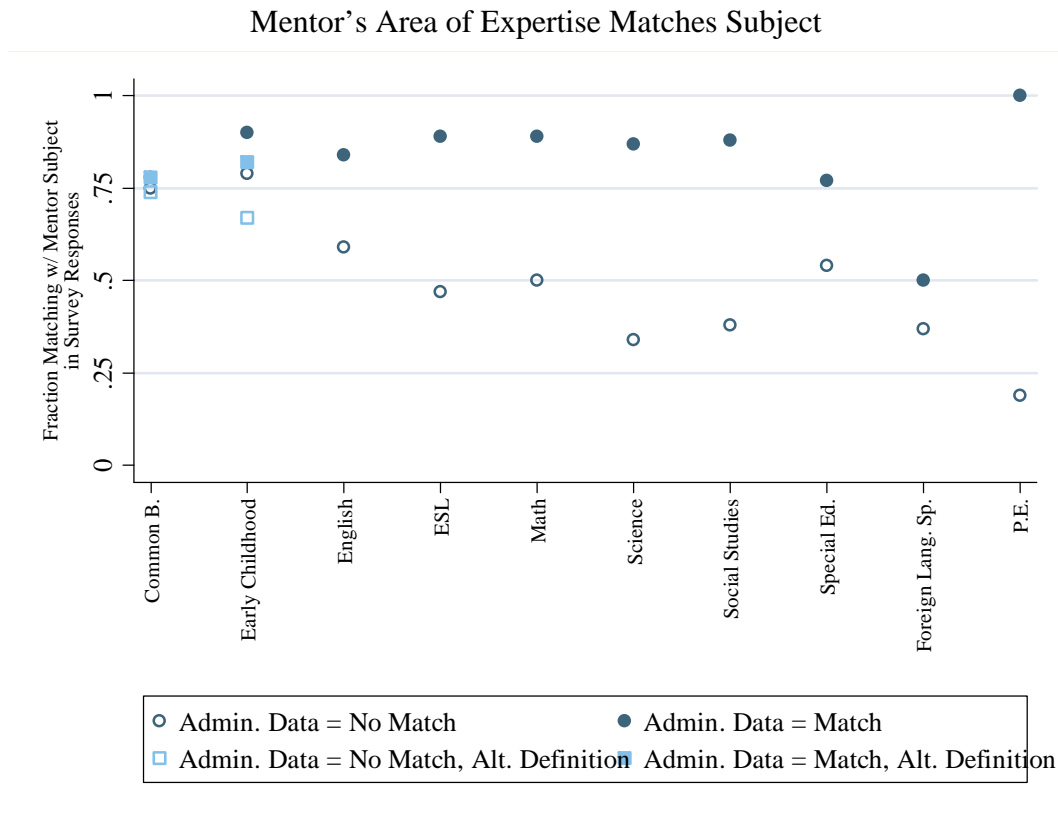
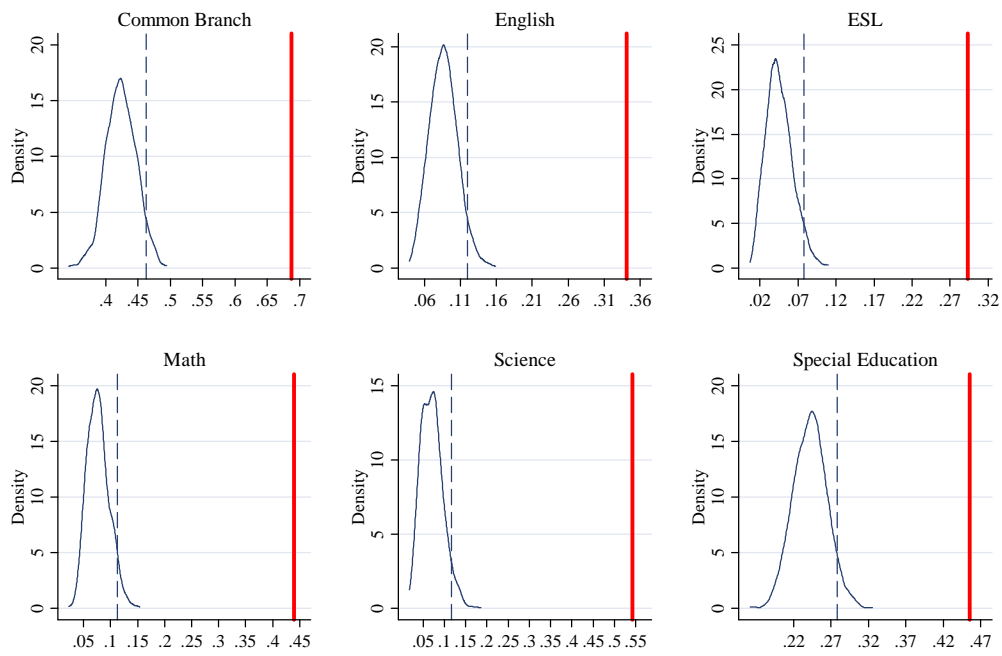
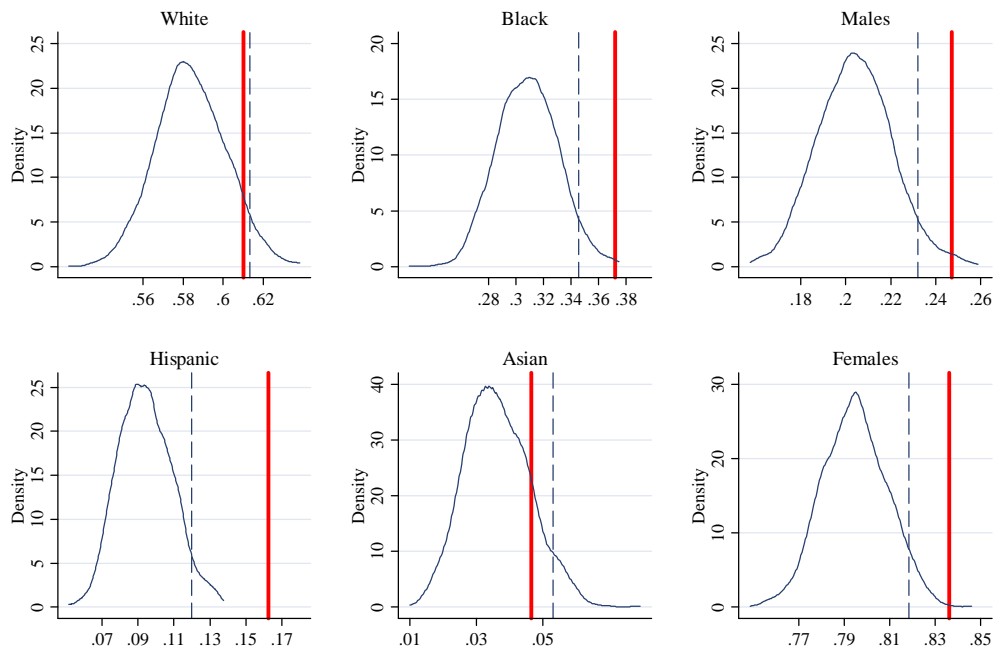


Figure A3: Actual Match Rates vs. Randomized Simulations
Subject within Regions



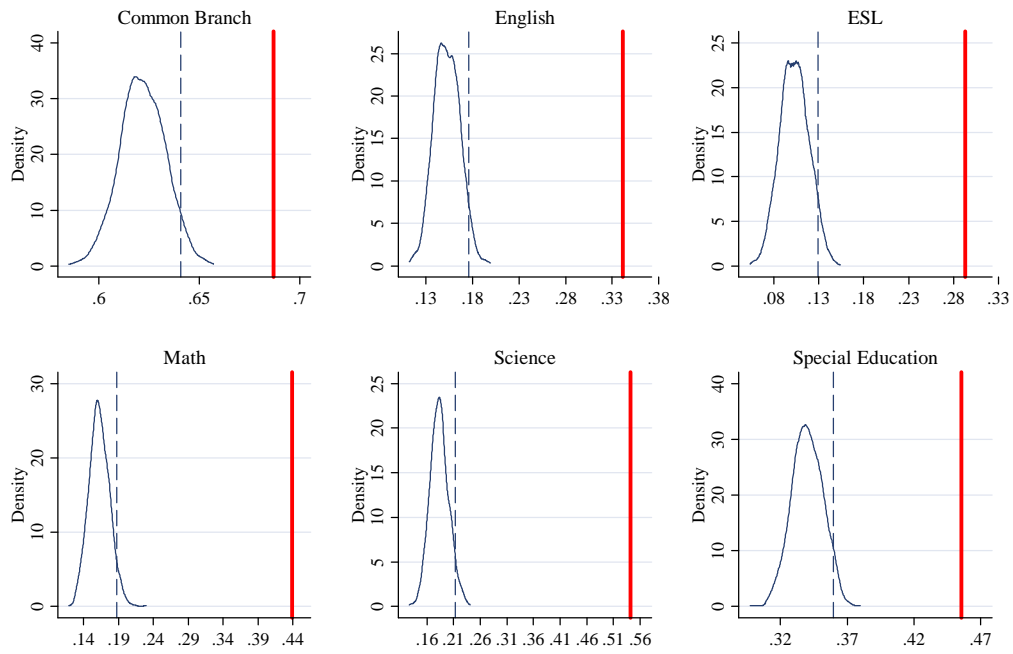
Note: Solid line represents actual match rate; dashed line is the 95th percentile of match rates from Monte Carlo simulations of mentor subject. For details on this procedure see text.

Demographics within Regions



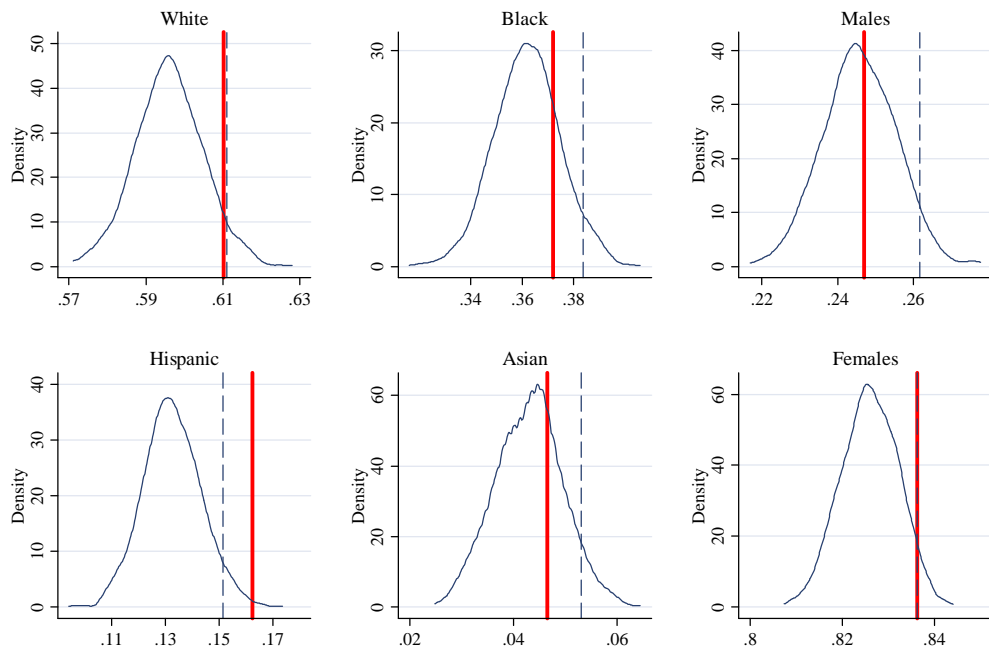
Note: Solid line represents actual match rate; dashed line is the 95th percentile of match rates from Monte Carlo simulations of mentor demographics. For details on this procedure see text.

Figure A4: Actual Match Rates vs. Randomized Simulations
Subject within Schools



Note: Solid line represents actual match rate; dashed line is the 95th percentile of match rates from Monte Carlo simulations of mentor subject. For details on this procedure see text.

Demographics within Schools



Note: Solid line represents actual match rate; dashed line is the 95th percentile of match rates from Monte Carlo simulations of mentor demographics. For details on this procedure see text.