CALDER

National
Center for Analysis of Longitudinal Data in Education Research

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

Urban Institute

A program of research by the Urban Institute with Duke University, Stanford University, University of Florida, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

Teacher Preparation and Student Achievement

Donald Boyd, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff

WORKING PAPER 20 • AUGUST 2008

Teacher Preparation and Student Achievement

Donald Boyd*, Pamela Grossman, Hamilton Lankford*, Susanna Loeb**, and James Wyckoff*****

**** University at Albany, ** Stanford University, *** University of Virginia***

We are grateful to the New York City Department of Education and the New York State Education Department for the data employed in this paper. We benefited particularly from insights and help from Carla Asher, Vicki Bernstein, Marilyn Cochrane-Smith, David Cohen, Ronald Ehrenberg, Jennifer Bell-Ellwanger, Robert Floden, Nicholas Michelli, Richard Murnane, Edward Silver, and Ana Maria Villegas. We appreciate financial support from the Carnegie Corporation of New York, City University of New York, the National Science Foundation, the Spencer Foundation and the National Center for the Analysis of Longitudinal Data in Education Research (CALDER), supported by Grant R305A060018 to the Urban Institute from the Institute of Education Sciences, U.S. Department of Education. We are also grateful for the exceptional research assistance provided by Marsha Ing, Luke Miller, Maria Perez, and Matt Ronfeldt, as well as Karen Hammerness, Morva McDonald, and Michelle Reininger. The views expressed in the paper are solely those of the authors and may not reflect those of the Urban Institute, the Institute of Education Sciences, the U.S. Department of Education, or any other sponsors of the research. Any errors are attributable to the authors.

Teacher Preparation and Student Achievement
Donald Boyd, Pamela Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff
CALDER Working Paper No. 20

Abstract

There are fierce debates over the best way to prepare teachers. Some argue that easing entry into teaching is necessary to attract strong candidates, while others argue that investing in high quality teacher preparation is the most promising approach. Most agree, however, that we lack a strong research basis for understanding how to prepare teachers. This paper is one of the first to estimate the effects of features of teachers' preparation on teachers' value-added to student test score performance in Math and English Language Arts. Our results indicate variation across preparation programs in the average effectiveness of the teachers they are supplying to New York City schools. In particular, preparation directly linked to practice appears to benefit teachers in their first year.

I. Introduction

There are fierce debates over the best way to prepare teachers to improve outcomes for the students they teach. Some argue that easing entry into teaching is necessary to attract strong candidates (U.S. Department of Education, 2002). Others argue that investing in high quality teacher preparation will better serve our nation's children (NCTAF, 1996). Even among those who believe that high quality preparation is important, there are sharp contrasts concerning the best approach (Levine, 2006). Most agree, however, that we lack a strong research basis for understanding how to prepare teachers to meet the challenges of urban schools (c.f. Cochran-Smith & Zeichner, 2005; Wilson, Floden, Ferrini-Mundy, 2001). Lack of evidence creates the opportunity for a myriad of potential "solutions" regarding teacher preparation and little way to evaluate their promise. This study is a first step towards developing evidence to inform these debates, looking carefully at the ways in which teachers are prepared and the consequences of that preparation for pupil learning.

Teachers in New York City enter teaching through a variety of pathways, including both more traditional and alternate routes. Even within these pathways, teachers can receive quite different preparation opportunities, with this variation existing both between and within institutions of higher education (Boyd, et al., 2008). Do these differences in the experiences of teachers in teacher education programs affect the achievement of the students taught by program graduates? If so, are there aspects of programs that are associated with greater improvements in student achievement? We explore these questions employing a unique database on teachers, their preparation, and the students they teach. We combine administrative data on individual teachers and students in New York City with detailed information about the components of teacher preparation programs as identified by an analysis of over 30 programs and a survey of all first-year teachers in New York City. Taken together, these data allow us to explore how the preparation of teachers who staff a large, diverse, urban school district influences student achievement.

II. Background

A large extant research literature on teacher preparation provides some useful information with which to evaluate effective preparation practices.[11] However, much of the research is limited in scope, focuses on inputs to the preparation process rather than outcomes, uses data that are only loosely connected to the concepts being examined, or employs case-study methodologies from which it is difficult to determine causal relationships or generalize to other populations. As a result, there is still much to learn about effective preparation practices. In their review of the literature, Wilson, Floden, and Ferrini-Mundy (2001) propose four research elements that would allow future research to address important gaps in our knowledge regarding teacher preparation.

- Studies should compare practices across institutions as a way of identifying effective practice.
- Studies should examine the relationship between specific components of teacher preparation programs and specific outcomes, such as student achievement.
- Research should include measures that are sensitive to program content and quality.
- Research should have a longitudinal component and examine impacts over time.

This study addresses each of these suggestions. First, we employ a detailed analysis of 31 elementary teacher preparation programs, each of which contribute a significant number of teachers to New York City public schools. We include both traditional pathways to teaching and alternate pathways so as to allow for comparisons within and between each of these routes. Using a survey of first-year

[1] For a very useful summary of the teacher preparation literature, see Wilson, Floden and Ferrini-Mundy (2001). For other relevant work see Ball and Cohen (1999), Carnegie Forum on Education and the Economy (1986), Cochran-Smith and Zeichner (2005), Darling-Hammond (2000), Darling-Hammond & Bransford (2005), Feiman-Nemser (1983, 1990), Goodlad (1990), Holmes Group (1986), Levine (2006), Allen (2003), and Wayne and Youngs, (2003).

teachers, we also compare the experiences of teachers across all routes serving New York City public schools, not just those routes for which we collected information from the program directly.

To address Wilson, Floden, and Ferrini-Mundy's second point, our analysis includes a detailed description of the policies and practices of teacher preparation programs. We: 1) analyze documents describing the structure and content of each preparation program, 2) interview program directors, directors of field experiences, and other administrative staff of these programs, 3) survey instructors of math and reading methods courses, and 4) survey program participants and graduates of these programs. We link this information to each program participant, their career decisions, and the outcomes for students they teach.

To address the third point concerning the need to employ outcome measures that are sensitive to program quality and content, we use the extensive information on program content through our analysis of program documents, interviews, observations, and survey data. We then link features of program content to the change in elementary school students' achievement in math and reading. Finally, to address Wilson, Floden, and Ferrini-Mundy's final point about the need for longitudinal analysis, we follow program participants through their first two years of teaching and link them to longitudinal data on student achievement.

A labor market perspective: This study of teacher education observes programs that prepare teachers for New York City (NYC) schools from what we might term an aerial perspective (c.f. Boyd, et al., 2006). Such a vantage point has its obvious disadvantages, particularly when it comes to portraying nuances of individual programs. Our goal, however, is to develop a broader picture of the terrain of teacher education in a single, large district, portraying, in general, how teachers are prepared to teach in NYC public schools and how variation in this preparation affects student learning.

Most prior studies of teacher education have produced case studies of individual programs, taking a ground-level view of programs that prepare teachers (c.f. Darling-Hammond, 2000; Goodlad, 1990). Such studies provide detailed analyses of what individual programs, often chosen on the basis of their reputations, offer students and how they organize opportunities for learning to teach. However, each program is situated in a broader labor market for teachers; the ability of one program to attract participants, as well as the effectiveness of the teachers it produces, is likely to be a function of aspects of the market as well as that program's offerings. Very few studies of teacher education have focused on a labor market, investigating the array of preparation programs that provide teachers to a specific locale.

The structure of pay; teachers' preferences for the characteristics of a school's students; the geographic segmentation of students by income and race, teachers, leadership, community, and facilities; and hiring practices including the post-and-fill system of seniority transfers all affect teachers. Studies that compare programs across the United States might consider how the different contexts or labor markets affect the preparation programs, but it is quite difficult to adjust for all the differences. In addition, if we looked at a small handful of programs or programs scattered across multiple markets, we would not be able to understand how pathways interact to fill the demand for teachers. How effective one teacher is relative to others in the school depends not only on that teacher's skills and preparation, but the skills and preparation of the other teachers. By looking at all pathways into teaching in New York City and by doing an in-depth analysis of the largest programs and pathways, we are able to address these interactions.

We also know that teacher labor markets are small geographically. In 2000, ninety percent of New York teachers went to high school within 40 miles of their first job and most of these teachers also attended college very close by (Boyd, Lankford, Loeb and Wyckoff, 2005). This confirms anecdotal accounts that most of New York City's teachers attended New York City K-12 public schools, underscoring the importance of understanding and improving the quality of teacher education received by those going through programs in and around NYC.

Features of Teacher Education Programs: We look quite broadly at teacher preparation, guided by the existing research literature in our selection of features of teacher education to study (Boyd, et al., 2006; Boyd, et al., 2008). We have collected information from a broad variety of sources on five areas identified as important indicators of program quality: program structure; subject specific preparation in reading and math; preparation in learning and child development; preparation to teach racially, ethnically, and linguistically diverse students; and the characteristics of field experiences (c.f. Cochran-Smith & Zeichner, 2005; Darling-Hammond, Bransford, LePage, Hammerness, & Duffy, 2005; Valli, Reckase, & Raths, 2003; Wilson, Floden, & Ferrini-Mundy, 2001).

For this paper, we first estimate differences in the average effectiveness of teachers from each program as measured by student learning gains in math and English language arts (ELA). We then look at the relationship between teachers' value-added in these subjects and the features of their programs and their experiences. For this later analysis we focus primarily on elements of preparation that are closely linked to the daily work of teachers in the classroom; reflecting the perspective that effective professional education is grounded in the practices of the profession (c.f. Ball & Cohen, 1999). This focus is clearly just a first step in understanding all elements of preparation, and we do assess the effects of other measures, largely as a comparison. The scope of possible preparation characteristics is too great to address all in a single paper.

The use of value-added methodologies to assess teacher effectiveness has both advantages and disadvantages. Student learning is a logical metric with which to measure the effectiveness of teaching. However, available measures of student achievement are never perfect indicators of what students know or what teachers have taught. Researchers have raised concerns about whether these tests are valid measures of the domains of knowledge that we care about, whether they reliably measure student learning, and, even if they do, whether they reliably measure the aspects of learning that teachers affect (see, for examples, Feldt & Brennan, 1989; Messik, 1989). An alternative would be to analyze how preservice preparation affects teacher behaviors, such as instructional practices and career decisions, instead of student outcomes. One benefit of this approach is that it eliminates the need to match teachers to the students they teach. It has the clear disadvantage of not actually measuring student progress; linkages between teacher behaviors and students needs to be established separately. A second alternative would be to study student progress employing measures other than test performance. Unfortunately, such measures typically are not available.

The three questions driving the analysis are as follows.

1. What is the distribution of the average value-added of teachers from different preparation programs?
2. How do features of those preparation programs affect teachers' value-added to student achievement gains in math and ELA?
3. How do teachers' reported experiences in teacher preparation affect their value added?

Establishing causality is rarely easy, especially with non-experimental data. The analyses in this report are just a first step in this direction, using regression analysis to account for possible biases; we see our study as the beginning of a larger exploration of the impact of teacher preparation. Despite the challenges of establishing causal linkages, the results provide evidence that focusing more on preparation directly linked to practice can produce teachers who are more effective in their first year of teaching.

III. Methods

A number of factors complicate the assessment of the effects of teacher preparation. First, teaching candidates select their teaching pathway, preparation institution and program. This selection is important both because of the need to account for it in our assessment of program effects and because by identifying the features of pathways that attract individuals with the potential to be great teachers we can recruit more effective teachers.

Second, different pathways into teaching can lead teachers into schools and classrooms with different characteristics. For example, some alternative route programs place teachers exclusively in high-poverty, underachieving schools. Again, this is important for several reasons; first, we must account for these differences in the matching of teachers to schools if we are to accurately assess the affect of pathway and program features; and second, if a policy goal is to improve teaching particularly in these high-needs schools, then it useful to understand the features of programs that are most effective for supplying good teachers specifically to these schools.

The study comprises three separate analyses. The first analysis estimates differences in the average value-added to student learning of teachers from different childhood teacher education programs providing a substantial number of entering teachers to New York City schools. For this, we look at value added to student achievement in math and ELA separately, netting out student, classroom and school influences. The second analysis explores the relationship between student outcomes and features of those teacher preparation programs, using data collected from programs. The third analysis examines the relationship between student achievement and teachers' own reports of their preparation experiences. Information on teachers' experiences come from a survey administered to all first-year NYC public school teachers in the spring of 2005 and, as such, this analysis is limited to the respondents from this single cohort of teachers.

The model for estimating program effects is based on the following equation:

$$A_{ijst} = \beta_0 + \beta_1 A_{ijs(t-1)} + X_{it}\beta_2 + C_{ijst}\beta_3 + T_{jst}\beta_4 + \Pi_j + \nu_s + \varepsilon_{ijst} \qquad (1)$$

Here, the achievement (*A*) of student *i* in year *t* with teacher *j* in school *s* is a function of his or her prior achievement, time-varying and fixed student characteristics (*X*), characteristics of the classroom (*C*), characteristics of the teacher (*T*), indicator variables (fixed effect) for the childhood preparation program the teacher completed (Π), a fixed-effect for the school (ν), and a random error term (ε). Student characteristics include race and ethnicity, gender, eligibility for free or reduced-price lunch, whether or not the student switched schools, whether English is spoken at home, status as an English language learner, the number of school absences in the previous year, and the number of suspensions in the previous year. Classroom variables include the averages of all the student characteristics, class size, grade, and the mean and standard deviation of student test scores in the prior year.

Whether or not to include teacher characteristics depends upon the question at hand. If we want to know whether teachers from one program are more effective than teachers from another program then there is no reason to include fixed teacher characteristics, such as certification exam scores. In fact, the benefit of one program or pathway may come from its ability to recruit and select high quality candidates. However, if we want to separate selection from preparation aspects of programs, then it is important to control for teachers' initial characteristics. These controls are particularly important for the parts of our analysis that look at the effects of program characteristics on preparation, as opposed to programs overall. Unfortunately, we have only weak controls for these initial characteristics, though it is unclear how well any program can do in distinguishing and then selecting individuals who will be particularly excellent teachers. The teacher characteristics that we include are age, gender, race and ethnicity, whether they passed their general knowledge certification exam on the first attempt, and their score on that exam.

We estimate Equation 1 on multiple samples of childhood education teachers: 2004-05 and 2005-06 first-year teachers, 2000-01 through 2005-06 first-year teachers, and 2000-01 through 2005-06 second-year teachers. We also estimate models for each of these samples using two definitions of programs. The first examines childhood-education teachers aggregated by pathway and institution.

For example, teachers who obtained childhood-education certification through the college-recommended pathway after attending CUNY Brooklyn would be in one group; those from Teachers College, in another group; and those from Teach for America, in a third group. Because programs within institutions may differ in characteristics, the second definition of program expands these categories so that within institutions teachers who attend a master's program at one institution are categorized as in a different group than those who attended a bachelor's degree program at that same institution; those who received their preparation in neither a master's or bachelor's program (e.g., a certificate program) are in a third group.

The model for estimating the effects of program characteristics is very similar to the model described above. As shown in Equation 2, the only difference is that in place of program fixed effects, we include program characteristics (*P*) with standard errors for the estimated effects clustered at the program level and we include pathway into teaching (college-recommended, individual evaluation, New York City Teaching Fellows, Teach for America, and other) as an additional teacher-level control.

$$A_{ijst} = \beta_0 + \beta_1 A_{ijs(t-1)} + X_{it}\beta_2 + C_{ijst}\beta_3 + T_{jst}\beta_4 + P_{jst}\beta_5 + \nu_s + \varepsilon_{ijst} \qquad (2)$$

The program characteristics, described in detail below, include: (1) the number of math (subject-matter content) courses required for program entry or exit; (2) the number of English language arts (subject-matter content such as English, writing or communication) courses required for program entry or exit; (3) the percent of the instructors for courses in math methods, learning and development, and English language arts methods who are tenure-line faculty; (4) program oversight of student teaching; and (5) whether or not the program requires some sort of capstone project (portfolio, research paper, action research project, etc.). The final two of these measures capture a link between preparation and practice while the other measures may capture content requirements and stability of the program.

The model for estimating the effects of teachers' experiences in their teacher preparation programs, as reported in survey responses, is again similar to the models described above. As shown in Equation 3, the only difference is that instead of including program fixed effects or characteristics we include teacher reports of their experiences (*E)*. Standard errors of these estimates are clustered at the teacher level.

$$A_{ijst} = \beta_0 + \beta_1 A_{ijs(t-1)} + X_{it}\beta_2 + C_{ijst}\beta_3 + T_{jst}\beta_4 + E_{jst}\beta_5 + \nu_s + \varepsilon_{ijst} \qquad (3)$$

The measures of self-reported features of experiences in teacher preparation come from responses to the survey of first-year teachers described above. They include: (1) the extent to which there was an emphasis on opportunities to engage in aspects of teaching practice during coursework, (2) the extent to which coursework covered the New York City curriculum in math and ELA, (3) whether the teacher had student teaching experience, (4) whether the grade and subject for which the teacher did student teaching are similar to their current assignments, (5) opportunities to learn about teaching math, (6) opportunities to learn about teaching ELA, and (7) opportunities to learn about English language learners, and (8) opportunities to learn about handling student misbehavior. Again, the first four measures, in particular, capture some aspect of the link between preparation and practice.

IV. Data

We estimate Equations 1 through 3 using extensive data on individuals during their education and their professional careers; information about the schools in which these teachers work; and student data including test scores. Of particular note, we constructed the variables characterizing teachers' preparation using detailed descriptions of the 31 childhood-education preparation programs whose graduates produce the vast majority of new teachers for New York City public schools and through a survey of all first-year teachers in the spring of 2005. Twenty six of these programs are more traditional programs in which teachers complete both coursework and student teaching prior to becoming a teacher of record; the remaining five programs are alternative-route programs in which teachers enter the classroom after approximately six weeks of preservice preparation and complete their coursework while teaching full-time. Four of these programs are associated with the New York City Teaching Fellows; and one is Teach for America.

Administrative Data on Students, Teachers and Schools: The dependent variables in our models come from annual student achievement exams given in grades four through eight to almost all New York City students. The student data, provided by the New York City Department of Education (NYCDOE), consists of a demographic data file and an exam data file for each year from 2000-01 through 2005-06. Demographic files include measures of gender, ethnicity, language spoken at home, free-lunch status, special-education status, number of absences, and number of suspensions for each student who was active in any of grades three through eight that year.

For most years, the data include scores for approximately 65,000 to 80,000 students in each grade. An exception is that the files contain no scores for seventh grade English language arts in 2002, because the New York City Department of Education is not confident that exam scores for that year and grade were measured in a manner that was comparable to the seventh grade English language arts exam in other years. Using these data, we construct a set of records with a student's current exam score and his or her lagged exam score. For this purpose, a student is considered to have value added information in cases where we had a score in a given subject (ELA or math) for the current year and a score for the same subject in the immediately preceding year for the immediately preceding grade. We do not include cases in which a student took a test for the same grade two years in a row, or where a student skipped a grade.[2]

[2] We also exclude observations for classrooms with less than ten or more than 50 students.

While NYCDOE does not maintain an identifier linking students directly to their teachers, in most cases we were able to create our own links using school and course identifiers, because the NYCDOE's data systems track the courses taken by each student and the courses taught by each teacher. Based on advice from NYCDOE staff, we matched students in grades three through five, and grade six if in an elementary school, to teachers based on the homeroom identifier. We matched other sixth grade students and students in grades seven and eight based on the section of a course being taught. Unfortunately, some middle schools do not participate in NYCDOE's middle-school performance assessment system (MSPA) and in those cases the course section identifier is not linked centrally to teachers. Because of this, we have a lower match rate for grades six through eight than for grades three through five, but never less than two-thirds.[3] However, the focus of this analysis is on teachers certified in childhood education, the large majority of whom teach in elementary schools, not in middle schools.

To enrich our data on teachers, we match New York City teachers to data from New York State Education Department (NYSED) databases, using a crosswalk file provided by NYCDOE that links their teacher file reference numbers to unique identifiers employed by NYSED. We draw variables for NYC teachers from New York State data files as follows:

- Teacher Experience: For teacher experience, we use transaction-level data from the NYCDOE Division of Human Resources to identify when individuals joined the NYCDOE payroll system in a teaching position. When this information is missing or when the value is less than the value in the NYSED personnel master files, we use the NYSED data.

- Teacher Demographics: We draw gender, ethnicity, and age from a combined analysis of all available data files, to choose most-common values for individuals.

- Test performance: We draw information regarding the teacher certification exam scores of individual teachers and whether they passed on their first attempts from the NYS Teacher Certification Exam History File (EHF).

[3] The average attributes of 6th through 8th grade students who are matched to teachers compared to those who are not matched are substantially the same with a few exceptions.

• Pathway: Initial pathway into teaching comes from an analysis of teacher certification data plus separate data files for individuals who participated in Teach for America or the Teaching Fellows Program.

• College Recommended: We obtain indicators for whether an individual had completed a college-recommended teacher preparation program and, if so, the level of degree obtained (bachelor's or master's) from NYSED's program-completers data files.

Using these data, we construct our indicator of the program and pathway into teaching as follows. Any individual who is separately identified as participating in Teach for America or the Teaching Fellows program is coded as entering teaching through that pathway, as appropriate. For the remaining teachers, we examine certification licensure records to determine the *earliest* pathway for which they had approval from NYSED prior to their first teaching job in New York State public schools, with those pathways defined as: (1) traditional college-recommended; (2) individual evaluation; (3) temporary license;[4] (4) Other certificates, including internship certificates, other Transitional B teachers, and those with certification through reciprocity agreements with other states. Teachers classified as entering through the college-recommended pathway are assigned to the program they completed based on information from the program completers file.

New York State changed teacher certification program requirements, which took effect for teachers receiving certification beginning in September 2004, but many programs had phased in changes over the previous two years. The 2004-05 cohort is the first group of program completers subject to these new requirements and thus our program and survey data collections are most relevant to these teachers. However, because we are concerned about controlling for the unobserved attributes of schools that affect the sorting of teachers to schools, we also estimate the models with longer panels of teachers to allow for better statistical controls.

---

[4] Temporary license signifies those individuals who failed to complete one or more requirements for a teaching certificate, but were allowed to teach under the temporary license provisions, whereby a school district can request NYSED to allow a specific individual to teach in a specific school for a temporary period.

Table 1 provides descriptive statistics for the main variables used in the analyses. The first panel gives student-level variables. The achievement scores of students in math and ELA are standardized by grade and year to have a zero mean and a standard deviation of one. The negative means reflect the fact that first-year teachers, on average, teach somewhat lower-performing students than do other teachers. New York City serves a diverse group of students: 43 percent of the students of the childhood education teachers are Hispanic, 28 percent are black, and 14 percent are Asian. Fifty-three percent of students speak English at home, and 62 percent are eligible for free lunch. The second panel of Table 1 reports summary statistics for the class-average measures used in the regression models.

The administrative data also provides information on teachers. For our sample, 61 percent obtained certification through the college-recommended pathway, a higher percentage than for New York City as a whole. This larger proportion is the result of limiting the sample to the programs for which we collected information, which excludes the very large number of temporary license teachers hired in New York City prior to 2004-05. The teacher population differs from the student population in both race and gender. Eighty-eight percent of the teachers are female, while only 11 percent are black and 11 percent are Hispanic. Eighty-seven percent of these new teachers passed their general knowledge certification exam on their first attempt.

Data on programs: The information on preparation programs comes from a data collection effort in the spring and summer of 2004 designed to characterize the preparation received by individuals entering teaching in 2004-05. We focused specifically on the 18 institutions that prepare about two-thirds of the college recommended teachers hired in NYC schools in recent years. Within these institutions, we concentrated on the pre-service preparation at 26 college-recommending childhood certification programs, as well as two large alternative route programs: the New York City Teaching Fellows and Teach for America. Those enrolled in the NYC Teaching Fellows program completed their preservice coursework at one of four institutions; we treat these as separate programs in the analysis, as the requirements differed by institution. Teach for America runs its own summer preservice program, so we count this as one program. Altogether, the analysis includes the preparation received by participants in 31 programs.

We rely on a number of data sources to document information about programs: state documents, institutional bulletins and program descriptions, NCATE documents when available, and institutional websites to find information about requirements and course descriptions. In documenting information about courses, whenever possible we use the information that is closest to what is actually taught. For example, we ask programs for the names of instructors who taught reading and math methods for the cohorts completing programs in 2004, and use this list rather than the list of faculty included in the state documents. We also conduct faculty surveys and collect course syllabi, and use this information to supplement course descriptions in catalogues and in state documents. In addition, we interview program directors and directors of field experiences about the curriculum, structure, and field experiences in their programs.

From this program information we create a large number of variables. For this particular analysis we choose to focus on measures that capture the link between the work teachers do in their preparation and the day-to-day work in the classroom. The program data is not ideal for doing this because of the rather general nature of much of the program information, but we identify whether or not the program requires a capstone project as one measure, as these projects generally involve connection to classroom experience, through teacher research or teaching portfolios. We also create a composite measure of the extent to which the program maintains oversight over student teaching experiences. In addition, for comparison to other features of the program that could influence student outcomes, we create variables measuring the math and English content course requirements and the percent of the program instructors in these courses who were tenure-line faculty. Table 2 provides a description of the variables.

The first panel of Table 2 gives the descriptive statistics for these variables. The capstone project measure indicates whether or not a final capstone project was required for program completion. Of these childhood programs, 13 of the 26 college-recommending programs require a final capstone project. We also collect data on the nature of the project. In most instances, the capstone project is either a portfolio, which captures prospective teachers' work both in courses and in the field over time, or an action research project, which requires prospective teachers to collect data in their field experience around a particular question related to their practice. Both of these options have the potential of helping prospective teachers link their work in classrooms to what they are learning at the university and focusing their attention on issues related to classroom practice.

The oversight-of-student-teaching variable combines three sub-measures: whether the program requires that cooperating teachers have a minimum number of years of teaching experience (32 percent of 24 programs), whether the program picks the cooperating teacher as opposed to selection by the K-12 school or the student teacher (42 percent of 28 programs), and whether a program supervisor observes their participants at least five times during student teaching (27 percent of 30 programs). Because these measures are highly correlated, we combine these binary variables into a single sum to measure the program's oversight of student teaching.

Finally, for math and ELA course requirements, programs range from no course requirements during preservice preparation to four in math and from zero to eight in ELA. We also measure the percentage of those teaching classes in math or ELA methods and learning and development who are tenure-line faculty. On average 45 percent of the instructors in these areas are tenure-line at the programs we studied. However, this varies greatly across programs (from zero to 88 percent).

Survey of First-Year Teachers: In the spring of 2005 we conducted a survey of all first-year New York City teachers in which we ask detailed questions about their preparation experiences, the mentoring they received in their first year, and their teaching practices and goals.[5] The response rate for the survey was over 70 percent. While again we focus on the extent to which programs emphasize preparation related to classroom practice, we also create other measures. For this analysis we have more degrees of freedom because we are not limited to teachers from the programs for which we collected detailed program information, and because individuals' experiences within programs to some extent differ. Because of this, we can control for other aspects of programs, when assessing the effects of the variables in question. For this purpose, we create measures of opportunities to learn about teaching math; opportunities to learn about teaching ELA; opportunities to learn about handling student misbehavior; and opportunities to learn about teaching English language learners. We also measure the extent to which preparation included links to practice through, for example, assignments that involve working with students; opportunities to study the New York City curriculum; whether or not the teacher had student-teaching experiences, not as the teacher-of-record in the classroom; and the congruence between their student-teaching placement and their current job assignment in terms of subject matter or grade level.

The second panel in Table 2 summarizes these variables. The ELA and Math measures are both composites. The ELA measure (alpha = 0.96) includes opportunities to: learn about characteristics of emergent readers; learn ways to teach students meta-cognitive strategies for monitoring comprehension; learn ways to teach decoding skills; learn ways to encourage phonemic awareness; learn ways to build student interest and motivation to read; learn how to help students make predictions to improve comprehension; learn how to support older students who are learning to read; learn ways to organize classrooms for students of different reading ability; study, critique, or

[5] The survey instrument is available at http://www.teacherpolicyresearch.org/portals/1/pdfs/Survey_of_04-05_NYC_First_Year_Teachers.pdf

adapt student curriculum materials; learn how to activate students' prior knowledge; listen to an individual child read aloud for the purpose of assessing his/her reading achievement; plan a guided reading lesson; discuss methods for using student reading assessment results to improve your teaching; and practice what you learned about teaching reading in your field experiences. The answer choices were (a) none, (b) touched on it briefly, (c) spent time discussing or doing, (d) explored in some depth, and (e) extensive opportunity. We standardize the composite variable to have a mean of zero and a standard deviation of one.

The math variable (alpha = 0.97) includes opportunities to: learn typical difficulties students have with place value; learn typical difficulties students have with fractions; use representations (e.g., geometric representation, graphs, number lines) to show explicitly why a procedure works; prove that a solution is valid or that a method works for all similar cases; study, critique, or adapt math curriculum materials; study or analyze student math work; design math lessons; learn how to facilitate math learning for students in small groups; adapt math lessons for students with diverse needs and learning styles; practice what you learned about teaching math in your teacher preparation program in your field experience. Unfortunately, this composite variable is not normally distributed. Instead, a group of participants had very little opportunity to learn math methods. As a result we split the composite variable into four groups: those teachers with a 1.0 ranking of opportunities (no opportunity), from 1.0 to 2.5 (little opportunity), 2.5 through 3.5 (some opportunity), greater than 3.5 (extensive opportunity).

Our measure of the link to practice is a composite of teachers' responses to three survey questions from the math and ELA composites: In your teacher preparation program, prior to September 2004, how much opportunity did you have to do the following? (1) Listen to an individual child read aloud for the purpose of assessing his/her reading achievement, (2) Plan a guided reading lesson, and (3) Study or analyze student math work. For each of the three elemental measures, we create a difference between the teacher's response to that question and his or her average response to all questions asking about opportunities to learn in their preparation program. The measures then reflect the relative emphasis of individuals' opportunities, rather than the level of the response to the relevant question alone. We then simply average these difference measures and standardize the result. As a specification check, we use the average of the elemental measures (not differenced). Whether we use the average response or the differenced response makes little difference in the effects observed in the analyses.

Our measure of the focus on the New York City curriculum comes from two questions similar to the ones above. The survey asks teachers about their opportunities to: (1) Review New York City's reading curriculum, and (2) Review New York City mathematics curriculum. We similarly difference the responses to these questions from each teacher's average response to questions about opportunities to learn about teaching reading and math, respectively; and then sum them to create the variable used in the analysis. The measure not differenced, again, provides similar results.

The two variables addressing student-teaching experience also come from teachers' responses to the survey. All teachers responded to these questions about field experience, unlike the curriculum and practice specific questions, which were directed only at elementary school teachers. One measure assesses whether the teacher participated in student teaching: (1) How much actual time did you spend student teaching as part of your teacher preparation prior to becoming a fulltime classroom teacher (assume one day is equivalent to 6 hours)? *Student teaching is a type of field experience involving taking full or partial responsibility for the classroom under the guidance of a full-time classroom teacher or supervisor.* Only 11.6 percent of the sample did not. A second set of questions measures the congruence between the teacher's current job and his or her field experience: *My experiences in schools were similar to my current job in terms of grade level and my experiences in schools were similar to my current job in* *terms of subject area*. Responses for both questions are on a five point scale, from strongly disagree to strongly agree. For all teachers, we average these two measures and standardize the composite to have a mean of zero and a standard deviation of one. For this sample of childhood teachers, the mean is slightly higher (0.07); and the standard deviation slightly lower (0.80), which is not surprising because they are all childhood teachers and so their subject area is less likely to be far out of field compared to high school teachers.

Finally, we construct a measure of opportunities to learn about learning and the *relative* emphasis placed on (1) opportunities to study how to handle student misbehavior and (2) opportunities to study teaching of English language learners, as perceived by program completers. The learning composite is made up of opportunities to: (a) study stages of child development and learning, (b) develop strategies for handling student misbehavior, (3) develop specific strategies for teaching English language learners, (4) develop specific strategies for teaching students identified with learning disabilities, (5) develop specific strategies for teaching students from diverse racial and ethnic backgrounds, and (6) develop strategies for setting classroom norms.

V. Results

Program Effects: Programs vary in the effectiveness of the teachers they prepare, as measured by student test-score gains. Figure 1a plots the average value-added to student achievement of institutions that produced at least 40 different first-year New York City teachers with value-added measures between 2000-2001 and 2005-2006.[6] The point (0, 0) is for the average of all institutions. Three results emerge from the figure. First, there is meaningful variation across institutions in average value-added. In all models the indicator variables for preparation institutions are jointly significant at traditional levels. The difference between the average of the institutions and the

[6] Institutions might include an undergraduate program, a graduate program and/or an alternative route program.

highest value-added institution is approximately 0.05 standard deviations in math and 0.04 standard deviations in ELA. This magnitude is about the same size as difference in average learning between students eligible for free- or reduced-price lunch and those who are not. It is also about the same size as the difference in effectiveness between first-year and second-year teachers. Second, the variation in average teacher effectiveness across institutions is approximately the same in math and ELA. Finally, on average, institutions that produce teachers who are more effective at increasing student learning in math are also more effective in ELA (correlation of 0.60).

Figure 1b replicates Figure 1a but instead of institution effects show program effects. For this analysis, institutions are separated into childhood bachelor's programs, childhood master's program, and other childhood programs (e.g., certificate programs leading to certification). The differences in effects across programs are somewhat larger in math with a range of approximately 0.18 standard deviations than in ELA with a range of 0.10. Again, programs that produce effective teachers in ELA also, on average, produce effective teachers in math (correlation 0.73).

Programs are likely to change over time, particularly with the recent focus on standards and aligning teacher education to state goals. As a result, a program that was effective in 2000 may be more or less effective in 2005. Figure 2 plots the institution effects for first-year teachers in the years 2004-05 and 2005-06 only. Again, similar patterns are evident (correlation between math and ELA of 0.52). The correlation between the point estimates for the fixed-effects in the current period and the full period is 0.65 for math and 0.42 for ELA.

The figures so far are based on models that do not include measured characteristics of teachers. The logic of this approach is that pathways and programs can supply high quality teachers by a combination of recruitment and selection of potentially excellent teaching candidate and by adding value to the teaching ability of its participants. By controlling for teacher characteristics we would understate the effects of those programs that put effort into, and are successful at, effective recruitment and selection. However, we are also interested in the variation across programs in value-added to teaching ability and for that we control for teachers' background characteristics. Moreover, in the analyses that follow, we want to identify the influence of particular aspects of teacher preparation on teaching. For that, we also will want to control for teacher characteristics. Figure 3 plots the estimates of program effects in Math controlling and not controlling for teacher age, gender, race/ethnicity, whether they passed the general knowledge certification exam on the first attempt, and the score on the exam. Our controls for teacher attributes make little difference. The correlation between observations with and without measured teacher attributes is 0.98. The data on ELA test performance produces the same patterns; there is little difference in the distribution of programs effects estimated with and without controls for other measured teacher characteristics.

Program Features: Figures 1-3 demonstrate some systematic differences across programs and institutions in the average value-added of their program completers. Table 3 reports estimates of the relationship between particular features of those preparation program and teachers' value-added to student achievement in math using the 2000-01 through 2005-06 and the 2004-05 and 2005-06 samples, respectively. Because of the small number of programs, the estimation relies on a small number of degrees of freedom. As a result we estimate the models entering one program feature at a time. Thus, Table 3 reports the coefficients and standard deviations for the program feature from 30 different estimations (five program features and three samples each for Math and ELA). We do not include models with all the features entered together because we do not have the degrees of freedom to support that analysis; however, while less stable across specifications, the estimated effects in most cases are similar when the measures of program characteristics are included in the same model.

As described above, the five measures of program features are the extent to which the program oversees the field experiences of its students; whether or not the program requires a capstone project, which is often a portfolio of work done in classrooms with students; content knowledge requirements as measured by content courses in math and ELA; and the percent of tenure-line faculty, a potential proxy for program stability and the extent to which institutions value teacher preparation. We collapse individual variables to these five measures because some of the component features are highly correlated, and thus measure very similar concepts. For example, this is why we use one measure of program oversight of teacher education, instead of entering each of the element measures separately. Doing so results in point estimates that are statistically significant, but we cannot tell which of the components of the constructed index drives the effect since the three are correlated.

The two measures of the link between program experiences and the practice of teaching are significant for first-year teachers for both math and ELA, for both the 2001-2006 sample and the 2005-2006 sample. The coefficients are also quite large – at least .04 for the capstone project in both ELA and math, and 0.04 in math and 0.01 in ELA for oversight of student teaching. Caution in the interpretation of these results is warranted. Since estimates for program features are estimated singly, the coefficients may reflect these variables and any omitted but correlated variables. As we show below, we find similar results in models that employ teacher survey data, which allow us to control for many preparation attributes simultaneously. As a result, we believe the results presented here warrant attention. However, the positive estimates do not hold for either outcome for second year teachers. This result is not surprising given that teachers are likely to learn quite a bit about practice during their first year of teaching and thus first-year differences converge as teachers acquire relevant knowledge and skills on the job.

Interestingly, the content-specific coursework requirements work in a different way. For math, math coursework is positively associated with teachers' value-added in the second year, but not consistently in the first year with small effects (about .02). Similarly for ELA student achievement, ELA coursework has a small positive and significant effect in the second year, but not in the first year. This is consistent with some qualitative research on the effect of methods coursework, which also found a one-year lag in the impact of methods courses (e.g. Grossman, Valencia, Evans, Thompson, Martin, & Place, 2000). Tenure status does not appear to be important for either first or second-year teachers in math or reading.

Teacher's Reports of Experiences / Survey Results: Tables 4a and 4b give the results for the survey analysis for the 2004-2005 cohort of New York City teachers in their first year for math and ELA, respectively. Because the variables are at the teacher level instead of the program level, we have more degrees of freedom, even though we are now working with only one cohort of teachers (instead of two and six in the program features analyses).

The first two variables in Table 4a, practice and New York City curriculum, are measures of how closely the preparation links to the work that teachers do in their first year. For a description of the components of each variable see Appendix Table 2. They are both positively and significantly related to value-added in math in all specifications, both for the full sample and for a sample limited to teachers who obtained their initial certification through a traditional preparation program. The magnitude of the practice effect suggests that a standard deviation increase in the focus on practice is associated with value-added being higher by 0.03 to 0.06 standard deviations, approximately the same effect as the gain from the first year of teaching experience. A similar increase in emphasis on the NYC curriculum is associated with value-added being higher by approximately 0.03 standard deviations.

The two measures of field experience – whether or not they student taught and the congruence between the context in which they had their field experiences and their current teaching position – are also positive in most models, though the student teaching measure is not stable. The 0.02 to 0.06 point estimates for congruence are similar in magnitude, again, to the first year of teaching experience. None of the other measures show consistent effects for first-year teachers and value-added in math.

Table 4b provides similar results for ELA. Here the findings are less clear. The full sample shows no consistent results. However, when the sample is limited to college recommended teachers, the practice and curriculum measures, again, are positive in all specifications. It is not uncommon in recent estimations of the effects of teacher characteristics on student learning to find larger effects in math than in ELA. The difference may be driven by schools having a greater effect on math learning than on reading learning. Students are probably more likely to be involved in activities outside school that contribute to reading learning than to math learning.

Tables 5a and 5b present similar results for second year teachers. The patterns are similar to those found in the program feature analysis. With the exception of studying curriculum used in New York City, none of variables that characterize the work of teachers are consistently significant. Some, in fact, have perverse signs in some specifications, but these unexpected results never are found in both the full and College Recommended samples. However, there is some evidence that second-year teachers who have additional courses in math content and math pedagogy have students with higher math value-added. This, too, echoes the results from the program features analysis presented in Table 3. No such evidence exists for ELA.

A new paper by Boyd, Grossman, Lankford, Loeb and Wyckoff (2008) shows that effect sizes as typically measured, including those reported here, understate the extent to which teacher attributes and other factors affect actual gains in student achievement. Judging such effects relative to the dispersion in achievement *gains* instead of relative to the dispersion of achievement *levels* and netting out that portion of the dispersion in test score gains attributable to measurement error results in effect sizes that are larger by a factor of four. Rather than having an effect size of 0.01 to 0.04 relative to the standard deviation in student test scores, as reported above, program attributes have an effect of four to 16 percent of a standard deviation of the true gain in students' achievement over the course of a school year.

VI. Conclusions

In summary, the results suggest that there is variation across programs in the average effectiveness of the teachers they are supplying to New York City schools, with some programs graduating teachers who have significantly greater impact on student achievement. On average, programs that produce childhood certified teachers who are more effective in math also produce teachers who are more effective in ELA; though, there are some programs that are stronger in one area than in the other. The results also suggest that features of teacher preparation can make a difference in outcomes for students. One factor stands out. Teacher preparation that focuses more on the work of the classroom and provides opportunities for teachers to study what they will be doing produces teachers who are more effective during their first year of teaching. This finding holds up across various model specifications and both for measures created from data on the requirements of programs and for measures created from surveys of teachers. Thus, similar measures created from two independent data collection efforts reach a shared conclusion.

As an example, programs that provide more oversight of student teaching experiences or require a capstone project supply significantly more effective first-year teachers to New York City schools. Teachers who have had the opportunity in their preparation to engage in the actual practices involved in teaching (e.g., listening to a child read aloud for the purpose of assessment, planning a guided reading lesson, or analyzing student math work) also show greater student gains during their first year of teaching. Similarly, teachers who have had the opportunity to review curriculum used in New York City perform better in terms of student test score gains in both math and ELA. Student teaching and the congruence of the student teaching placement are also positively associated with student learning in ELA and math, for first-year teachers.

The estimated effects of many of the measures of teacher preparation are educationally important, about the same size as the effect of the first year of teaching experience. As noted in Boyd et al. (2008) effect sizes estimated relative to the standard deviation of overall student achievement and with measurement error are roughly one quarter as large when measured relative to student achievement gains adjusting for measurement error. Thus, making such an adjustment increases estimated effect sizes presented in this paper by a factor of four.

We also find some support for the hypothesis that math content preparation improves the outcomes of students of second-year teachers, but not first-year teachers. This result is supported by statistically significant and meaningful estimates across the measures created from the program requirements and from the teacher surveys, but the effects in some specifications are estimated imprecisely. Taken with the findings on the actual work of teachers, these estimates suggest that inexperienced teachers may make use of their preparation sequentially. Teachers with stronger preparation in day-to-day issues are relatively more effective in their first year, while those with stronger content knowledge are able to make use of that knowledge by their second year.

Finally, we fail to find consistent support for any of our other teacher preparation hypotheses. For example, our results do not support the hypothesis that greater opportunities to learn how students learn influences student achievement among first-year or second-year teachers.

We urge caution in interpreting these results as they represent only the first stage of research exploring the relationships between preparation programs and the subsequent impact of graduates on pupil achievement. Research analyzing such relationships is still in its infancy. Our study suggests that programs may indeed affect the quality of teachers; however, it also points to some of the challenges of trying to make such linkages. We put substantial effort into collecting information on programs but we may not have collected the right information. In addition, some of the measures may be proxies for underlying characteristics or correlated unmeasured features. For example, the requirement of a capstone project may simply be a proxy for a program's rigor or of the engagement of its faculty, just as the percent of tenure-line faculty teaching core courses may be a proxy for institutional commitment to professional preparation.

Similarly, if features did not have significant effects in our analysis, it may not mean those features are not important in the preparation of teachers. We may not have sufficient variation in some of these features for them to emerge as significant. Teacher certification requirements in New York State are among the most demanding in the U.S., particularly for alternative route programs, and thus our study does not include individuals who have low absolute levels of many preparation attributes (Boyd et al., 2008). It is also possible that we simply measure the features of teacher education poorly. Well-tested instruments for describing preparation did not exist when we began this study, requiring us to develop the instruments used in this analysis. While we piloted the measures, they have not been validated for this purpose. In addition, the results presented here focus on teachers from childhood education programs, who typically teach elementary students. Some preparation attributes may be important for middle or high school teachers but not for elementary teachers.

Finally, our measures of student learning deserve the same caveats as exist for all such studies. We are not sure the extent to which the value-added measures of student achievement are actually good measures of either the range of student learning that we care about or of teachers' impact on learning. First, so many other things affect student learning that we have to be careful to adjust for other factors. However, removing all this variation, may also remove the variation in actual effectiveness. This would happen, for example, if teachers sorted perfectly by effectiveness across schools, and we then identified our results from only within-school variation. Second, the tests themselves may be misleading measures of the learning that policymakers desire. Nonetheless, the results presented here are an initial indication that pre-service preparation can influence teacher effectiveness, particularly the effectiveness of first-year teachers.

References

Ball, Deborah L., and David K. Cohen. 1999. "Developing Practice, Developing Practitioners." In *Teaching as the Learning Profession: Handbook of Policy and Practice,* edited by Linda Darling-Hammond and Gary Sykes (3-32). San Francisco: Jossey Bass Publishers.

Boyd, Donald J., Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2005. "The Draw of Home: How Teachers' Preferences for Proximity Disadvantage Urban Schools." *Journal of Policy Analysis and Management* 24 (1): 113-132.

Boyd, Donald J., Pamela L. Grossman, Hamilton Lankford, Susanna Loeb, Nicholas Michelli, and James Wyckoff. 2006. "Complex by Design: Investigating Pathways into Teaching in New York City schools. *Journal of Teacher Education* 57(2): 155-166.

Boyd, Donald J., Pamela L. Grossman, Karen Hammerness, Hamilton Lankford, Susanna Loeb, Morva McDonald, Michelle Reininger**,** Matthew Ronfeldt, and James Wyckoff. 2008. "Surveying the Landscape of Teacher Education in New York City: Constrained Variation and the Challenge of Innovation." *Educational Evaluation and Policy Analysis* 30(4): 319-343.

Boyd Donald J., Pamela L. Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2008. "Measuring Effect Sizes: The Effect of Measurement Error." CALDER Working Paper No. 19. Washington, D.C.: National Center for Analysis of Longitudinal Data in Education Research. June 2008.

Cochran-Smith, Marilyn, and Kenneth M. Zeichner, eds. 2005. *Review of Research in Teacher Education.* Washington, D.C.: American Educational Research Association.

Darling-Hammond, Linda. ed. 2000. *Studies of Excellence in Teacher Education*. Washington D.C.: American Association of Colleges of Teacher Education.

Darling-Hammond, Linda, John Bransford, Pamela LePage, Karen Hammerness, and Helen Duffy. eds. 2005. *Preparing Teachers for a Changing World*: *What Teachers Should Learn and Be Able to Do.* San Francisco: Jossey-Bass.

Feldt, Leonard S., and Robert L. Brennan. 1989. "Reliability." In *Educational Measurement*, edited by Robert L. Linn (3rd ed., 105-146). Washington, D.C.: American Council on Education; Macmillan.

Goodlad, John I. 1990.*Teachers for Our Nation's Schools.* San Francisco: Jossey Bass.

Grossman, Pamela L., Sheila W. Valencia, Kate Evans, Clarissa Thompson, Susan Martin, and Nancy Place. 2000. "Transitions into Teaching: Learning To Teach Writing in Teacher Education and Beyond." *Journal of Literacy Research* 32: 631-662.

Levine, Arthur. 2006. "*Educating School Teachers.*" The Education Schools Project Report. Washington, D.C. : The Education Schools Project.

Messick, Samuel. 1989. "Validity." In *Educational Measurement*, edited by Robert L. Linn (3rd ed., 13-103). Washington, D.C.: American Council on Education; Macmillan.

National Commission on Teaching and America's Future. 1996. "What Matters Most: Teaching for America's Future." Woodbridge, VA: National Commission on Teaching and America's Future. http://www.nctaf.org/documents/WhatMattersMost.pdf. (Accessed August 6, 2008.)

U.S. Department of Education. 2002. *Meeting the Highly Qualified Teachers Challenge.* Washington, D.C.: U.S. Department of Education, Office of Postsecondary Education, Office of Policy Planning and Innovation. http://www.ed.gov/about/reports/annual/teachprep/2002title-ii-report.pdf. (Accessed August 6, 2008.)

Valli, Linda, Mark D. Reckase, and James D. Raths. 2003. "Teacher Education, Program Outcomes, Teaching Practice, and Pupil Achievement on State Tests." Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, April 2003.

Wilson, Suzannne M., Robert E. Floden, and Joan Ferrini-Mundy. 2001. "Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations." Center for the Study of Teaching and Policy Report R-01-3. Center for the Study of Teaching and Policy, University of Washington. http://depts.washington.edu/ctpmail/PDFs/TeacherPrep-WFFM-02-2001.pdf. (Accessed August 6, 2008.)

Table 1: Descriptive Statistics: For 2001-2006 program features (Math sample)

Students	**Mean**	**Stan Dev**	**# of student**
ELA Standardized Score	-0.14	0.93	23549
Math Standardized Score	-0.12	0.96	27027
Female	0.50		27048
Hispanic	0.43		27048
Black	0.28		27048
Asian	0.14		27048
Other Non-White Race/Ethnicity	0.01		27048
Home Language English	0.53		27048
Receive Free Lunch	0.62		27048
Receive Reduced-Price Lunch	0.08		27048
Lunch Missing	0.19		27048
Entitled to ELL per lab	0.15		27048
Days absent in previous year	11.17	10.84	17858
Days suspended in previous year	0.01	0.13	17858
Classroom Averages	**Mean**	**Stan Dev**	**# of student**
Asian	0.14	0.21	27048
Black	0.28	0.30	27048
Hispanic	0.43	0.29	27048
Other	0.01	0.02	27048
Class Size	23.91	4.81	27048
Entitled to ELL per lab	0.15	0.22	27048
Receive Free Lunch	0.62	0.28	27048
Receive Reduced-Price Lunch	0.08	0.09	27048
Home Language English	0.53	0.29	27048
Days absent in previous year	12.05	6.41	22744
Days suspended in previous year	0.02	0.08	22744
Math scores from previous year	-0.04	0.53	18425
English scores from previous year	-0.08	0.56	17956
Standard dev: prior Math scores	0.72	0.20	18425
Standard dev: prior ELA scores	0.69	0.19	17822
Teachers	**Mean**	**Stan Dev.**	**# Teachers**
Path - College Recommended	0.61		773
Path - IE	0.08		773
Path - TFA	0.05		773
Path - NYCTF	0.19		773
Path - Other	0.06		773
Black	0.10		762
Hispanic	0.11		762
Other	0.06		762
Female	0.87		784
Age	29.16	7.05	784
Liberal Arts and Sciences Test Passed	0.87		784
Liberal Arts and Sciences Test Score	250.67	27.30	771

Table 2: Program Characteristics and Education Experiences of Teachers

For Program Features	**Mean**	**Standard Deviation**
Number of Math courses	1.16	1.13
Number of ELA courses	1.29	1.74
Proportion with capstone project	0.50	0.51
Proportion tenure track	0.45	0.23
Oversight of Student Teach.	0.95	1.07
For survey analysis	**Mean**	**Standard Deviation**
Practice	0.07	0.51
NYC Curriculum	-0.42	0.85
Congruence with Job	0.07	0.80
No Student Teaching	0.09	0.29
Math	0.19	0.96
ELA	0.12	0.91
Exp handling misbehavior	0.22	0.74
Exp to teach ELs	-0.55	0.83
Basic Skills	0.11	0.31

Table 3: The Effects of Program Characteristics

	Math			**ELA**		
	2001-06 1st Year	**2005&06 1st Year**	**2001-06 2nd Year**	**2001-06 1st Year**	**2005&06 1st Year**	**2001-06 2nd Year**
Capstone project	0.0410**	0.1216**	-0.0077	0.0496***	0.1019*	-0.0271
	(0.0159)	(0.0545)	(0.0221)	(0.0112)	(0.0501)	(0.0178)
Oversight	0.0324***	0.1240***	-0.0145	0.0122~	0.1038**	0.0022
	(0.0075)	(0.0345)	(0.0125)	(0.0073)	(0.0387)	(0.0138)
Math courses	0.0239***	0.0098	0.0225**	-0.0034	0.0014	0.0011
	(0.0062)	(0.0174)	(0.0091)	(0.0084)	(0.0200)	(0.0088)
ELA courses	-0.0026	-0.0272***	0.0087	-0.0091**	-0.0060	0.0113**
	(0.0050)	(0.0085)	(0.0056)	(0.0039)	(0.0096)	(0.0051)
Percent Tenure	0.1184**	0.0614	0.0857	0.0184	-0.0478	0.0077
	(0.0503)	(0.1242)	(0.0805)	(0.0338)	(0.0874)	(0.0548)

Table 4a: Effects of First-Year Teachers' Experiences in Teacher Preparation, Math

	Full Sample			College Recommended		
	Fixed-effects	Random-effects	OLS	Fixed-effects	Random-effects	OLS
Practice	0.061	0.044	0.027	0.122	0.053	0.033
	(0.011)***	(0.011)***	(0.007)***	(0.016)***	(0.012)***	(0.008)***
Curriculum	0.025	0.028	0.026	0.029	0.025	0.044
	(0.012)**	(0.011)**	(0.007)***	(0.017)*	(0.015)*	(0.009)***
No Student Teaching	-0.088	-0.015	0.056	-0.026	0.052	0.116
	(0.039)**	(0.038)	(0.024)**	(0.044)	(0.052)	(0.033)***
Congruence	0.072	0.038	0.024	0.059	0.050	0.042
	(0.013)***	(0.011)***	(0.007)***	(0.017)***	(0.016)***	(0.010)***
Math 2	-0.072	-0.023	-0.016	0.022	0.033	-0.012
	(0.046)	(0.045)	(0.030)	(0.083)	(0.079)	(0.047)
Math 3	-0.114	0.000	0.034	0.013	0.015	0.010
	(0.060)*	(0.053)	(0.032)	(0.093)	(0.081)	(0.048)
Math 4	-0.114	0.010	0.014	-0.123	0.022	-0.010
	(0.062)*	(0.056)	(0.034)	(0.085)	(0.085)	(0.049)
Learning	0.011	-0.005	-0.001	0.044	-0.012	0.007
	(0.014)	(0.013)	(0.008)	(0.017)***	(0.017)	(0.010)
ELL	0.032	0.005	0.001	0.086	0.029	0.013
	(0.014)**	(0.012)	(0.008)	(0.021)***	(0.017)*	(0.010)
Misbehavior	0.019	0.016	0.017	-0.007	0.017	0.012
	(0.012)	(0.012)	(0.007)**	(0.030)	(0.018)	(0.011)
Observations	7037	7037	7037	4482	4482	4482
Number of schools	233	233		162	162	
R-squared	0.526		0.629	0.524		0.622

Table 4b: Effects of First-Year Teachers' Experiences in Teacher Preparation, ELA

	Full Sample			College Recommended		
	Fixed-effects	Random-effects	OLS	Fixed-effects	Random-effects	OLS
Practice	0.001	0.010	0.009	0.037	0.021	0.022
	(0.013)	(0.009)	(0.007)	(0.020)*	(0.010)**	(0.008)***
Curriculum	-0.010	0.015	0.019	0.036	0.027	0.030
	(0.012)	(0.011)	(0.008)**	(0.024)	(0.013)**	(0.009)***
No Student Teaching	-0.062	-0.028	-0.006	-0.111	0.027	0.066
	(0.051)	(0.033)	(0.024)	(0.073)	(0.039)	(0.033)**
Congruence	0.004	-0.005	-0.005	-0.018	-0.000	0.003
	(0.015)	(0.011)	(0.007)	(0.021)	(0.014)	(0.009)
ELA	0.001	-0.012	-0.020	-0.033	-0.022	-0.035
	(0.021)	(0.013)	(0.010)**	(0.034)	(0.016)	(0.012)***
Learning	-0.004	0.011	0.013	-0.015	0.010	0.024
	(0.014)	(0.012)	(0.009)	(0.021)	(0.015)	(0.011)**
ELL	0.031	0.004	-0.005	0.024	0.012	0.002
	(0.015)**	(0.012)	(0.009)	(0.029)	(0.015)	(0.011)
Misbehavior	0.025	0.014	0.010	-0.022	0.010	0.006
	(0.015)*	(0.012)	(0.007)	(0.026)	(0.015)	(0.010)
Observations	7112	7112	7112	4735	4735	4735
Number of schools	238	238		167	167	
R-squared	0.479		0.617	0.494		0.623

Table 5a: Effects of Second-Year Teachers' Experiences in Teacher Preparation, Math

	Full Sample			College Recommended		
	Fixed-effects	Random-effects	OLS	Fixed-effects	Random-effects	OLS
Practice	-0.016	-0.009	-0.012	0.027	-0.024	-0.032
	(0.023)	(0.015)	(0.008)	(0.033)	(0.019)	(0.010)***
Curriculum	0.053	0.035	0.024	0.073	0.046	0.045
	(0.025)**	(0.017)**	(0.009)***	(0.037)**	(0.021)**	(0.011)***
No Student Teaching	0.129	0.106	0.102	-0.110	0.069	0.073
	(0.058)**	(0.040)***	(0.024)***	(0.124)	(0.050)	(0.033)**
Congruence	-0.031	-0.025	-0.036	0.006	-0.026	-0.029
	(0.015)**	(0.015)*	(0.008)***	(0.024)	(0.018)	(0.010)***
Math 2	0.323	0.163	0.040	0.071	0.154	0.081
	(0.098)***	(0.063)***	(0.034)	(0.148)	(0.082)*	(0.056)
Math 3	0.312	0.188	0.087	-0.067	0.113	0.092
	(0.089)***	(0.067)***	(0.037)**	(0.142)	(0.077)	(0.058)
Math 4	0.377	0.197	0.040	0.051	0.174	0.100
	(0.103)***	(0.072)***	(0.038)	(0.154)	(0.086)**	(0.060)*
Learning	-0.063	-0.024	0.017	0.007	-0.016	-0.000
	(0.023)***	(0.018)	(0.009)*	(0.037)	(0.021)	(0.012)
ELL	-0.043	-0.028	-0.036	-0.004	-0.031	-0.034
	(0.020)**	(0.015)*	(0.009)***	(0.035)	(0.018)*	(0.011)***
Misbehavior	0.061	0.013	-0.015	0.092	0.026	-0.011
	(0.029)**	(0.021)	(0.010)	(0.032)***	(0.027)	(0.014)
Observations	6119	6119	6119	4126	4126	4126
Number of group(sdbn4)	215	215		155	155	
R-squared	0.553		0.628	0.547		0.622

Table 5b: Effects of Second-Year Teachers' Experiences in Teacher Preparation, ELA

	Full Sample			College Recommended		
	Fixed-effects	Random-effects	OLS	Fixed-effects	Random-effects	OLS
Practice	-0.011	-0.006	-0.009	-0.092	-0.014	-0.012
	(0.024)	(0.014)	(0.009)	(0.032)***	(0.017)	(0.011)
Curriculum	-0.024	0.022	0.026	-0.143	0.016	0.028
	(0.028)	(0.014)*	(0.009)***	(0.040)***	(0.019)	(0.011)**
No Student Teaching	-0.046	0.025	0.041	0.206	0.074	0.041
	(0.053)	(0.027)	(0.026)	(0.080)**	(0.040)*	(0.036)
Congruence	0.020	-0.008	-0.010	-0.032	-0.017	-0.013
	(0.016)	(0.011)	(0.008)	(0.017)*	(0.014)	(0.010)
ELA	-0.010	0.009	0.009	0.035	-0.001	0.002
	(0.025)	(0.016)	(0.012)	(0.039)	(0.022)	(0.015)
Learning	0.012	0.014	0.010	-0.033	0.021	0.009
	(0.026)	(0.014)	(0.010)	(0.024)	(0.018)	(0.013)
ELL	-0.013	-0.024	-0.018	0.017	-0.027	-0.021
	(0.026)	(0.015)	(0.010)*	(0.023)	(0.018)	(0.012)*
Misbehavior	0.060	0.014	0.008	0.101	0.015	0.024
	(0.024)**	(0.015)	(0.010)	(0.034)***	(0.021)	(0.013)*
Observations	6560	6560	6560	4462	4462	4462
Number of schools	221	221		164	164	
R-squared	0.486		0.587	0.493		0.587

Figure 1a: Institution effects in Math (x-axis) and ELA (y-axis) for first-year teachers 2000-01 through 2005-06 (institutions with 40 or more teachers with value-added estimates).

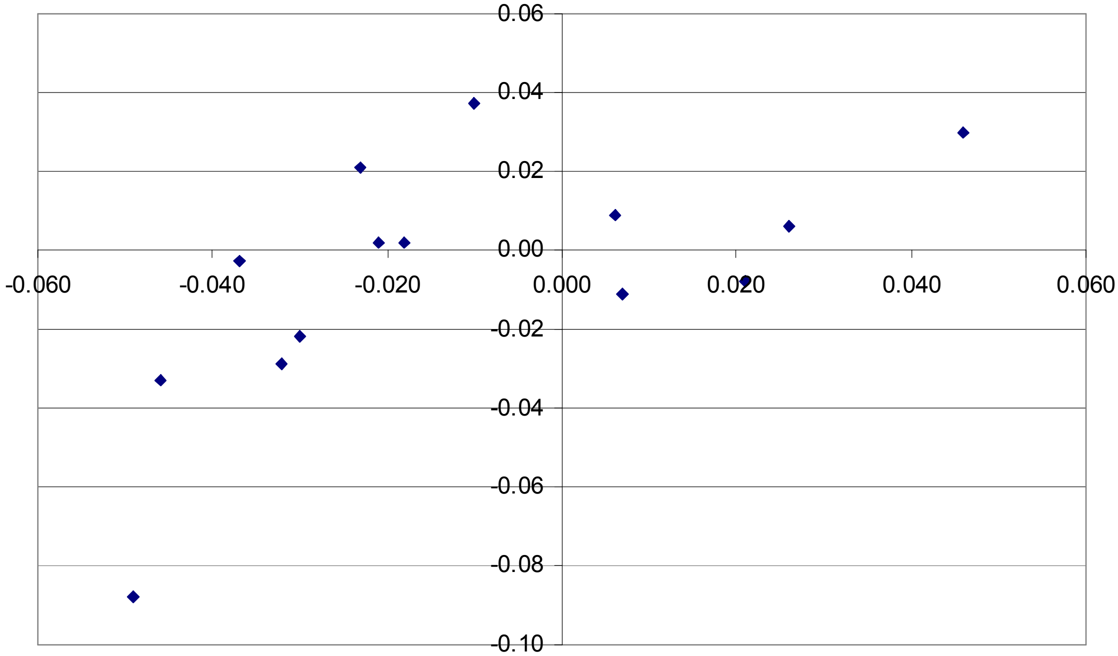


Figure 1b: Program effects in Math (x-axis) and ELA (y-axis) for first-year teachers 2000-01 through 2005-06 (institutions with 40 or more teachers with value-added estimates).

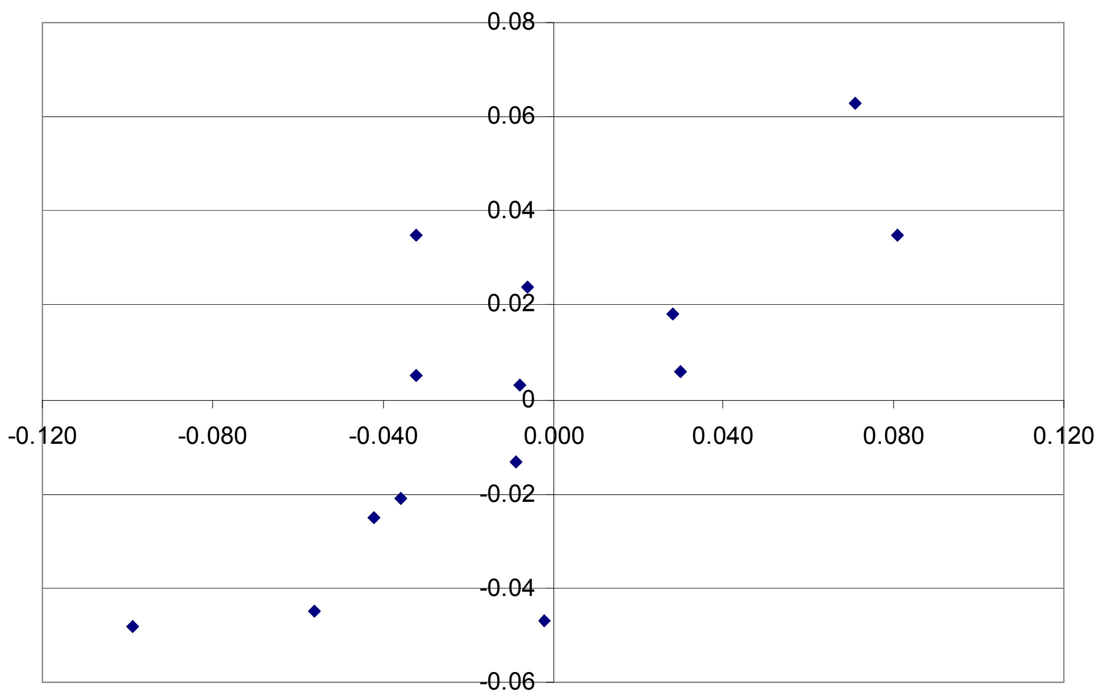


Figure 2. Institution effects in Math (x-axis) and ELA (y-axis) for first-year teachers 2004-05 and 2005-06 (institutions with 20 or more teachers with value-added estimates).

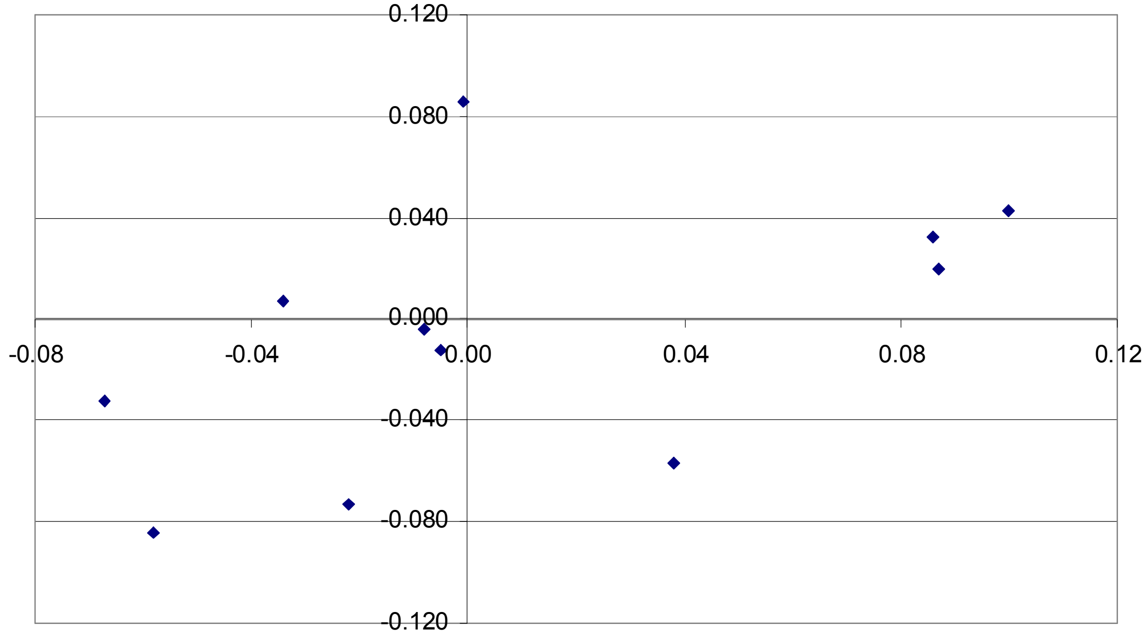
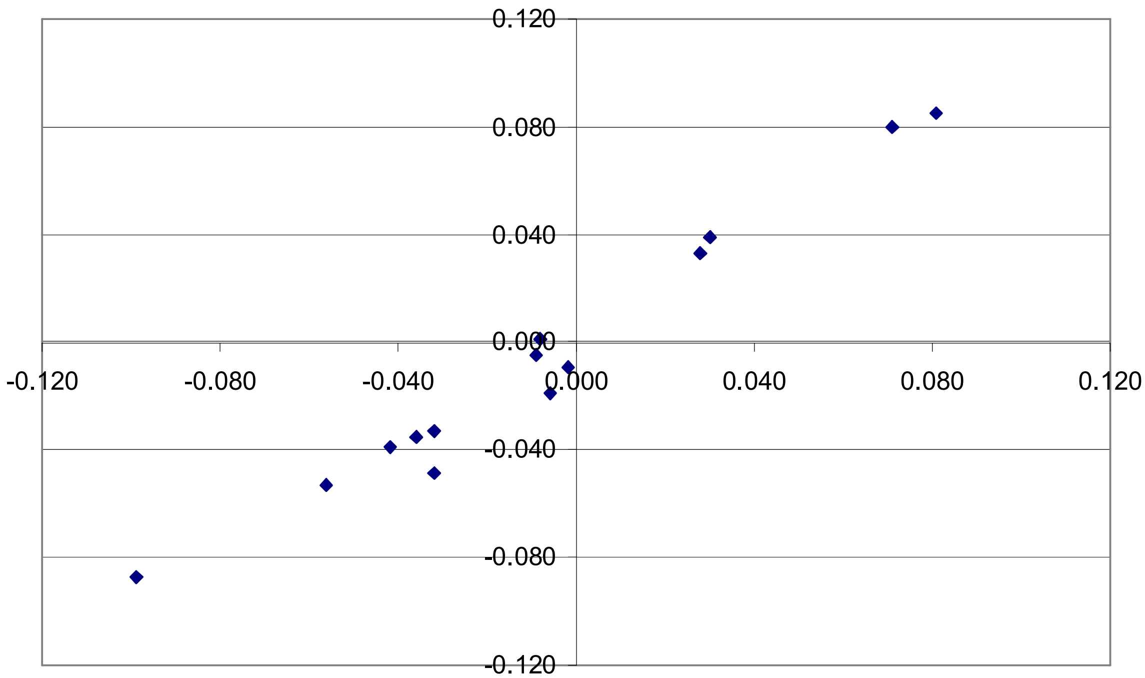


Figure 3. Program effects in Math with no controls for teacher characteristics (x-axis) and controls for teacher characteristics (y-axis) first-year teachers 2000-01 through 2005-06 (programs with 40 or more teachers with value-added estimates).



Appendix

Table A1: Sample Results for Math with Pathway / Institution Effects

Lagged value of standardized math score	6.09E-01 [126.15]	grade 5	1.02E-01 [11.30]
lagstdmscore2	-2.50E-02 [7.02]	grade 6	2.13E-01 [10.08]
Lagged value of standardized ELA score	1.43E-01 [40.46]	grade 7	2.59E-01 [10.57]
lagstdescore2	9.32E-03 [4.94]	grade 8	1.38E-01 [5.15]
changed schools	-2.71E-02 [3.63]	pathinst==0	-7.14E-03 [0.24]
female	-4.11E-02 [10.79]	pathinst==1	7.80E-02 [2.23]
Hispanic	-5.89E-02 [7.31]	pathinst==2	-4.71E-02 [0.83]
African American	-7.66E-02 [8.92]	pathinst==3	4.93E-02 [0.82]
Asian	1.29E-01 [13.22]	pathinst==4	5.38E-02 [1.78]
Other	3.66E-02 [1.30]	pathinst==5	2.21E-02 [0.57]
home language is English	-6.36E-02 [13.08]	pathinst==6	1.99E-03 [0.05]
received free lunch	-5.23E-02 [6.90]	pathinst==7	1.10E-02 [0.31]
received reduced lunch	-1.78E-02 [1.98]	pathinst==8	4.82E-02 [0.74]
Missing information for free/reduced lunch	-5.57E-02 [4.84]	pathinst==9	3.93E-02 [1.03]
Entitled per IEP or lab exam	-5.71E-02 [3.92]	pathinst==10	9.74E-03 [0.33]
NOT entitled to ELL, per IEP or category U	-8.85E-02 [0.71]	pathinst==11	-1.18E-02 [0.26]
ELL-entitled per the school	-6.04E-01 [2.12]	pathinst==12	-1.66E-02 [0.42]
days absent in previous year	-2.91E-03 [15.30]	pathinst==13	3.81E-02 [0.96]
days suspended in previous year	-1.96E-02 [1.52]	pathinst==14	1.42E-02 [0.49]
math class Asian	1.54E-02 [0.22]	pathinst==15	5.83E-02 [1.22]
math class African American	-2.16E-01 [3.35]	pathinst==16	8.94E-03 [0.37]
math class Hispanic	-1.92E-01 [3.24]	pathinst==17	4.40E-01 [6.32]
math class other ethnicity	-4.53E-01	pathinst==18	1.19E-03

	[1.93]		[0.05]
average math class size	-8.09E-04	pathinst==19	-1.41E-02
	[0.86]		[0.55]
math class entitled to IEP or lab exam	6.29E-03	pathinst==20	-5.04E-03
	[0.15]		[0.16]
math class free lunch	-3.57E-02	pathinst==21	-1.15E-03
	[1.60]		[0.05]
math class reduced lunch	9.30E-02	pathinst==22	5.85E-03
	[1.65]		[0.22]
math class english as home language	-2.38E-02	2002	6.77E-03
	[0.59]		[0.57]
math class absent in previous year	-4.18E-03	2003	3.41E-02
	[3.30]		[2.62]
math class suspended in previous year	-5.59E-02	2004	3.17E-02
	[0.58]		[2.28]
math class ELA standard score from previous year	6.98E-02	2005	9.62E-03
	[6.40]		[0.66]
SD of prior-year ELA scores for math class	2.01E-02	2006	2.14E-02
	[1.08]		[1.32]
Observations	89221	Constant	2.38E-01
Number of group(sdbn4)	857		[3.57]
R-squared	0.54		
Robust t statistics in brackets			

Table A2: Description of Variables

For Program Features	
Math courses	Number of math courses the program required for entry or exit in math (subject matter content)
ELA courses	Number of English/Language arts courses the program required for entry or exit in reading or language arts (English, writing communication)
Capstone project	Whether the program required some sort of capstone project (portfolio, research paper, action research project, etc.) for exit
Percent Tenure	Percent Math, English, Learning/Development faculty who were listed as tenure line faculty
Oversight of Student Teach.	Whether the program requires a minimum number of years of teaching experience for its cooperating teachers, picks the cooperating teacher as opposed to the K-12 school or the student teacher selecting, supervisor observes their participants a minimum of five times during student teaching
For Survey Analysis	
Practice	In teacher preparation program, prior to September 2004, the amount of opportunity for practical coursework (listen to individual child read aloud for the purpose of assessing his/her reading achievement, planning a guided reading lesson, study or analyze student math work)
NYC Curriculum	In teacher preparation program, prior to September 2004, the amount of opportunity to learn about New York City's curriculum (review reading and math curriculum)
Basic Skills	Whether teacher placed high amount of emphasis on basic skills (reading, writing, math, speaking) and mastery of subject matter/academic excellence
Misbehavior	Prior to becoming a teacher, the amount of opportunity to develop strategies for handling student misbehavior
Exp to teach ELs	Prior to becoming a teacher, the amount of opportunity to develop specific strategies for teaching English language learners (those with limited English proficiency)
ELA	In teacher preparation program, prior to September 2004, the amount of opportunity to learn how to teach reading/language arts. This is factor created by responses to the following questions: learn about characteristics of emergent readers, learn ways to teach student meta-cognitive strategies for monitoring comprehension, learn ways to teach decoding skills, learn ways to encourage phonemic awareness, learn ways to build student interest and motivation to read, learn how to help students make predictions to improve comprehension, learn how to support older students who are learning to read, learn ways to organize classrooms for students of different reading ability, study, critique, or adapt student curriculum materials, learn how to activate students' prior knowledge, listen to an individual child read aloud for the purpose of assessing his/her reading achievement, plan a guided reading lesson, discuss methods for using student reading assessment results to improve your teaching, and practice what you learned about teaching reading in your field experiences
Math	In teacher preparation program, prior to September 2004, the amount of opportunity to learn how to teach mathematics. This is factor created by responses to the following questions: learn typical difficulties students have with place value, learn typical difficulties students have with fractions, use representations to show explicitly why a procedure works, prove that a solution is valid or that a method works for all similar cases, study, critique, or adapt math curriculum materials, study or analyze student math work, design math lessons, learn how to facilitate math learning for students in small groups, adapt math lessons for students with diverse needs and learning styles, and practice what you learned about teaching math in your teacher preparation program in your field experience
Congruence with Job	Degree of similarity between supervision and feedback received during experience in schools as part of preparation to become a teacher and prior to becoming a full-time classroom teacher; and experience in schools in terms of grade level and subject area
No Student Teaching	No actual time spent student teaching as part of teacher preparation prior to becoming a full-time classroom teacher.

CALDER