Teacher Preparation and Student Achievement

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

EDUCATIONAL EVALUATION AND POLICY ANALYSIS 2009; 31; 416
DOI: 10.3102/0162373709353129

Educational Evaluation and Policy Analysis
December 2009, Vol. 31, No. 4, pp. 416–440
DOI: 10.3102/0162373709353129

Teacher Preparation and Student Achievement

Donald J. Boyd
University at Albany, State University of New York

Pamela L. Grossman
Stanford University

Hamilton Lankford
University at Albany, State University of New York

Susanna Loeb
Stanford University

James Wyckoff
University of Virginia

There are fierce debates over the best way to prepare teachers. Some argue that easing entry into teaching is necessary to attract strong candidates, whereas 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 article is one of the first to estimate the effects of features of teachers' preparation on teachers' value added to student test score performance. 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 1st year.

Keywords: *teacher preparation, teacher quality, teacher effectiveness, achievement gains, program evaluation*

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

This research was supported by funding from the City University of New York (through a grant from the Corporation for National and Community Service under AmeriCorps), the Carnegie Corporation of New York, the National Science Foundation (REC-0337061), the New York State Department of Education, and the Spencer Foundation. We are grateful to the New York City Department of Education and the New York State Education Department for the data employed in this article. We benefited from suggestions by participants at sessions at the APPAM, AERA, AEFA, and AACTE meetings as well as those at university seminars at NYU, Stanford, and the University of Virginia. We benefited particularly from insights and help from Carla Asher, Vicki Bernstein, Marilyn Cochrane-Smith, David Cohen, Ronald Ehrenberg, Jennifer Bell-Ellwanger, Robert Floden, Joseph Frey, 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). CALDER is supported by IES Grant No. R305A060018 to the Urban Institute. We are also grateful for the exceptional research assistance provided by Karen Hammerness, Marsha Ing, Morva McDonald, Luke Miller, Maria Perez, Michelle Reininger, and Matt Ronfeldt. The views expressed in the article are solely those of the authors and may not reflect those of the funders. Any errors are attributable to the authors.

preparation will better serve our nation's children (National Commission on Teaching and America's Future, 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 (cf. Cochran-Smith & Zeichner, 2005; Wilson, Floden, & Ferrini-Mundy, 2001). Lack of evidence creates the opportunity for a myriad of potential "solutions" with regard to teacher preparation and little way to evaluate their promise. This study is a first step toward developing evidence to inform these debates, by looking carefully at the ways in which teachers are prepared and the consequences of that preparation for pupil learning.

Teachers in New York City (NYC) 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, Grossman, et al., 2008). Do these differences in the experiences of teachers in teacher preparation 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 by employing a unique database on teachers, their preparation, and the students they teach. We combine administrative data on individual teachers and students in NYC with detailed information about the components of teacher preparation programs as identified by an analysis of more than 30 programs and a survey of all 1st-year teachers in NYC. Taken together, these data allow us to explore how the preparation of teachers who staff a large, diverse, urban school district influences student achievement.

Background

A large extant research literature on teacher preparation provides some useful information with which to evaluate effective preparation practices.[1] However, much of the research is limited in scope, focuses on inputs to the preparation process rather than outcomes, uses data that are connected only loosely 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 et al. (2001) propose four research elements that would allow further research to address important gaps in our knowledge concerning 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 effects over time.

This study addresses each of these suggestions. First, we employ a detailed analysis of 31 elementary teacher preparation programs, each of which contributes a significant number of teachers to NYC 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 1st-year teachers, we also compare the experiences of teachers across all routes that prepare teachers for NYC 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 (a) analyze documents describing the structure and content of each preparation program; (b) interview program directors, directors of field experiences, and other administrative staff of these programs; (c) survey instructors of math and reading methods courses; and (d) survey program participants and graduates of these programs. We link this information to the program participants, their career decisions, and the outcomes for students they teach.

To address the third point concerning the need to employ measures that are sensitive to program quality and content, we use student achievement data linked to extensive information on program content that we obtain through our analysis of program documents, interviews, observations, and survey data. These measures of preparation experiences provide detailed descriptions of similarities and differences across programs that help us to examine the relationship between specific components of teacher preparation programs and student achievement outcome measures. 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 2 years of teaching and link them to longitudinal data on student achievement.

Our study is most similar to Harris and Sass (2007) in that it focuses on teacher education and employs rich administrative data that follow students and teachers over time. Harris and Sass find that students learn more during the course of a year when their teacher has participated in content-focused professional development; however, they find no relationship between preservice training and student outcomes. Our study differs from the Harris and Sass study both in location—this study focuses on New York, whereas the Harris and Sass study is based on Florida data—and in the measures of preservice preparation. The Florida study uses course taking credits and hours of in-service to measure preparation, whereas our study uses data on the characteristics of programs, courses, and field experiences as collected directly from programs and from surveys of program completers.

A labor market perspective. In this study of teacher education, we observe programs that prepare teachers for NYC schools from what one might term an aerial perspective (cf. Boyd et al., 2006). Such a vantage point has its obvious disadvantages, in particular 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 (cf. 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 available to teachers within a specific locale.

A number of factors affect teacher labor markets, including the structure of pay; teachers' preferences for the characteristics of a school's students; the geographical distribution of students by income and race, teachers, leadership, community, and facilities; and hiring practices including the post-and-fill system of seniority transfers. 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 also on the skills and preparation of the other teachers. By looking at all pathways into teaching in NYC and by doing an in-depth analysis of the largest programs and pathways, we are able to address these interactions.

Teacher labor markets are small geographically. In 2000, 90% of NYC 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, & Wyckoff, 2005). This local focus confirms anecdotal accounts that most of NYC's teachers attended NYC K–12 public schools; it also underscores the importance of studying a labor market in depth when aiming to improve the quality of teacher education and student outcomes. Our focus on this particular school district provides insights into the teacher labor market of the largest school district and affords us the opportunity to explore the nuanced relationships between teacher programs and student outcome measures.

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, Grossman, et al., 2008). Prior research, with the exception of Harris and Sass (2007), which focused on credit hours, has not linked the features of teacher preparation programs to student outcome measures while controlling for teacher and school characteristics. Without this type of analysis, we cannot identify the characteristics or features of the programs that are most likely to influence student achievement and to prepare effective teachers. We have collected information from a broad variety of sources on five areas identified in the literature 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 (cf. Cochran-Smith & Zeichner, 2005; Darling-Hammond, Bransford, LePage, Hammerness, & Duffy, 2005; Valli, Reckase, & Raths, 2003; Wilson et al., 2001). These data help us identify the features of teacher education programs that are related to student achievement gains and fill a gap in prior research on teacher preparation.

Characteristics of teachers. Although our primary concern is to examine the effects of teacher preparation programs, which include both the selection of preparation students and their preparation, we are also interested in separating the selection of teachers from teacher preparation program features and include teacher controls in our models. Some research suggests that teacher characteristics such as race or ethnicity, gender, and achievement as measured by examination scores are related to student learning gains (for review, see Wayne & Youngs, 2003). It is often the case that highly selective teacher preparation programs attract and prepare teachers with different initial characteristics. These controls include gender, ethnicity, age, and certification exam scores. Although we have some additional data on teacher prior achievement including performance on the SAT math and SAT verbal sections and a measure of the competitiveness of the undergraduate institution, these data are more frequently missing. Given the small sample sizes and the strong correlation with the licensure exam score, we choose to use only the licensure scores in the models.

Modeling preparation effects. For this article, 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 linked closely to the daily work of teachers in the classroom, reflecting the perspective that effective professional education is grounded in the practices of the profession (cf. Ball & Cohen, 1999). It is clear that this focus is 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 in a single article.

The use of value-added methodologies to assess teacher effectiveness using statistical adjustments has both advantages and disadvantages (McCaffrey, Koretz, Lockwood, & Hamilton, 2003). Student achievement gains are a logical metric with which to measure the effectiveness of teaching, in that we care most about how teachers affect students, and these measures are widely available. As examples, Goldhaber and Anthony (2007) use data from North Carolina to relate teachers' certification by the National Board for Professional Teaching Standards and math and reading performance of 3rd, 4th, and 5th graders. Clotfelter, Ladd, and Vigdor (2007) use similar data from North Carolina to relate to teacher credentials, years of experience, test scores, and licensure type and certification by subject with math and reading performance of four cohorts of 10th graders. Data from Florida allow Burke and Sass (2008) to link peer effects with student achievement while controlling for teacher and school characteristics, and data from a nationally representative sample allow Figlio and Kenny (2007) to link teacher incentives to student performance. However, although the link to student outcomes is central to measuring teacher effectiveness, there is no consensus on the particular methodology that best captures the relationship between teachers and student performance. Student test scores are never perfect indicators of what students know or of 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 (e.g., Feldt & Brennan, 1989; Messick, 1989).

An alternative would be to analyze how preservice preparation affects more proximal outcomes, including 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; the linkage between teacher behaviors and student outcomes would need to be established separately. A second alternative would be to study student progress, employing measures other than standardized test performance. It is unfortunate that such measures typically are not available. We therefore use value-added measures that rely on achievement on standardized tests, while recognizing their limitations.

The three questions driving the analysis are as follows:

1. What is the distribution of the average value-added scores 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 scores?

Establishing causality is rarely easy, especially with nonexperimental data. The analyses included in this study are just a first step in this direction. We use regression analysis to account for possible biases including those stemming from selection of teachers into programs and then into schools. We see our study as the beginning of a larger exploration of the effect of teacher preparation. Despite the challenges of establishing causal linkages between teacher preparation and student outcomes, the results provide evidence that focusing more on preparation directly linked to practice can produce teachers who are more effective in their 1st year of teaching.

Method

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 alternate-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 assess accurately the effect of pathway and program features. Second, if a policy goal is to improve teaching, in particular in these high-needs schools, then it is 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 (programs to prepare elementary school teachers) providing a substantial number of entering teachers to NYC 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 comes from a survey administered to all 1st-year NYC public school teachers in spring 2005 and, as such, this analysis is limited to the participants from this single cohort of teachers.

For most of the analyses, we fix the effects of schools instead of allowing random effects and assuming a specific distribution of these effects, which is the common approach in hierarchical linear modeling. Our approach is one of many different ways of measuring teacher effects on student achievement. These approaches are referred to as value-added analysis and differ along a number of dimensions including what variables are included and how the error structure is defined (for reviews, see Ballou, Sanders, & Wright, 2004; Lissitz, 2005; McCaffrey et al., 2003; Meyer, 1997). There is no consensus on which approach is optimal, and a number of researchers have expressed concern about the use of value-added methods to link teachers to student performance (e.g., Rothstein, 2009). We present multiple specifications of these models—including fixed school effects, random school effects, and ordinary least squares (OLS) specifications—to test the robustness of our findings.

The first analyses examine teachers aggregated by program and institution. The model for estimating program effects is based on the following equation:

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

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 on 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 features of the preparation itself, 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. It is unfortunate that we have only weak controls for these initial characteristics, although 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 teachers: 2004–2005 and 2005–2006 first-year teachers, 2000–2001 through 2005–2006 first-year teachers, and 2000–2001 through 2005–2006 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 from those who attended a bachelor's degree program at that same institution; those who received their preparation in neither a master's nor 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:

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

The childhood education program characteristics, described in detail below, include (a) the number of math (subject-matter content) courses required for program entry or exit; (b) the number of ELA (subject-matter content such as English, writing, or communication) courses required for program entry or exit; (c) the percentage of the instructors for courses in math methods, learning and development, and ELA methods who are tenure-line faculty; (d) program oversight of student teaching; and (e) whether 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, whereas 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:

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

The measures of self-reported features of experiences in teacher preparation come from responses to the survey of 1st-year teachers described above. They include (a) the extent to which there was an emphasis on opportunities to engage in aspects of teaching practice during coursework; (b) the extent to which coursework covered the NYC curriculum in math and ELA; (c) whether the teacher had student teaching experience; (d) whether the grade and subject at which the teacher did student teaching are similar to his or her current teaching assignments; (e) opportunities to learn about teaching math; (f) opportunities to learn about teaching ELA; (g) opportunities to learn about English language learners; and (h) opportunities to learn about handling student misbehavior. Again, the first four measures, in particular, capture some aspect of the link between preparation and practice.

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 preparation programs whose graduates produce the vast majority of new teachers for NYC public schools and through a survey of all 1st-year teachers in spring 2005. Twenty-six of these programs are more traditional programs in which teachers generally complete both coursework and student teaching prior to becoming a teacher of record; the remaining five programs are alternate-route programs in which teachers enter the classroom after approximately 6 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 fourth through eighth grades to almost all NYC students. The student data, provided by the New York City Department of Education (NYCDOE), consist of a demographic data file and an exam data file for each year from 2000–2001 through 2005–2006. 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 ELA in 2002, because the NYCDOE is not confident that exam scores for that year and grade were measured in a manner that was comparable with the seventh-grade ELA 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. We do not include cases in which a student took a test for the same grade 2 years in a row, or where a student skipped a grade.[2]

The method we used to link students to teachers is based on advice from the NYCDOE. The NYCDOE data do not allow direct linkage of students to their teachers, but because their data systems track the courses taken by each student and the courses taught by each teacher, students can be linked to their courses, which in turn can be linked to the course teacher. For sixth through eighth grades, we used a course-section identifier that indicates the actual teacher of the class, whenever that identifier was available. For third through fifth grades, we used the homeroom identifier. We also used the homeroom identifier for sixth graders who were missing a course-section identifier but were in an elementary school, based on advice from the NYCDOE indicating that in these cases, the homeroom teacher likely was to be the teacher of the class in question. It was not practical to go into classrooms to verify the accuracy of these identifiers. Because some middle schools do not participate in the NYCDOE's middleschool performance assessment system (MSPA) and therefore do not have the course-section identifier linked centrally to teachers, we have a lower match rate for sixth through eighth grades than for third through fifth grades, but never less than two thirds.[3] 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 NYC teachers to data from New York State Education Department (NYSED) databases, using a crosswalk file provided by the NYCDOE that links their teacher file reference numbers to unique identifiers employed by the NYSED. We draw variables for teacher test performance, teacher's initial pathway to entry in NYC, and the teacher's preparation program from the NYSED. The test performance measure comes from the Liberal Arts and Science Test (LAST), the general knowledge exam that teachers must pass to earn certification. This is a 4-hour exam that consists of multiple-choice questions and a written assignment that measure conceptual and analytical skills, critical-thinking and communication skills, and multicultural awareness (Pearson Education, 2008).

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 the NYSED prior to their first teaching job in New York State public schools, with those pathways defined as (a) traditional college-recommended; (b) individual evaluation; (c) temporary license[4]; or (d) 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 that they completed based on information from the program completers file (NYSED).

New York State changed teacher certification program requirements, and these changes took effect for teachers receiving certification beginning in September 2004. Prior to September 2004, teachers could achieve certification in New York State through one of six paths: completing an approved program registered by the NYSED, individual evaluation of educational background, interstate agreement on qualifications of educational personnel, completing an alternate certification program, possession of Northeast regional credential, or emergency regulations for out-of-state applicants. Teachers also needed to pass the New York State teacher certification examinations (LAST, Assessment of Teaching Skills, Content Specialty Test). After September 2004, teachers achieve certification in New York State through one of five paths in addition to passing the certification exams: one of the first four paths above or a certificate issued by the National Board for Professional Teaching Standards. In addition, there are requirements for general education coursework, content coursework, and pedagogical coursework that were not required prior to 2004. As described in Ing and Loeb (2008), the 2004–2005 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.

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 1st-year teachers, on average, teach somewhat lower performing students than do other teachers. NYC serves a diverse group of students: 43% of the students of the childhood education teachers are Hispanic, 28% are Black, 14% are Asian, and 15% are White or other. Fifty-three percent of students speak English at home, and 62% 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 provide information on teachers. For our sample, 61% obtained certification through the college-recommended pathway, a higher percentage than for NYC 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 NYC prior to 2004–2005. The teacher population differs from the student population in both race and gender. Eighty-seven percent of the teachers are female, whereas only 11% are Black and 11% 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 spring and summer 2004 designed to characterize the preparation received by individuals entering teaching in 2004–2005. 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 preservice preparation at 26 college-recommending childhood certification programs, as well as two large alternate-route programs: the New York City Teaching Fellows and Teach for America. Those enrolled in the New York City 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 childhood education programs across 18 institutions.

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 actually is 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

TABLE 1
Descriptive Statistics: For 2001–2006 Program Features (math sample)

Students	*M*	*SD*	# of Students
ELA standardized score	–0.14	0.93	23,549
Math standardized score	–0.12	0.96	27,027
Female	0.50		27,048
Hispanic	0.43		27,048
Black	0.28		27,048
Asian	0.14		27,048
Other non-White race/ethnicity	0.01		27,048
Home language English	0.53		27,048
Receive free lunch	0.62		27,048
Receive reduced-price lunch	0.08		27,048
Lunch missing	0.19		27,048
Entitled to ELL per lab	0.15		27,048
Days absent in previous year	11.17	10.84	17,858
Days suspended in previous year	0.01	0.13	17,858
Classroom Averages	***M***	***SD***	**# of Students**
Asian	0.14	0.21	27,048
Black	0.28	0.30	27,048
Hispanic	0.43	0.29	27,048
Other	0.01	0.02	27,048
Class size	23.91	4.81	27,048
Entitled to ELL per lab	0.15	0.22	27,048
Receive free lunch	0.62	0.28	27,048
Receive reduced-price lunch	0.08	0.09	27,048
Home language English	0.53	0.29	27,048
Days absent in previous year	12.05	6.41	22,744
Days suspended in previous year	0.02	0.08	22,744
Math scores from previous year	–0.04	0.53	18,425
English scores from previous year	–0.08	0.56	17,956
SD: prior math scores	0.72	0.20	18,425
SD: prior ELA scores	0.69	0.19	17,822
Teachers	***M***	***SD***	**# of Teachers**
Path—college recommended	0.61		773
Path—individual evaluation	0.08		773
Path—Teach for America	0.05		773
Path—New York City Teaching Fellows	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

Note. ELA = English language arts; ELL = English language learner.

analysis, we choose to focus on measures that capture the link between the work that teachers do in their preparation and the day-to-day work in the classroom. The program data are 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 percentage of the program instructors in these courses who were tenure-line faculty. Table 2 provides a description of the variables. Appendix B provides a more detailed 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 submeasures: whether the program requires that cooperating teachers have a minimum number of years of teaching experience; whether the program picks the cooperating teacher, as opposed to selection by the K–12 school or the student teacher; and whether a program supervisor observes the participants at least five times during student teaching. 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.

Survey of 1st-year teachers. In spring 2005, we conducted a survey of all 1st-year NYC teachers in which we ask detailed questions about their preparation experiences, the mentoring they received in their 1st year, and their teaching practices and goals.[5] Our overall response rate is 71.5%. We can identify the initial pathway of survey participants but cannot as easily identify the pathway of all individuals in the sampling frame, and thus response rates by pathway are somewhat less precise. Our analysis of the available data suggests that the response rates for the major pathways considered in this article—including college recommended, Teaching Fellows, and Teach for America—are all approximately the same and above 70%. The response rate for the other pathway also just exceeds 70%. Although, 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,

TABLE 2
Program Characteristics and Education Experiences of Teachers

For Program Features	*M*	*SD*
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 teaching	0.95	1.07
For Survey Analysis	***M***	***SD***
Practice	0.03	1.00
NYC curriculum	–0.05	1.02
Had student teaching	0.88	
Congruence with job	0.09	0.98
Math 1	0.09	
Math 2	0.37	
Math 3	0.28	
Math 4	0.26	
ELA	0.12	0.94
Learning	–0.03	0.98
Experienced to teach ELLs	–0.05	1.05
Experience handling misbehavior	–0.01	0.98

Note. ELA = English language arts; NYC = New York City; ELL = English language learner.

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 NYC 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 a range of topics related to teaching ELA, including, for example, characteristics of emergent readers, ways of teaching students meta-cognitive strategies for monitoring comprehension, and how to critique and adapt student curriculum materials (see Appendix B for a complete list of items). 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 composite math variable (alpha = 0.97) includes opportunities related to the teaching of mathematics, including, for example, learning about typical difficulties that students have with place value and fractions, math curriculum materials, and designing of math lessons (see Appendix B for a complete list of items). It is unfortunate that this composite variable is not distributed normally. 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*), from 2.5 to 3.5 (*some opportunity*), and 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? (a) listen to an individual child read aloud for the purpose of assessing his or her reading achievement, (b) plan a guided reading lesson, and (c) 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 the 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 NYC curriculum comes from two questions similar to the ones above. The survey asks teachers about their opportunities to (a) review NYC's reading curriculum and (b) review NYC's 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: *How much actual time did you spend student teaching as part of your teacher preparation prior to becoming a full-time classroom teacher (assume 1 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% 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

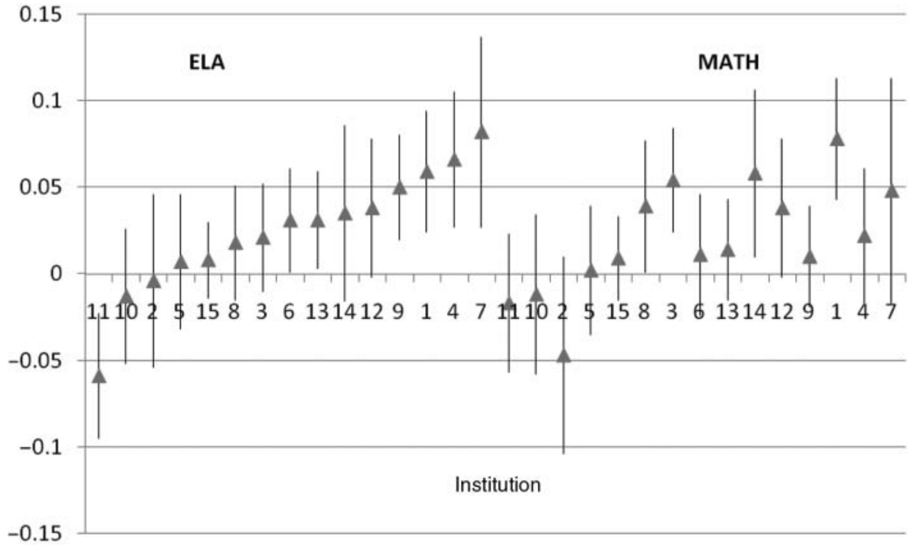

FIGURE 1. *Institution effects in math and English language arts (ELA) for 1st-year teachers, 2000–2001 through 2005–2006.*
Note. Institutions with 40 or more teachers with value-added estimates. Point estimates and standard error ranges.

questions are on a 5-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 with high school teachers.

Finally, we construct a measure of opportunities to learn about learning and the relative emphasis placed on (a) opportunities to study how to handle student misbehavior and (b) 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, (c) develop specific strategies for teaching English language learners, (d) develop specific strategies for teaching students identified with learning disabilities, (e) develop specific strategies for teaching students from diverse racial and ethnic backgrounds, and (f) develop strategies for setting classroom norms.

Results

Program effects. Programs vary in the effectiveness of the teachers they prepare, as measured by student test-score gains. Figure 1 plots the point estimates for each institution from which at least 40 teachers with value-added measures were recommended for certification (see Appendix A for the full results)[6]. The institutions are sorted by the average effectiveness of their teachers in ELA. Three results emerge from our analyses. First, there is some variation across institutions in the average value-added of their graduates, although the standard errors are large and, thus, differences are difficult to detect. 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 highest value-added institution is approximately 0.07 standard deviations in both math and ELA. This magnitude is about the same size as the 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 average learning between Black students and White students. Figure 1 also shows the standard error of each estimate. Although these standard errors are quite large, we can see that there is more than a two-standard error difference between the higher and lower value-added programs. 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. The correlation between math and ELA value added is approximately 0.60.

An alternate approach to presenting the results is to calculate the Empirical Bayes Shrinkage estimates for the point estimates, which adjusts for the variance in the estimates due to measurement error and allows us to look at all the estimates, instead of just those with a large number of teachers. The standard deviation of the Empirical Bayes Estimates for math is approximately 0.05. Using the same process for teacher fixed effects with these same data results in a standard deviation of teacher fixed-effects estimates with Empirical Bayes Shrinkage of 0.13 (Boyd, Lankford, Loeb, & Wyckoff, 2009). Thus, a one standard deviation change in the effectiveness of the preparation institution corresponds to more than a third of a standard deviation in the teacher effect for new teachers.

We see similar patterns when looking at program effects instead of institution effects. For this analysis, institutions are separated into bachelor's programs, master's programs, and other 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). Program effects are also jointly significant across models.

Programs are likely to change over time, in particular 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. We see similar patterns when we model institution effects for 1st-year teachers in the years 2004–2005 and 2005–2006 only (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.

Thus far, our models 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 candidates 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 also are 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. We estimate the 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; that is, including these teacher controls does very little to explain the differences across programs. One possible explanation for the consistency of these findings with and without controls is that the set of controls is incomplete. We would, for example, like to have controls for teachers' experience working with children prior to entering their program and for their leadership skills. However, as one indication that these controls are not without meaning, in another study, similar controls do explain quite of bit of the variation between pathways into teaching for teachers certified in high school mathematics (Boyd, Grossman, et al., 2009).

Program features. There are some systematic differences across programs and institutions in the average value added of their program completers; we next look at how features of these programs predict the value added of their graduates. Table 3 reports estimates of the relationship between particular features of those preparation programs and teachers' value added to student achievement in

TABLE 3
The Relationship Between Program Characteristics and Student Test Performance

	Math			English Language Arts (ELA)		
	2001–2006	2005 & 2006	2001–2006	2001–2006	2005 & 2006	2001–2006
	1st Year	1st Year	2nd Year	1st Year	1st Year	2nd Year
Capstone	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)
Percentage 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)

Note: ~p< .10, *p < .05, **p < .01, ***p < .001.

math using the 2000–2001 through 2005–2006 and the 2004–2005 and 2005–2006 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 features 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, although 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 whether or not the program requires a capstone project, which is often a portfolio of work done in classrooms with students; the extent to which the program oversees the field experiences of its students; content knowledge requirements as measured by content courses in math and ELA; and the percentage 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 because the three are correlated.

The two measures of the link between program experiences and the practice of teaching are significant for 1st-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 0.04 for the capstone project in both ELA and math, and 0.03 in math and 0.01 in ELA for oversight of student teaching. Caution in the interpretation of these results is warranted. Because 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 2nd-year teachers. This result is not surprising given that teachers are likely to learn quite a bit about practice during their 1st year of teaching, and thus, 1st-year differences converge as teachers acquire relevant knowledge and skills on the job.

TABLE 4A

The Relationship Between 1st-Year Teachers' Reported Experiences in Teacher Preparation and Student Test Performance in 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)***
NYC 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)***
Had 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 with job	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^2	0.526		0.629	0.524		0.622

Note. OLS = ordinary least squares; NYC = New York City; ELL = English language learner. *p < .05, **p < .01, ***p < .001.

It is interesting that the content-specific coursework requirements work in a different way. For math, coursework in mathematics is positively associated with teachers' value added in the 2nd year, but not consistently in the 1st year with small effects (about 0.02). Similar for ELA student achievement, ELA coursework has a small positive and significant effect in the 2nd year, but not in the 1st year. This is consistent with some qualitative research on the effect of methods coursework, which also found a 1-year lag in the effect of methods courses (e.g., Grossman et al., 2000). Tenure status does not appear to be important for either 1st- or 2nd-year teachers in math or reading.

Teachers' reports of experiences/survey results. Tables 4a and 4b give the results for the survey analysis for the 2004–2005 cohort of NYC teachers in their 1st 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, respectively, in the program features analyses).

The first two variables in Table 4a, practice and NYC curriculum, are measures of how closely the preparation links to the work that teachers do in their 1st year. For a description of the components of each variable, see Appendix B. 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 1st 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.

TABLE 4B

The Relationship Between 1st-Year Teachers' Reported Experiences in Teacher Preparation and Student Test Performance in English Language Arts (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)***
NYC 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)***
Had 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 with job	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^2	0.479		0.617	0.494		0.623

Note. OLS = ordinary least squares; NYC = New York City; ELL = English language learner. *p < .05, **p < .01, ***p < .001.

One of the two measures of field experience—the congruence between the context in which they had their field experiences and their current teaching position—is also positive across models, although the other measure—whether or not they had student-teaching experience—is not stable. The 0.02 to 0.06 point estimates for congruence are similar in magnitude, again, to the 1st year of teaching experience. None of the other measures show consistent effects for 1st-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 achievement. Students are probably more likely to be involved in activities outside school that contribute to reading achievement than to math learningTables 5a and 5b present similar results for 2nd-year teachers. The patterns are similar to those found in the program feature analysis. With the exception of studying curriculum used in NYC, none of the 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 2nd-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 article by Ing and Loeb (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. The standard deviation that creates the effect size for these estimates is the standard deviation in achievement levels, not the standard deviation

TABLE 5A

The Relationship Between 2nd-Year Teachers' Reported Experiences in Teacher Preparation and Student Test Performance in 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)***
NYC 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)***
Had 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 with job	–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^2	0.553		0.628	0.547		0.622

Note. OLS = ordinary least squares; NYC = New York City; ELL = English language learner. *p < .05, **p < .01, ***p < .001.

in gains, which is substantially smaller. In addition, the standard deviation in test-score gains is inflated artificially by measurement error, in particular at scores that are farther from the mean. If we use the standard deviation in gains and adjust for measurement error in the tests, 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 4% to 16% of a standard deviation of the true gain in students' achievement over the course of a school year.

Conclusion

In summary, the results suggest that there is variation across programs in the average effectiveness of the teachers they are supplying to NYC schools, with some programs graduating teachers who have a significantly greater effect 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, although 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. Our data and methods are imperfect and the results are suggestive rather than clearly establishing cause. In particular, some programs may appear stronger not because they provide better opportunities for students to learn to teach but because they are able to attract better teacher candidates. This ability to attract good teachers is a program characteristic, and when we assess the effectiveness of different programs, we want to include this ability to attract talent as one of its features. However, when trying to identify what aspects of the preparation contribute to teacher effectiveness in the classroom, we do need to control for differences in teachers' entering characteristics. In this study, we make these adjustments in a common regression framework, including controls for background characteristics.

TABLE 5B

The Relationship Between 2nd-Year Teachers' Reported Experiences in Teacher Preparation and Student Test Performance in English Language Arts (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)
NYC 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)**
Had 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 with job	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^2	0.486		0.587	0.493		0.587

Note. OLS = ordinary least squares; NYC = New York City; ELL = English language learner. *p < .05, **p < .01, ***p < .001.

These controls may not be sufficient to fully account for entering differences in teachers.

With this concern in mind, we find one particular aspect of program preparation consistently related to student outcomes. Teacher preparation that focuses more on the work of the classroom and provides opportunities for teachers to study what they will be doing as 1st-year teachers seems to produce teachers who, on average, are more effective during their 1st 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 1st-year teachers to NYC 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 1st year of teaching. Similarly, teachers who have had the opportunity to review curriculum used in NYC 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 1st-year teachers.

Learning that is grounded in the practice of teaching—such as that proxied by the capstone project, studying curricula, and oversight of student teaching—is associated positively with student achievement gains in the 1st year, and content learning—as proxied by disciplinary coursework requirements—is associated positively with learning in the 2nd year. Although our study is not designed to identify the mechanisms for these effects, there are possible explanations. For example, practice in the day-to-day work of teaching may facilitate teachers' transition into the classroom during their 1st year, a typically challenging time. Content knowledge is likely important for teaching but may not distinguish more and less effective teachers until the 2nd year, when teachers are more comfortable with the basic practices of teaching. Our understanding both of the mechanisms of these effects and of characteristics of effective preparation would benefit from additional research that takes a closer look at these features linked to practice and to content preparation.

The estimated effects of many of the measures of teacher preparation are educationally important, about the same size as the effect of the 1st year of teaching experience. As noted in Boyd, Lankford, Loeb, and Wyckoff (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 article by a factor of four.

We also find some support for the hypothesis that math content preparation improves the outcomes of students of 2nd-year teachers, but not 1st-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 1st year, whereas those with stronger content knowledge are able to make use of that knowledge by their 2nd 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 influence student achievement among 1st-year or 2nd-year teachers. This lack of findings does not necessarily mean that these components of teacher education are not important; there may be less variation across programs in some of these areas, or we may not have measured these features effectively, as we discuss below.

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 effect 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 the engagement of its faculty, just as the percentage of tenure-line faculty teaching core courses in teacher education 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 that 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 United States, in particular for alternate-route programs, and thus our study does not include individuals who have low absolute levels of many preparation attributes (Boyd, Grossman, 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. Although 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 either of the range of student learning that we care about or of teachers' effect 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 preservice preparation can influence teacher effectiveness, at least the effectiveness of 1st- and 2nd-year teachers.

This study also suggests a way of moving beyond research that tries to compare alternate pathways with more traditional pathways into teaching. This distinction between alternate and traditional pathways is often not helpful in this policy debate because of the great variety of programs within each group and the overlap of many features between programs of different types (e.g., Boyd, Grossman, et al., 2008; Humphrey, Wechsler, & Hough, 2008). These categories rarely capture consistent differences in recruitment, selection, or preparation of teachers. Our research pushes the policy discussion forward by exploring particular features of programs, whether classified as alternate or traditional, that contribute to gains in student achievement. Our goal is to provide information that is useful for designing and implementing effective teacher preparation programs. There is a wide variety of individual features and combinations of features that make up teacher preparation programs, and this study is just a first step in linking these program features to student learning gains.

Appendix A
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-price lunch	–1.78E–02 [1.98]	pathinst==8	4.82E–02 [0.74]
Missing information for free or reduced-price 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	–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 [1.93]	pathinst==18	1.19E–03 [0.05]
Average math class size	–8.09E–04 [0.86]	pathinst==19	–1.41E–02 [0.55]
Math class entitled to IEP or lab exam	6.29E–03 [0.15]	pathinst==20	–5.04E–03 [0.16]
Math class free lunch	–3.57E–02 [1.60]	pathinst==21	–1.15E–03 [0.05]
Math class reduced-price lunch	9.30E–02 [1.65]	pathinst==22	5.85E–03 [0.22]
Math class English as home language	–2.38E–02 [0.59]	2002	6.77E–03 [0.57]
Math class absent in previous year	–4.18E–03 [3.30]	2003	3.41E–02 [2.62]
Math class suspended in previous year	–5.59E–02 [0.58]	2004	3.17E–02 [2.28]
Math class ELA standard score from previous year	6.98E–02 [6.40]	2005	9.62E–03 [0.66]
SD of prior-year ELA scores for math class	2.01E–02 [1.08]	2006	2.14E–02 [1.32]
Observations	89221	Constant	2.38E–01
Number of group (school id)	857		[3.57]
R^2	0.54		
Robust *t* statistics in brackets			

Note. ELA = English language arts; IEP = Individualized Education Program; ELL = English language learner.

Appendix B
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 ELA 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
Percentage tenure	Percentage math, English, learning/development faculty who were listed as tenure-line faculty
Oversight of student teaching	The oversight-of-student-teaching variable combines three submeasures: whether the program requires that cooperating teachers have a minimum number of years of teaching experience, whether the program picks the cooperating teacher as opposed to selection by the K–12 school or the student teacher, and whether a program supervisor observes the participants at least five times during student teaching. Because these measures are highly correlated, we combine these binary variables into a single sum to measure the program's oversight of 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 or her reading achievement, plan 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 NYC's curriculum (review reading and math curriculum). This variable comes from two questions about teachers' opportunity to (a) review NYC's reading curriculum and (b) review NYC's mathematics curriculum. We 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 provides similar results.
Congruence with job	This variable reflects the 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.
Had student teaching	This variable reflects whether the teacher spent any time student teaching as part of teacher preparation prior to becoming a full-time classroom teacher.
Math	In teacher preparation program, prior to September 2004, the amount of opportunity to learn how to teach mathematics. This is a factor (alpha = 0.97) created by responses concerning the following items: 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. 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. It is unfortunate that this composite variable is not normally distributed. As a result, we split the composite variable into four groups: *Math 1* (0–1.0 ranking of opportunities), *Math 2* (1.0–2.5 ranking of opportunities), *Math 3* (2.5–3.5 ranking of opportunities), and *Math 4* (greater than 3.5 ranking of opportunities).

Appendix B (Continued)

ELA	In teacher preparation program, prior to September 2004, the amount of opportunity to learn how to teach reading/language arts. This is a factor (alpha = 0.96) created by responses to the following questions: 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 adapt student curriculum materials; learn how to activate students' prior knowledge; listen to an individual child read aloud for the purpose of assessing his or 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.
Learning	Prior to becoming a teacher, the amount of opportunity to develop student stages of child development and learning
Experienced to teach ELLs	Prior to becoming a teacher, the amount of opportunity to develop specific strategies for teaching ELLs (those with limited English proficiency)
Misbehavior	Prior to becoming a teacher, the amount of opportunity to develop strategies for handling student misbehavior

Note. ELA = English language arts; NYC = New York City; ELL = English language learner.

Notes

[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, LePage, Hammerness, and Duffy (2005), Feiman-Nemser (1983, 1990); Goodlad (1990), Holmes Group (1986); Levine (2006), Allen (2003); and Wayne and Youngs (2003).

[2]We also exclude observations for classrooms with fewer than 10 or more than 50 students.

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

[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 the New York State Education Department to allow a specific individual to teach in a specific school for a temporary period.

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

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

References

Allen, M. (2003). *Eight questions on teacher preparation: What does the research say?* Denver, CO: Education Commission of the States.

Ball, D. L., & Cohen, D. (1999). Developing practice, developing practitioners. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco: Jossey-Bass.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, *29*, 37–65.

Boyd, D., Grossman, P., Hammerness, K., Lankford, H., Loeb, S., McDonald, M., et al. (2008). Surveying the landscape of teacher education in New York City: Constrained variation and the challenge of innovation. *Education Evaluation and Policy Analysis*, *30*(4) , 319-343.

Boyd, D., Grossman, P., Hammerness, K., Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2009). *Recruiting effective math teachers, how do math immersion teachers compare? Evidence from New York City*. Working paper.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., Michelli, N., & Wyckoff, J. (2006). Complex by design: Investigating pathways into teaching in New York City schools. *Journal of Teacher Education*, *57*(2), 155–166.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). The draw of home: How teachers' preferences for proximity disadvantage urban schools. *Journal of Policy Analysis and Management*, *24*(1), 113–132.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2008, April 22–24). *Measuring effect sizes: The effect of measurement error*. Working paper prepared for the National Conference on Value-Added Modeling, University of Wisconsin-Madison.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2009, March). *Who leaves? Teacher attrition and student achievement* (CALDER Working Paper 23).

Burke, M. A., & Sass, T. R. (2008). *Classroom peer effects and student achievement* (CALDER Working Paper 18). Retrieved March 3, 2009, from http://www.urban.org/UploadedPDF/1001190_peer_effects.pdf

Carnegie Forum on Education and the Economy. (1986). *A Nation Prepared: Teachers for the 21st century*. Washington, D.C.: The Task Force on Teaching as a Profession.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). *How and why do teacher credentials matter for student achievement?* (CALDER Working Paper 2). Retrieved March 3, 2009, from http://www.caldercenter.org/PDF/1001058_Teacher_Credentials.pdf

Cochran-Smith, M., & Zeichner, K. (Eds.). (2005). *Review of research in teacher education*. Washington, DC: American Educational Research Association.

Darling-Hammond, L. (Ed.). (2000). *Studies of excellence in teacher education*. Washington, DC: American Association of Colleges of Teacher Education.

Darling-Hammond, L., Bransford, J., LePage, P., Hammerness, K., & Duffy, H. (Eds.). (2005). *Preparing teachers for a changing world*. San Francisco: Jossey-Bass.

Feiman-Nemser, S. (1983). Learning to teach. In L. Shulman & G. Sykes (Eds.), *Handbook of teaching and policy* (pp. 150-170). New York: Longman.

Feiman-Nemser, S. (1990). Teacher preparation: Structural and conceptual alternatives. In W. R. Houston, M. Huberman, & J. Sikula (Eds.), *Handbook of research in teacher education* (pp. 212-233). New York: Macmillan.

Feldt, L., & Brennan, R. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105–146). New York: Macmillan.

Figlio, D. N., & Kenny, L. W. (2007). Individual teacher incentives and student performance. *Journal of Public Economics*, *91*, 901–914.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching. *The Review of Economics and Statistics*, *89*(1), 134–150.

Goodlad, J. I. (1990). *Teachers for our nation's schools*. San Francisco: Jossey-Bass.

Grossman, P. L., Valencia, S. W., Evans, K., Thompson, C., Martin, S., & Place, N. (2000). Transitions into teaching: Learning to teach writing in teacher education and beyond. *Journal of Literacy Research*, *32*, 631–662.

Harris, D. N., & Sass, T. (2007). *Teacher training, teacher quality and student achievement* (CALDER Working Paper 3).

Holmes Group. (1986). Tomorrow's teachers: A report of the Holmes Group. East Lansing, MI: Author.

Humphrey, D., Wechsler, M., & Hough, H. (2008). Characteristics of effective alternative teacher certification programs. *Teachers College Record*, *110*(1), 1–63.

Ing, M., & Loeb, S. (2008). Assessing the effectiveness of teachers from different pathways: Issues and results. In P. Grossman & S. Loeb (Eds.), *Alternative routes to teaching: Mapping the new landscape of teacher education* (pp. 157–185). Cambridge, MA: Harvard Education Press.

Levine, A. (2006). *Educating school teachers*. New York: The Education Schools Project.

Lissitz, R. (Ed.). (2005). *Value-added models in education: Theory and applications*. Maple Grove, MN: JAM Press.

McCaffrey, D. F., Koretz, D. M., Lockwood, J. R., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.

Meyer, R. (1997). Value-added indicators of school performance: A primer. *Economics of Education Review*, *16*(3), 183–301.

National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future*. Woodbridge, VA: Author. Retrieved August 6, 2008, from http://www.nctaf.org/documents/WhatMattersMost.pdf

Pearson Education, Inc. (2008). *New York State Teacher Certification Exam Registration Bulletin*. Retrieved April 28, 2009, from http://www.nystce.nesinc.com/NY13_teachercertification.asp

Rothstein, J. (2009). *Teacher quality in educational production: Tracking, decay, and student achievement* (Working paper). Retrieved April 28, 2009, from http://www.princeton.edu/~jrothst/working papers/rothstein_teacherfx_20090325.pdf

U.S. Department of Education. (2002). *Meeting the highly qualified teachers challenge*. U.S. Department of Education, Office of Postsecondary Education, Office of Policy Planning and Innovation. Retrieved August 6, 2008, from http://www.ed.gov/about/reports/annual/teachprep/2002title-ii-report.pdf

Valli, L., Reckase, M., & Raths, J. (2003, April). *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.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, *73*, 89–122.

Wilson, S., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations*. Center for the Study of Teaching and Policy, University of Washington. Retrieved August 6, 2008, from http://depts.washington.edu/ctpmail/PDFs/TeacherPrep-WFFM-02-2001.pdf

Authors

DONALD J. BOYD, Ph.D., is the deputy director of the Center for Policy Research at the University at Albany, State University of New York, Albany, NY 12222; boydd@rockinst.org. His work includes research on teacher labor markets and on school district finances. He is part of a research team that is examining the relationships between teacher preparation and student academic performance and teacher labor market choices.

PAMELA L. GROSSMAN, Ph.D., is the Nomellini-Olivier Professor of Education at Stanford University, 485 Lasuen Mall, Stanford, CA 94305-3096; pamg@stanford.edu. Her research interests include teacher education and professional education more broadly, teacher knowledge, and the teaching of English in secondary schools.

HAMILTON LANKFORD, Ph.D., is a professor of educational administration and policy studies in the School of Education at the University at Albany, State University of New York, Education 317, 1400 Washington Avenue, Albany, NY 12222; hamp@albany.edu. His areas of specialization are education policy, the economics of education, and quantitative methods.

SUSANNA LOEB, Ph.D., is a professor of education at Stanford University, 520 Galvez Mall, CERAS Building, 5th Floor, Stanford, CA 94305-3084; sloeb@stanford.edu. She also is the director of the Institute for Research on Education Policy and Practice at Stanford, a codirector of Policy Analysis for California Education, a faculty research fellow at the National Bureau of Economic Research, a member of the Policy Council of the Association for Policy Analysis and Management, a coeditor of *Educational Evaluation and Policy Analysis*, a member of the editorial board of *Education Finance and Policy*, and president-elect of the American Education Finance Association.