Available online at www.sciencedirect.com
ScienceDirect
Economics of Education Review 27 (2008) 615–631
Economics of Education Review
www.elsevier.com/locate/econedurev

What does certification tell us about teacher effectiveness? Evidence from New York City

Thomas J. Kane[a,*], Jonah E. Rockoff[b], Douglas O. Staiger[c]

[a]Harvard Graduate School of Education, Gutman Library, Room 455, Appian Way, Cambridge, MA 02138, USA
[b]Columbia Business School, 3022 Broadway, Uris Hall 603, New York, NY 10027, USA
[c]Department of Economics, Dartmouth College, Hanover, NH 03755, USA

Received 15 May 2006; accepted 19 May 2007

Abstract

We use six years of panel data on students and teachers to evaluate the effectiveness of recently hired teachers in the New York City public schools. On average, the initial certification status of a teacher has small impacts on student test performance. However, among those with the same experience and certification status, there are large and persistent differences in teacher effectiveness. Such evidence suggests that classroom performance during the first two years is a more reliable indicator of a teacher's future effectiveness. We also evaluate turnover among teachers by initial certification status, and the implied impact on student achievement of hiring teachers with predictably high turnover. Given modest estimates of the payoff to experience, even high turnover groups (such as Teach for America participants) would have to be only slightly more effective in each year to offset the negative effects of their high exit rates (I2, J24).

Keywords: Teacher quality; Teacher certification; Alternative certification; Teacher experience; Valve-added; Student achievement

1. Introduction

Traditionally, federal and state governments have regulated teacher quality with *ex ante* certification requirements. To gain legal permission to teach, prospective teachers have been required to study full-time for one or two years in an approved education program. However, recruiting difficulties have forced many districts to hire large numbers of uncertified or alternatively certified teachers. Despite the ubiquity of alternative teacher certification (AC) programs, there is little research on the impacts on student achievement. We examine the relationship between teachers' certification status and student achievement in New York City (NYC), using students' test scores in math and reading in grades four through eight.

Besides having the largest enrollment in the United States, NYC is a major employer of certified, uncertified, and alternatively certified teachers. During the school years 1999–2000 to 2004–2005, New York hired more than 50,000 teachers, of which 46% were certified, 34% uncertified, and 20% AC teachers. The vast majority of AC teachers in New York are recruited through the NYC Teaching Fellows program, while the

*Corresponding author. Tel.: +1 617 496 4359; fax: +1 617 496 2614.
E-mail address: kaneto@gse.harvard.edu (T.J. Kane).

0272-7757/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2007.05.005

remainder comes primarily from international recruitment and the Teach for America (TFA) program (a non-profit entity that recruits and sends AC teachers to districts throughout the nation).

When studying their impacts on math achievement, we find no difference between Teaching Fellows and certified teachers or between uncertified and certified teachers. Classrooms of students assigned to internationally recruited teachers scored 0.02 standard deviations lower in math than similar classrooms assigned to certified teachers, while classrooms of students assigned to TFA corps members scored 0.02 standard deviations higher relative to certified teachers. (We measure teacher effectiveness in terms of test scores among NYC students, where test scores have been normalized by year and grade level to have a mean of zero and standard deviation of one.) In reading, students assigned to Teaching Fellows underperformed students assigned to certified teachers by 0.01 standard deviations. All of the above reflect average differences in student impacts between groups of teachers, controlling for years of teaching experience. These are the only instances in which we find that a teacher's initial certification status has statistically significant implications for student achievement.

Consistent with other studies, we also find that both certified and AC teachers' effectiveness improves with the first few years of experience. We examine teacher turnover and its implications for student achievement. Critics of alternative certification programs argue that such programs actually harm student achievement by bringing in candidates with less commitment to teaching as a career and, as a result, have higher turnover rates (Darling-Hammond, 2007). However, while turnover was indeed high among TFA corps members—reflecting their two-year commitment—our results suggest that Teaching Fellows and traditionally certified teachers had very similar retention rates. Moreover, even the higher prevalence of novice teachers among TFA participants has only a small negative effect on student achievement—about 0.02 standard deviations of achievement in math and reading. On net, the modest negative impact of higher turnover is roughly offset by the slightly higher initial effectiveness of TFA participants.

Although initial certification status provides little predictive power, there are large differences in teacher effectiveness within all of these groups. We estimate that the average value-added among the top quartile of elementary school math teachers is 0.33 standard deviations greater than that generated by the bottom 25%—almost 10 times the magnitude of any difference associated with initial certification status! Thus, although shifting the mix of teachers with different types of certification does not appear to be a useful tool for improving student achievement, selectively retaining only the most effective teachers appears to be a much more promising strategy (Gordon, Kane, & Staiger, 2006).

There exist only a few high-quality studies of AC teachers, most notably Decker, Mayer, and Glazerman (2004), who conducted a randomized evaluation of the TFA program. They find that teachers recruited through TFA are significantly more effective than both uncertified and certified teachers at math instruction and statistically indistinguishable in reading instruction.[1] Although this is an extraordinarily important study for evaluating the impact of TFA corps members in the districts and schools where they are operating, it offers few conclusions that can be generalized to other AC programs. TFA is unique among AC programs in that it is highly selective, draws from a national pool of applicants, assigns teachers to schools nationwide, and recruits individuals with an explicit two year teaching commitment. Moreover, TFA corps members are typically placed in a small number of high needs schools.

Although our study does not have the benefit of being able to randomly assign teachers to classrooms, our results allow us to compare impacts for those entering teaching from a number of different routes and for teachers working in a large number of schools. Boyd, Grossman, Lankford, Loeb, and Wyckoff (2005a, 2005b) also use data from NYC to evaluate differences in teacher effectiveness by initial certification status. Our work differs from theirs in several ways. We use an additional year of testing data; we incorporate application data from the Teaching Fellow program to study how the fellows are selected; and we estimate the signal variance in teacher effectiveness within each certification group. We view the last distinction as important, since we interpret between-group differences in effectiveness in light of these within-group differences. Although there are statistically significant differences between groups, such differences

[1] Raymond, Fletcher, and Luque (2001) and Darling-Hammond, Holtzman, Gatlin, and Heilig (2005) also report on the impact of TFA corps members on student achievement.

Table 1
Characteristics of teachers by certification and program

	Regular certified	Regular uncertified	Teaching fellow	Teach for America	International teacher
Number of teachers	23,306	15,910	8976	1544	2052
Black (%)	11.1	30.9	18.9	8.9	40.1
Hispanic (%)	8.9	17.7	11.2	9.1	7.9
Female (%)	79.8	66.7	66.7	72.6	72.9
Median age at hire	27	29	27	23	36
Graduate education at hire (%)	35.5	15.0	13.9	3.6	60.1
College SAT math pctile	59	55	68	74	n/a
College SAT verbal pctile	63	59	73	79	n/a

Note: This table includes data for teachers hired during the 1999–2000 to 2004–2005 school years. Age at hire is calculated as the difference between school year and birth year. College data is unavailable for over 99% of international teachers, so we do not report means of SAT percentiles for this group.

are generally dwarfed by the differences in effectiveness within groups of teachers.

2. Alternative teacher certification in NYC

In the school year 1999–2000, approximately 60% of all new teachers hired by the NYC Department of Education (DOE) were uncertified. Recruiting difficulties were more severe in schools with low average achievement levels; 73% of new hires in 1999–2000 were uncertified in both elementary and high schools in the lowest deciles of pass-rates on end of year math examinations. Motivated by changes in New York State law that made certification requirements more stringent, the DOE cut hiring of uncertified teachers and expanded its recruitment of alternatively certified teachers. Over the school years 1999–2000 to 2004–2005, the fraction of uncertified hires fell from 60% to 7%, while the fraction of AC hires rose from 2% to 36%. It is unlikely that this shift was just a re-labeling of individuals who would have otherwise become uncertified teachers. The populations of uncertified teachers and AC teachers differ on a number of characteristics (see Table 1).

New York City Teaching Fellows Program (NYCTF), created in the summer of the year 2000, has accounted for most of the growth in alternative certification in the district. The number of Teaching Fellows hired grew from 350 in the school year 2000–2001 (less than 5% of new hires) to 2500 in the school year 2003–2004 (more than 30% of new hires) and 2000 in the school year 2004–2005 (more than 25% of new hires).

New York recruits AC teachers through several other sources: TFA, the Peace Core Fellows Program, and the Teaching Opportunity Program Scholars. The number of TFA corps members and international teachers is substantial and we present results for them in our basic analysis. In contrast, there are only a handful of Peace Corps Fellows and Teaching Opportunity Program Scholars in our data, and we therefore do not discuss them further. The DOE also recruits teachers from other countries. International teachers have been certified in their home country and are not AC teachers. However, given their non-traditional recruitment, we consider them separately from certified teachers in our analysis.

In our analysis, we classify teachers based on their certification status in their year of hire. Certified, uncertified, international, and AC teachers differ along a number of observable dimensions (Table 1). For example, the fraction of teachers who are black or Hispanic is lower among regularly certified teachers and TFA corps members (about 20%) than among Teaching Fellows (30%), uncertified teachers (49%), or international teachers (48%). Not surprisingly, certified teachers and international recruits are more likely to have graduate education than other groups. There are also differences across groups in the selectivity of undergraduate institution. As measured by median SAT scores, certified and uncertified teachers attended significantly less selective colleges than Teaching Fellows or TFA corps members.[2]

[2]We have data from the DOE on the undergraduate institution attended for about 25% of teachers hired during our sample who are not Teaching Fellows or international recruits. We have this data for 95% of Teaching Fellows (98% of Teaching Fellows who attended a US institution) because we have data from applications. We have this data for less than 1% of international

Table 2
School characteristics initial teacher certification status

	Teachers hired 1999–2000 to 2003–2004				
	Regular certified (%)	Regular uncertified (%)	Teaching fellow (%)	Teach for America (%)	International program (%)
Black or Hispanic	72.8	78.4	88.1	96.2	89.1
Free lunch	70.5	72.4	80.3	88.9	77.4
English language learner	13.4	13.6	15.3	19.0	14.2
Special education	10.3	10.3	10.9	13.2	10.2
Math pass (ES)	78.1	69.8	69.2	68.6	66.3
English pass (ES)	83.7	79.3	76.5	75.9	77.4
Math pass (HS)	68.9	66.8	65.6	n/a	63.5
English pass (HS)	75.3	75.3	72.5	n/a	71.6
Graduation rate	58.8	57.2	55.4	n/a	53.5

Note: "Math/English pass (ES)" denotes the fraction of students scoring above level 1 on city and state exams for grades 3–8. "Math/English pass (HS)" denotes the fraction of students scoring above a 55 on the state Regents examinations. High school information is not given for Teach for America corps members because very few of these teachers work in high schools; less than 3% of all TFA corps members teach in schools that do not serve grades 3–8.

There are also substantial differences in the characteristics of students taught by different groups of teachers (Table 2).[3] Uncertified teachers, Teaching Fellows, TFA corps members, and international teachers all tend to teach in schools that—relative to those employing certified teachers—have a higher fraction minority students, higher fraction eligible for free lunch (our best measure of household poverty), bilingual education, special education, lower fraction passing elementary and high school tests in reading and math, and lower high school graduation rates.

2.1. Teaching fellow selection

Teaching Fellows are selected in a three-step process. First, applicants submit information on their demographic characteristics, their previous work experience, their academic history, and their qualifications for teaching particular subject areas. The application also requires a personal essay. Approximately 60% of applicants are then invited for an interview. The interview process lasts approximately 4–5 h and has multiple parts: applicants give a five minute lesson on a subject of their choosing, participate in a guided discussion, write an essay on a topic not given out in advance, and sit through a one-on-one interview. Approximately 50% of those interviewed have their applications forwarded to the final stage. These applications are reviewed by a committee, and about 85% are offered positions. Roughly 75% of offers are accepted, and 85% of those accepting offers complete the necessary training to gain alternative certification. Overall, less than 15% of all applicants become Teaching Fellows.

On average, the selection process favors those with stronger academic credentials.[4] Among those screened out before the interview, the median quantitative SAT score of applicants' undergraduate institution was at the 50th percentile of the national distribution. For applicants making it to the interview before being rejected and for those eventually offered a position, the median quantitative SAT score was at the 57th and 69th percentile

(*footnote continued*)
recruits because so few attended US institutions. The distribution of median SAT scores of Teaching Fellows with and without data from the DOE is very similar, leading us to believe that the availability of this data is not strongly related to the selectivity of undergraduate institution.

[3]We present school level averages, which do not take into account any sorting of students within schools. In addition, school level data for the school year 2004–2005 were not yet available, so teachers in this year are assigned the characteristics of the students in the school from the prior year.

[4]We have data on the Teaching Fellow selection process for the school years 2003–2004 and 2004–2005. Though we cannot be certain that selectivity was constant across time, NYCTF officials estimate that the fraction of applicants who became Teaching Fellows was 14% in the school year 2000–2001. This is about the same as the fraction who became Teaching Fellows in 2003–2004 and 2004–2005, though the program was more than five times as large.

of the national distribution, respectively.[5] The selection process also favored those with higher verbal SAT scores and undergraduate grade point averages.

Following selection, Teaching Fellows attend a seven-week summer training course and assist in a summer school classroom. They must also pass New York State certification examinations and begin coursework in a certification program in an approved graduate school of education. The fellows receive a tuition grant to help pay for their coursework. Otherwise, Teaching Fellows are no different than their colleagues with respect to union membership, salary, and general rights and privileges afforded to teachers in New York.

3. Data on students and their teachers in NYC

For students, we assembled data on demographic background, attendance, suspensions from school, test performance, eligibility for free lunch, special education and bilingual education, and a student identification number. The dataset also contains teacher identification numbers for students' math and reading teachers, which were often the same teacher for elementary school students. We capture data on teachers using information from the DOE payroll system. This gives us information on teachers' certification and their position on a salary schedule. We use the salary schedule variables to construct measures of teachers' education and experience.

Our study focuses on teachers of math and reading in grades four through eight in 1998–1999 through 2004–2005. In these grades, 95% of elementary school students and 82% of middle school students were matched to their math and reading teacher(s). Middle school students had a lower match rate because roughly 20% of middle schools did not enter information on teacher codes. We found no statistically significant relationship between schools' reporting of course information and school average student characteristics (Results available upon request). We dropped classrooms in which fewer than seven or more than 45 students are tested (corresponding to the 1st and 99th percentile), because we were concerned that classes with extremely low or high numbers of students may be incorrectly identified in our data. We dropped classrooms where the teacher did not work in the school during the entire year and classrooms taught by teachers listed as working in more than one school per year. We also dropped classrooms in any school-year cell for which less than 75% of students were successfully matched to a teacher. Finally, we dropped classrooms where 25% or more of the students receive special education. This eliminates over 95% of special education students from our analysis, most of whom attend special education only classrooms. Finally, the regression sample does not include students who were not tested in the prior year, since prior test scores are used as a control variable. These sample restrictions dropped 25% of students from our sample. The vast majority of these deletions were due to either never being matched with a teacher for a full year (12%), always missing observable characteristics such as prior test scores (12%), or only attending special education only classrooms (7%). Our sample restrictions also result in the deletion of 25% of teachers matched with students. The vast majority of these deletions were due our exclusion of those teaching exclusively special education students (19%) and those who did not remain in any school for a full year (5%).

4. Estimation of teacher effectiveness

To generate estimates of teachers' effectiveness in raising student achievement, we estimate the following regression with student-level data:

$$A_{it} = \beta_g X_{it} + \gamma_g \bar{X}^c_{it} + \zeta_g \bar{X}^s_{it} + \delta W_{it} + \pi_{gt} + \varepsilon_{it}, \quad (1)$$

where A_{it} represents the math or reading test score of student i in year t, X_{it} represents student characteristics (prior-year math and reading scores, gender, ethnicity, receipt of free/reduced price lunch, special education and ELL services, and prior-year absences and suspensions), $\bar{X}^c_{it}$ and $\bar{X}^s_{it}$ are the mean characteristics of the students in student i's classroom and school, respectively, in year t, W_{it} represents characteristics of the teacher to which the student is assigned in year t, and π_{gt} is a fixed effect for the grade in which student i is enrolled and the year t in which we observe him/her.[6] The coefficients on individual, class, and

[5] Among those offered a position, we find no relationship between academic credentials and initial acceptance or academic credentials and employment in the district. Thus, it does not appear that the selectivity of the Teaching Fellows program is being undone after offers have been given out.

[6] Analyses of teacher quality have typically used one of four different empirical specifications. Our specification falls under the

school characteristics (β_g, γ_g, and ζ_g) are allowed to vary by the grade level of the test being taken. The class-level variables also include class size, and school-level variables include average class size in the school. All test scores are normalized within grade and year to have a mean of zero and a standard deviation of one. According to data from the National Assessment of Educational Progress (NAEP), a standard deviation of student test scores in NYC is comparable to that of students nationally.

We calculate standard errors allowing for clustering at the school-level (Rogers, 1993). This method is robust to heteroskedasticity in the student-level errors as well as non-zero covariance of ε_{it} among those attending the same school over time (such as would be produced by random effects at the student, teacher, classroom, and school level). As Sanders, Ashton, and Wright (2005) show, standard error estimates of between-group differences in teacher impacts can be dramatically understated when not allowing for random effects. They could also be sensitive to student effects which could lead to correlations in ε_{it} within school over time.

We also estimated specifications that replace controls for school average characteristics with either school fixed effects or fixed effects for permutations of school, grade and year. These fixed effects absorb variation at the school level in factors that we cannot observe (e.g., the effectiveness of the school administration). When fixed effects for school or school by grade by year permutations are included, the coefficients on teacher characteristics are identified only from variation among teachers working within the same school or the same school, grade level, and year. The inclusion of school fixed effects implicitly controls for other unobservable student or school characteristics that vary by school. However, doing so also washes out any differences in teacher quality across schools. Fortunately, our results are very similar whether we use school level mean characteristics as regressors or fixed effects by school and by school, grade and year.

4.1. Baseline results on teacher certification and teacher effectiveness

In Table 3 we report estimates of differences in students' math and reading achievement by teachers' initial certification status. In all specifications, we adjust for individual years of teaching experience. For illustrative purposes, in columns (1) and (6) we show the results of regressions that omit student, classroom, or school characteristics. There are very large differences in test scores for students assigned to different types of teachers. The math scores of students assigned to Teaching Fellows and TFA corps members were 0.20 and 0.28 standard deviations below those of students assigned to regularly certified teachers. Students assigned to international recruits and uncertified teachers scored 0.48 and 0.13 standard deviations, respectively, below students in classrooms taught by regularly certified teachers. These differences were similar for reading test scores.

However, much of the apparent difference in student performance for those assigned to different groups of teachers is simply due to differences in students' prior year test performance and demographics. When we include test scores from the prior year and other student level covariates in the regressions (columns (2) and (7)), the coefficients change dramatically. In math, students assigned to Teaching Fellows performed no differently than similar students assigned to traditionally certified teachers, while those assigned to TFA corps members outperformed by 0.01 standard deviations. Students assigned to international recruits and uncertified teachers continue to underperform relative to those assigned to certified teachers, but the differences are much smaller (−0.05 and −0.005 standard deviations, respectively). In reading, when controlling for baseline characteristics of students, students assigned to Teaching Fellows, TFA corps members, international recruits, and uncertified teachers underperformed those assigned to regularly certified teachers, but these differences are a small fraction of those reported in columns (1) and (6).

(*footnote continued*)
"quasi-gains" category: the dependent variable is the level of current student achievement, and prior achievement levels are used as control variables. In a "gains" specification, the dependent variable is the change in student test scores between the current period and some base period, usually the prior school year (see Rivkin, Hanushek, & Kain, 2005). This is essentially a quasi-gains specification where past achievement is restricted to have a one-to-one relationship with current achievement. In a "levels with fixed effects" specification, the level of current student achievement is the dependent variable and student fixed effects—not prior achievement levels—are used as control variables (see Rockoff, 2004). A "gains with fixed effects" specification, as the name implies, combines both of these features. Our results are not sensitive to using any of these three other approaches.

Table 3
Differences between teacher certification groups in math and reading value-added

	Math					Reading				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Teaching fellow	−0.199	0.004	0.007	0.004	0.000	−0.238	−0.017	−0.012	−0.016	−0.012
	(0.024)	(0.006)	(0.006)	(0.005)	(0.005)	(0.025)	(0.006)	(0.005)	(0.005)	(0.004)
Teach for America	−0.280	0.012	0.018	0.031	0.023	−0.311	−0.018	−0.003	−0.000	0.005
	(0.035)	(0.012)	(0.010)	(0.010)	(0.009)	(0.039)	(0.010)	(0.007)	(0.008)	(0.008)
Internat'l programs	−0.476	−0.054	−0.027	−0.029	−0.023	−0.445	−0.030	0.004	0.002	0.004
	(0.048)	(0.010)	(0.010)	(0.010)	(0.008)	(0.042)	(0.011)	(0.009)	(0.009)	(0.008)
Other uncertified	−0.126	−0.005	0.002	0.001	−0.000	−0.128	−0.010	−0.000	0.002	0.005
	(0.020)	(0.005)	(0.004)	(0.004)	(0.004)	(0.020)	(0.004)	(0.004)	(0.004)	(0.004)
Student covariates		√	√	√	√		√	√	√	√
Class average covariates			√	√	√			√	√	√
School average covariates			√					√		
School FE				√					√	
School ∗ grade ∗ year FE					√					√
Grade ∗ year FE	√	√	√	√		√	√	√	√	
Sample size	1,462,100	1,462,100	1,462,100	1,462,100	1,462,100	1,366,479	1,366,479	1,366,479	1,366,479	1,366,479
Number of teachers	18,856	18,856	18,856	18,856	18,856	19,083	19,083	19,083	19,083	19,083
Number of students	623,482	623,482	623,482	623,482	623,482	607,563	607,563	607,563	607,563	607,563
R^2	0.03	0.67	0.68	0.68	0.70	0.03	0.63	0.64	0.64	0.66

Note: All specifications include dummy variables for teacher experience, a dummy for those missing experience and a dummy for those hired before the 1999–2000 school year. Student-level covariates include a cubic polynomial in both prior-year math and reading scores, gender, six categories for race and ethnicity, an indicator for Free/Reduced Price Lunch status, an indicator for special education status, and an indicator for English Language Learners. Each of these were also interacted with grade level. The school average and class average covariates included the school-level and classroom-level means of the student-level covariates and class size, each also interacted with grade level. Coefficients on indicator variables for TOP scholars and Peace Corps fellows were also estimated but were not reported given large standard errors. Standard errors (in parentheses) allow for clustering within school.

In columns (3) and (8), we include classroom level and school level average characteristics as control variables. The addition of these group level control variables has little impact on the coefficient estimates for Teaching Fellows and uncertified teachers. Relative to the students of regularly certified teachers, TFA corps members' students are now estimated to score 0.02 standard deviations higher, and international recruits' students score 0.03 standard deviations lower. In reading, students assigned to Teaching Fellows scored 0.01 standard deviations below those assigned to regularly certified teachers. This is the only between-group difference that is statistically significant for reading test scores.

In the remaining columns, we report the results of regressions that include school fixed effects or fixed effects for each school by grade by year permutation. The addition of these controls does not substantially change our findings.

4.2. Additional estimates of the relative effectiveness of groups of teachers

In Table 4, we present estimates for various subgroups of schools and teachers. All specifications also control for single year of teaching experience, the full set of student, classroom level, and school level baseline characteristics, and fixed effects by grade and year. The first column of Table 4 redisplays our baseline estimates from Table 3. Results for math and reading are shown in the top and bottom panels, respectively. First we separately examine elementary and middle school grades.[7] Then, we separately examine schools above and below the median average test score. Splitting the sample produces little change in our estimates of the impact of certification status. Last, we separately examine teachers with zero, one, and two years of experience. Here the identification of between-group differences comes only from direct comparisons between teachers with the same amount of experience. While the differences in the estimates between teachers with zero, one or two years of experience are not significantly different from each other, the pattern of results is suggestive: Teaching Fellows, TFA corps members, and uncertified teachers tend to fare worse relative to traditionally certified teachers as rookie teachers than after they have gained a year or two of experience. For example, estimated achievement impacts for Teaching Fellows (relative to certified teachers) with zero, one and two years of experience are −0.010, 0.005 and 0.018, respectively, for math and −0.018, −0.007, and −0.003, respectively, for reading. One possibility for this is that AC teachers and uncertified teachers may have higher returns to experience than other teachers on average. Another possibility is that there is negative selective attrition of these teachers, relative to traditionally certified teachers. In Section 4.4 we focus directly on the return to experience and explore these potential sources of dynamic heterogeneity across groups of teachers.

[7] Elementary grades are 4 and 5, middle grades are 7 and 8. Sixth graders are considered elementary if they attend a school where the maximum grade is 6, otherwise they are considered middle.

4.3. Selection of teaching fellows and teacher effectiveness

Prior research has found a relationship between teacher effectiveness and the selectivity of the college a teacher attended (Summers & Wolfe, 1977), tests of teachers' verbal ability (Hanushek, 1971), or a teacher's own ACT (American College Testing program) scores when applying to college (Ferguson & Ladd, 1996). However, as Hanushek and Rivkin (2004) argue in summarizing the research on teacher impacts on student achievement, the association between teacher test scores and student outcomes is relatively weak. Moreover, the literature on teacher effectiveness has consistently failed to find that those holding master's degrees are more effective, despite the fact that most teacher pay scales reward higher educational attainment (Aaronson, Barrow, & Sander, 2003; Clotfelter, Ladd, & Vigdor, 2006; Ehrenberg & Brewer, 1994; Murnane, 1975; Summers & Wolfe, 1977).

Teaching Fellows attended more selective colleges than traditionally certified teachers. Yet we find that the average Teaching Fellow is no more effective in the classroom than traditionally certified teachers. To investigate further the relationship between student achievement and teacher academic credentials, we focus on differences in effectiveness among the Teaching Fellows—for whom we have more complete data on undergraduate institution and undergraduate GPA. Columns (1) and (4) of Table 5 show the relationship between math and reading value-added and the median math

Table 4
Additional estimates of between-group differences in teacher value-added

	Baseline estimates from Table 6 Column 3	*Students in elementary grades*	*Students in middle grades*	*Schools w/ above median avg test scores*	*Schools w/ below median avg test scores*	*Teachers with zero years of experience*	*Teachers with 1 year of experience*	*Teachers with 2 years of experience*
Math								
Teaching fellow	0.007	0.001	0.003	0.006	0.000	−0.010	0.005	0.018
	(0.006)	(0.008)	(0.007)	(0.009)	(0.007)	(0.008)	(0.009)	(0.012)
	[1439]	[905]	[559]	[386]	[1076]	[911]	[829]	[443]
Teach for America	0.018	0.015	0.027	0.038	0.016	−0.006	0.016	0.021
	(0.010)	(0.018)	(0.012)	(0.021)	(0.012)	(0.013)	(0.017)	(0.030)
	[301]	[172]	[140]	[49]	[257]	[221]	[179]	[55]
Internat'l programs	−0.027	0.002	−0.034	−0.028	−0.036	−0.006	−0.034	−0.024
	(0.010)	(0.021)	(0.011)	(0.018)	(0.012)	(0.029)	(0.030)	(0.024)
	[284]	[99]	[192]	[59]	[229]	[27]	[35]	[29]
Uncertified teachers	0.002	−0.005	−0.001	0.000	0.001	−0.017	−0.015	−0.000
	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.008)	(0.008)
	[2603]	[1502]	[1171]	[1086]	[1591]	[1008]	[1222]	[1140]
Number of regularly certified teachers	3628	2886	788	2049	1636	1595	1563	1237
Number of students	623,482	453,632	404,700	371,888	319,012	114,961	143,182	126,135
R^2	0.68	0.66	0.70	0.66	0.60	0.65	0.65	0.66
Reading								
Teaching fellow	−0.012	−0.018	−0.009	−0.009	−0.009	−0.018	−0.007	−0.003
	(0.005)	(0.007)	(0.007)	(0.010)	(0.006)	(0.007)	(0.008)	(0.011)
	[1413]	[893]	[553]	[330]	[1104]	[936]	[835]	[473]
Teach for America	−0.003	−0.012	−0.001	0.002	0.001	−0.015	0.005	−0.025
	(0.007)	(0.013)	(0.009)	(0.017)	(0.008)	(0.010)	(0.012)	(0.022)
	[347]	[171]	[187]	[48]	[302]	[260]	[201]	[57]
Internat'l programs	0.004	−0.008	0.006	0.045	−0.011	0.062	−0.018	−0.025
	(0.009)	(0.019)	(0.011)	(0.021)	(0.009)	(0.031)	(0.026)	(0.034)
	[245]	[91]	[156]	[57]	[192]	[21]	[28]	[23]
Uncertified teachers	−0.000	0.002	−0.006	−0.002	0.004	−0.013	−0.010	−0.012
	(0.004)	(0.006)	(0.005)	(0.005)	(0.005)	(0.007)	(0.006)	(0.008)
	[2417]	[1473]	[1025]	[982]	[1498]	[857]	[1109]	[1044]
Number of regularly certified teachers	4062	2875	1243	2300	1818	1836	1769	1373
Number of students	607,563	451,105	370,489	368,328	295,799	112,085	140,035	126,621
R^2	0.64	0.63	0.66	0.62	0.57	0.61	0.62	0.63

Note: The number of teachers from each group (in brackets for groups other than traditionally certified) refer to teachers hired within our sample period (starting with school year: 1999–2000). All specifications (excluding those where experience is used to restrict the sample) include dummy variables for years of teaching experience and a dummy for those missing experience. They also include the same student, class, and school-level covariates as the regressions reported in Table 3. Elementary grades are 4 and 5, middle grades are 7 and 8. 6th graders are considered elementary if their school's terminal grade is 6, otherwise they are considered middle. Standard errors (in parentheses) allow for clustering within school.

Table 5
Teaching fellows' academic credentials and value-added

	Math			Reading		
	(1)	(2)	(3)	(4)	(5)	(6)
Teaching fellow	0.005	0.004	0.004	−0.013	−0.013	−0.012
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
College SAT math decile	0.003		0.003	0.000		0.000
	(0.002)		(0.002)	(0.002)		(0.002)
College GPA		−0.008	−0.009		−0.008	−0.008
		(0.012)	(0.013)		(0.011)	(0.011)
SAT math * GPA			−0.006			−0.001
			(0.007)			(0.006)
Teaching fellows with non-missing credentials	1269	1293	1269	1227	1249	1226
Sample size	1,462,100	1,462,100	1,462,100	1,366,479	1,366,479	1,366,479
Number of teachers	18,856	18,856	18,856	19,083	19,083	19,083
Number of students	623,482	623,482	623,482	607,563	607,563	607,563
R^2	0.68	0.68	0.68	0.64	0.64	0.64

Note: Standard errors (in paretheses) allow for clustering within school. All specifications include dummy variables for participation in other recruitment programs, being an uncertified teacher, teacher experience, a dummy for those missing experience, a dummy for those hired before the school year 1999–2000, and student, classroom, and school-level covariates. Student-level covariates include a cubic polynomial in both prior-year math and reading scores, gender, six categories for race and ethnicity, an indicator for Free/Reduced Price Lunch status, an indicator for special education status, an indicator for English Language Learners. Each of these were also interacted with grade level. The classroom and school covariates included the classroom-level and school-level means of all the student-level covariates, each also interacted with grade level. Controls are also included for a teaching fellow missing either GPA or SAT scores when those variables are included in the regressions.

SAT score of Teaching Fellows' undergraduate institution, measured in deciles. There is a coefficient of 0.003 standard deviations in math and a coefficient of less than 0.0005 in reading, neither of which are statistically significant at conventional levels.[8] We also find no statistically significant relationship between academic credentials and value-added when we examine undergraduate GPA or when we add an interaction between GPA and median math SAT score. Thus, there is little evidence that academic credentials are a strong predictor of teacher effectiveness among Teaching Fellows. This is consistent with our failure to find large differences in mean effectiveness between Teaching Fellows and traditionally certified teachers.

[8]These regressions control for year-grade fixed effects, student characteristics and classroom- and school-level average student characteristics (as in Table 3 columns (3) and (8)). Regressions using reading SAT decile show no significant predictive power for either math or reading value-added, so we do not present results for them here.

4.4. Group differences in the returns to teaching experience

There are several reasons why one might think that AC and uncertified teachers should have higher returns to experience than traditionally certified teachers. Unlike traditionally certified teachers who are required to have some supervised student teaching experience in graduate school, the AC and uncertified teachers typically have little prior classroom experience. Also, AC teachers are usually enrolled in night classes during their first two years of teaching, and that might also lead to greater improvements when they complete their coursework. However, it is equally plausible that returns to experience are higher for traditionally certified teachers. Traditionally certified teachers, because of their training, may be better equipped to learn and improve their teaching skills. Furthermore, the coursework AC teachers are required to take at night and on weekends in their first two years of teaching may well be a hindrance that stops them from learning as much as possible at work.

Table 6
Returns to teacher experience by initial teacher certification

	Math	Reading
Second year teachers (relative to 1st yr)	0.033 (0.007)	0.018 (0.006)
Third year	0.047 (0.008)	0.027 (0.008)
Fourth year	0.069 (0.010)	0.037 (0.009)
Fifth + year	0.082 (0.012)	0.048 (0.012)
Teaching fellow *2nd year	0.011 (0.011)	0.010 (0.012)
TF*3rd year	0.030 (0.015)	0.036 (0.016)
TF*4th year	0.014 (0.020)	0.015 (0.020)
TF*5th + year	0.014 (0.028)	0.023 (0.033)
TFA corps member *2nd year	−0.006 (0.020)	0.022 (0.017)
TFA*3rd year	−0.001 (0.041)	0.012 (0.028)
TFA*4th year	−0.041 (0.054)	0.037 (0.030)
TFA*5th + year	−0.135 (0.084)	0.024 (0.046)
Uncertified teacher *2nd Year	0.016 (0.010)	−0.007 (0.011)
Uncertified*3rd year	0.029 (0.011)	−0.007 (0.013)
Uncertified*4th year	0.015 (0.012)	−0.007 (0.013)
Uncertified*5th + year	0.006 (0.015)	−0.019 (0.016)
Teaching fellow experience interactions = 0 (p-val)	0.387	0.265
TFA experience interactions = 0 (p-val)	0.581	0.641
Uncertified experience interactions = 0 (p-val)	0.096	0.840
Sample size	1,462,100	1,366,479
Number of teachers	18,856	19,083
Number of students	623,482	607,563
R^2	0.71	0.66

Note: Standard errors (in parentheses) allow for clustering within school. All specifications include dummy variables for teacher certification status and a dummy for those missing experience, as well as the student-, classroom-, and school-level covariates regressions shown in Table 3.

The differences in impacts by year of experience reported in Table 4 above may have been confounded by teachers' and schools' decisions about who will remain in teaching—which may be different for AC teachers and traditionally certified teachers. To isolate the return to experience for a given teacher (as opposed to this composition effect), we estimated models including teacher fixed effects.[9] We allowed for differential returns to experience across teacher certification groups by including interactions between group dummy variables (e.g., whether a teacher is a Teaching Fellow, TFA corps member, etc.) and experience level dummy variables. We also restrict our sample to teachers with five years of experience or less, since AC teachers are almost all in this population. The results are reported in Table 6. While the overall pattern of estimates in Table 6 suggests higher returns for AC and uncertified teachers, the only differences that were statistically significant (based on a joint test of the experience interactions) were those for uncertified teachers in math (p-value = 0.03) and Teaching Fellows in reading (p-value = 0.07).

4.5. Attrition and steady-state differences in experience

Suppose that there were two groups of teachers with identical impacts on student achievement after controlling for experience, but with different retention rates. A school district would be better off hiring the group of teachers with higher retention rates, since they would be less likely to forfeit the benefits of on-the-job learning when a teacher leaves. However, the magnitude of the preference depends upon two things: the payoff to teaching experience and the proportion of teachers at each experience level in steady state.[10]

[9] If less effective teachers are more likely to remain teaching, we will understate the returns to experience; if more effective teachers are more likely to remain, we will overstate the returns to experience. Estimates of the return to experience that are identified using variation within teachers over time are not susceptible to these sources of bias.

[10] Although we focus on the direct effects on student achievement, lower turnover would also reduce hiring and training costs.

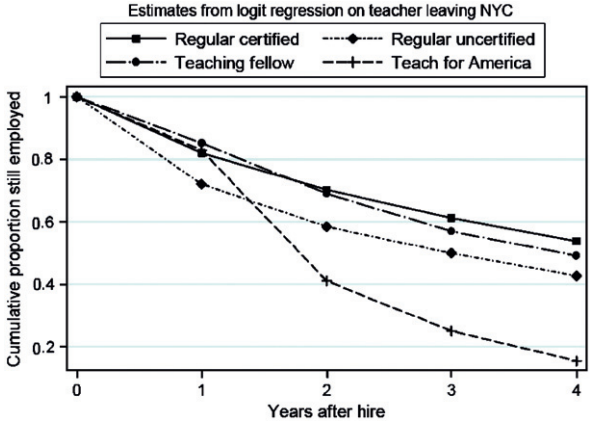


Fig. 1. Cumulative retention estimates for certification and program groups.

Fig. 2. Steady-state distributions of teacher experience by group.

To study retention for different groups of teachers, we estimated logistic regressions of hazard rates for certified, uncertified, and AC teachers hired since the school year 1999–2000 with 0–7 years of experience, while controlling for age (in 5 year categories) and year dummies.[11] We use the full sample of teachers in the payroll files.[12] We exclude international recruits, given the difficulty of measuring their experience prior to joining the district. To account for differences in retention rates, we included interactions between years of experience and initial certification status at time of hiring (Teaching Fellow, TFA corps member and uncertified teachers). For uncertified teachers, we also included an interaction with the year dummy in 2003, since starting in the fall of the school year 2003–2004, school districts in New York State were no longer permitted to employ uncertified teachers without special permission.

Fig. 1 reports the cumulative retention rates adjusting for age and year for the different groups of teachers. Teaching Fellows have very similar retention rates to regular certified teachers (with Teaching Fellows having slightly higher retention rates in the first two years). By their fifth year in teaching (with four years of experience), approximately 50% of both groups are still with the district. Uncertified teachers have somewhat lower retention rates, with 45% remaining with the district in their fifth year.

In contrast, TFA corps members have much lower cumulative retention rates. By the fifth year, only about 18% of corps members remain with the district. Presumably, this reflects the fact that TFA corps members sign up for a two-year teaching commitment.[13]

Assuming a constant hazard rate between the fifth year and the 30th year of experience, Fig. 2 reports the steady-state proportions of all four groups by year of experience implied by the above analysis. The only large difference is between TFA corps members and the other three groups. In steady state, TFA corps members would be roughly twice as likely to be in their first year of teaching (25.6%) than certified teachers (11.1%).

Suppose that $e_{j1}, e_{j2}, \ldots, e_{jT}$ represents the proportion of teachers from a given group j that would be in their first through Tth year of teaching in steady state, given group-specific retention rates. And suppose δ_j represents the mean value-added of group j teachers in their first year of teaching and that $r_{j2}, \ldots, r_{jT}$ represent the average returns to experience for the same group of teachers in their second through Tth year of teaching. In steady state, the average value-added of the jth group of teachers in steady state would be equal to $\delta_j + \sum_{t=2}^{T} e_{jt} r_{jt}$.

Many supporters of the TFA program argue that corps members have larger impacts on student

[11] We found no substantive impact on the results using logistic regression versus OLS, nor adding school fixed effects.

[12] The results are similar if we limit the sample to those teaching in grades 4–8.

[13] It is worth noting that TFA corps members leave teaching for other positions in education. TFA alumni surveys indicate that more than 50% of its corps members are working in education even ten years after they enter the program, 18 alumni are currently working as principals or assistant principals in NYC's schools, and another 10 alumni are working for the DOE in other district level jobs.

achievement than regularly certified teachers, particularly among those certified teachers willing to work in low-income schools. Critics of the program argue that because of high turnover rates, school districts are constantly having to replace TFA corps members with novice teachers (i.e., they argue that the program has low $e_{j2},\ldots,e_{jT}$). Both groups could be correct in their assertions, but the net impact on student achievement could be positive or negative. Assuming that returns to experience are similar between TFA corps members and certified teachers (which is consistent with our findings), we can calculate the difference between δ_{TFA} and δ_{Cert} that would be required to ensure that TFA corps members had larger steady-state impacts than certified teachers:

$$\delta_{\mathrm{TFA}}+\sum_{t=2}^{T}e_{\mathrm{TFAt}}r_t>\delta_{\mathrm{Cert}}+\sum_{t=2}^{T}e_{\mathrm{Certt}}r_t$$

$$\Rightarrow(\delta_{\mathrm{TFA}}-\delta_{\mathrm{Cert}})>\sum_{t=2}^{T}(e_{\mathrm{Certt}}-e_{\mathrm{TFAt}})r_t.$$

Using the estimates of the returns to experience that include teacher fixed effects, $\delta_{\mathrm{TFA}}-\delta_{\mathrm{Cert}}$ would have to be greater than 0.019 in math and 0.012 in reading in order for TFA corps members to have greater steady-state value-added. This is quite a modest difference. In other words, despite large differences in retention rates, the differences in returns to experience and retention rates are not enough to generate large differences in steady-state impacts between TFA and other groups of teachers.[14] Even the small positive difference in value-added for TFA corps members in teaching math would be sufficient to compensate for their higher turnover.

5. Variation in value-added within groups of teachers

The above results suggest that there are little or no differences in average value-added by initial certification status. In this section, we derive estimates of the variation in teacher effectiveness among those with the same initial certification status. First, we calculate classroom average residuals from a regression of student test scores on student, classroom, and school characteristics, as well as grade-by-year fixed effects. If the residual for teacher j is the sum of a persistent component (μ_j) (i.e., teacher effectiveness) and a non-persistent component (ξ_{ij}) that is independent from class to class, then the covariance in the residuals between any two classrooms taught by the same teacher is an unbiased estimate of the variance in the persistent component of teacher effectiveness.

Table 7
Variation of teacher effectiveness within certification status

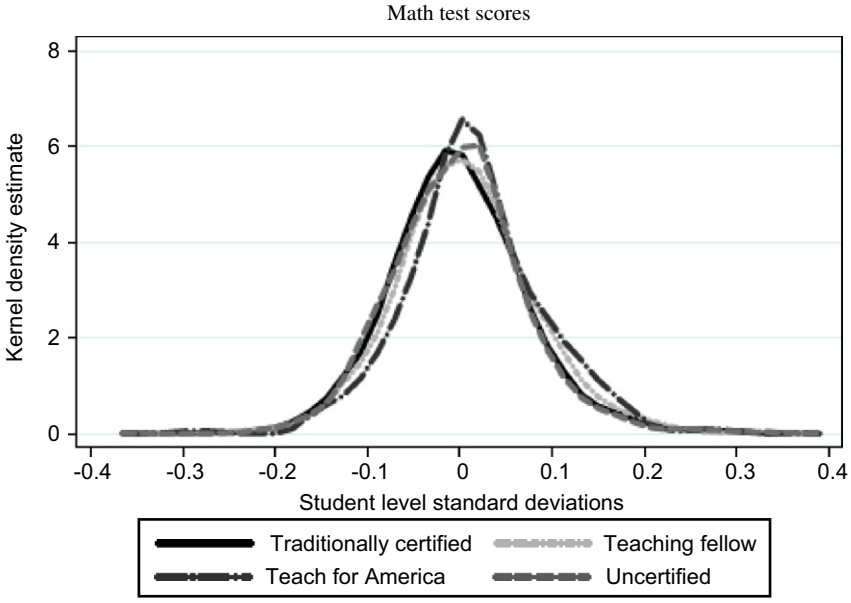
Standard deviation	Math		Reading	
	Total	Signal	Total	Signal
(A) *Elementary school*				
Full sample	0.21	0.13	0.20	0.10
Novices				
All novices	0.21	0.12	0.19	0.08
Traditionally certified	0.20	0.12	0.19	0.08
Uncertified	0.22	0.12	0.19	0.09
Teaching fellows	0.22	0.12	0.20	0.07
(B) *Middle school*				
Full sample	0.17	0.08	0.17	0.06
Novices				
All novices	0.16	0.09	0.16	0.07
Traditionally certified	0.16	0.08	0.16	0.07
Uncertified	0.16	0.09	0.15	0.07
Teaching fellows	0.16	0.09	0.17	0.08
(C) *Middle school teaching same course*				
Full sample	0.17	0.09	0.16	0.07
Novices				
All novices	0.16	0.10	0.16	0.08
Traditionally certified	0.16	0.10	0.16	0.09
Uncertified	0.17	0.10	0.16	0.08
Teaching fellows	0.16	0.10	0.16	0.09

Note: "Total" is the variation among classroom-mean residuals in a regression of student test scores on student, classroom, teacher, and school characteristics. "Signal" is the covariance among mean residuals for classrooms taught by the same teacher.

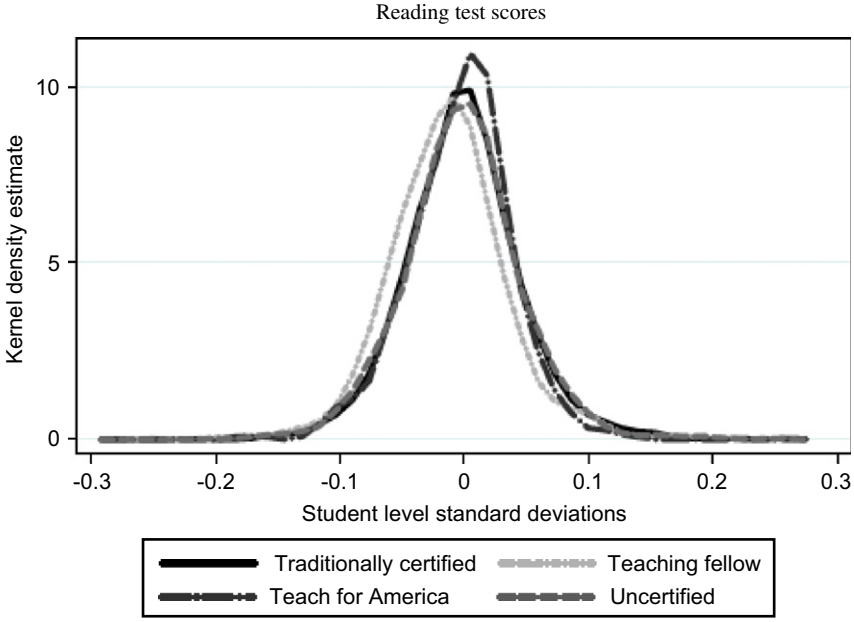
In Table 7, we report estimates of the standard deviation of the total variation in estimated teacher effectiveness as well as of the persistent component (μ_j). In elementary schools, the standard deviation of the persistent component is estimated to be 0.13 in math and 0.10 in reading. There is slightly less variation (0.12 and 0.08) in teacher performance among novice teachers—those in their first 3 years of teaching—with similar amounts of variation found among certified, uncertified, and Teaching Fellows.

In middle school, we see less variation, with the standard deviation of the persistent component of performance being about two-thirds as large. Part

[14] Ballou (1996) makes this point more generally using a simulation. Large differences in turnover and modest experience effects do not generate large costs to hiring high turnover teachers in terms of value added.



Note: Shown are estimates of teachers' Impacts on average student performance, controlling for teachers' experience levels and students' baseline scores, demographics and program participation; Includes teachers of grades 4-8 hired since the 1999-2000 school year.



Note: Shown are estimates of teachers' Impacts on average student performance, controlling for teachers' experience levels and students' baseline scores, demographics and program participation; Includes teachers of grades 4-8 hired since the 1999-2000 school year.

Fig. 3. Variation in value-added within and between groups of teachers.

of the explanation for this result is that performance is less persistent when a middle school teacher is teaching different courses (e.g., advanced versus remedial math): when we focus on teachers teaching the same course in multiple classrooms (the bottom panel of Table 7), the persistent component grows.

To the extent that teachers in middle school often change the courses they teach from year to year, this will make it more difficult to reliably identify performance differences for middle school teachers.

For both math and reading in elementary and middle school, our estimates imply there are large and persistent performance differences across teachers. If we were able to rank teachers by their value added, the difference in average value added between the top 25% and the bottom 25% of teachers would be approximately 2.5 times the standard deviation across teachers (assuming that teacher performance is normally distributed). Thus, the estimates in Table 7 imply that the average value added among the top quarter of elementary school math teachers is 0.33 standard deviations greater than the value added among the bottom 25% of teachers. For middle school teachers, this difference is somewhat smaller, but there is still at least a 0.20 standard deviation difference in value added between the top and bottom 25% of teachers. In other words, the impact of assigning a student to a bottom quarter teacher rather than a top quarter teacher is roughly three times the impact of being assigned to a novice teacher rather than an experienced teacher, and more than 10 times the impact of being assigned to a teacher with a particular kind of certification or from a particular program!

Fig. 3 shows the variation in teacher effectiveness (value-added) within and among four groups of teachers—certified, uncertified, Teaching Fellows, and TFA. This figure plots kernel density estimates of the distribution of the estimated persistent component (μ_j) of teacher effectiveness (value-added).[15] While the differences *between* the four groups observed in Fig. 3 are small, the differences *within* the four groups are quite dramatic. In other words, there is not much difference between certified, uncertified, and alternatively certified teachers overall, but effectiveness varies substantially among each group of teachers. To put it simply, teachers vary considerably in the extent to which they promote student learning, but whether a teacher is certified or not is largely irrelevant to predicting their effectiveness.

6. Conclusion

State and federal efforts to regulate teacher effectiveness focus almost entirely on *ex ante* qualifications of prospective teachers. For example, under the federal No Child Left Behind act, states and districts are required to hire certified teachers or those enrolled in an alternative certification program. However, our results suggest that the emphasis on certification status may be misplaced. We find little difference in the average academic achievement impacts of certified, uncertified and alternatively certified teachers.

The fact that we find little or no differences in the average teacher effectiveness of certified, uncertified, and AC teachers does not imply that selection of teachers is unimportant. The standard deviation in value-added among teachers is roughly 0.10 student level standard deviations. Our estimates are remarkably similar to estimates from other contexts. For example, using data from two school districts in New Jersey, Rockoff (2004) reports that one standard deviation in teacher effects is associated with a 0.1 student-level standard deviation in achievement. Using data from Texas, Rivkin et al. (2005) report very similar estimates—suggesting that a standard deviation in teacher effectiveness is associated with 0.11 student-level standard deviations in math and 0.095 standard deviations in reading. Using data on high school students in Chicago Public Schools, Aaronson et al. (2003) find that a standard deviation in teacher effectiveness is associated with a 0.09–0.16 student-level standard deviation difference in performance.

To put these estimates in context, raising the effectiveness of novice teachers in New York by one standard deviation would have a similar impact on student achievement as the expected improvement of novices who spend 8 years teaching in the district! Thus, policies that enable districts to attract and retain high quality teachers (or screen-out less effective teachers) have potentially large benefits for student achievement.

The large observable differences in teacher effectiveness *ex post* suggest that districts should use performance on the job, rather than initial

[15] For each teacher with observations from t classrooms, we estimated their persistent component using an empirical Bayes (shrinkage) estimator of the form:

$$E(\mu_j|\varepsilon_{j,1},\ldots,\varepsilon_{j,t}) = \bar{\varepsilon}\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\xi^2/t} \quad \text{where } \bar{\varepsilon} = \frac{1}{t}\sum_{s=1}^{t}\varepsilon_{j,s}.$$

In other words, teacher effectiveness is measured by their average residual multiplied by a scaling factor equal to the proportion of variance in the average residual that is due to signal variance (i.e., the reliability).

certification status to improve average teacher effectiveness. Recent research on teachers certified by the National Board for Professional Teaching Standards (NBPTS) suggests that *ex post* measurement of teacher practice can identify some small differences in practice among experienced teachers that are associated with student achievement.[16] It is also important to investigate other *ex post* measures, such as statistical estimates of "value-added". Raising the stakes associated with value-added measures could lead to narrowing of the curriculum or teacher malfeasance (Figlio, 2005; Figlio & Winicki, 2002; Jacob & Levitt, 2003; Koretz, 2002). However, value-added measures have other appealing properties, such as objectivity and low cost. Districts and states have yet to develop a system of evaluating teacher performance, which balances the risks of failing to promote effective teachers, mistakenly promoting ineffective teachers and providing perverse incentives leading to teacher cheating and narrowing of the curriculum. We intend to pursue these issues in future work.

Acknowledgments

We are grateful to Betsey Arons, Vicki Bernstein, Bruce Groneveldt, Julie Horowitz, Dominic Mancello, Lori Mei, Cathi Schoenberger, Henry Solomon, LaVerne Srinivasan, and Harvey Weinstein at the NYC DOE, who generously assembled data and patiently answered questions. A number of others in the NYC DOE, including Robert Gordon, Joel Klein and James Liebman, provided helpful comments. Timothy Daly, Michelle Rhee, and Andrew Sokatch of The New Teacher Project and Abigail Smith and Cynthia Skinner of TFA were extremely helpful in answering our questions regarding these programs. We thank Caroline Hoxby and Andrew Leigh for providing us with data on college median SAT scores. Dan Obus provided excellent research assistance. Seminar participants at the NYC DOE, Columbia University and the NBER summer institute provided many helpful comments and suggestions.

[16]Recent research by Goldhaber and Anthony (2004), Cavalluzzo (2004), Vandevoort, Amrein-Beardsley, and Berliner (2004), Clotfelter, Ladd, and Vigdor (2007), Sanders et al.(2005) and Harris and Sass (2007) has suggested that National Board certified teachers are more effective than comparison teachers. The NBPTS, a non-profit organization, was created in 1987 to provide an objective means for recognizing and rewarding effective teaching. When applying for certification by NBPTS, teachers provide a videotape of their work in front of class, submit examples of written assignments and the feedback they provided to students, and answer a number of essay questions in a testing center.

References

Aaronson, D., Barrow, L., & Sander, W. (2003). *Teachers and student achievement in the Chicago public schools*. Chicago: Federal Reserve Bank of Chicago WP-2002-28.

Ballou, D. (1996). Do public schools hire the best applicants? *Quarterly Journal of Economics*, *111*(1), 97–133.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2005a). *How reduced barriers to entry into teaching changes the teacher workforce and affects student achievement*, Manuscript.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2005b). *How reduced barriers to entry into teaching changes the teacher workforce and affects student achievement*. National Bureau of Economic Research working paper 11844, Cambridge.

Cavalluzzo, L. C. (2004). *Is national board certification an effective signal of teacher quality*? CNA Corporation working paper, Alexandria, VA.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). *Teacher–student matching and the assessment of teacher effectiveness*. National Bureau of Economic Research working paper 11936, Cambridge.

Clotfelter, C. T., Ladd H. F., & Vigdor, J. L. (2007). *How and why do teacher credentials matter for student achievement*? National Bureau of Economic Research working paper 12828, Cambridge.

Darling-Hammond, L. (2007). A marshall plan for teaching: What it will really take to leave no child behind. *Education Week*, *26*(18), 48, 28.

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives*, *13*(42), 1–48.

Decker, P. T., Mayer, D. P., & Glazerman, S. (2004). *The effects of teach for America on students*: *Findings from a National evaluation*. Mathematica policy research report 8792-8750, New York.

Ehrenberg, R., & Brewer, D. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review*, *13*(1), 1–17.

Ferguson, R., & Ladd, H. (1996). How and why money matters: An analysis of Alabama Schools. In H. Ladd (Ed.), *Holding schools accountable* (pp. 265–298). Washington, DC: Brookings Institution.

Figlio, D. N. (2005). *Testing, crime and punishment*. National Bureau of Economic Research working paper 11194, Cambridge.

Figlio, D. N. & Winicki, J. (2002). *Food for thought*: *The effects of school accountability plans on school nutrition*. National Bureau of Economic Research working paper 9319.

Goldhaber, D. & Anthony, E. (2004). *Can teacher quality be effectively assessed*? University of Washington and Urban Institute working paper, Seattle.

Gordon, R., Kane, T. J., & Staiger, D. O. (2006). *Identifying effective teachers using performance on the job*. Hamilton project discussion paper. Washington, DC: Brookings Institution.

Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review*, *61*(2), 280–288.

Hanushek, E., & Rivkin, S. G. (2004). How to improve the supply of high-quality teachers. In D. Ravitch (Ed.), *Brookings papers on education policy 2004*. Washington, DC: Brookings Institution.

Harris, D. N., & Sass, T. R. (2007). The effects of NBPTS-certified teachers on student achievement. Center for the Analysis of Longitudinal Data in Education Research (CALDER) working paper no. 4, Washington, DC.

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, *118*(3), 843–877.

Koretz, D. M. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources*, *37*(4), 752–777.

Murnane, R. (1975). *The impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger.

Raymond, M., Fletcher, S. H., & Luque, J. (2001). *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas*. Stanford, CA: The Hoover Institution, Center for Research on Education Outcomes.

Rivkin, S. G., Hanushek, E., & Kain, J. F. (2005). Teachers, schools and academic achievement. *Econometrica*, *73*(2), 415–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*(2), 247–252.

Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin*, *13*, 19–23 (Reprinted in *Stata Technical Bulletin* Reprints, *3*, 88–94).

Sanders, W. L., Ashton, J. J., & Wright, W. P. (2005). Comparison of the effects of NBPTS certified teachers with other teachers on the rate of student academic progress. Report by the SAS Institute to the National Board on Professional Teaching Standards, Arlington, VA.

Summers, A., & Wolfe, B. (1977). Do schools make a difference? *American Economic Review*, *67*(4), 639–652.

Vandevoort, L. G., Amrein-Beardsley, A., & Berliner, D. (2004). National board certified teachers and their students' achievement. *Education Policy Analysis Archives*, *12*(46), 1–117.