NBER WORKING PAPER SERIES

DOES PRACTICE-BASED TEACHER PREPARATION INCREASE STUDENT ACHIEVEMENT?
EARLY EVIDENCE FROM THE BOSTON TEACHER RESIDENCY

John P. Papay
Martin R. West
Jon B. Fullerton
Thomas J. Kane

Does Practice-Based Teacher Preparation Increase Student Achievement? Early Evidence
from the Boston Teacher Residency
John P. Papay, Martin R. West, Jon B. Fullerton, and Thomas J. Kane
NBER Working Paper No. 17646
December 2011
JEL No. I20

ABSTRACT

The Boston Teacher Residency is an innovative practice-based preparation program in which candidates
work alongside a mentor teacher for a year before becoming a teacher of record in Boston Public Schools.
We find that BTR graduates are more racially diverse than other BPS novices, more likely to teach
math and science, and more likely to remain teaching in the district through year five. Initially, BTR
graduates for whom value-added performance data are available are no more effective at raising student
test scores than other novice teachers in English language arts and less effective in math. The effectiveness
of BTR graduates in math improves rapidly over time, however, such that by their fourth and fifth
years they out-perform veteran teachers. Simulations of the program's overall impact through retention
and effectiveness suggest that it is likely to improve student achievement in the district only modestly
over the long run.

John P. Papay
Brown University
Education Department
Providence, RI 02912
john_papay@mail.harvard.edu

Martin R. West
Harvard Graduate School of Education
6 Appian Way, Gutman 454
Cambridge, MA 02138
martin_west@gse.harvard.edu

Jon B. Fullerton
Harvard Graduate School of Education
Center for Education Policy Research
50 Church Street, 4th Floor
Cambridge, MA 02138
jon_fullerton@gse.harvard.edu

Thomas J. Kane
Harvard Graduate School of Education
Center for Education Policy Research
50 Church St., 4th Floor
Cambridge, MA 02138
and NBER
kaneto@gse.harvard.edu

1. Introduction

A central challenge in American education is expanding the supply of effective teachers in urban school districts, especially in high-demand subject areas (Jacob 2007). Yet there is considerable controversy over how best to recruit, prepare, develop, and retain teachers who will be effective in those settings. Alternative certification programs, which now exist in 47 states and the District of Columbia, aim to ease entry into teaching by delaying or bypassing some requirements of traditional university-based preparation programs (NAE 2008). Critics of alternative certification, however, contend that the training offered by most programs is inadequate to prepare new entrants for the challenges of teaching in urban schools; they call instead for renewed efforts to improve the quality and rigor of traditional preparation programs (NCTAF 1996, NCATE 2010).

Urban Teacher Residencies (UTRs) represent a "third way" that combines elements of traditional and alternative teacher preparation programs. Typically run by a school district independently or in partnership with an intermediary organization, residency programs select uncertified teaching candidates through a competitive process to work alongside a mentor for a full year before becoming a teacher of record. Residents also complete a streamlined set of coursework leading to both state certification and a master's degree from a partner university. In exchange for tuition remittance and a residency-year stipend, they commit to teaching in the district for a specified period, generally three to five years.

The UTR model has spread rapidly since the first programs were launched in Chicago, Boston, and Denver between 2002 and 2004, attracting substantial public and philanthropic investment. Urban Teacher Residency United, a national support organization for residency programs, lists 12 programs that meet its quality standards in districts including Los Angeles,

Philadelphia, and New York City.[1] The federal government has devoted more than $143 million

since 2009 to establishing or expanding 26 residencies (Sawchuck 2011), and proposals have

been made to include additional funding for UTRs in the next reauthorization of the Elementary

and Secondary Education Act (Berry et al. 2008). The practice-based training model developed

by UTRs has also influenced broader conversations about the reform of university-based teacher

preparation programs, with a 2010 panel commissioned by the National Center for the

Accreditation of Teacher Education endorsing a move to "programs that are fully grounded in

clinical practice and interwoven with academic content and professional courses" (2010, p. ii).

Despite their growing influence, however, UTRs have not yet been formally evaluated.

In particular, no studies have compared graduates of residency programs to other newly hired

teachers in terms of their effectiveness in raising student achievement (NAE 2008). The Boston

Teacher Residency and Chicago-based Academy for Urban School Leadership report three-year

retention rates of greater than 80% (Berry et al. 2008, Solomon 2009), a figure which far exceeds

the comparable rates for newly hired teachers in urban districts nationwide (Ingersoll 2001).

Improving teacher retention in urban schools could in theory improve student achievement by

reducing both staff churn and the reliance on novice teachers, who tend to be less effective in the

classroom (Ronfeldt et al. 2010). Yet the degree to which these benefits materialize hinges

largely on the relative effectiveness of those teachers who are retained.

This paper offers the first independent evaluation of the effectiveness of graduates of the

Boston Teacher Residency (BTR), one of the nation's longest-running UTR programs. Started in

2003 as a partnership between the Boston Public Schools (BPS) and the non-profit Boston Plan

for Excellence, BTR provides a large and growing share of the new teachers hired each year in

Boston. Since 2008-09, for example, BTR graduates accounted for about one third of all newly

---

[1] http://www.utrunited.org/ [Accessed July 20, 2011].

hired teachers in the district. We combine district administrative data with internal records obtained from BTR to track the careers of all BTR graduates placed in BPS. For the roughly 20% of BTR graduates who teach math and English language arts (ELA) classrooms in grades four to eight, we use value-added methods to compare their effectiveness in raising student test scores to that of other teachers with the same level of experience.

Our analysis confirms that the retention rate of BTR graduates through year five exceeds that of other new hires in the district during the same period by roughly 20 percentage points. BTR graduates are more racially and ethnically diverse than other new teachers in BPS and the program provides a disproportionate share of the district's math and science teachers. In these respects, BTR is accomplishing key purposes for which it was established.

We also find, however, that BTR graduates for whom value-added performance data are available are, on average, no more effective at raising student test scores than other teachers with the same level of experience in ELA and substantially *less* effective in math. These results are clearly disappointing in light of the program's expectations. A closer inspection of the math results reveals a more nuanced pattern in which individual BTR graduates improve their performance more rapidly during their first five years in the classroom. Our estimates indicate that students assigned to BTR graduates in their first year as teachers gain as much as 9 percent of a standard deviation less on math tests over the course of an academic year than do students assigned to the district's other novice teachers – a quite substantial difference equivalent to roughly two months of instruction in the relevant grades (Hill et al. 2008). By their fourth and fifth years, however, BTR graduates outperform not only other teachers at the same level of experience but also veteran teachers in the district by as much as 7 percent of a standard deviation.

The fact that BTR graduates to date have been less effective in raising math achievement as early-career teachers is noteworthy, especially in light of the year they spend in classrooms prior to becoming teachers of record. However, the greater retention of BTR graduates and their more rapid improvement through year five suggest that, if these patterns were to persist, the negative average differences in effectiveness documented in this study would diminish over time and the program could raise student achievement in the district over the long run.

Several points are important to keep in mind in interpreting our results concerning the effectiveness of BTR graduates. First, as noted above, this analysis is based on the relatively small group of BTR graduates working in tested grades and subjects. Given that the program is new, the results for more experienced BTR teachers are based on particularly small samples. It is also possible that the relative effectiveness of BTR graduates differs in high school and untested subjects, where performance data are currently unavailable. Second, consistent with other studies of pathways into teaching (e.g., Boyd et al. 2006; Kane, Rockoff, and Staiger 2008), we find that the variation in performance among BTR graduates and other Boston teachers is far larger than the differences in average performance between the two groups. In other words, knowing whether or not an individual teacher is a BTR graduate is hardly a reliable indicator of their effectiveness in raising student test scores. Finally, our data do not allow us to assess directly whether the differences in effectiveness we document reflect BTR's recruitment process (i.e., who applies and which applicants they ultimately select) or the quality of the training they receive. Even so, the mixed success of BTR to date in producing highly effective teachers leads us to conclude that more research is needed to establish whether the practice-based residency model represents a superior alternative to traditional forms of teacher preparation.

The remainder of the paper is organized as follows. The next section describes the

Boston Teacher Residency and the questions addressed by our research in greater detail, while section three introduces the data we use for our analysis. Section four provides descriptive evidence concerning the demographic composition of BTR graduates, the grades and subjects they teach, and the rate at which they have been retained in Boston Public Schools. Section five then presents the methods and results of our analysis of these BTR graduates' classroom effectiveness, while section six concludes.

2. The Boston Teacher Residency: Origins and Aims

The Boston Teacher Residency was founded in 2003 with a mission to "recruit, prepare, and sustain excellent teachers" (Solomon 2009). More specifically, BTR was designed as a partnership between the Boston Public Schools and the Boston Plan for Excellence (a local education fund) to help the system address three challenges in its recruitment of new teachers:

1. The district was having trouble attracting teachers in a number of subject areas (math, science, special education, and teachers of English Language Learners).
2. The district sought to diversify its teaching force (a majority of whom are white) to better match its majority-minority student population.
3. The district sought to reduce the turnover of new teachers.

Over time the program's leadership has come to emphasize as the program's overarching goal the preparation of teachers who are highly effective in raising student achievement. As Jesse Solomon, BTR's co-founder and original director, puts it:

> As BTR has developed, it has reached the conclusion that it should ultimately be measured by the academic achievement of the students in its graduates' classrooms. BTR wants its teachers to produce measurable and significant gains in student learning. If BTR is to be part of the answer in Boston to help all students reach academic proficiency, it must graduate teachers that can support students to make steady and significant gains. If BTR does not produce teachers who support those academic gains on a reliable basis, the program's effects will have fallen short (Solomon 2009, p. 486).

As is evident from its name, the program's approach is based more upon the medical residency model than upon either traditional teacher preparation programs or alternative

certification programs such as Teach for America or The New Teacher Project. The differences between BTR and other programs are best described in terms of recruitment, preparation, and ongoing induction.

- **Recruitment.** BTR recruits candidates to teach specifically in the Boston Public Schools. The BTR admissions staff therefore screens candidates considering both BPS needs and candidate qualities. This is quite unlike traditional teacher education programs, which do not place graduates in any particular district and cannot recruit for district needs. In addition, BTR candidates commit to teach in BPS for at least three years after their residency year. Those who leave have to repay the program prorated tuition fees up to $10,000. Thus there is a greater upfront commitment to both teaching and a specific district than is typical in many other programs (Solomon 2009)

- **Preparation: the residency year.** Instead of spending most of their training time taking courses, as is typical in traditional preparation programs, BTR residents are assigned a mentor teacher in the Boston Public Schools and spend four days a week in that teacher's classroom. Residents gradually assume more responsibility for leading instruction and ordinarily co-teach with their mentor during the second half of the year. In addition to their time in the classroom, residents take coursework designed and taught by BTR program staff which leads to a Massachusetts Initial Teacher License and a masters' degree from the University of Massachusetts Boston. During their training year, residents receive a small stipend of just over $10,000, or approximately 25% of an initial Boston teacher's salary, from the AmeriCorps program. Thus, unlike participants in Teach for America or The New Teacher Project, residents do not earn full teacher salaries from the start.

- **Ongoing support.** After the residency year, BTR graduates are eligible to be hired as first-

year "teachers of record" with full classroom responsibilities. Because BPS principals have control over the hiring process in their schools, BTR cannot place program graduates in schools or even guarantee them a teaching position in the district. However, through informal networks and active placement efforts, BTR has succeeded in placing roughly 90% of program graduates in the district. BTR initially provided its graduates with ongoing training and network support for three years after candidates graduate the residency. In 2006, however, BPS developed its own district-wide induction program for all novice teachers and assumed control of the first year of induction. BTR continues to provide the next two years of support. The extent of the ongoing support provided to BTR teachers early in their careers is another of the program's distinguishing features.

BTR continues to modify its strategy based upon the school district's needs and insights developed over its eight years of experience. For example, there is a new focus on clustering several program graduates within low-performing schools in order to create a stronger support network and maximize the program's impact. We do not evaluate these recent strategic changes.

In this paper, we examine how well the BTR program has achieved the goals for which it was established. In particular, we address the following questions:

1) Does the BTR program prepare more teachers in hard-to-staff subject areas (i.e., math and science) than BPS's other recruitment sources?

2) Are BTR recruits more racially and ethnically diverse than teachers who enter the district through other pathways?

3) Do BTR recruits remain in the district longer than other new hires?

4) Are BTR teachers more effective than teachers with the same level of experience entering the district through other routes?

The BTR program's effect on student achievement in Boston depends primarily on the answers to the third and fourth questions. It is well documented that teachers improve rapidly in their ability to raise student achievement in their first few years in the profession (Rockoff 2004).

7

Higher retention rates will thus improve student achievement by reducing the number of inexperienced teachers hired each year. However, these retention effects could be either multiplied or offset by differences in the average effectiveness of those who are retained. We therefore conclude our analysis with a simulation intended to model the overall effects of the BTR program on student achievement in the Boston Public Schools in light of the program's effects on both retention and effectiveness.

In all of these analyses, we compare BTR graduates to other new teachers hired by Boston Public Schools. Thus, our evaluation examines the BTR program as a whole and does not explore the precise mechanisms responsible for its effects. For example, candidates who are attracted to BTR may well differ on many dimensions from teachers who choose to enter BPS through other routes, even before these candidates enter teacher preparation programs. BTR also chooses candidates intentionally, selecting teachers who its staff believes will succeed in the classroom, demonstrate a strong attachment to the district, and align with the recruiting goals established by the district. Finally, BTR trains these teachers and offers them ongoing support. We cannot disentangle whether our results reflect the teachers who choose BTR, BTR's selection process, or BTR's training and support. We must rather interpret our findings as the end product of all of these program elements.

3. Data

The Boston Public Schools enrolls 56,000 students in 135 schools. Of these students, 39% are Hispanic, 37% are African-American, 13% are White, and 9% are Asian. Nearly three-quarters of all students are eligible for federal free or reduced price lunches; 21% are enrolled in special education; and 38% are classified as English Language Learners. Boston is an interesting laboratory for research because it has become a relatively successful urban district over the past

8

decade. The district won the Broad Prize in 2006 and scores significantly above the large district average on the National Assessment of Educational Progress in math and reading in both 4th and 8th grades. In fact, the district's students have shown the most rapid growth in NAEP math scores since 2003 among all districts participating in the Trial Urban District Assessment.[2]

Our analysis draws on administrative data provided by BPS that includes student and teacher records for the 2001-02 to 2010-11 school years. The results presented here are based primarily on the period starting in 2004-05, the first school year in which BTR graduates served as teachers in their own classrooms. Our information about BPS teachers, including BTR graduates, comes from the BPS Human Resources database. This dataset contains demographic information on all BPS employees, including their self-identified race/ethnicity and gender. It also indicates whether a BPS employee served as a teacher in the district each year, the definition we use in our analyses of the diversity of BTR graduates and the rate at which they are retained as teachers.[3]

For our analyses of teacher effectiveness, we focus on classroom "teachers of record" whom we can match to students in grades 4 to 8 in the district's student dataset. We define the teacher of record in each subject as the teacher who provides the student their course grade in that subject. We also use these student-teacher links to create grade and subject area indicators for each teacher of record regardless of grade level.

These two definitions identify a different population of BTR graduates and other BPS employees as teachers for several reasons. First, specialist teachers, including Special Education resource teachers, may be identified as teachers in the Human Resources files but not identified

[2] http://nationsreportcard.gov/math_2009/district_summ.asp [Accessed July 20, 2011].
[3] We also examine teacher diversity and retention using only the subset of teachers who are actually "teachers of record" and obtain very similar results. As a result, we present results from the more comprehensive sample of "teachers."

as the teacher of record for any students. Second, some students are missing a teacher indicator for certain subjects; most of these missing student-teacher links come in non-academic subjects such as health and physical education. Finally, a few schools do not report student-teacher links at all. These schools are necessarily excluded from the student achievement analysis. Despite these complexities, a comparison with internal BTR records from the relevant years indicated only minor discrepancies between the sample of BTR graduates included in our value-added analysis and those the organization believed to be teaching in relevant grades and subjects.[4]

In addition to an identifier for the student's teacher in each course, our student dataset includes test scores in math and ELA along with basic demographic information. The test scores come from the Massachusetts Comprehensive Assessment System (MCAS), through which the state tests all students annually in grades 3 through 8. The MCAS test carries high stakes for schools but not for individual teachers or students in the grades we examine. We standardize MCAS scores by grade, subject, and year to have a mean of zero and standard deviation of one. The available demographic information for students includes indicators of race or ethnicity, gender, free and reduced price lunch status, language proficiency, special education status, and whether the student was new to their school or the district that year.

Because BTR is a fairly new program, the total number of BTR graduates (and therefore the sample size for our analyses) remains relatively small. As seen in Table 1-A, 296 BTR graduates in total entered the district as teachers between the 2004-05 and 2010-11 school years. The district employed 4,916 teachers in 2010-11, of whom 262 were BTR graduates. Thus, BTR graduates comprise approximately 5% of all BPS teachers. The program has been growing over

---

[4] In ELA, only 2 teachers listed as grade 4-8 teachers in the BTR records did not appear in our value-added sample, while 2 teachers that the BTR records indicated were not teaching ELA in grades 4-8 were linked to students in those grades (and were therefore included in the value-added analysis). In math, 4 teachers listed as grade 4-8 teachers in the BTR records did not appear in our value-added sample, while 2 of the teachers included in our sample were not listed as grade 4-8 teachers in the BTR records.

time, however, and provided approximately one-third of all first-year teachers hired by the district since 2008-09.[5]

Our analysis of the relative effectiveness of BTR graduates uses a more restricted sample of teachers and students. We only have student test score data in math and ELA, so teachers of other subjects are necessarily excluded. The models we use to estimate teachers' effectiveness in raising student achievement also require students to have consecutive years of testing data, which restricts our analysis to certain grades. Beginning in 2006-07, we have such data for teachers in grades 4 through 8. Importantly, this means that we cannot examine the effectiveness of BTR program graduates in early elementary grades or high school. As seen in Table 1-B, our results on teacher effectiveness derive from approximately 50 BTR graduates (and approximately 100 teacher-year observations) in each subject. This represents roughly 20% of all BTR graduates.

4. Evidence on Recruitment and Retention

We first examine whether the BTR program is achieving its goals of recruiting teachers in hard-to-staff subject areas and increasing teacher retention in the Boston Public Schools. Table 2, which provides the distribution of BTR and other BPS teachers of record across subjects, shows that BTR teachers hired through 2009-10 were indeed more likely to fill math and science classrooms than both other BPS teachers overall and other first-year teachers in particular.[6] Thirty percent of BTR teachers in their first year taught at least one math course and

---

[5] The district does not keep clear records of teachers who are new to BPS. In particular, the available information does not allow us to distinguish easily teachers who are new from those who left the district and returned (e.g., for maternity leave or to teach in another district) or from teachers with experience in another school district. As a result, we base our comparisons on the teacher's official position on the salary scale. Here, we define first-year teachers as those at the first step on the district's salary scale who did not teach in the district in previous years. BTR teachers are considered to be first-year teachers in the year immediately following the residency. We examined several different ways of categorizing new teachers, including examining only teachers who were assigned a district mentor in their first year. All of our results are consistent across these definitions of "new" teachers.

[6] As explained in Solomon (2009), BTR also seeks to address the district's needs for qualified teachers of special education students and English Language Learners. All BTR residents therefore earn partial credit toward a dual certification in special education or English as a Second Language, which they are then expected to complete the

20% taught a science course, compared to 22% of other first-year teachers in math and 11% in science. A similar pattern is evident if we focus on individuals who teach math or science, exclusively. The proportion of new math and science positions filled by BTR teachers has grown over the duration of the program. By 2009-10, 62% of Boston's new math teachers and 42% of new science teachers came through the BTR program, which provided only 34% of the total pool of new teachers BPS hired that year.

Table 3 confirms that BTR graduates were more racially/ethnically diverse than other new BPS teachers hired from 2004-05 to 2010-11. BTR graduates were substantially less likely to be white (52% vs. 68%) than their non-BTR novice peers. In turn, they were 8 percentage points more likely to be African-American and 4 percentage points more likely to be Hispanic, suggesting that the program is increasing diversity among the district's teachers along these dimensions. BTR graduates were quite similar to other novice teachers in terms of gender (both groups were about 70% female).

The right side of Table 3, which provides data on all BPS and BTR teachers regardless of their experience, makes it possible to interpret this evidence concerning diversity among novices in the context of recent trends in the composition of the BPS teacher workforce. The district, apparently, has been less successful in hiring minority teachers than in the past. In particular, we see that non-BTR novice teachers are substantially less likely than BPS teachers overall to be African-American. BTR graduates overall are 13 percentage points less likely than other BPS teachers to be white. They are modestly more likely to be African-American, and nearly twice as likely to be Hispanic or Asian.

BTR residents commit to teach for three years in BPS and face a financial penalty if they

---

following year. Because we do not have data on teacher certification, we are unable to compare BTR graduates to other BPS teachers in this respect.

renege on this commitment, creating strong incentives for program graduates to remain in BPS early in their careers. In Figure 1, we plot retention rates up to year five for the six cohorts of BTR graduates and other BPS teachers hired as novices between 2004-05 and 2009-10. Because we can follow teachers only through the 2010-11 school year, we observe the cohorts of teachers entering after 2006-07 for fewer than five years. The size of the samples on which each data point is based therefore grows smaller with each successive year.

The figure shows that other BPS novices leave the district after their first year at much greater rates than BTR graduates (17% vs. 12%). By year 5, 49% of all BPS novices have left teaching in the district, which aligns closely with retention figures for urban districts nationwide (Ingersoll 2001). However, three-quarters of BTR graduates remain. It is interesting to note that the retention rate of BTR graduates does not decline suddenly when their commitment to BPS (and financial obligation to BTR) has been fulfilled at the end of their third year.

In sum, the BTR program is accomplishing its goals of supplying teachers in "hard-to-staff" subjects and increasing the diversity of the district's teaching force. Program graduates also have substantially lower turnover through year 5 than other novice teachers. Whether the increased retention of BTR graduates leads to higher levels of student achievement in the district, however, depends on their effectiveness in the classroom.

5. Evidence on Teacher Effectiveness

Our analysis of the relative effectiveness of BTR graduates in raising student achievement focuses on math and ELA teachers in grades 4 through 8. We exploit our data linking these teachers to individual students by classroom to estimate what are commonly referred to as value-added models. These models attempt to isolate the portion of each student's growth in test scores from one year to the next that is attributable to that student's teacher from

13

such other factors as achievement in the previous year, demographic characteristics, and the effects of classroom and school peers. We describe the details of our preferred value-added specification below before presenting the results of our analysis.

5.1 Model Specification

All of our results derive from covariate-adjusted value-added models that include prior year test scores as a right-hand-side control variable when predicting the outcome test score. For each subject (math and ELA), we fit a series of models that represent the relationship between a student's standardized test score and a variety of predictors, including a cubic polynomial of their previous year's test scores in both math and ELA ($Y_{i,t-1}$).[7]

Our preferred model is as follows:

$$(1) \qquad Y_{ijsgt} = \alpha_g * f(Y_{ig,t-1}) + \beta * BTR_{jt} + \gamma * f(EXPER_{jt}) + X_{ijst}'\delta + \overline{X_{jst}}'\varsigma + \overline{X_{st}}'\varphi + \theta_{gt} + \varepsilon_{ijsgt}$$

for student i with teacher j in school s, grade g, and year t. We include controls for the teacher's experience level, specified as a full set of dummy variables.[8] We allow the effects of baseline test score to vary by the student's grade and include grade-by-year (θ_{gt}) fixed effects. The individual-level control vector, X_{ijst}, includes all available student demographic characteristics described above in section 3. For each student, we also generate a vector of classroom-level means ($\overline{X}_{jst}$) and school-level means ($\overline{X}_{st}$) of these demographic characteristics, as well as the class size and average student baseline test scores for each subject, to control for peer and classroom or school composition effects. The predictor BTR_{jt} indicates whether the teacher is a

---

[7] In all cases we cluster our standard errors at the classroom level to account for the fact that students in the same classroom are likely to share common unmeasured influences on their achievement.
[8] The BPS salary scale reaches its maximum at 10 years of experience, so we treat all teachers with more than 10 years in the district as having the same level of teaching experience. Fortunately, this limitation does not affect our analysis because no BTR program graduates have been teaching for 10 years. We also include a dummy variable to indicate teachers who are missing salary experience data, allowing us to include them in the analysis. Obviously, we omit one of the experience indicators as a reference category.

14

BTR program graduate. Our parameter of interest is β. If its estimated value is statistically significant and positive, we can conclude that BTR graduates are more effective at raising student achievement than other BPS teachers at the same level of teaching experience.

Model (1) provides an overall comparison of the BTR program and the district's other sources of new teachers, but it constrains any differences in effectiveness between BTR and non-BTR teachers to be the same regardless of their level of experience. Thus, we also fit a more flexible model that separately compares the performance of BTR and non-BTR teachers within three experience categories:

$$(2) \quad Y_{ijsgt} = \alpha_g * f(Y_{i,t-1}) + \beta_1 * NOVICE_{jt} + \beta_2 * NOVICExBTR_{jt} + \beta_3 * EXPER2to3_{jt}$$
$$+ \beta_4 * EXPER2to3xBTR_{jt} + \beta_5 * EXPER4to5_{jt} + \beta_6 * EXPER4to5xBTR_{jt}$$
$$+ X_{ijst}'\delta + \overline{X_{jst}}'\varsigma + \overline{X_{st}}'\varphi + \theta_{gt} + \varepsilon_{ijsgt} .$$

Here, we include predictors that indicate whether the teacher is a novice ($NOVICE_{jt}$), is in their 2nd or 3rd year ($EXPER2to3_{jt}$), or is in their 4th or 5th year ($EXPER4to5_{jt}$), as well as the interactions of these predictors with BTR status.[9] This specification enables us to compare explicitly BTR and other BPS teachers at different stages in the careers.[10] Specifically, parameter β_2 now represents how much more effective BTR teachers are in their first year compared to other novice teachers, while parameter β_4 represents the relative effectiveness of BTR teachers in their 2nd and 3rd years and parameter β_6 represents the same relative effectiveness for teachers in years 4 and 5.

In each analysis, we include only students with non-missing baseline and outcome test

---

[9] In these specifications, we include predictors for teachers with missing experience and for the very few teachers who are non-novices but are at the first step on the district's salary scale (in most cases, these teachers were in the district the previous year). This approach enables us to include the largest possible sample in our models and prevents us from mis-categorizing these teachers as part of the reference category.

[10] We focus on these two-year bins rather than individual experience levels to increase the precision of our estimates given our limited sample size.

data. We also exclude student-year records with missing teacher links (approximately 3% of the sample). In each year, we drop students who are enrolled in very small or large classes or who have multiple teachers in the subject in that year.[11] Thus, our final sample consists exclusively of teachers of traditionally structured math and ELA classes.[12]

Although they are now widely used by researchers, the validity of this kind of value-added model of teacher effectiveness using non-experimental data continues to be debated (e.g., Kane and Staiger 2008; Rothstein 2010). The primary threat to the validity of our inferences is that the model may not fully correct for the sorting of BTR graduates to schools, classroom assignments, or students. To examine the robustness of our results against these threats, we fit a variety of supplementary models. First, we replace our school-level average characteristics with school fixed effects, restricting our inferences to comparisons between BTR graduates and other teachers within the same school. This eliminates any bias that could result from BTR graduates being placed in more (or less) effective schools and fully accounts for any non-random sorting of students and teachers to schools.

Second, we replace our student-level demographic predictors with student fixed effects. As Rothstein (2010) has shown, different types of sorting of teachers to students can create different types of biases. If students are sorted based on pre-existing characteristics, models that do not fully capture these differences (as in (1)) may be biased. Including student fixed effects fully accounts for any of these time-invariant student characteristics. However, models with student fixed effects have several shortcomings. First, they rely exclusively on a non-representative group of students who remain in the sample for several years. Second, to the

---

[11] We exclude students in classes below the 1st or above the 99th percentile of class size (fewer than 5 or more than 39 students), or with fewer than 5 students with pre- and post-test scores.

[12] In supplementary analyses, we fit models that excluded students in classes with majority LEP or special education students. Our results were nearly identical to those reported here.

extent that schools sort students to teachers based on changes in their performance trajectories over time, these models may make the bias due to student sorting worse. Although these models are not our preferred approach, the fact that our main results are generally robust to the inclusion of student fixed effects suggests that differential student sorting is not driving our results.[13]

Finally, we consider whether any differences in the effectiveness of BTR graduates and other teachers are driven by demographic differences in the types of teachers BTR recruits. For example, if male teachers were generally less effective than female teachers and BTR recruited additional male candidates in response to a district priority, BTR program graduates would appear to be less effective. To account for this possibility, we add teacher race/ethnicity and gender to the models in (1) and (2).

5.2 Comparing BTR Graduates and other BPS Teachers

Table 4 presents our main results on the relative effectiveness of BTR graduates and other BPS teachers in grades 4 through 8 assigned to math (top panel) and ELA (bottom panel) classrooms. The first row of each panel presents multiple specifications of model (1) comparing all BTR graduates to other BPS teachers with the same level of experience. The remaining rows present results of model (2), which makes these comparisons separately among teachers within three distinct experience categories.

In math, our preferred specification of model (1), controlling for student, classroom, and school characteristics as well as teacher experience, indicates that other BPS teachers raise student achievement by 5% of a standard deviation more in math than do BTR graduates. Using

---

[13] We also test the sensitivity of our results in several other ways. First, to account for any implementation effects or other anomalies, including the possibility that teachers in early cohorts were particularly ineffective while the program was first being developed, we conducted the analysis by removing each cohort individually. We also examined whether BTR program graduates were relatively more effective with students in special education or those with limited English proficiency. In all cases, the results of these analyses confirmed the general patterns presented here.

school fixed effects to limit comparisons to teachers within the same school increases the estimate to 7% of a standard deviation. The parameter estimate becomes negligible when we fit a model with student fixed effects, a pattern which could reflect systematic sorting of students to teachers or simply the limitations of student fixed-effects models described above. The point estimate also becomes slightly smaller when we introduce controls for teacher demographic characteristics but remains statistically significant at the 90 percent confidence level, suggesting that the estimate in our preferred model is not driven by observed differences in the demographic characteristics of the teachers recruited by BTR.

Model (2), which allows differences in effectiveness between BTR and non-BTR teachers to vary depending on their experience level, provides a more nuanced picture. In their first year as teachers of record, BTR graduates underperform other novice teachers by 9% of a standard deviation in math. This gap is substantial in magnitude, representing approximately one-tenth of the black-white test score gap on the National Assessment of Educational Progress or roughly 2 months of instruction for a typical 6th-grade student (Hill et al. 2008). The performance gap appears to narrow modestly for those in their second or third year as teachers. And, by years 4 and 5, we estimate that the small number of BTR graduates who have reached this stage of their career actually outperform other BPS teachers with the same level of experience by 7% of a standard deviation.

Importantly, the other columns of Table 4 confirm that this general pattern remains the same when we compare teachers in the same school, include student fixed effects, or control for teachers' demographic characteristics. The robustness of our results across these four models suggests that the sorting of teachers to schools, the sorting of students to teachers, and the different demographic make-up of BTR graduates are likely not driving these results.

18

The top panel of Figure 2 illustrates these findings by plotting our preferred estimates (and their 90% confidence intervals) of the average effectiveness of BTR graduates and non-BTR teachers within experience categories, relative to BPS teachers with six or more years of experience. Although both groups underperform veteran BPS teachers by a considerable amount, BTR graduates lag far behind other teachers as novices. By their second and third years, all teachers are much better than novices, and the gap between BTR graduates and other teachers has narrowed modestly. By years four and five, however, BTR graduates are estimated to outperform both non-BTR teachers with the same experience and veteran BPS teachers.

In contrast to the math results, we find that BTR graduates and other BPS teachers do not differ systematically in their ability to raise student test scores in English language arts. As shown in Table 2, each specification of model (1) comparing the performance of teachers with the same level of experience yields estimates that are very close to zero and not statistically significant. Furthermore, the results for model (2) show no statistically significant differences in the performance of BTR graduates and other early-career teachers, regardless of their experience category. The bottom panel of Figure 2, which again plots our estimates of the effectiveness of both groups relative to the teachers with at least six years of experience, illustrates that both groups are relatively ineffective as novices. This gap is essentially eliminated for both groups by years two and three, but there is no evidence that the performance of BTR graduates is worse initially or improves more substantially by years four and five than that of other teachers. These patterns are consistent across model specifications.

As noted above, the differences in math effectiveness documented above are quite large, especially for novice teachers. However, it is important to note that there is far more variation in effectiveness among BTR and non-BTR teachers separately than there is between the two groups

of teachers as a whole in both math and ELA. This is evident in Figure 3, which plots the

distribution of estimated teacher effects separately for first-year BTR graduates and non-BTR

novice teachers hired between 2006-07 and 2009-10.[14] In both math and ELA, the center of the

distribution of novice BTR teachers is to the left (more negative) of the distribution for non-BTR

novices. The distributions are almost entirely overlapping, however, indicating that some BTR

graduates are nonetheless among the district's most effective new teachers in both subjects.

One hypothesis about the performance of teachers from alternative certification programs

holds that, because they are more likely than traditional teacher preparation candidates to have a

disciplinary college major, they may be relatively more effective with older students (Xu,

Hannaway, & Taylor, 2011; Goldhaber & Brewer, 2000). Although we are not able to examine

the performance of BTR graduates in high school classrooms, Table 5 presents results from

specifications of model (1) in which we allow the effect of BTR participation to vary by grade.[15]

In math, there is little evidence to suggest that the relative effectiveness of BTR graduates varies

across grades. Although only the result for grade 4 is statistically significant (at the 90 percent

confidence level) when examined separately, our estimate of the relative effectiveness of BTR

graduates is negative at each grade level and none of the estimates differs by a statistically

significant amount. In ELA, we find suggestive evidence that BTR graduates are relatively less

effective in elementary grades (where our estimates indicate large negative effects in grades 4

and 5) than they are in middle school grades (where we estimate null effects in grades 6 and 8

---

[14] More specifically, we fit a version of model (1) that includes teacher-year fixed effects (τ_{jt}), as

follows: $Y_{ijsgt} = \alpha_g * f(Y_{i,t-1}) + X_{ijst}'\delta + \overline{X_{jst}}'\varsigma + \overline{X_{st}}'\varphi + \theta_{gt} + \tau_{jt} + \varepsilon_{ijsgt}$. We then extracted estimates of
teacher-year effects, which represent the average performance of a teacher's students in a given year, taking into
account their initial level of achievement, their demographic characteristics, and the characteristics of their
classroom and school peers. Plotting the distribution of these effects for all first-year teachers in the district allows
us to compare the variation in effectiveness among BTR graduates and other novices.

[15] We do not present results from model (2) because the number of teachers in each of the experience categories
considered is generally too small within specific grade levels to yield precisely estimated results.

and a large positive effect in grade 7). Although these patterns are intriguing, the small sample

sizes on which these estimates are based again suggest that they should be interpreted cautiously.

5.3 Differential Attrition vs. Returns to Experience

Our primary results in math indicate that the performance of BTR graduates who remain

in the district through years four and five improves more rapidly than that of other new teachers.

There are two possible explanations for this pattern. On the one hand, BTR graduates may

simply improve more rapidly in the early stages of their careers, perhaps leveraging the support

provided by the program or its emphasis on the value of being reflective about their practice even

after their formal training has concluded. On the other hand, this pattern could emerge if BTR

graduates who leave the district are relatively less effective than other teachers who leave.

The results presented in Table 6 attempt to disentangle whether differential attrition or

differential returns to experience are driving these results. In the top panel, we present results

from a model that compares the relative effectiveness of teachers who leave the district to those

who stay.[16] Consistent with recent evidence on early-career attrition in other settings (Hanushek

et al. 2005; West and Chingos 2009; Goldhaber, Gross, and Player 2010), we see that teachers

who leave the district are less effective on average than those who stay. However, this pattern is,

if anything, weaker for BTR graduates than for other teachers, suggesting that differential

attrition is not driving their more rapid improvement by years four and five.

Instead, we find strong evidence that the effectiveness of individual BTR graduates

improves much more rapidly with experience than that of the district's other new teachers. In

the bottom panel of Table 6, we present results from a version of model (1) that includes teacher

---

[16] Technically, we include in model (1) an indicator for whether the teacher leaves the district and the interaction
between this indicator and BTR.

fixed effects. We specify teacher experience as a logarithmic function,[17] and include the interaction between this experience predictor and BTR status.[18] The results confirm that the coefficient on this experience function is much greater for BTR teachers than for non-BTR teachers, suggesting that the within-teacher returns are much larger for BTR program graduates.

5.4 Examining the Effectiveness of BTR Mentors

One possible explanation for why BTR program graduates are relatively ineffective in their first year is that the mentor teachers are not effective "trainers". These mentors may not be effective teachers themselves and thus mentor the residents in ineffective techniques. Or, they may not know the specific reasons they are effective or how to transfer this knowledge. We cannot study the second issue directly, but we can use the same value-added methods described above to examine whether mentors are more effective than other BPS teachers in raising student achievement. In addition, the data on mentors allows us to explore another possible mechanism through which BTR could impact student achievement, by affecting the mentors directly. The presence of residents during the residency year could make mentors more effective (by providing an additional adult in the classroom on a regular basis) or less effective (by providing a distraction to the primary task of classroom instruction). Furthermore, serving as a mentor could improve teachers' effectiveness in the future, as they gain insights into their practice by mentoring a resident.

---

[17] Obviously, constraining the relationship between experience and teacher effectiveness to follow a logarithmic functional form has some limitations. We use this approach for several reasons. First, the overall pattern of within-teacher career growth in BPS appears to follow this form, at least for the period early in the career on which we focus. Second, with our limited sample of BTR teachers, we need to impose some structure on this model to provide reliable estimates. Third, we are primarily interested in the degree of the relative difference in within-teacher returns between BTR and other teachers, not in the shape of those returns in general.

[18] Because the salary scale in BPS is capped at 10 years of experience, all teachers with more than 10 years in the district have the same value on this predictor. As a result, we do not face the issue of collinearity of year and experience within teacher (see Rockoff [2004] for a detailed discussion). This approach likely introduces a downward bias in the overall magnitude of the returns to experience, but we are interested only in the relative difference between BTR and non-BTR teachers.

We examine these questions in two different ways. First, we fit an analogous version of model (1), but we replace the BTR indicator with two predictors: one that indicates whether the teacher was ever a mentor and the other that indicates whether the teacher was a mentor in a given year. We present the results from this model in the second and fourth columns of Table 7, labeled "Overall". Here, we see that BTR mentors are considerably more effective than other BPS teachers, even controlling for experience, in both math and ELA. Thus, we can reject the hypothesis that mentors are simply ineffective teachers and teach these ineffective practices to their residents. However, we find no clear evidence that having a resident in the classroom improves student outcomes. That the parameter estimating the effect of being a mentor in a given year is not statistically significant suggests that their performance in the mentor year is no different than their performance when they do not have a resident in their classroom.

We also examine this question by fitting a model with teacher fixed effects, looking at whether participating as a mentor improves teachers' effectiveness in the mentor year or beyond. Here, we essentially use each teacher as their own control group, comparing their performance during and after their mentor year to their performance before serving as a mentor. Again, we find no evidence that participating as a mentor improves teacher effectiveness, either in the mentor year or beyond. An important caveat to these analyses is that we have no information on which classrooms have teachers' aides or other adults present on a regular basis. As a result, these estimates of the effectiveness of mentors when they are paired with a resident may understate the true effect of having an additional adult in the classroom.

5.5 The Net Impact of Differences in Retention and Effectiveness

Understanding the overall effect of the program on student outcomes in the system is challenging. On the one hand, BTR novice teachers appear to be initially much less effective

than their novice peers in BPS, at least in math. On the other hand, these BTR teachers appear to become more effective faster than other teachers, such that they eventually outperform not only other teachers at the same level of experience but also veteran teachers in the district. Furthermore, BTR teachers stay in the district much longer than their peers. Because novice teachers in general are less effective, improved retention is associated with better student achievement. Any evaluation of the program's overall effect needs to incorporate both of these considerations.

We therefore examine two hypothetical staffing models: In Staffing Model A we start with 50 classrooms filled with traditional BPS new teachers and in Staffing Model B we start with 50 classrooms filled with BTR graduates. We use our estimated retention rates to determine turnover each year within each staffing model and assume that any vacancies within each model are filled by novice teachers from traditional routes in Model A and BTR graduates in Model B.[19] We track the predicted performance of the fifty classrooms over time.

Because of higher retention rates in the early years, the BTR teaching force grows more experienced more quickly than the teaching force constructed from novices entering through traditional routes. This higher retention rate is important for understanding the possible systemic effects of BTR. In mathematics, BTR residents initially perform much worse than their peers. As a result, staffing Model A dominates in the early years. However, BTR teachers' superior retention and more rapid improvement leads staffing Model B to become more effective in aggregate in math after year 4. In the long run, the average student under the BTR staffing model would gain roughly 0.04 standard deviations each year more than their peers under the

---

[19] Importantly, because we have no information about BTR retention after year 5, we assume a constant year-to-year retention rate of 90% for BTR and non-BTR teachers. To the extent that BTR graduates do indeed remain in the district at greater rates even after year 5, these results will understate the true program effects. Conversely, to the extent that the overall retention rate for BTR graduates is actually lower than that of other BPS teachers after year 5, true program effects will be overstated.

non-BTR model. This effect is driven both by the improved performance of BTR graduates in years 4 and 5 and by their higher retention rates. In ELA, our effectiveness analysis showed no statistically significant differences in the performance of BTR graduates and of BPS teachers from traditional routes. The higher early-career retention rate of BTR graduates implies that Model B would raise student achievement more than Model A by year 2. However, the lack of substantial performance differences means that these two models would be quite similar in their effect on ELA student achievement over time.

It is important to note that this analysis of alternative staffing models is only illustrative. First, it takes literally the point estimate differences between BTR teachers and other teachers in retention and in effectiveness at different levels of experience. In particular, the mathematics analysis rests on our estimate of improved BTR performance among more experienced teachers, which is based on a sample of just 11 teachers. We also have no information about either performance or retention for these teachers after five years in the classroom. Furthermore, it is only a partial equilibrium analysis and focuses only on these two mechanisms. To the extent that improved teacher effectiveness may have spillover effects on colleagues, as suggested by Jackson & Bruegmann (2010), these results may be understated. Furthermore, there is mounting evidence that teacher turnover not only causes students to have less experienced teachers, but it can also have other disruptive effects. For example, turnover can inhibit schools' efforts to build instructional capacity and develop strong organizational cultures. Ronfeldt et al. (2011) find that turnover reduces student achievement above and beyond changes in the composition of teachers. Because our attempt to quantify these effects only focuses on compositional changes, it again likely understates any differences. However, these results are important because they suggest that over time the BTR program would not have a negative impact on math achievement in the

district and may even have a positive effect. This is the case despite the fact that its graduates are initially less effective than other new teachers hired by the district in math.

6. Conclusion

The Boston Teacher Residency has been successful in recruiting teachers in hard-to-staff subjects, in increasing the diversity of the teaching workforce, and in reducing turnover among novice teachers in Boston Public Schools. However, our analysis found that BTR graduates in the grades and subjects for which effectiveness data are available do not outperform their novice colleagues in terms of raising student test scores in English Language Arts, and in math BTR graduates initially underperform their colleagues by a large margin. The initial struggles of BTR novices come despite the intensive residency year training. Over time, however, we find that BTR graduates "catch up" with their colleagues and go on to outperform them in math.

Regardless of their initial route into teaching, classroom effectiveness rises dramatically during the first few years of teaching for both groups. We do not have any evidence to determine whether or not the BTR graduates benefited from the residency year. It could be that their first-year effectiveness would have been even worse without it. However, given the cost of the program, especially in terms of the resident's time, it will be important to determine if the BTR residents would have demonstrated the same pattern without the residency year. For example, a recent analysis of the New York City Teaching Fellows program, an alternative route program which does not include a residency year, shows a similar pattern: slightly lower impacts the first year and more rapid payoffs to experience (Kane, Rockoff, and Staiger 2008).

A number of important caveats to the achievement findings should be considered before coming to the conclusion that BTR and its practice-based model of teacher preparation either works or does not work in raising student achievement. First, the achievement analysis could

only be conducted for a minority of BTR graduates. Except for middle-school math, the teachers

in our sample were not teaching in the hard-to-staff subject areas on which BTR targets its

recruitment efforts and where we might expect them to have a comparative advantage. Even in

the grades and subjects for which performance data are available, the total number of BTR

graduates remains quite small, especially when we compare teachers in years 4 and 5.

Second, understanding the appropriate counterfactual when assessing BTR's impact on

Boston Public Schools is a challenge. We would like to compare BTR graduates to the novice

teachers who would otherwise be teaching in their classrooms. In our analysis, we assume that

these teachers would perform as well as the average novice in the district. However, the teachers

who would have been hired to fill these classrooms in the absence of BTR could have been

worse (or better) than the average novice teacher. For instance, if we believe that in the absence

of BTR, BPS schools would have had to dig deeper in their applicant pool to fully staff, we

might expect for these additional, last hires to be of lower quality than the current average

novice. In this way, BTR may serve to keep the quality of non-BTR novices high.

Third, it may be that Boston Public Schools in general faces a more favorable labor

market for hiring new teachers than many of the urban school districts in which residencies now

operate. BPS is a relatively small urban system in a desirable location – and it pays relatively

well compared to other large urban districts. A recent NCTQ report finds that "Boston enjoys

clear advantages in the quality of teachers it attracts," hiring teachers with stronger academic

credentials than an average school district (NCTQ, 2010, p. 5). As a result, BPS may be more

selective in which novices it hires and these novices may not reflect those in other systems.

Finally, it is important to emphasize that our effectiveness analysis cannot distinguish the

roles played by selection, training, and on-going support. BTR provides substantial on-going

27

induction and mentoring support to its graduates in all but their first year. All first-year teachers receive a mentor provided through a district-wide program distinct from BTR. After the first year, BTR graduates once again receive coaching from BTR mentors. Thus, a BTR graduate goes through the BTR program, receives a different mentor in their first year as a teacher of record, and returns to a BTR mentor afterwards. This discontinuity in the support program could create confusion for BTR graduates during their first year, thus lowering effectiveness.

Nonetheless, the early achievement effects are disappointing. We can only speculate on why the patterns we find occur, but understanding them is critical for policymakers seeking to decide about expanding the Urban Teacher Residency model. For example, BTR graduates may underperform their colleagues initially because being in a classroom with an expert mentor makes the resident reliant and does not teach them how to build an effective classroom culture. When the resident gets her own classroom, she may not have built the skills to effectively manage it. On the other hand, it may be that the district's induction program for new teachers conflicts with key messages from the BTR preparation, leaving teachers confused. Of course, there are many other hypotheses that could explain these patterns.

There is clearly the need for further research on the Urban Teacher Residency model before it is widely adopted. This research should examine not only the mechanisms behind these patterns, but a variety of other questions. We do not have any information on the longer-term retention and achievement impacts of teacher residencies, and the number of even intermediate-term graduates in this study is quite small. Furthermore, we examine only one specific program, the Boston Teacher Residency. Program design and implementation, and therefore program effectiveness, likely varies across sites.

Future evaluations of the UTR model should pay careful attention to program costs.

Although we did not have access to this information, it would be useful to examine the full costs of teacher residencies relative to other teacher preparation pathways, including the costs of turnover, training, and opportunity costs for the candidate. In the case of BTR, external funding covers much of these costs, while individuals bear the brunt of program expenses in traditional preparation programs. These analyses should pay careful attention both to programmatic costs and to whether districts, individuals, or other entities bear these costs.

This study provides mixed results concerning the effectiveness of the BTR program. On the one hand, new BTR graduates are much less effective than their peers in mathematics. On the other, these teachers improve much more rapidly and are more likely to stay in the district. These findings highlight two key lessons for similar evaluations of teacher preparation programs. First, it is important to assess the dynamics of programs over time, especially with a new and evolving program. With teachers, whose performance in the classroom changes rapidly over time, understanding not only initial performance but also development throughout the career is critical. Similarly, evaluations should not simply examine average differences in teacher effectiveness, but should simultaneously consider the cumulative effects of differences in retention rates on student achievement in the district. Doing so will provide richer and more comprehensive information about a program's overall effectiveness.

References

Barry, Barnett, Diana Montgomery, Rachel Curtis, Mindy Hernandez, Judy Wurtzel, and Jon Snyder. 2008. "Creating and Sustaining Urban Teacher Residencies: A New Way to Recruit, Prepare, and Retain Effective Teachers in High-Needs Districts." The Aspen Institute and Center for Teaching Quality.

Boyd, Donald J., Pamela L. Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2006. "How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement." *Education Finance and Policy* 1(2): 176-216.

Boyd, Donald J., Pamela L. Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2009. "Teacher Preparation and Student Achievement." *Education Evaluation and Policy Analysis* 31(4): 416-440.

Goldhaber, Dan, and Dominic Brewer. 2000. "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement." *Educational Evaluation and Policy Analysis* 22(2): 129–145.

Goldhaber, Dan, Betheny Gross, and Daniel Player. 2010. "Teacher Career Paths, Teacher Quality, and Persistence in the Classroom: Are Public Schools Keeping their Best?" *Journal of Policy Analysis and Management*, 30(1): 57–87.

Hanushek, Eric A, John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. 2005. "The Market for Teacher Quality." NBER Working Paper No. 11154. Cambridge, Mass: National Bureau of Economic Research.

Hill, Carolyn J., Howard S. Bloom, Alison Reback Black, and Mark W. Lipsey. 2008. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives* 2(3): 172-177.

Ingersoll, Richard. 2001. "Teacher turnover and teacher shortages: An organizational analysis." *American Educational Research Journal* 38(3): 499-534.

Jacob, Brian A. 2007. "The Challenges of Staffing Urban Schools with Effective Teachers." *Future of Children* 17(1): 129-153.

Kane, Thomas J. and Douglas O. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." NBER Working Paper No. 14607. Cambridge, Mass: National Bureau of Economic Research.

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger. 2008. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27(6): 615-631.

National Academy of Education. 2008. "Teacher Quality." Education Policy White Paper.

Washington, DC: Author.

National Commission on Teaching and America's Future. 1996. *What Matters Most: Teaching for America's Future*. Woodbridge, VA: Author.

National Council for Accreditation of Teacher Education. 2010. "Transforming Teacher Education Through Clinical Practice: A National Strategy to Prepare Effective Teachers."

National Council on Teacher Quality. 2010. *Human Capital in Boston Public Schools: Rethinking How to Attract, Develop, and Retain Effective Teachers.* Washington, DC: Author.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94(2): 247-52.

Ronfeldt, Matthew, Hamp Lankford, Susannah Loeb, and Jim Wyckoff. 2011. "How Teacher Turnover Harms Student Achievement." Unpublished manuscript.

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 94(2): 247–252.

Sawchuk, Stephen. 2011. "Teacher Residencies Make Strides, Encounter Obstacles," *Education Week* 30(36, July 13): 12-13.

Solomon, Jesse. 2009. "The Boston Teacher Residency: District-Based Teacher Education." *Journal of Teacher Education* 60(5): 478-488.

West, Martin R. and Matthew M. Chingos. 2009. "Teacher Effectiveness, Mobility, and Attrition in Florida," in Matthew G. Springer, ed., *Performance Incentives: Their Growing Impact on American K-12 Education*. Washington DC: Brookings Institution Press, pp. 251–271.

Xu, Zeyu, Jane Hannaway, and Colin Taylor. 2011. "Making a Difference? The Effects of Teach For America in High School." *Journal of Policy Analysis and Management* 30(3): 447-469.

Figure 1. Percentage of first-year BTR and non-BTR teachers who remain teaching in BPS to years 2, 3, 4 and 5, with asterisks to denote statistically significant between-group differences.
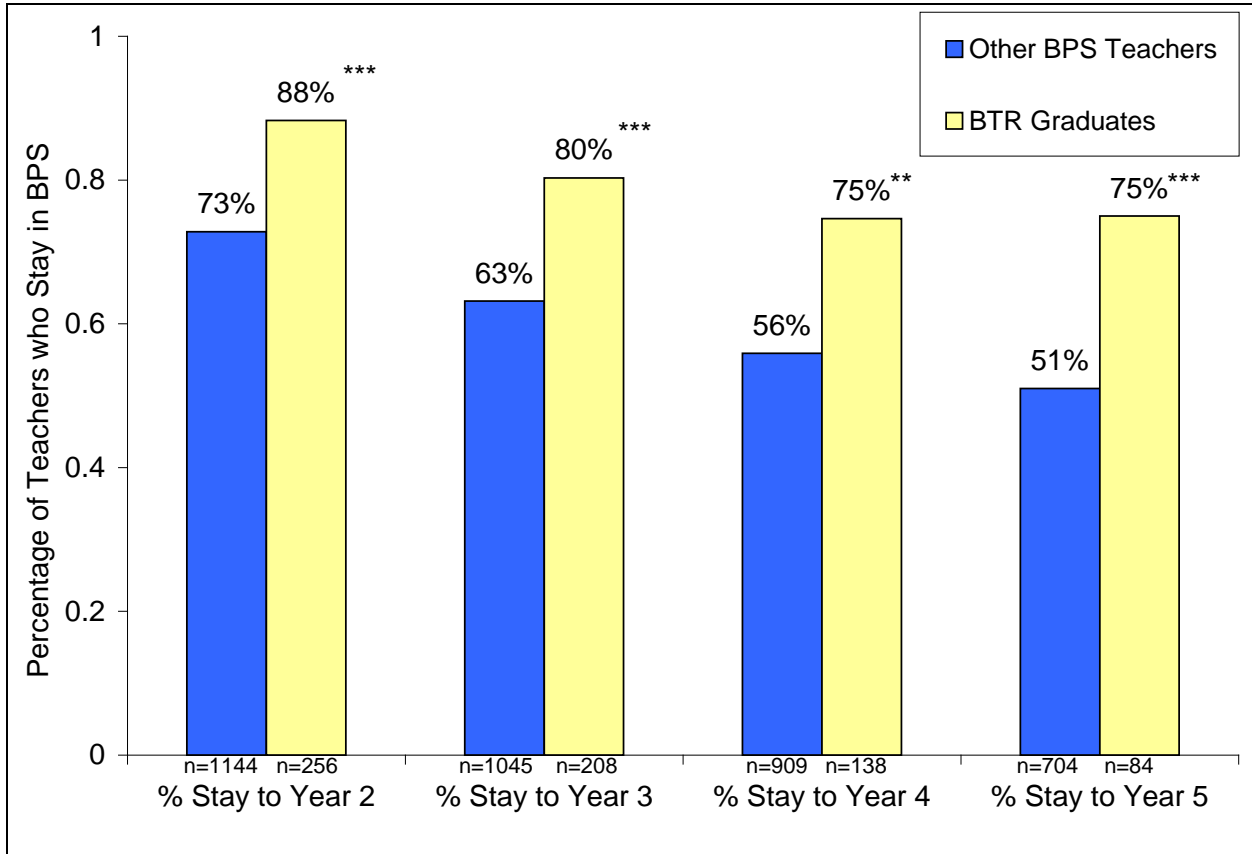
NOTE: *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$

Figure 2. Average value-added for BTR and non-BTR teachers, relative to BPS teacher with 6+ years of experience, with 90% confidence bands in mathematics (top) and ELA (bottom panel).
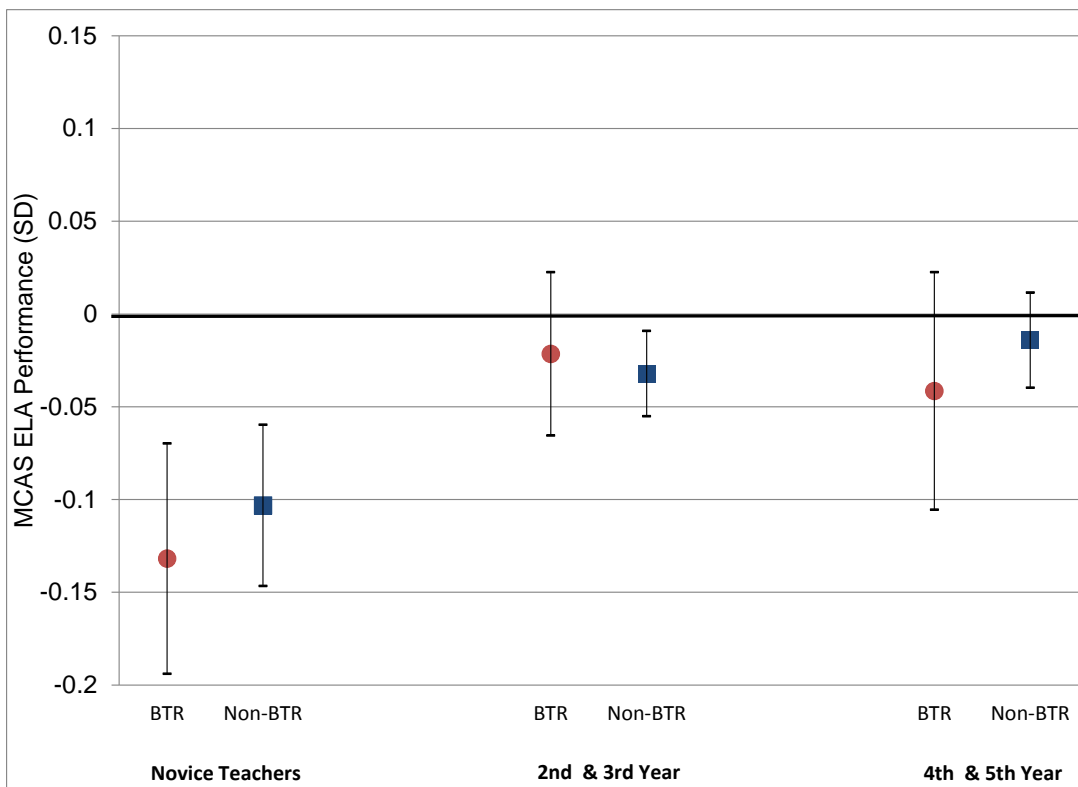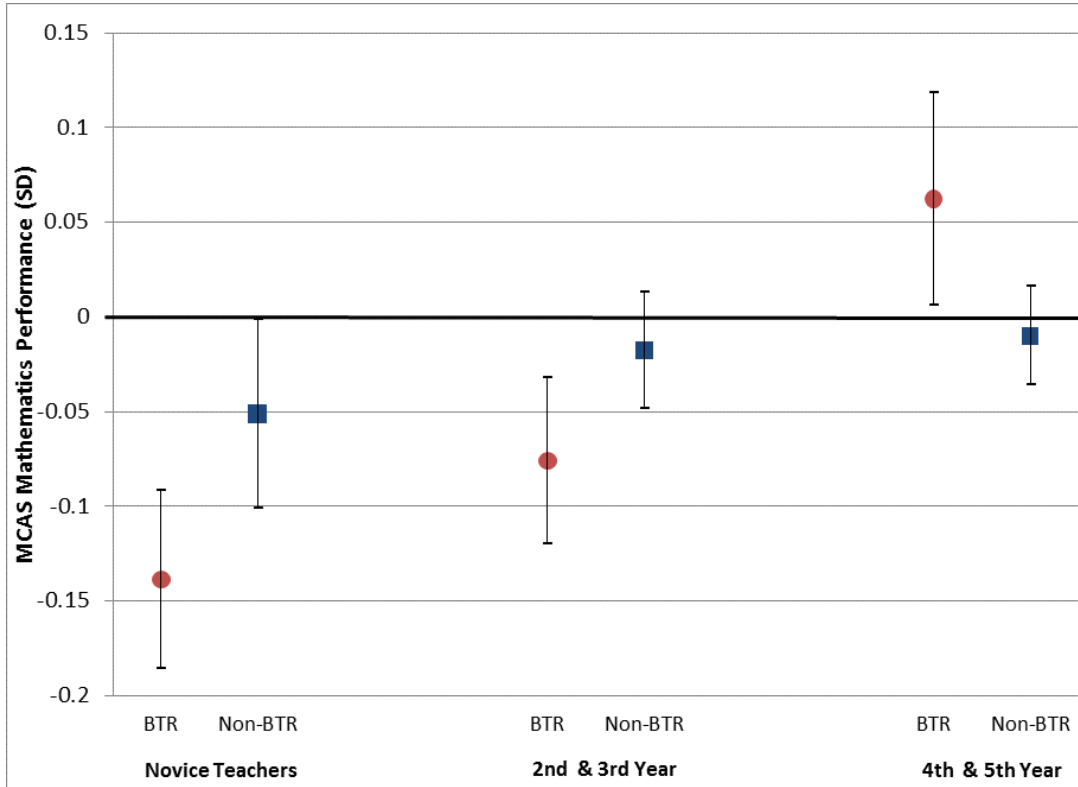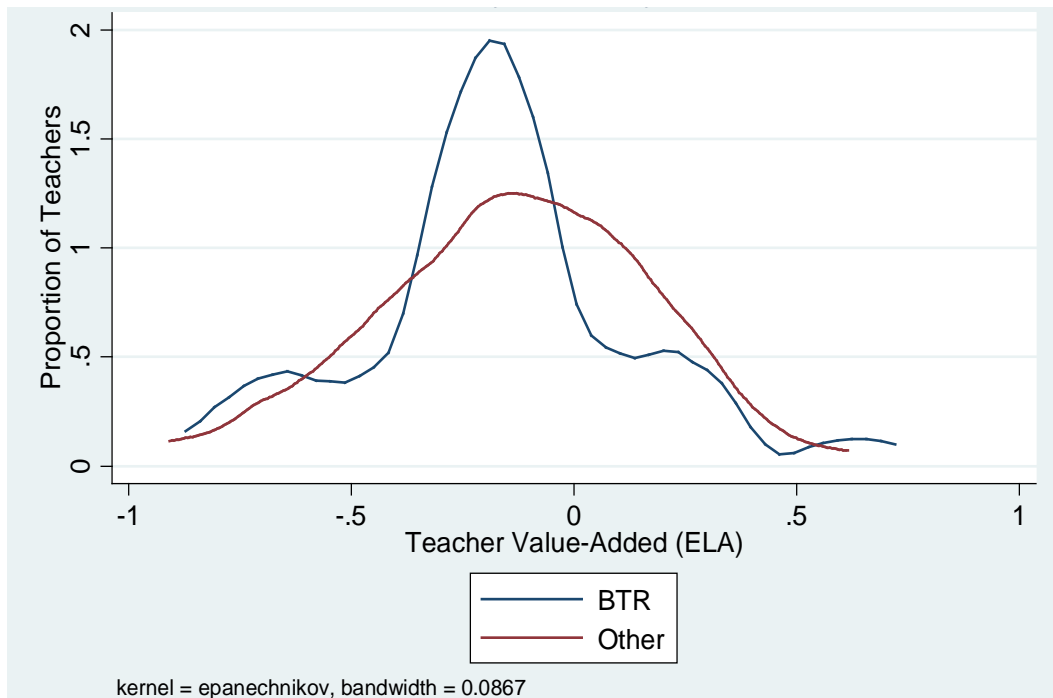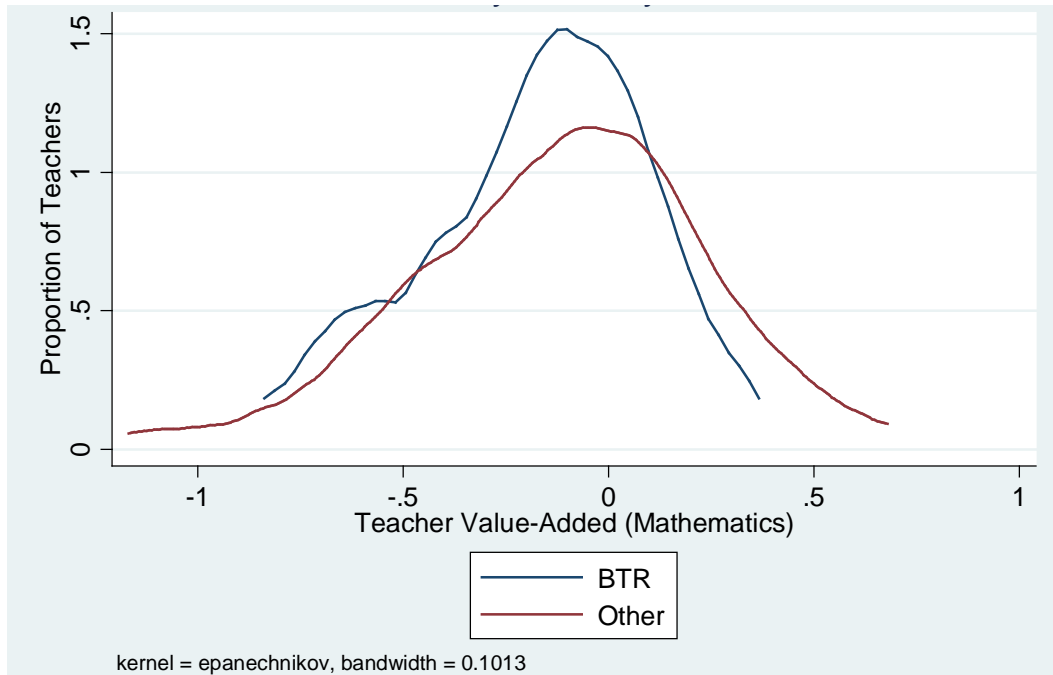
Figure 3. Distribution of teacher value-added estimates for novice teachers, for BTR and non-BTR teachers in mathematics (top panel) and ELA (bottom panel).

kernel = epanechnikov, bandwidth = 0.1013

kernel = epanechnikov, bandwidth = 0.0867

Table 1-A. Total number of BTR and non-BTR teachers in Boston Public Schools, overall and among first-year teachers, by year.

Category	First-year Teachers		Overall	
	BTR	Non-BTR	BTR	Non-BTR
2004-05	11	244	11	4,803
2005-06	31	228	43	4,759
2006-07	42	229	84	4,893
2007-08	54	205	123	4,939
2008-09	70	136	180	4,828
2009-10	48	98	220	4,595
2010-11	56	193	262	4,654

NOTE: District employees are considered to be teachers if they are listed as a teacher in the Human Resources data files.

Table 1-B. Number of BTR and non-BTR teachers in the grades 4-8 value-added samples, overall and by year.

Category	Math VA Sample		ELA VA Sample	
	BTR	Non-BTR	BTR	Non-BTR
Overall (teachers)	52	995	50	1003
By Year (teacher-years)				
2006-07	21	640	19	631
2007-08	25	599	21	600
2008-09	30	585	31	556
2009-10	28	542	28	536

NOTE: The value-added analysis includes only teachers who are listed as the "teacher of record" for students in math or ELA classrooms in grades 4 through 8.

Table 2. Subject areas taught by BTR and other BPS teachers linked to students, overall and for first-year teachers, with row percentages.

	Total	Elementary School	Middle and High School			
			Math only	Any Math	Science only	Any Science
All Teacher-years						
BTR	573	182	123	167	68	105
		31.8%	21.5%	29.1%	11.9%	18.3%
Non-BTR	24,870	8,766	3,087	4,942	1,790	3,255
		35.2%	12.4%	19.9%	7.2%	13.1%
All First-year teachers						
BTR	208	60	47	63	26	42
		28.8%	22.6%	30.3%	12.5%	20.2%
Non-BTR	900	242	125	197	96	96
		26.9%	13.9%	21.9%	10.7%	10.7%

Table 3. Proportion of BTR and other BPS teachers with selected demographic characteristics, overall and among novice teachers.

	Novices 2004-05 to 2010-11		All BPS Teachers	
	Non-BTR Average	BTR Average	Non-BTR Average	BTR Average
% Female	0.727	0.691	0.729	0.676 *
% African-American	0.163	0.246 **	0.220	0.257
% Asian-American	0.062	0.090	0.046	0.083 **
% Hispanic	0.089	0.133 *	0.088	0.141 **
% White	0.682	0.523 ***	0.642	0.511 ***

NOTE: Asterisks denote statistically significant differences between BTR and other BPS teachers: ~ $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$. District employees are considered to be teachers if they are listed as a teacher in the Human Resources data files. Novices include all BTR teachers in their first year and all other BPS teachers at Step 1 on the district's salary schedule.

Table 4. Average difference in value-added between BTR and other BPS teachers, overall (model 1) and by teacher experience level (model 2) in mathematics (top panel) and ELA (bottom panel).

| | BTR vs. other BPS teachers | | | |
Specification	Preferred Model	School Fixed Effects	Student Fixed Effects	Teacher Demographics
Mathematics				
M1: BTR vs All teachers, controlling for experience (BTR n= 52)	-0.052 *	-0.072 **	0.000	-0.042 ~
	(0.023)	(0.026)	(0.025)	(0.024)
M2: BTR vs. Other novice teachers (BTR n= 37)	-0.088 *	-0.099 *	-0.064	-0.089 *
	(0.040)	(0.042)	(0.045)	(0.040)
BTR vs. Other teachers years 2-3 (BTR n= 27)	-0.058 ~	-0.081 *	0.012	-0.045
	(0.031)	(0.033)	(0.032)	(0.031)
BTR vs. Other teachers years 4-5 (BTR n= 11)	0.072 *	0.066 ~	0.081	0.083 *
	(0.036)	(0.039)	(0.052)	(0.038)
English language arts				
M1: BTR vs All teachers, controlling for experience (BTR n= 50)	-0.009	-0.006	0.016	-0.014
	(0.023)	(0.024)	(0.026)	(0.024)
M2: BTR vs. Other novice teachers (BTR n= 31)	-0.029	-0.020	-0.030	-0.034
	(0.045)	(0.043)	(0.048)	(0.047)
BTR vs. Other teachers years 2-3 (BTR n= 31)	0.011	0.007	0.037	0.005
	(0.028)	(0.029)	(0.035)	(0.029)
BTR vs. Other teachers years 4-5 (BTR n= 13)	-0.027	-0.010	0.030	-0.033
	(0.041)	(0.041)	(0.046)	(0.040)
School-level averages	Yes	No	Yes	Yes
School fixed effects	No	Yes	No	No
Student fixed effects	No	No	Yes	No
Teacher demographic characteristics	No	No	No	Yes

NOTE: Cell entries contain point estimates and robust standard errors (clustered at the classroom level), with asterisks to denote inference: ~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001. All models include lagged mathematics and ELA test scores, student demographic characteristics, class-level demographic characteristics, and grade-by-year fixed effects. Total numbers of BTR teachers included in the sample are presented at the end of each row label. Estimates refer only to teachers in grades 4-8.

Table 5. Average difference in value-added between BTR and other BPS teachers, from our preferred specification of model (1) controlling for teacher experience, in mathematics and ELA, by teacher grade level.

Grade	BTR vs. other BPS teachers	
	Mathematics	**ELA**
Grade 4 (BTR n= 14M, 14E)	-0.100 ~	-0.128 *
	(0.054)	(0.056)
Grade 5 (BTR n= 13M, 13E)	-0.033	-0.150 **
	(0.062)	(0.049)
Grade 6 (BTR n= 15M, 13E)	-0.018	0.025
	(0.046)	(0.037)
Grade 7 (BTR n= 18M, 14E)	-0.046	0.136 ***
	(0.030)	(0.032)
Grade 8 (BTR n= 9M, 13E)	-0.062	-0.027
	(0.040)	(0.040)

NOTE: Cell entries contain point estimates and robust standard errors (clustered at the classroom level), with asterisks to denote inference: ~ $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All models include lagged mathematics and ELA test scores, student demographic characteristics, class-level demographic characteristics, and grade-by-year fixed effects. Total numbers of BTR teachers included in the sample are presented at the end of each row label.

Table 6. Estimates from variations of model (1) that examine the relative effectiveness of BTR and non-BTR teachers who leave the district (top panel) and the relative within-teacher returns to experience for BTR and non-BTR teachers, in mathematics and ELA.

	Non-BTR	BTR	Relative Difference
Attrition: Relative performance of teachers who leave (compared to stayers)			
Mathematics	-0.092 ***	-0.052	0.040
	(0.021)	(0.096)	(0.099)
English language arts	-0.080 ***	-0.087	-0.007
	(0.019)	(0.089)	(0.091)
Returns to Experience: Within-teacher growth (coefficient on ln(exper))			
Mathematics	0.015	0.109 ***	0.094 *
	(0.025)	(0.032)	(0.037)
English language arts	0.072 **	0.118 *	0.046
	(0.026)	(0.051)	(0.053)

NOTE: Cell entries contain point estimates and robust standard errors (clustered at the classroom level), with asterisks to denote inference: ~ $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$. All models include lagged mathematics and ELA test scores, student-level, class-level, and school-level demographic characteristics, and grade-by-year fixed effects. Estimates refer only to teachers in grades 4-8.

Table 7. Average difference in value-added between BTR mentor teachers and other BPS teachers, overall and in the mentor year.

Predictor	Mathematics		English Language Arts	
	Overall	Within Teacher	Overall	Within Teacher
Ever Mentor	0.054 **	--	0.065 **	--
	(0.021)		(0.021)	
Ever Mentor X Mentor Year	0.027	0.030	-0.001	-0.015
	(0.026)	(0.024)	(0.029)	(0.032)
After Mentor Year	--	0.017	--	-0.013
		(0.036)		(0.041)

NOTE: Cell entries contain point estimates and robust standard errors (clustered at the classroom level), with asterisks to denote inference: ~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001. All models include lagged mathematics and ELA test scores, student-level, class-level, and school-level demographic characteristics, and grade-by-year fixed effects. N= 50 BTR mentors in math, 45 in ELA. Estimates refer only to teachers in grades 4-8.