Different Skills? Identifying Differentially Effective Teachers of English Language Learners

Benjamin Master (bmaster@rand.org)

RAND Corporation

Susanna Loeb (sloeb@stanford.edu)

Camille Whitney (cwhitney@stanford.edu)

Stanford University

James Wyckoff (Wyckoff@virginia.edu)

University of Virginia

Abstract

Nationwide, K-12 students designated as English language learners (ELLs) must learn both language and content simultaneously, and ELLs score far below the national average in math achievement. Many educators have suggested that identifying or developing teachers with skills specific to ELLs' instructional needs may be critical to addressing this challenge. However, the evidence base to test this assertion is sparse.

This study seeks to identify the characteristics and learning experiences of general education teachers who are differentially effective at promoting math achievement among ELLs compared to non-ELLs. Our analyses indicate that individual teachers can learn specific skills that make them more effective with ELL students. In particular, while generic years of teaching experience do not differentially affect gains for ELLs compared to non-ELLs, specific prior experience teaching ELLs predicts improvements in novice teachers' differential instructional effectiveness with ELLs. We also find that both in-service and pre-service training focused on ELL-specific instructional strategies are associated with higher gains for a teacher's ELLs relative to their non-ELLs. The findings of this study provide valuable new evidence in support of the notion that general education teachers can develop useful ELL-specific instructional skills, informing efforts by educators and policymakers to improve the quality of instruction that ELLs receive.

Introduction

The number of English language learners (ELLs) in K-12 schools has grown rapidly in recent years. Roughly 4.6 million students were estimated to be ELLs in school year (SY) 2009-10, representing approximately 10 percent of all public school students (U.S. Department of Education, 2010), with a growth rate of 27 percent over the previous decade, compared to a 4 percent growth rate for all other students. As the number of ELLs grows, the question of how schools can best meet these students' academic needs also grows in importance.

English Language Learners and Math Achievement

ELLs perform quite poorly on math achievement exams, relative to other groups of students (U.S. Department of Education, 2011). For instance, 72 percent of students classified as ELLs in 8th grade scored “below basic” on NAEP math exams in 2010-11, compared to 17 percent of white students, 40 percent of Hispanic students, and 41 percent of low-income students (U.S. Department of Education, 2011). Research indicates that some ELLs understand more math content than they are able to demonstrate given the linguistic demands of the tests (Abedi, 2002). However, scholars have argued that understanding language is important for learning new math concepts, that these concepts cannot be reduced to a universal language of “symbols,” and that teachers need to consider how to make new concepts more accessible to ELLs in particular (Garrison & Mora, 1999). In the near future, teachers will also need to support ELLs’ access to the higher-order problem solving and argumentation required by the Common Core State Standards, such as being able to “construct viable arguments and critique the reasoning of others” (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010.) Thus, in order to mitigate the math achievement disparities between ELLs and their peers, teachers need to provide support for ELLs to learn math concepts, participate successfully in classroom activities, and demonstrate knowledge on assessments.

Teacher Quality and English Language Learners

Teacher quality is one of the most important school-related factors that researchers have identified for improving student academic performance (Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). However, relatively little research has examined teacher effectiveness specifically for ELLs. Greater attention to identifying the characteristics of teachers who are effective with ELLs may help schools to support ELL’s math achievement by informing teacher training, development, and classroom assignments.

In this paper, we investigate two different types of teacher characteristics that we hypothesize will relate to their effectiveness with ELLs. First, we expect that characteristics and experiences associated with more effective teachers in general - such as knowledge and teaching experience - will be associated with roughly similar achievement gains for both ELLs and non-ELLs. Second, we expect that some characteristics of teachers' training or practice will support differential effectiveness with their ELLs, relative to non-ELLs. These characteristics include having attended pre-service training or professional development related to instructional strategies specific to ELLs, or having prior experience working with ELLs in the classroom.

General teacher quality and student achievement. A large body of research has assessed the characteristics of effective teachers for student math achievement. While many of the measured characteristics of teachers, such as whether they have a master's degree, do not predict greater effectiveness, a variety of identifying characteristics do. One of the characteristics examined in this study is overall knowledge, as measured by the Liberal Arts and Science Test (LAST) which is required for prospective teachers in New York. LAST scores as well as college-entrance exam (SAT) scores are positively associated with teachers' performance in the classroom, though relatively weakly (Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2008). This study also examines teaching experience, which has been shown to predict greater effectiveness, especially during the first few years of teaching, both overall (Clotfelter, Ladd & Vigdor, 2007; Nye, Konstantopoulos, & Hedges, 2004; Rockoff, 2004) and in particular teaching assignments (Ost, 2009). Some studies have also found beneficial effects of particular teacher preparation experiences. For instance, high quality field experiences and preparation directly linked to the practice of teaching predict instructional effectiveness (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Ronfeldt, 2012). Teachers' effectiveness may also improve as a result of high-quality school-based interventions and professional development (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007).

Our first hypothesis is that certain teacher characteristics that are associated with general teaching effectiveness will predict performance with ELLs to a similar degree as with non-ELLs. Specifically, we examine failing the LAST exam on the first try as well as years of teaching experience in New York state (as a proxy for total years of teaching experience). We do not examine the particular mechanisms such as specific teaching practices that explain why certain characteristics are associated with greater effectiveness. However, prior literature suggests that many teaching practices that support learning for students in general also support learning for ELLs, such as having clear objectives and frequent assessments (see Goldenberg & Coleman, 2010 for a review). Prior research has found that, while not perfectly correlated, there is a high correlation between a teacher's effectiveness with their ELLs and their non-ELLs. In Miami, effectiveness in math with the two groups of students is highly correlated at .89 (Loeb et al., 2014). We theorize that teachers with more knowledge and experience are more proficient in engaging in those practices that benefit ELLs and non-ELLs alike.

Identifying the characteristics of teachers that are associated with general effectiveness and assigning these teachers to teach ELLs can help to support ELL academic achievement. In this vein, many studies have called attention to the potentially problematic tendency for ELLs in some regions to be taught by less skilled or credentialed teachers, in large part due to the schools that they attend (Gandara, Rumberger, Maxwell-Jolly & Callahan, 2003; Grunow, 2011; Lankford, Loeb, & Wyckoff, 2002; Peske & Haycock, 2006).

Specific teacher quality and ELL achievement. Our second hypothesis is that certain teacher characteristics and training experiences will predict differential teacher effectiveness with ELLs, meaning that ELLs will gain more ground on math test scores relative to their non-ELL peers when assigned to such teachers. A growing body of theory and qualitative evidence has suggested particular instructional practices and frameworks that may be specifically helpful for teachers of ELLs (Echevarria, Vogt, & Short, 2010; Faltis, Arias, & Ramirez-Marin, 2009; Genesee, Lindholm-

Leary, Saunders, & Christian, 2006). In particular, teachers who recognize and explicitly teach the semiotics and vocabulary used in math, and who successfully relate the ‘language’ of math to ELLs’ existing knowledge of English and of their first language, may be more effective (Garrison & Mora, 1999; Gutierrez 2002; Janzen, 2008; Lyster, 2007; Moschkovich, 2012; Schleppegrell, 2007; Zwiers, 2013).

There is, however, relatively little corresponding research that explicitly tests whether teacher characteristics or learning experiences relate to improved academic outcomes for ELLs (Lucas, Villegas, & Freedson-Gonzalez, 2008). In this vein, a recent study by Loeb et al. (2014) indicates that teachers who speak Spanish or who possess bilingual teaching certification have a smaller achievement gain gap between their ELLs and non-ELLs in math, all else equal. In contrast, Betts, Zau, and Rice (2003) explore ELL student achievement gains as a function of teacher characteristics, and find little evidence that teachers with ELL-specific certifications spur greater ELL achievement gains. They do not, however, explicitly examine differential effectiveness between ELLs and non-ELLs taught by the same teacher.

In addition to research on teachers’ certification status and fluency in students’ home language, there is some evidence that specific instructional techniques may be differentially relevant for ELLs. Randomized controlled trials have found that ELLs benefit from explicit teaching of academic vocabulary (Carlo et al., 2004; Lesaux, Kieffer, Kelley, & Harris, 2014) and from practicing oral and written English within the context of science or social studies instruction (Brown, Ryoo, & Rodriguez, 2010; Vaughn et al., 2009), which suggest that these practices might be beneficial for math achievement as well (Baker et al., 2014).Collectively, this research offers some evidence that there are specific skills that teachers can learn and employ that are especially helpful to ELLs’ achievement. However, much of the extant research on ELL-specific instruction has used tests of student language proficiency as outcomes, rather than students’ academic achievement gains. Math outcomes in particular have received scant attention (Genesee et al., 2006).

Teacher quality and novice teachers. A large proportion of students learn math with novice teachers: about 25 percent of the students in our district sample learn math with teachers in their first two years of experience, for example. Teaching is considered a profession where much learning occurs "on the job," which corresponds to a steep slope of improvement in their first few years of experience (Clotfelter et al., 2007; Nye et al., 2004; Rockoff, 2004). Many interventions target professional development and support in particular to their novice teachers in an effort to help these teachers gain skills more quickly (Ingersoll & Strong, 2011). Thus, a focus on novice teachers may be of particular interest to policy-makers. In addition, our analyses of novice teachers are motivated by a theory that these teachers should be more responsive to training and experiences working with ELLs than more experienced teachers who already have more established practices and thus may not feel a need or desire to change their practices with ELLs.

Investigating differential teacher effectiveness with ELLs. Based on the prior literature, we hypothesize that certain characteristics associated with more effective teachers for students in general - such as general knowledge and teaching experience – are associated to a similar degree with effectiveness for ELLs. Second, we hypothesize that specific instructional skills and strategies that teachers can learn from training or practice may support differential effectiveness with ELLs, relative to non-ELLs, in their math instruction. We examine these hypotheses by considering the following research questions:

1. Do teacher characteristics that predict higher math achievement gains for students in general – including (a) teachers' own test performance on the LAST or (b) generic teaching experience – predict similar math achievement gains for ELLs as for non-ELLs?
2. Do specific teacher characteristics and experiences that may support learning to teach ELLs – including (a) experiences teaching ELLs in the classroom, (b) reported pre-service or in-service training during a teacher's first year specific to ELL instruction, (c) reported readiness to teach

ELLs following pre-service training, or (d) ELL-specific teaching certifications –predict differential math achievement gains for ELLs compared to non-ELLs?

Our analyses indicate that individual math teachers with specific characteristics and experiences are especially effective with their ELL students. While generic years of teaching experience predict similar-sized gains for ELLs and non-ELLs, specific experience teaching ELLs predicts differential improvement in novice teacher effectiveness with ELLs. We also find that both in-service and pre-service teacher training focused on ELL-specific instructional strategies are associated with differential ELL achievement gains. Thus, this paper addresses an important gap in the literature by presenting evidence that teachers can improve their relative effectiveness with ELLs, and that certain teacher training experience may be helpful for identifying teachers who are relatively more effective with ELLs. These findings have implications for teacher selection, assignment, and training.

Data

Context and overview of the study sample. The data for this study come from the New York City (NYC) public school system, which includes a large and diverse population of ELLs. We utilize rich administrative information about teachers and students in NYC, as well as unique survey data about teachers' preparation experiences, to predict differences in ELLs' math achievement gains in grades four through eight.[1] The New York City Department of Education (NYCDOE) provided us with student and teacher demographic data files and a student exam data file for each school year from 2004-2005 through 2007-2008 (we will refer to these school years by the spring year: e.g. 2007-2008 is SY 2008). We also match NYC teachers to employee data from New York State Education Department (NYSED) databases. Finally, we leverage a survey of all first-year NYC teachers, conducted in the spring of 2005, which asked teachers detailed questions about their pre-service preparation experiences and their in-service training in their first year of teaching.

Administrative data on students, teachers, and schools. Our administrative data include a range of demographic characteristics of students, teachers, and students' classroom peers. In addition, the combination of NYC and NYCDOE databases yields data on teachers' certification status, their performance on licensing exams, and their experience level within the NYC school system. Our primary student outcome measures consist of annual student achievement scores in math, based on exams given in third through eighth grades to most NYC students. For each year, the data include scores for approximately 65,000 to 80,000 NYC students in each grade. Using these data, we construct a set of records with a student's current math exam score standardized within each grade level and year, as well as his or her lagged exam score, and match these student records to their primary math teachers in each year.

Identifying ELLs. In NYC, the vast majority (98 percent) of ELLs are initially classified as ELLs based on a home survey that determines whether English is the primary language spoken at home, followed by a Language Assessment Battery (LAB) exam administered to students whose home language was not English, in order to determine their level of English proficiency. Our data include an indicator for ELL status in each year in which a student was present in the data set. [2] The district was unable to provide the scores that students received on their LAB exams.

Survey of first-year teachers. We use data from a survey of all first year NYC teachers in the spring of 2005. This survey was originally conducted to study pathways into teaching for New York City teachers. The survey asked detailed questions about teacher preparation experiences, in-service training in their first year of teaching, teaching practices, and preferences.[3] Of particular interest for this analysis, the survey included questions related to the quantity and efficacy of teachers' ELL-specific pre-service and in-service training. The overall response rate for this survey was 71.5 percent, representing 4303 teachers across all grades and subjects. Of these, 702 were math teachers present in our analytical sample.

Descriptive information about students and teachers. We describe the characteristics and distribution of ELLs across the district in Tables 1 and 2. The percentage of students in our sample who are ELLs ranges from 11.7 percent in 2006 to 13 percent in 2008. In line with national trends in ELL performance, district-wide ELL math achievement is substantially lower than that of non-ELLs in NYC. The bulk of ELLs (69 percent) are classified as Hispanic, while the second largest subgroup is those of Asian descent (18 percent). In comparison, 35 percent of non-ELLs are Hispanic and 14 percent are Asian. ELLs are more likely to receive free or reduced price lunch (75 percent) than non-ELLs (66 percent).

The majority of ELLs in NYC (80 percent) attend math classrooms composed of both ELL and non-ELL peer students. Thus, while this study focuses exclusively on teachers of both ELLs and non-ELLs, those mixed classrooms reflect the experiences of most ELLs in New York in grades 4-8. However, as shown in Table 3, ELLs are not distributed evenly across classrooms. In the median or 50th percentile classroom, only 3 percent of students are ELLs, compared to a mean of 12 percent across all classrooms. The classrooms in the top five percent in terms of proportion of ELLs in the classroom serve a large proportion of ELLs: forty-three percent of all ELLs attend classrooms above the 95th percentile, meaning their classmates are predominantly (94 percent or more) ELLs. [4] Classrooms in which most students are ELLs also tend to serve larger proportions of economically disadvantaged and Hispanic students than other classrooms. In classrooms at the 95th percentile, 71 percent of students are eligible for subsidized lunch and 77 percent are Hispanic.

While we do not find substantial differences between ELLs and non-ELLs in terms of assignment to novice teachers, we do see some assignment differences according to teachers' test performance and ELL-specific training. The first three columns of Table 4 show that, on average in the district, 23.7 percent of non-ELLs in grades 4-8 were taught by a first or second year teacher, compared to 23.4 percent for ELLs. Liberal Arts and Sciences Test (LAST) scores, New York's general knowledge certification exam, were somewhat lower for the teachers of ELLs, 239 on average, compared to 246 for non-ELLs, and initial failure rate was notably higher for teachers of ELLs (23.3 percent compared to 15.6 percent for non-ELLs). ELLs are more likely to learn math with a teacher who is certified as an ESL or bilingual education instructor, with 7.5 percent of ELLs receiving math instruction from an ESL-certified teacher, and 21.0 percent from a bilingual education certified teacher, in comparison to 1.0 and 1.8 percent, respectively, for non-ELLs. [5] In addition, we see that ELLs are much more likely to be taught by Hispanic teachers (32.4 percent compared to 9.3 percent for non-ELLs) and less likely to be taught by white (45.7 versus 59.0 percent) or black teachers (13.2 versus 24.2 percent).

Unsurprisingly, given that we would expect shifts in the teacher workforce over time, the sample of first-year teachers surveyed in 2005 differs somewhat from the district-wide sample. Survey respondents had higher scores and lower failure rates on LAST, were more likely to be white or list their race as “other” and were less likely to be black or Hispanic compared to math teachers district-wide. In addition, ELL students were more likely to have teachers with ESL certification and less likely to have teachers with bilingual certification in the survey sample compared to the district-wide sample.

Table 5 describes the questions of interest from our survey of new teachers that addressed ELLs and provides the distribution of teacher responses for each question. In line with previous research on teacher preparation to teach ELLs (Boyd et. al., 2009), the proportion of teachers reporting substantial training to teach ELLs is low relative to the proportion likely to be supporting ELLs in their classrooms. Although we cannot identify the specific content of the professional development provided to teachers, it is notable that this survey was administered during a period when there was much more ELL-related professional development and coaching across NYC than in previous years, particularly in elementary and middle school grades (Horwitz et al., 2009).

Methods

This section details the empirical approach that we use to answer each of our research questions about how teachers' characteristics and learning experiences relate to differential effectiveness with ELLs in math. In particular, we address two key challenges to generating accurate results. The first challenge is that unobserved characteristics may, in fact, be driving a relationship that we observe between a certain teacher characteristic and student learning. For example, if teachers who participated in a substantial amount of professional development were more motivated to improve their teaching than other teachers, then we might find a positive relationship between professional development and teacher effectiveness even if professional development itself wasn't the reason teachers were more effective. Second, teachers with a given characteristic may be assigned to teach students who have different propensities to learn, and thus this student selection and not the teacher characteristic could be driving the results. For example, administrators might choose to assign students who are struggling to engage in school to teachers who take advantage of professional development opportunities because they think these teachers will be more motivated and effective in helping these students become more engaged. If students with low engagement also have lower achievement gains on average than other students, then we would observe a negative relationship between professional development and teacher effectiveness even though, again, the professional development of teachers is not the cause of lower student learning.

In order to address these potential biases, we include in our models a wide range of theoretically appropriate controls that may relate to students' achievement gains. In addition, we evaluate effects by comparing the performance of ELLs and non-ELLs taught by the same teacher, thus minimizing bias associated with students' assignment to teachers or with unobserved teacher characteristics that influence both groups of students.

In our analyses, we measure learning by student achievement in a given year adjusted for their achievement in the prior year as well as other characteristics of the student and his/her classmates.

We are interested in whether ELLs learn more relative to their non-ELL peers when they have teachers with particular characteristics. To address this question, we interact the teacher characteristic of interest with ELL status to see whether the characteristic predicts differential effectiveness with ELLs, while controlling for several other potential confounds.

Specifically, we estimate effects based on the following equation:

$$(1) \quad A_{itjs} = \beta_0 + \beta_1 A_{ijs(t-1)} + \beta_2 X_{it} + \beta_3 C_{ijst} + \beta_4 T_{jst} * ELL_{it} + \mu_s * ELL_{it} + \tau_g + \delta_t + \omega_j + \varepsilon_{ijst}$$

Here, the standardized achievement (A) of student i in year t with teacher j in school s is a function of his or her prior achievement (A at t-1), time varying and fixed student characteristics (X), and characteristics of the classroom (C). In order to identify differential ELL effects associated with specific fixed teacher characteristics of interest (β_4), we run separate regressions for each teacher characteristic of interest in which we include an interaction of ELL student status with the teacher characteristic (T) of interest in that model, to measure how the relative performance of ELLs to their non-ELL peers varies for teachers with those characteristics (some models have controls for additional teacher characteristics interacted with ELL status, as described below). In order to control for contextual differences in school-wide ELL versus non-ELL math achievement gaps where each teacher works, we also include an interaction of school fixed effects with ELL student status ($\mu_s * ELL_{it}$).[6] We also include a fixed-effect for the grade level of the student (τ), a fixed effect for the year (δ), a fixed effect for each teacher in the sample (ω), and a random error term (ε).

When controlling for prior achievement, we include both a linear and quadratic term to represent the student's standardized prior achievement result. Also at the student level, we include observable characteristics that tend to predict differential achievement, including race and ethnicity, gender, eligibility for free or reduced-price lunch, the number of school absences in the previous year, and the number of suspensions in the previous year. At the classroom level, we control for potentially relevant peer effects by including the average of all the student characteristics already mentioned, as well as the percentage of students in the classroom who are designated ELLs, and the mean and standard deviation of students' test scores in the prior year. At the teacher level, we include observable teacher characteristics that tend to be associated with instructional efficacy and that vary within individual teachers over time, including years of teaching experience in New York state.

By including the teacher fixed-effect, we investigate how characteristics that vary across different teachers predict relative outcomes for the ELL versus non-ELLs within their classrooms. The clear advantage of this approach is that bias related to unobserved sorting of students to teachers that may be associated with the teachers' overall effectiveness is no longer of concern. The focus of our investigation is not to examine teachers' overall effectiveness with their students, which is captured by the teacher fixed effect, but rather their relative effectiveness across different students that they teach. [7]

Applying study methodology to specific questions of interest. For each teacher characteristic of interest, we test for associations with differential effectiveness with ELLs. This requires us to tailor each model to account for the specific characteristic and research question. We provide a description of model variants addressing our research questions below. The specific covariates included in each model are also detailed in Appendix Table 1.

Q1: Teachers' own test performance and teaching experience. For this question, we use the approach described above. First we look at teachers' own test performance. We include an interaction between teachers' initially failing the LAST exam and ELL student status to identify any differential effects of failing the exam on ELL versus non-ELL students. For experience, we similarly enter indicator variables for each year of experience up to nine and an indicator variable for ten or greater years, and interact each of these experience indicator variables with ELL student status. For example, the interaction term for three years of experience would indicate any differential effect on ELLs compared to non-ELLs from the teacher having three years of experience compared with being in their first year of experience.

Q2: Teachers' prior experience teaching ELLs, and ELL-specific preparation or certification. We investigate how experience teaching a substantial number of ELLs in the prior year predicts current year differential effectiveness with ELLs.[8] For simplicity, the models we present define prior experience with ELLs as the experience of teaching more than six ELLs in the prior school year. The results are not especially sensitive to this cutoff number, but the relationship between number of prior ELLs and effects does not appear to be linear so we do not use the continuous measure of the number of ELLs taught. We choose six because it is close to the estimated mean number of ELLs taught by teachers in mixed classrooms across NYC in each year, and represents a sufficient quantity of ELLs to reasonably be expected to challenge a teacher to modify his/her instruction in response. Note that the teacher fixed effect means that we are examining the effects of prior ELL teaching experiences on individual teachers' subsequent differential performance with ELL students.

Experience teaching a substantial number of ELLs might be particularly important among novice teachers. Teachers with more years of experience have likely accumulated knowledge and expertise to support their effectiveness with ELLs and thus would be less affected by teaching a substantial number of ELLs in the prior year than would novice teachers. For this reason, we conduct separate analyses that examine effects of prior experience teaching ELLs for novice teachers only (the sample of teachers in their second year of teaching who did or did not teach at least six ELLs the prior year), versus a model that includes teachers with more than two years of experience.

Teachers who are assigned more ELLs may differ from other teachers in ways that predict greater ELL-specific instructional effectiveness, and this may have led to their assignment to ELL-populated classrooms in the first place. To address this potential bias, when examining experience effects, we control for the number of ELLs taught in the current year; this control is in addition to the classroom-level control for proportion of ELLs taught in the year of instruction, which is present in all of our models.

We also examine how ELL-specific pre-service preparation and in-service PD relate to differential effectiveness with ELLs using responses on the 2005 survey of first-year teachers. We create dichotomous variables that identify those with a substantial amount of training and preparedness compared to those who reported lower levels of training and preparedness. Table 5 reports the responses on each of three survey questions about pre-service training, PD, and preparedness to teach ELLs that we used as a threshold for those dichotomous variables.

When investigating teachers' certification for teaching English as a second language (ESL) or bilingual education, we first compare effects for certified and non-certified teachers in the full population of teachers for whom certification status is available. We then examine whether this relationship differs for novice teachers, defined as those with three or fewer years of experience, to test whether these certifications might represent a temporary early advantage relative to other new teachers.

Specification checks. The within-teacher modeling approach described above may not completely eliminate bias associated with non-random and unobserved student and teacher sorting. For example, there may be sorting of students to teachers within schools that is different for ELL and non-ELLs. To better gauge this potential bias, we examine each teacher characteristic of interest for evidence of whether teachers possessing that trait are assigned ELL and non-ELLs that differ on "pre-treatment" observable characteristics. Our specification section in the appendix offers a more detailed investigation of these results, which we also refer to in the limitations section.

Results

Do Teacher Characteristics that Predict Math Achievement Gains for Students in General Predict Similar Gains for ELLs versus non-ELLs?

We find that the relationship between a teacher's test scores and student math learning is slightly weaker for ELLs than for non-ELLs but that the effects of generic teaching experience are similar.

Table 6 shows that teachers who failed the LAST general knowledge exam had higher gains with their ELLs than non-ELLs, all else equal, though the size of this differential relationship is small (.016 standard deviations).

In contrast to the differential effects of test performance, Table 6 shows additional results from the same regression indicating that each year of teacher experience yields similar math achievement gains for ELLs and non-ELLs over most of the range of teacher experience. The point estimates and standard errors are small (less than .01 standard deviations), such that not only are the point estimates insignificant but we also can rule out meaningfully different effects of experience on ELLs and non-ELLs. We do observe that very experienced teachers, with more than 10 years of teaching experience, are differentially effective (0.022) with their ELLs than their non-ELLs.

Do Teacher Experiences that Support Learning to Teach ELLs Predict Differential Effectiveness with ELLs?

Next we look at teacher characteristics and learning experiences that we expect may differentially predict instructional efficacy with ELLs. First we explore teachers' experience teaching ELLs, then, pre-service preparation, in-service professional development, and teacher certification status.

Prior experience teaching ELLs. We find that prior-year experience teaching ELLs matters for novice teachers, but not for other teachers. In Table 7, we compare teachers' own performance between the current and prior year. Column 1 is a model restricted to second-year teachers, and Column 2 is a model restricted to teachers in their third year or higher. The last row of Table 7 shows that having taught more than six ELLs in the prior school year predicts significantly higher ELL-specific student learning gains in math in a teacher's second year (.042 standard deviations) compared to their first year, controlling for the number of ELLs taught in the current year. This coefficient is substantial compared to an average ELL versus non-ELL within-teacher learning gap of -0.092 standard deviations for all teachers.

Reported preparation to teach ELLs and professional development. While teacher learning experiences relevant to ELL instruction may happen informally "on the job," significant investment and attention has been focused on formal teacher preparation to support ELLs, through both pre-service and in-service training experiences. To examine teacher training, we use responses of first-year teachers on the 2005 survey. Thus, our sample includes only teachers whose first year in the sample was 2004-05. As shown in Table 8, we find that reported training experiences that address specific instructional strategies for teaching ELLs predicts significant differential efficacy in ELL math instruction. Each column of Table 8 shows results from a separate model. As shown in the first column, the differential ELL achievement gain was 0.091 standard deviations in size among teachers who reported learning ELL-specific instructional strategies "in some depth" or "extensively" in pre-service training, relative to those who did not.

The third column of Table 8 shows that teachers who reported receiving more than nine hours of in-service professional development (PD) focused on ELL instructional strategies in the first half of their first year of teaching, when compared to teachers who did not receive such PD, had greater differential efficacy with ELLs during the same year (0.226 standard deviations in size), controlling for pre-service experience and training with ELLs. However, this effect appears to be limited to the year in which the PD was received, with smaller and non-significant effects when considering all years subsequent to teachers' receiving PD as shown in the second column.

The final column of Table 8 shows that teachers' reported readiness to teach ELLs as of the start of the school year was not associated with differential effectiveness with ELLs.

Certification status. Existing research on the effects of specialized certification to teach ELLs has yielded mixed results. We find that ESL certification for math teachers in NYC predicts differential effectiveness with ELLs, particularly among teachers with three or fewer years of teaching experience in New York. The first column of Table 9 shows that ESL certification is associated with a .053 standard deviation differential learning effect for ELLs across all teachers in the sample, while the second column shows a larger point estimate for ESL certification when the sample is limited to novice teachers (.106). In contrast, for bilingual certification, we only see a significant differential effect among novice teachers in elementary school grades (.072 standard deviations).[9]

Limitations

This study provides valuable new evidence regarding the potential learning experiences and characteristics associated with teachers' differential performance with ELL students. Our findings include both longitudinal evidence regarding ELL-specific teacher skill development over time, as well as time-invariant associations between fixed teacher characteristics and differential ELL achievement gains. By including teacher fixed effects, we account for unobserved teacher characteristics and some types of sorting of students to teachers that might bias our results. For example, if teachers with certain characteristics are assigned both ELLs and non-ELLs that are equally high in motivation, and motivation leads to higher achievement gains for both types of students, we would *not* have omitted variable bias because our treatment effects are estimated by comparing ELLs and non-ELLs within the same teacher.

However, while our methods and available data improve in important ways on previous research on this topic, we should be cautious about interpreting results of the study as causal effects because there may still be omitted variable bias. In particular, while we control for a range of observable student characteristics, there may be other student characteristics that predict learning that we do not observe, and if ELLs and non-ELLs are assigned to teachers differently based on these characteristics, then our results would be biased. For example, if some teachers are assigned more motivated ELLs than their peers, but are assigned equally motivated non-ELLs, this could bias our results. While by definition we cannot test for omitted variables, we do find some evidence of differential sorting of ELL and non-ELL students according to our teacher characteristics of interest (see Appendix Table 2). However, for the most part the differential sorting that we observe is in a direction that is more likely to lead us to understate, rather than overstate, the differential instructional effects on ELLs that we observe. In other words, to the extent that we observe differential within-teacher sorting, the ELLs taught by teachers with our characteristics of interest tend to look more disadvantaged than their non-ELL peers.

In a similar vein, it is possible that the differential teacher performance that we observe as a function of fixed teacher characteristics, such as pre-service training or certification, are merely proxies for other unobserved teacher characteristics, such as an affinity for working with ELLs. Thus, our results indicate that these characteristics may be useful for identifying differentially effective teachers, but are not necessarily the underlying cause of their effectiveness. That said, in contrast to our analysis of fixed teacher characteristics of interest, we are more confident about attributing the differential results for prior experience with ELLs to novice teacher learning, since these analyses identify improvements in differential effectiveness for teachers over time.

Another limitation of this study is that we cannot directly identify the program of math instruction for ELLs or their classes. A majority of ELLs in NYC opt into ESL programs, while others participate in transitional bilingual programs. Our analyses focus on grades 4-8, when most students who participated in bilingual programs will have transitioned out of the programs. However, we would like to control for students' past program participation, especially in order to account for student selection into programs and any relationship this student selection has with assignment to teachers with certain characteristics (for example, ELL students with more educated parents might select into bilingual programs, and these students might be more likely to take classes with bilingual-certified teachers than ELLs who opted into ESL programs, which could lead to bias in our estimation regarding differential effectiveness for bilingual-certified teachers). In addition, we cannot test hypotheses such as whether bilingual-certified teachers are differentially effective with ELLs within bilingual-education classrooms. Such questions deserve consideration in future studies that include information on program type.

Finally, as in any study seeking to isolate a teacher's effect on student test score gains, there may be concerns regarding the specification of the model. For instance, we control for student tests scores and those of their peers in the prior year, rather than include a model that attempts to measure growth for students in test scores across all years that they are present in the dataset. Our specification thus assumes that observations of the same student across teachers and years are independent, conditional on covariates. While this may seem to be a strong assumption, this type of model is common in the value-added literature. Our chosen model, while straightforward with regard to interpretation, does not fully address longitudinal patterns in student achievement growth, or possible variability in individual teacher effects over time.

Discussion and Conclusions

This study provides valuable new evidence about which characteristics and experiences of teachers are associated with differential effectiveness with ELLs. We find that novice teachers improve their relative effectiveness with ELLs when they have experience teaching ELLs in their first year. We also find that pre-service and in-service training experiences focused on ELLs, as well as ESL certification, predict differentially greater math learning gains with ELLs. Taken together, these findings constitute suggestive evidence for the notion that teachers can (and do) learn practices on the job and from training that specifically support greater effectiveness in math with their ELL students. Thus, the findings are coherent with a theory that ELLs have some specific learning needs that differ from non-ELLs. In addition, the findings of this study can inform district and school decisions about teacher assignment, since assigning ELLs to teachers with ELL-specific experiences and training has the potential to help to mitigate the math achievement gap between ELLs and non-

ELLs. They also provide suggestive evidence that increased teacher training specific to ELLs can also help mitigate the achievement gap.

We hypothesized that teachers' own test performance and generic years of experience would be associated with similar-sized gains for a teacher's ELLs compared to their non-ELLs, and the results are consistent with the hypothesis for experience but not teacher test performance. We find that the relationship between failing the teacher certification exam and student math learning is weaker for ELLs than for non-ELLs but that the effects of generic teaching experience are similar. These results should be considered in light of what we know about the relative importance of these characteristics in predicting achievement gains. General knowledge is only weakly associated with gains (in supplementary analyses, we find an association of about .01 standard deviations in student achievement gains when comparing teachers who initially failed the LAST with other teachers at their school), and thus is not a critical tool to guide assignment of students to teachers. On the other hand, teacher experience is a much stronger predictor of achievement gains (for example, students of second-year teachers have gains about .06 standard deviations higher than students of first-year teachers). Thus, just like non-ELLs, ELLs can benefit from being assigned to more experienced teachers, and policies that encourage this practice hold promise in reducing the achievement gap between ELLs and non-ELLs.

We also find evidence consistent with our second hypothesis, that certain teacher characteristics and experiences would predict differential effectiveness with ELLs. Our findings indicate that prior experience teaching ELLs and specialized training and certification hold promise as meaningful indicators of differential ELL instructional impact, particularly among novice teachers. For example, we find that assigning ELLs in mixed classrooms to novice teachers who have substantial pre-service training related to ELL instruction predicts a reduction in the resulting ELL learning gap by around 0.09 standard deviations in student achievement. This differential effect size is comparable in magnitude to the impact of assigning ELLs to a teacher whose generic "value added" quality is a full standard deviation higher than the average teacher (Kane and Staiger, 2008).

We also find that novice teachers who gain experience teaching ELLs in their first year improve their ELL-specific performance in their second year by nearly half of the standard deviation in typical teacher quality (0.042). However, for teachers with more than two years of experience teaching in NYC, having taught more than six ELLs in the prior year does not predict differential effectiveness with ELLs. It could be that second-year teachers who did not teach many ELL students in their first year lag behind their colleagues in terms of knowing how to teach ELLs, while more experienced teachers have already learned some methods for teaching ELLs or may be more skilled at adjusting to student learning needs based on their experience teaching. It is also important to remember that these are two different samples of teachers (novices versus more experienced teachers) and thus the difference in the pattern of results could also have to do with unobserved differences in characteristics of these two groups of teachers.

Similarly, while receiving a substantial amount of professional development focused on ELLs is associated with differential effectiveness with ELLs in the same year as the PD, the differential effectiveness is not sustained in subsequent years. The reason for this fadeout is not clear. We would not expect teachers to abandon practices they used in their first year that benefited ELLs. One possibility is that the PD itself coincided with external instructional supports that differentially targeted ELLs. Alternately, the PD may have encouraged teachers to temporarily shift their instructional attention towards their ELL students, rather than non-ELLs.

Overall, these findings suggest that increasing ELL math achievement may require not only a focus on enlisting or training generically 'better' teachers for ELLs, but also greater attention to those instructional skills and characteristics most relevant to ELL instruction. In particular, the improvement in ELL effectiveness among teachers who gain experience teaching ELLs provides credible evidence that a distinct skill-set is valuable for teaching ELLs and that these skills can be learned through practice. In the cases of specialized certification or pre-service and in-service training, we cannot definitively distinguish between those preparation experiences that may have increased teachers' instructional expertise with regard to ELL instruction from those that may simply help to sort teachers with differential ability, motivation, or programmatic supports to teach ELLs. That said, in line with our findings about on-the-job experience, it is certainly plausible that these types of ELL-specific training experiences also directly support teacher learning in this vein.

This is the second paper that we know of to find evidence of an association between teacher characteristics and differential effectiveness with ELLs. Loeb et al. (2014) used data from Miami to find that speaking Spanish and bilingual teaching certification were associated with a teacher's differential effectiveness with ELLs in elementary school, though not in middle or high school. Our findings regarding certification are relatively similar in that ESL certification predicts differential effectiveness and bilingual certification predicts differential effectiveness among novice elementary school teachers,[10] though this effect size of .07 is smaller than the bilingual elementary effect observed in Miami (.18).

While our results suggest that ELL-specific instructional practice and training may be worthwhile investments, more research is needed to better understand which specific skills are most relevant for supporting academic achievement among ELLs. A significant body of theory recommends elements of effective ELL instruction, but little evidence exists to test these assertions. Which important teacher practices are developed through experience teaching ELLs? What, if any, key skills can be reliably developed through training to improve instructional efficacy with ELLs? The lack of evidence about what specific skills teachers should learn in order to be effective with ELLs could explain some of the mixed evidence we find for training and certification (e.g., in-service PD is not associated with differential effectiveness beyond the same year of training; bilingual certification is not associated with differential effectiveness except among novice elementary school teachers).

Training and certification programs could be improved if they had a stronger research base describing the particular skills that help teachers to be more effective with ELLs.

Research that attends to the learning gains of ELLs in particular and that examines specific instructional interventions over time to directly assess improvement in teacher effectiveness with ELLs could inform instructional decisions and reduce the disparity in achievement between ELLs and other students. With the adoption of the Common Core State Standards, math classes may have higher linguistic demands on students, since students are expected to demonstrate higher-order problem solving skills and argumentation. As a result, examining how to select and train teachers who can adapt instruction to fit the needs of ELLs takes on even greater importance. This study helps to lay the groundwork for additional exploration of these topics, and our results indicate that such investigation may yield valuable insights for supporting ELL student achievement.

Notes

1. We focus on math outcomes in this study in part due to limitations in the available ELA data. ELL-designated students were not consistently tested in ELA in NYC until 2007. In addition, ELA exams were administered at mid-year during this period, resulting in reduced precision of teacher effect estimates. In preliminary analyses we found that ELA teacher results were directionally similar to results for math teachers, but markedly attenuated in size.

2. As a specification check, we also considered a more expansive definition of ELL classification that identified students as ELLs if they had been classified as ELLs in either the current or the prior school year. Results were similar, though attenuated, when using this alternative ELL classification.

3. For detailed information about the survey data collection, questionnaire, and related publications, see http://cepa.stanford.edu/tpr/overview and Boyd et al. (2009).

4. One limitation of this study is that we cannot directly identify the program of math instruction for ELLs or their classes. A majority of ELLs in NYC opt into ESL programs, while others participate in transitional bilingual programs. However, our analyses focus on grades 4-8, when most students who participated in bilingual programs will have transitioned out of the programs.

5. In NYC, ESL certification applies to all grades and can be a teacher's sole certification, while bilingual certification is a certification extension, meaning that all teachers must have an elementary or secondary certification in addition to bilingual certification.

6. As noted in Appendix Table 1 which lists all covariates, we do not include a school-by-ELL status fixed effect in the models examining differential effects of experience with ELLs (Table 7), since that treatment variable varies within teachers over time (who typically remain in the same school).

7. In preliminary work, we found a very similar pattern of results with respect to our research questions when utilizing models that controlled for school fixed effects in lieu of teacher fixed effects. While less conservative with respect to potential bias, these models provided a measure of

the absolute effect sizes associated with each teacher characteristic. Results from this specification are available upon request.

8. In preliminary analyses, we investigated but did not find evidence of differential effectiveness with ELLs associated with teachers' cumulative exposure to ELLs over multiple prior years.

9. Some caution is warranted in interpreting this finding for bilingual certification. As detailed in the Appendix , our specification checks identify the possibility of substantial conservative bias in differential assignment of ELLs and non-ELLs to bilingual certified classrooms. This may serve to minimize any true differential effectiveness of these teachers.

10. Bilingual-certified teachers in elementary school may engage in practices that affect differential ELL math achievement to a greater extent than bilingual-certified teachers in middle school, which likely has to do with official bilingual programming being more common in elementary than in middle school.

References:

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometric issues. *Educational Assessment, 8*(3), 231-257.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., Gersten, R., Haymond, K., Kieffer, M. J., Linan-Thompson, S., & Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE 2014-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from the NCEE website: http://ies.ed.gov/ncee/wwc/publications_reviews.aspx.

Betts, J., Zau, A., & Rice, L. (2003). *Determinants of Student Achievement: New Evidence from San Diego.* San Francisco, CA: Public Policy Institute of California.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The Narrowing Gap in New York City Teacher Qualifications and its Implications for Student Achievement in High-Poverty Schools. *Journal of Policy Analysis and Management, 27*(4), 793-818.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher Preparation and Student Achievement. *Education Evaluation and Policy Analysis, 31*(4), 416-440.

Brown, B. A., Ryoo, K., & Rodriguez, J. (2010). Pathway towards fluency: using "disaggregate instruction"to promote science literacy. *International Journal of Science Education*, *32*(11), 1465–1493.

Carlo, M. S., August, D., Mclaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., … White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, *39*(2), 188–215.

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2007). *How and why do teacher credentials matter for student achievement?* (Working Paper No. 2) Retrieved from National Center for Analysis of

Longitudinal Data in Education Research website: http://www.caldercenter.org/PDF/1001058_Teacher_Credentials.pdf.

Echevarria, J., Vogt, M. E., & Short, D. (2010). *Making content comprehensible for secondary English learners: The SIOP Model.* Boston: Allyn and Bacon.

Faltis, C., Arias, M. B., & Ramirez-Marin, F. (2009). *Building consensus: Critical competencies for secondary teachers of English learners.* Paper presented at the annual meeting of the Language Minority Research Institute, Riverside, CA.

Gandara, P., Rumberger, R., Maxwell-Jolly, J. & Callahan, R. (2003). English Learners in California Schools: Unequal resources, unequal outcomes. *Education Policy Analysis Archives*, *11*(36). Retrieved from http://epaa.asu.edu/ojs/article/view/264

Garrison, L., & Mora, J. K. Adapting Mathematics Instruction for English Language Learners: The Language-Concept Connections. Changing the Faces of Mathematics: Perspectives on Latinos (March 1999): 35-48.

Genesee, F., Lindholm-Leary, K., Saunders, W.M., & Christian, D. (2006). *Educating English language learners: A synthesis of research evidence*. New York: Cambridge University Press.

Goldenberg, C., & Coleman, R. (2010). *Promoting Academic Achievement Among English Learners: A Guide to the Research*. Thousand Oaks, Calif: Corwin.

Grunow, A. J. (2011). *Access to English: resources for Spanish-speaking children's academic development*. Unpublished thesis submitted to the School of Education at Stanford University. Retrieved from: http://purl.stanford.edu/mj490dq2140

Gutiérrez, R. (2002). Beyond essentialism: The complexity of language in teaching mathematics to Latina/o students. *American Educational Research Journal*, *39*(4), 1047–1088.

Horwitz, A., Uro, G., Price-Baugh, R., Simon, C., Uzzell, R., Lewis, S., & Casserly, M. (2009). *Succeeding with English Language Learners: Lessons Learned from the Great City Schools*. Washington, DC: Council of the Great City Schools.

Ingersoll, R. M., & Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of Educational Research, 81*(2), 201-233.

Janzen, J. (2008). Teaching English language learners in the content areas. *Review of Educational Research, 78*(4), 1010–1038. doi:10.3102/0034654308325580.

Kane, T. J., & Staiger, D. O. (2008). *Estimating teacher impacts on student achievement: An experimental evaluation* (No. w14607). National Bureau of Economic Research.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. *Educational Evaluation and Policy Analysis, 24*(1), 37–62.

Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of Academic Vocabulary Instruction for Linguistically Diverse Adolescents Evidence From a Randomized Field Trial. *American Educational Research Journal*, 0002831214532165.

Loeb, S., Soland, J., & Fox, L. (2014). Is a Good Teacher a Good Teacher for All? Comparing Value-Added of Teachers With Their English Learners and Non-English Learners. *Educational Evaluation and Policy Analysis*.

Lucas, T., Villegas, A. M., & Freedson-Gonzalez, M. (2008). Linguistically responsive teacher education: Preparing classroom teachers to teach English language learners. *Teacher Education Quarterly, 59*(4), 361-373.

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach* (Vol. 18). John Benjamins Publishing.

Moschkovich, J. (2012). Mathematics, the Common Core, and language: Recommendations for mathematics instruction for ELs aligned with the Common Core. Understanding Language: Commissioned Papers on Language and Literacy Issues in the Common Core State Standards and Next Generation Science Standards, 17–31.

National Governors Association Center for Best Practices & Council of Chief State School Officers.

(2010). *Common Core State Standards for Mathematics.* Washington, DC: Authors.

Nye, B., Konstantopoulos, S., & Hedges, L. V., (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.

Ost, B. (2009). *How do teachers improve? The relative importance of specific and general human capital* [Electronic version]. Retrieved from Cornell University, School of Industrial and Labor Relations site: http://digitalcommons.ilr.cornell.edu/workingpapers/125/

Peske, H., & Haycock, K. (2006). *Teaching Inequality: How Poor and Minority Students are Shortchanged on Teacher Quality. A Report and Recommendations by the Education Trust.* Washington, DC: The Education Trust.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

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

Ronfeldt, M. (2012). Where should student teachers learn to teach? Effects of field placement school characteristics on teacher retention and effectiveness. *Educational Evaluation and Policy Analysis*, *34*(1), 3-26.

Schleppegrell, M. J. (2007). The linguistic challenges of mathematics learning and teaching: A research review. *Reading & Writing Quarterly*, *23*, 139–159.

U.S. Department of Education, National Center for Education Statistics (2011). *The Nation's Report Card: Mathematics 2011* (NCES 2012-458). Washington, D.C.: Author.

U.S. Department of Education, National Center for Education Statistics (2010). *The Condition of Education*. Washington, D.C.: Author.

Vaughn, S., Martinez, L. R., Linan-Thompson, S., Reutebuch, C. K., Carlson, C. D., & Francis, D. J. (2009). Enhancing social studies vocabulary and comprehension for seventh-grade English language learners: Findings from two experimental studies. *Journal of Research on Educational Effectiveness*, *2*(4), 297–324.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement. Issues & Answers. REL 2007-No. 033. *Regional Educational Laboratory Southwest (NJ1)*. Retrieved from http://eric.ed.gov/?id=ED498548

Zwiers, J. (2013). *Building academic language: Essential practices for content classrooms, grades 5-12*. John Wiley & Sons.

Tables

TABLE 1

Percentage of Students who are ELLs and Standardized Math Test Scores in New York City, by year

		Year			
	Total	2005	2006	2007	2008
Percentage of students in grades 4-8 who are ELLs	12.4	12.0	11.7	12.8	13.0
Percentage ELLs in each grade of study					
Grade 4	14.3	13.2	12.3	15.9	15.7
Grade 5	11.6	10.4	11.1	12.1	13.1
Grade 6	11.8	12.5	10.7	11.7	12.4
Grade 7	12.0	12.3	12.1	11.2	12.1
Grade 8	12.0	11.8	12.2	12.5	11.4
Standardized math test scores	-0.614	-0.649	-0.604	-0.614	-0.585

Note: ELL = English language learner.

Table 2

Race/ethnicity and Free and Reduced Price Lunch Status, by ELL status

	ELLs	Non-ELLs
White	7%	15%
Black	5%	35%
Hispanic	69%	35%
Asian	18%	14%
Other race/ethnicity	0%	1%
Free or reduced price lunch status	75%	66%

Note: ELL = English language learner. Data shown are for students in grades 4 through 8, from school years 2004-05 to 2007-08. All differences between means for ELLs compared to non-ELLs are significant at the p<.001 level.

TABLE 3

Distribution of ELLs across Math Classrooms

	% ELLs in the class	% FRPL status	% Hispanic	% Asian	% Black	% of all ELLs at or below percentile (cumulative)
Mean across all classrooms	12	58	39	15	31	n/a
% of ELLs in Class, Percentiles						
25th percentile classrooms	0	52	30	15	36	0
50th percentile classrooms	3	48	28	20	28	1
75th percentile classrooms	8	65	38	11	40	8
90th percentile classrooms	40	69	59	18	14	35
95th percentile classrooms	94	71	77	16	3	57

Note: ELL = English language learner; FRPL = Free or reduced price lunch Data shown are for students in grades 4 through 8, from school years 2004-05 to 2007-08.

TABLE 4

Characteristics of Teachers Serving ELL and Non-ELL Students in Math Classrooms, District-wide and First-year Survey Sample

	District-wide analyses			*Respondents to 2005 first-year teacher survey*		
	Mean across NYC	Mean for ELLs	Mean for non-ELLs	Mean across survey respondents	Mean for ELLs	Mean for non-ELLs
Experience:						
% Taught by Teacher in 1st or 2nd year of experience	23.7	23.4	23.7	100	100	100
General knowledge:						
Average Initial LAST Scores (Std. Dev.)	245 (29.9)	239 (33.2)	246 (29.4)	252 (27.2)	252 (27.6)	252 (27.1)
% Taught by teacher who Initially failed LAST exam	16.6	23.3	15.6	13.5	13.4	13.6
Certification						
% Taught by Teacher with ESL Certification	1.8	7.5	1.0	1.9	11.4	0.4
% Taught by Teacher with Bilingual Ed. Certification	4.1	21.0	1.8	4.1	17.4	2.1
Teacher demographics, % of students taught by:						
Female	73.7	71.1	74.1	78.2	80.3	77.9
White	57.5	45.7	59.0	67.9	62.9	68.6
Black	22.8	13.2	24.2	14.5	10.1	15.1
Hispanic	12.2	32.4	9.3	9.2	17.2	8.0
Other race	7.5	8.7	7.4	8.4	9.7	8.2

Note: LAST = Liberal Arts and Sciences Test. ESL = English as a Second Language. Data shown are for students in grades 4 through 8, from school years 2004-05 to 2007-08. Of all math teachers with either ESL or Bilingual Education certification, 6% possessed both types of certification. All differences between ELL and non-ELL means district-wide are significant at the $p<.05$ level. For the survey sample, differences are significant at the $p<.05$ level for ESL and bilingual certification, white, black, and Hispanic.

TABLE 5

Responses and Number of Respondents for First-year Teacher Survey Questions of Interest

	% Yes	Total N
1. Pre-service opportunity to learn instructional strategies for teaching ELLs: “explored in some depth” or “extensively”	14.1	702
2. Number of hours of in-service PD focused on ELL instruction by mid-year: more than 9 hours	13.9	697
3. Preparedness at the start of the year to teach ELLs: “prepared” or “very well prepared”	9.1	698

Note: ELL = English language learner. Data shown are from math teachers who completed a spring 2005 survey of NYC teachers in their first year of teaching.

TABLE 6

The Relationship Between Teachers' Experience and Teachers' Test Results and ELL versus non-ELL Relative Math Test Performance Within their Classrooms

	Relationship with Standardized Math Achievement Test Score		Relationship with Standardized Math Achievement Test Score
Teacher initially failed the LAST exam x ELLs	0.016*	6th year teaching in NYC x ELLs	-0.009
	(0.007)		(0.010)
2nd year teaching in NYC x ELLs	-0.007	7th year teaching in NYC x ELLs	-0.017
	(0.008)		(0.011)
3rd year teaching in NYC x ELLs	0.001	8th year teaching in NYC x ELLs	-0.011
	(0.009)		(0.012)
4th year teaching in NYC x ELLs	-0.012	9th year teaching in NYC x ELLs	-0.011
	(0.009)		(0.013)
5th year teaching in NYC x ELLs	-0.019~	10 or more years teaching in NYC x ELLs	0.022**
	(0.010)		(0.010)
Observations	841,928		
Number of Teachers	12,012		
R-squared	0.511		

Note: LAST = Liberal Arts and Science Test; NYC = New York City; ELL = English language learner. ~p<.1, *p < .05, **p < .01, ***p < .001. All results shown are from the same model, which also includes controls for student prior performance and demographic characteristics, comparable classroom average characteristics (including the percent of students designated as ELLs), and individual year, grade, teacher, and individual school fixed effects interacted with ELL status, as detailed in Appendix Table 1.

TABLE 7

The Relationship Between Teachers' Prior Experience Teaching ELLs and ELL versus non-ELL Relative Math Test Performance within their Classrooms

	Relationship with Standardized Math Achievement Test Score	
	2nd Year Teachers	3rd Year or Higher
ELL Student	-.092***	-.117***
	(0.008)	(0.005)
# of ELLs taught this year	-0.002***	-0.001**
	(0.000)	(0.000)
# of ELLs taught this year x ELL	0.002***	0.003***
	(0.000)	(0.000)
>6 ELLs taught in prior year	-0.016*	0.008*
	(0.008)	(0.003)
>6 ELLs taught in prior year x ELL	0.042***	-0.004
	(0.013)	(0.006)
Observations	246,512	782,005
Number of Teachers	5,356	11,126
R-Squared	0.502	0.509

Note: NYC = New York City; ELL = English language learner. ~p<.1, *p < .05, **p < .01, ***p < .001. Column 1 is a model restricted to second-year teachers, and Column 2 is a model restricted to teachers in their third year or higher. Models include controls for student prior performance and demographic characteristics, comparable classroom average characteristics (including the percent and number of students designated as ELLs), teacher experience, and individual year, teacher and grade fixed effects, as detailed in Appendix 1 Table 1.

TABLE 8

The Relationship Between First-Year Teachers' ELL-specific Training Experiences and ELL versus non-ELL Relative Math Test Performance Within their Classrooms

	Relationship with Standardized Math Achievement Test Score			
	Pre-service training	In-service training	In-service (same-year)	Reported readiness
ELL-specific pre-service training x ELL	0.091*	0.042	-0.062	
	(0.038)	(0.050)	(0.102)	
ELL-specific in-service PD x ELL		0.043		
		(0.056)		
ELL-specific PD x ELL, same year effects			0.226*	
			(0.129)	
Reported initial readiness to teach ELLs		-0.043	-0.216	0.084
		(0.072)	(0.157)	(0.056)
Observations	45,807	44,877	13,800	45,243
Number of Teachers	702	697	401	698
R-squared	0.538	0.536	0.532	0.537

Note: NYC = New York City; ELL = English language learner. ~p<.1, *p < .05, **p < .01, ***p < .001. Note: ELL = English language learner. ~p<.1, *p < .05, **p < .01, ***p < .001. Sample includes only teachers who first began teaching in NYC in SY 2005. Each column shows results from a separate regression model. Models include controls for student prior performance and demographic characteristics, comparable classroom average characteristics (including the percent of students designated as ELLs), teacher's experience, individual year, teacher, and grade fixed effects, and individual school fixed effects interacted with ELL status, as detailed in Appendix Table 1.

TABLE 9

The Relationship Between Teachers' ELL-specific Certification and ELL versus non-ELL Relative Math Test Performance Within their Classrooms

	Relationship with Standardized Math Achievement Test Score			
	All teachers	Novices (<=3 years teaching)	Novices, Elementary School	Novices, Middle School
ESL Certification x ELL	0.053**	0.106**	0.105**	0.095
	(0.016)	(0.036)	(0.039)	(0.126)
Bilingual Education Certification x ELL	-0.008	0.043	0.072*	0.019
	(0.015)	(0.026)	(0.031)	(0.060)
Observations	670,600	243,148	84,118	159,030
Number of Teachers	10,944	4,948	2,926	2,063
R-squared	0.518	0.515	0.500	0.527

Note: Novice teachers defined as individuals with 3 or fewer years of teaching experience in NYC. ES = Elementary School Grades 4 and 5, MS = Middle School Grades 6, 7, and 8; ESL = English as a Second Language; ELL = English language learner. ~p<.1, *p < .05, **p < .01, ***p < .001. Column 1 includes the full teacher sample; Column 2 includes only novice teachers (those with three or fewer years of teaching in NY); Column 3 includes novice elementary school teachers only; Column 4 includes novice middle school teachers only. Models include controls for student prior performance and demographic characteristics, comparable classroom average characteristics (including the percent of students designated as ELLs), teacher's experience, individual year, teacher, and grade fixed effects, and individual school fixed effects interacted with ELL status, as detailed in Appendix Table 1.

Appendix

Appendix TABLE 1

Summary of Control Covariates Included in Each Model

	Experience with ELLs	All other teacher characteristics	Specification Checks, Predicting Student Assignment
Student			
Prior-year math (z-score)	x	x	
Prior-year math score, squared	x	x	
Gender	x	x	
Free Lunch	x	x	
Prior-year absences	x	x	
Prior-year suspensions	x	x	
Race (black, Hispanic, Asian)	x	x	
Home language English (i.e., non-ELLs)	x	x	
Classroom			
Avg of prior-year math scores	x	x	
SD of Prior-year math scores	x	x	
Race and ELL proportions	x	x	
Prior year-absences	x	x	
Prior-year suspensions	x	x	
Free Lunch proportion	x	x	
Teacher			
Teacher IDs	x	x	x
Years of experience in NYC	x	x	
Current # of ELLs taught	x		
Cohort			
Year effects	x	x	x
Grade effects	x	x	x
Interactions			
School IDs by ELL status		x	x

Note: ELL = English language learner.

Specification Checks

In order to better assess the robustness of our findings in light of potential sorting biases, we investigate non-random sorting of ELLs and non-ELLs to teachers in ways that are associated with our teacher characteristics of interest. Given our lack of experimental design, we do not expect to see a randomized assignment of students to our “treatment” teachers. To account for non-random sorting, we explicitly control for a range of observable controls associated with student achievement, including prior achievement. However, by identifying whether observable student characteristics that we know to be associated with student ability are different between “treatment” and comparison teachers at the time of assignment, we can spotlight instances where our methodology is less likely to have eliminated bias and in which we are more reliant on our observable controls.

Our specifications eliminate student assignment bias that occurs similarly for both ELL and non-ELLs of treatment teachers. However, they do not eliminate bias in cases where the ELL and non-ELLs assigned to these teachers differ systematically in terms of their propensity to learn. Evidence of consistent assignment of higher-ability ELLs or lower-ability non-ELLs to teachers with our characteristics of interest would highlight a greater potential for unobserved biases that might invalidate our findings. In order to investigate assignment of students by teacher characteristics, we fit models predicting each teacher characteristic of interest (e.g. ESL certification, pre-service training, etc.) as a function of ability-related student characteristics that were determined prior to assignment to those teachers (i.e. prior-year test scores, free/reduced price lunch status, and prior-year absences).

In a number of specification checks, we do see evidence of differential assignment of ELL and non-ELL students to teachers according to our teacher characteristics of interest. In particular, teachers with prior experience with ELLs, ELL-specific PD, and bilingual certification all appear to be differentially assigned ELLs (i.e. rather than non-ELLs) with lower initial ability levels. If anything, this suggests that these teachers’ ELLs may have less propensity for achievement gains as a function of any unobserved confounders. If this is the case, then our estimated effect sizes for these teachers' instructional effects on ELLs achievement gains may be overly conservative. However, in the case of ESL certified novice teachers, we do see some evidence of differential assignment of non-ELLs with lower initial achievement and free/reduced lunch rate gaps. In this case, some caution is warranted in interpreting our within-teacher effect sizes, as we rely more explicitly on our observable controls to account for differences in initial student academic ability, and any lingering bias might inflate our estimated effect sizes.

Appendix TABLE 2

A Series of Specification Checks for Various Teacher Fixed Effects Models: Predicting ELL and non-ELL Pre-treatment Student Assignment

	ELL Versus non-ELL Relative Assignment		
	Prior-year test scores	FRPL status	Prior-year absences
Experience with ELLs in prior year, novices – (1)	-0.078***	0.022***	0.296
	(0.015)	(0.006)	(0.223)
ELL-specific pre-service training – (2)	0.025	-0.029	0.003
	(0.051)	(0.026)	(0.738)
ELL-specific in-service PD– (3)	-0.136*	-0.014	0.072
	(0.064)	(0.033)	(0.926)
ELL-specific PD, same year effects – (4)	-0.305*	-0.031	1.790
	(0.128)	(0.063)	(1.894)
Reported initial readiness to teach ELLs – (5)	-0.026	-0.027	0.246
	(0.075)	(0.038)	(1.086)
ESL certification, novice teachers – (6)	0.101*	-0.48*	0.369
	(0.049)	(0.024)	(0.753)
Bilingual Education certification, novice teachers – (7)	-0.102**	0,030~	1.504**
	(0.036)	(0,018)	(0.544)

Note: ELL = English language learner. ESL = English as a Second Language. *p < .05, **p < .01, ***p < .001.