ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-94

English Learner Labeling: How English Learner Status Shapes Teacher Perceptions of Student Skills & the Moderating Role of Bilingual Instructional Settings

Ilana Umansky
University of Oregon

Hanna Dumont
Leibniz Institute for Research and Information in Education

Prior research has shown that EL classification is consequential for students, however, less is known about how EL classification impacts students' outcomes. In this study, we examine one hypothesized mechanism: teacher perceptions. Using nationally-representative data (ECLS-K:2011), we use coarsened exact matching to estimate the effect of EL status on teachers' perceptions of students' skills in language arts, math, science, and social studies in kindergarten through second grade. We further explore whether that impact is moderated by instructional setting (bilingual versus English immersion). We find evidence that EL classification results in lower teacher perceptions across content areas and grade levels. This impact is, however, moderated by bilingual environments. This study adds to research on teacher perceptions and the effects of EL classification.

VERSION: June 2019

Suggested citation: Umansky, Ilana, and Hanna Dumont. (2019). English Learner Labeling: How English Learner Status Shapes Teacher Perceptions of Student Skills & the Moderating Role of Bilingual Instructional Settings. (EdWorkingPaper: 19-94). Retrieved from Annenberg Institute at Brown University: http://www.edworkingpapers.com/ai19-94

English Learner Labeling: How English Learner Status Shapes Teacher Perceptions of Student Skills & the Moderating Role of Bilingual Instructional Settings

Corresponding author:
Ilana Umansky, University of Oregon
ilanau@uoregon.edu
(541) 346-5244
102Q Lokey Education Building
University of Oregon
Eugene, OR 94703

Hanna Dumont, Leibniz Institute for Research and Information in Education
dumont@dipf.de
1-49-30-293360–64
Warschauer Strasse 34-38
10243 Berlin, Germany

The authors extend thanks to the following individuals: Richard Murnane, John Willett, Doug Ready, Joe Cimpian, David Liebowitz, & Lauren Lanahan. This work was supported by a grant from the Jacobs Foundation.

Abstract

Prior research has shown that EL classification is consequential for students, however, less is known about how EL classification impacts students' outcomes. In this study, we examine one hypothesized mechanism: teacher perceptions. Using nationally-representative data (ECLS-K:2011), we use coarsened exact matching to estimate the effect of EL status on teachers' perceptions of students' skills in language arts, math, science, and social studies in kindergarten through second grade. We further explore whether that impact is moderated by instructional setting (bilingual versus English immersion). We find evidence that EL classification results in lower teacher perceptions across content areas and grade levels. This impact is, however, moderated by bilingual environments. This study adds to research on teacher perceptions and the effects of EL classification.

Keywords: Teacher perceptions, English learners, Coarsened exact matching, Bilingual education

With large achievement and attainment gaps between students classified as English learners (ELs) and those that are not, scholarly and practitioner attention has turned to consider the extent to which these gaps and outcomes may, in part, be driven by the very services and treatments apportioned to ELs. Quasi-experimental studies on the effects of initial EL classification when students first enter school on later academic achievement have come to varied conclusions: Some show positive effects (Shin, 2018), while others show negative ones (Author, 2016). Likewise, studies measuring the effects of remaining an EL rather than exiting EL status while advancing in grade level have demonstrated a range of effects on achievement, course placement, behavioral outcomes, graduation, and post-secondary enrollment. These include neutral effects (Reyes & Hwang, 2019; Robinson, 2011), mixed effects (Cimpian, Thompson, & Makowski, 2017; Robinson-Cimpian & Thompson, 2016), and negative effects (Carlson & Knowles, 2016). Such studies illustrate that although educational ramifications may be varied, EL classification has tangible effects on students' experiences and opportunities in school, and as such, is consequential for students in both the short and the long term.

In order to maximize the beneficial effects of EL classification and subsequent services, and minimize harmful ones, it is necessary to understand the mechanisms that drive the educational effects of EL classification. Mechanisms associated with EL classification that may result in positive educational outcomes include access to direct instruction in the English language (Baker et al., 2014), instruction in students' home languages (Steele et al., 2017), and specially trained teachers (Master, Loeb, Whitney, & Wyckoff, 2016). Mechanisms associated with EL classification that may lead to damaging educational outcomes include linguistic isolation (Gifford & Valdés, 2006), tracking into low level classes (Estrada, 2014; Kanno &

Kangas, 2014), and placement into classes with less experienced teachers (Gándara, Rumberger, Maxwell-Jolly, & Callahan, 2003).

Drawing on labeling theory (Link & Phelan, 2013), scholars have highlighted how English learner classification is a deficit-oriented classification -- it identifies students by their lack of English proficiency (Gutiérrez & Orellana, 2006; Wiley & Lukes, 1996) -- which may trigger treatments that harm rather than benefit students (Flores, Kleyn, & Menken, 2015; Martínez, 2018). For example, an important, although infrequently examined, potential mechanism of negative EL classification effects relates to teacher perceptions and expectations. Ethnographic research has identified how some teachers of ELs may hold downwardly biased academic perceptions of their EL students and may interpret students' lack of English proficiency as a lack of academic skill or potential (Katz, 1999; Olsen, 1997; Valenzuela, 1999). The large body of teacher perceptions and expectancy research from the past 50 years (see Jussim & Harber, 2005) indicates that downwardly biased expectations could negatively impact EL student outcomes. While other minoritized and/or stigmatized groups have been studied in great detail (e.g., Ferguson, 2003; Rubie-Davies, 2010) teacher perceptions and expectations of EL students have received very little attention as far as large-scale quantitative research is concerned (for an exception, see Blanchard & Muller, 2015). This study begins to fill that gap by drawing on the Early Childhood Longitudinal Study – Kindergarten Cohort of 2010–2011 (ECLS-K:2011), a nationally representative dataset that asks teachers a series of questions about their perceptions of individual student skill levels across a range of academic content areas.

In an experimental study, to test the idea that EL classification negatively effects teachers' expectations of their students, we would randomly assign students to EL and non-EL status and then measure any subsequent differences in their teachers' expectations. However, as

is generally the case in education, such an experimental design would be neither ethical nor practical. In this study we are able to take advantage of a unique policy characteristic that creates what we will argue is a pseudo-experimental design. Specifically, states and districts not only use a range of different assessments to measure English proficiency, they also set and implement different English proficiency thresholds for English learner classification. As a result, in some locales, students with a given true English proficiency level are classified as ELs while, in other locales, students with the same true English proficiency level are not classified as ELs. There is, therefore, a set of students who fall into a band of English proficiency levels who are, in effect, randomly assigned to EL or non-EL status based on their district or state of enrollment. This idiosyncrasy of EL identification policy allowed us to estimate the causal effect of EL classification on teachers' perceptions of student academic skill levels. To do so, we used coarsened exact matching analysis. Importantly, because of the dataset used, the results here are generalizable at the national level.

In addition, we examined a factor that may moderate the impact of EL classification on teacher perceptions. By law, EL-classified students must be afforded both instruction in the English language and accessible grade-appropriate core content instruction ("Lau v. Nichols", 1974). However, schools and districts have enormous flexibility in how they structure services for ELs. As one example relevant for this study, most ELs are served in English instructional programs, i.e. programs where instruction, be it science, math, or other content, is provided in English. A much smaller proportion of EL students are served in whole, or in part, in bilingual programs, where content instruction is provided in their home language.

In thinking about teachers' perceptions of their EL-classified students, it is plausible that factors, such as the type of instructional program a teacher works in, moderate the impact of EL

classification on teacher perceptions. Specifically, a large body of research has found that bilingual instruction is beneficial for EL students (Steele et al., 2017; Takanishi & Le Menestrel, 2017). While relevant theory posits that this effect is likely due to the increased comprehension and accessibility of content, it also suggests this beneficial effect may be due to an asset orientation in which bilingual classroom teachers hold more positive beliefs about their EL students in bilingual settings (Baker, 2011; Ruiz, 1984). As such, we tested whether teacher perceptions of ELs' academic skill level differed depending on whether the teacher and student were in a bilingual or an English instructional classroom.

Conceptual Framework and Literature Review

Why Teacher Perceptions Matter

Scholarship addressing and debating the impact and importance of teacher perceptions on student outcomes and experiences has a long and rich history. Beginning with a seminal work that catalyzed teacher perception and teacher expectancy research (Rosenthal & Jacobson, 1968), hundreds of correlational and experimental studies, reviews, and meta-analyses have looked at factors that influence teachers' perceptions of their students, and how teachers' perceptions can impact important student outcomes such as test scores or measures of intelligence (Dusek & Joseph, 1983; Hinnant, O'Brien, & Ghazzarian, 2009; Jussim, Eccles, & Madon, 1996; Jussim & Harber, 2005 Sorhagen, 2013). These effects on student outcomes have been explained via mechanisms including grade retention (Burkam, LoGerfo, Ready, & Lee, 2007), track placement (Oakes, 2005), within-class ability grouping (Tach & Farkas, 2006), and instructional quality and characteristics (Page, 1987). Importantly, and particularly relevant for the present study, teacher perceptions and expectations have been found to be systematically lower for historically and currently minoritized and/or stigmatized groups of students, including African-American

students, Latinx students, and low-income students (Auwarter, & Aruguete, 2008; Ferguson, 2003; Meissel, Meyer, Yao, & Rubie-Davies, 2017; McKown, C., & Weinstein, R. S., 2008; Ready & Wright, 2011; Rubie-Davies, 2010; Tenebaum & Ruck, 2007). A core question has been whether and to what extent these differential perceptions reflect differences in skill level or whether they are the result of teacher stereotypes or bias. Taken together, the results of various studies examining this question have suggested that teachers' perceptions of students' skill and knowledge levels tend to be relatively accurate (Jussim et al., 1996; Jussim & Harber, 2005; Madon et al., 1998; Meisels, Bickel, Nicholson, Xue, & Atkins-Burnett, 2001; Ready & Wright, 2011) but that teachers' accuracy is lower (and bias is higher) when they do not share their students' background characteristics (Farkas, 2003) and when students come from more highly stigmatized groups (Downey & Pribesh, 2004; McKown & Weinstein, 2008; Ready & Wright, 2011; Tach & Farkas, 2006).

Official labels or classifications assigned by the school or school system have also been shown to impact teacher perceptions. In particular, research has shown that special education labels negatively impact teachers' expectations of students (Bianco, 2005). This problem of biased and inaccurate expectations and perceptions of stigmatized, minoritized, and/or labeled groups is compounded by the fact that these same groups of students have been found to be more vulnerable to the effects of negative teacher expectancy (Ferguson, 2003; Jussim et al., 1996; Jussim & Harber, 2005; Van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010).

Teacher Perceptions of EL-Classified Students

The research on teacher perceptions of English learner classified students is nascent. These early findings suggest that teachers have different expectation patterns depending on the specific population of interest such as immigrant students, students who speak a language other

than English at home, or EL-classified students. Findings also differ with regard type of perception, such as perceptions of students' personal attributes, academic knowledge, or future prospects. For example, using a nationally representative sample, Blanchard and Muller (2015) found that teachers' perceptions vary systematically for immigrant students compared to U.S.-born students whose home language is not English. They found that teachers are more likely to perceive immigrant students as hard working compared to nonimmigrant students. At the same time, teachers tended to believe that non-immigrant students who did not speak English at home were less likely to complete college than students whose primary language was English, a finding that is also reflected in qualitative research findings (Dabach, Suárez-Orozco, Hernandez, & Brooks, 2018). In another national study, Ready and Wright (2011) found that teachers underestimated the academic skills of students who did not speak English at home, although this underestimation changed by grade level and varied by student ethnicity. Perceptions of EL-classified students, the vast majority of whom are Latinx or Asian, are likely also tied to students' race and ethnicity, with research demonstrating that teachers often hold stereotypes of Asian students as 'model minorities' while holding stereotypes of Latinx students as 'under-achieving' (Lee & Zhou, 2015, López, 2003; Ochoa, 2013).

Research that has looked specifically at EL-classified students is sparse but indicates that, on average, teachers have low or negative perceptions of EL-classified students (Katz, 1999; Valenzuela, 1999; Walker, Shafer, & Iiams, 2004). This important research, largely ethnographic and qualitative, has not, however, accounted for measures of student skill level and therefore is not able to identify teacher bias or measure a negative effect of EL classification on teacher perceptions.

Context Matters: Bilingual Classrooms as a Moderator

The context in which teachers and students find themselves is associated with both the degree of bias or accuracy in teacher perceptions and expectations and the degree to which these factors influence students' outcomes. For example, teachers in classrooms serving lower socio-economic students and those in classes with lower average achievement rates are more likely to underestimate students' skills and knowledge (Ready & Wright, 2011). Likewise, younger students, students in settings with more differentiated services, and students in moments of transition are more vulnerable to teacher perception effects (Jussim & Harber, 2005). Research has also suggested that racial congruence or dissonance also moderates teacher perception effects (Oates, 2003).

Just as the broader literature has found that context matters for teachers' perceptions, context also likely matters in teachers' perceptions of EL students. Several studies have shown that teacher perceptions of ELs vary according to teacher characteristics, including how they understand their role as teachers, their education level, their training to work with ELs, and their level of experience with ELs (Byrnes, Kiger, & Manning, 1997; Dabach, 2011; Yoon, 2008; Youngs & Youngs Jr, 2001). Yoon (2008), for example, found stronger EL student-teacher relationships in classrooms where teachers considered themselves teachers of all students rather than of mainstream students only or of a given subject area.

Research has not examined how teacher perceptions and expectations may differ according to linguistic instructional environment and specifically depending on whether the classroom environment is bilingual versus exclusively English. Yet a robust body of work has identified beneficial effects of bilingual education (August & Shanahan, 2006; Takanishi & Le Menestrel, 2017), and many have theorized that at least part of this benefit may derive from a more asset-oriented environment in bilingual classrooms, which values students' linguistic,

familial, and cultural backgrounds as an educational resource (Baker, 2011; Ruiz, 1984). These findings on the beneficial effects of bilingual education, combined with the larger research indicating that teachers' perceptions are moderated by school and classroom context, suggest that teacher perceptions may be systematically different in bilingual versus monolingual English instructional environments.

Research Questions and Hypotheses

This study sought to fill two important gaps in the literature: First it attempted to estimate the causal effect of EL classification on teacher perceptions, and, second, it looked at whether teachers' perceptions of their non-English-dominant students differed systematically in bilingual versus English instructional environments. To do so, we drew on data from a nationally-representative sample of kindergartners of who spoke a primary language other than English at home, a group of students we refer to as multilingual students (Garcia, 2009). In order to estimate the causal effect of EL classification on teacher perceptions, we exploited the fact that states and districts use both a range of different assessments to measure English proficiency and implement different English proficiency thresholds for English learner classification. Our outcomes included teacher perceptions of students' skill levels across a range of different academic content areas at the end of kindergarten, first grade, and second grade.

Our research questions and hypotheses were as follows:

1) What is the estimated impact of EL status on teachers' perceptions of students' academic skills among multilingual students? We hypothesized that teachers would perceive EL-classified multilingual students as less academically skilled compared to non-EL-classified multilingual students with the same set of measured skills and characteristics.

2) Do bilingual classrooms operate as a moderator of the impact of EL classification on teachers' perceptions of students' academic skills? We hypothesized that teachers would be less biased toward EL-classified students in bilingual instructional environments.

Method

Data and Analytic Sample

This study drew on the Early Childhood Longitudinal Study, Kindergarten Class of 2010-2011 (ECLS-K:2011) dataset, a federally-collected, nationally-representative sample of students who entered kindergarten in the 2010–2011 school year. The dataset contains information on this cohort of students through the fifth grade. For the purposes of this study we only included data from kindergarten, first grade, and second grade. Our sample of interest included students who spoke a primary language other than English at home based on either or both teacher and parent reports in kindergarten (Garrett & Hong, 2015). Throughout this paper we will refer to these students as *multilingual* students. While many of these students were in the process of acquiring English (and others may have also been in the process of developing their home language), we call them multilingual because they were operating in, and developing, more than one language. We further limited the sample to those multilingual students who attended public schools, where identification of English learner students is mandated. This subsample of ECLS-K:2011 included 3,885 students. 1,719 students, 44% of the sample, were missing one or more of our variables of interest or control variables. For our main analyses, we did not impute missing data but used listwise deletion (Pepinsky, 2018), leaving an analytic sample of 2,166 students. Descriptive statistics for the analytic sample are shown in Table 1. From the table, it is evident that there were large differences between those in the sample of multilingual students who were classified as EL and those who were not. For instance, non-EL multilingual students had, on average,

higher English proficiency and measured academic skill levels, higher family socioeconomic status, and were less likely to be in bilingual classrooms. Because of these differences, it was important to identify a counterfactual group of non-EL students with similar characteristics to the EL students, as we did in the present study.

[TABLE 1 ABOUT HERE]

Key Variables

Outcome variables. The ECLS-K:2011 data collection included a host of questions in which teachers recorded their perceptions of students' academic skills and knowledge over time. We used these teacher perception variables from the spring of kindergarten, after teachers had been working with their students for approximately one full academic year, and then again at the end of first grade and the end of second grade. We focused on four outcomes of interest. These included teachers' perceptions of students' skills and knowledge in the areas of (1) language and literacy, (2) math, (3) social studies, and (4) science. Importantly, for students acquiring English, teachers were instructed to answer these questions based on their perception of student skill, independent of language: "Please answer the questions based on your knowledge of this child's skills. If the child does not yet demonstrate skills in English but does demonstrate them in his/her native language, please answer the questions with the child's native language in mind" (National Center for Education Statistics, no date).

In kindergarten, teacher perceptions for math and language/literacy were measured via multiple items (e.g. "This child uses complex sentence structures"), each of which were answered on a 5-point Likert scale ranging from one, which represented "not yet proficient" to five, "proficient." We created an overall score for each domain by taking the average of all the questions in that domain. Reliabilities of the average scores for both domains were high (math: 8

items, α = .89; language/literacy: 9 items, α = .94). Teacher perceptions in science and social studies were measured via a single question in which teachers were asked: "Overall, how would you rate this child's academic skills in each of the following areas, compared to other children of the same grade level?"; the five-point Likert scale for this item ranged from one, for "far below average", to five, "far above average".

In first grade, math, language/literacy, and science were measured via multiple items, which were each answered on the same 5-point Likert scale as the multiple items in kindergarten. Again, we created an average score for each domain. Reliabilities for each domain average were high (math: 8 items, α = .96; language/literacy, 9 items, α = .97; science: 8 items, α = .97). Teacher perceptions in social studies were measured on a five-point Likert scale via a single question, as in kindergarten.

In second grade, teacher perceptions were assessed via one question for math, science, and social sciences and three questions for language/literacy (one each on reading, writing, and oral language skills), each of which were answered on a three-point Likert scale, which ranged from one, for below grade level, to three, for above grade level. For language/literacy, we took the average score of the three questions (α = .87).

Because the scale for the second-grade perception variables (1-3) is different from the kindergarten and first grade scales (1-5), we standardize all outcome variables in kindergarten, first, and second grade. This allows us to compare effect sizes across grade levels. It also facilitates effect size interpretation by translating unique scales into standard measures of effect size (standard deviations). We standardize all outcome variables using their mean and standard deviation within the full ECLS-K:2010 dataset.

Predictor variables of interest. The primary predictor variable of interest is EL status. EL status was derived from a single question posed to teachers in a questionnaire in the spring of kindergarten.[1] Teachers were asked about each sample participant whose primary language was not English: "Does this child participate in an instructional program designed to teach English language skills to children with limited English proficiency?" While the question did not ask directly about whether a student was classified as an EL in school, it did ask whether the student was in an EL program. Thus, this measure may not have been a completely accurate measure of EL status, as some EL-classified students may not have been, in practice, receiving EL services. However, prior data suggests that the vast majority of EL-classified students are in some form of EL program ("D.J. et. al. v. State of California," 2015). In total, 1,221 out of 2,166 multilingual students (56%) were considered EL. The remaining 945 multilingual students were students who spoke a language other than English at home but were not receiving EL services at school. For most of these students, this was presumably because their kindergarten English proficiency scores on local assessments surpassed established EL thresholds. In other cases, schools may have been failing to provide EL services to eligible students or parents may have opted out of EL supports.

We used the kindergarten measure of EL status in order to have a stable treatment group. While students take an average of five to seven years to reach English proficiency (Takanishi & Le Menestrel, 2017), some of these EL-classified students might have exited EL status by the time they reached the first or second grade. If so, this would bias our estimates downward in

[1] As stated, this key variable is measured in the spring of Kindergarten. This timing is appropriate for our analytic strategy because in kindergarten (and elementary school grades, more generally), student classification is unlikely to change midyear (unless a student moves to a different district or state). In addition, schools have up to 30 days after a student enters a school for the first time to assess his/her English proficiency level and determine EL identification (Every Student Succeeds Act, 2015). Thus, if the fall ECLS-K:2011 surveys were conducted before 30 days of school had passed, they might not contain accurate EL classification data. For these reasons, spring measures are preferable to fall measures.

those later grades, as some of the students in the treatment group would no longer be receiving the treatment (i.e., EL classification).

Matching variables. Our primary matching variables were two variables that measured oral English proficiency level and English reading skill level in the fall of kindergarten. As described below, school districts make determinations about EL status by assessing individual multilingual students' English proficiency levels (typically the four domains of speaking, listening, reading, and writing) using state or local assessments. In the ECLS-K:2011 dataset, all students, including all multilingual students, were administered two measures of English proficiency: the Preschool Language Assessment Scale (PreLAS) and the English Basic Reading Skill (EBRS) assessment. Taken together, they served as a baseline measure of students' incoming English proficiency. The first assessment, the PreLAS, was used as a screener to assess each student's oral (speaking and listening) English proficiency and determine whether he/she should be given the rest of the ECLS-K:2011 assessments in English. The PreLAS consisted of 20 questions that assessed expressive vocabulary in English from picture prompts and whether students could follow simple instructions in English. Students who scored at or above 16 were considered English proficient and given the rest of the battery of direct assessments in English (including assessments in reading, math, science, and executive functioning). Students who scored below 16 took the EBRS but no other direct assessments in English. Spanish speaking students who did not meet the PreLAS threshold were administered baseline assessments in Spanish. The PreLAS distribution is skewed to the right. Among multilingual students, 17% scored the full 20 points, and the mean score was 14.9.

The second assessment was the EBRS. It consisted of 18 literacy questions in English covering topics including print familiarity, letter recognition, rhyming words, and word

recognition. Two questions from the PreLAS were added to the EBRS final score, for a total possible score of 20 (Tourangeau et al., 2015). The EBRS was relatively normally distributed, with an overall mean score among the analytic sample of 11.3 (only 1.7% of the sample scored the full 20 points).

In addition to these direct measures of English proficiency, including reading skills, we matched students on key variables that may be associated with both EL status and teacher perceptions. These included additional baseline assessments of reading, math, and executive functioning skills; student race, gender, and socioeconomic status; and school district urbanicity.

Non-Spanish speaking multilingual students who scored below 16 on the PreLAS did not undergo the math, reading, and executive functioning baseline assessments by design (Tourangeau et al., 2015). This is because no test translations were done beyond Spanish. As a result, roughly one in five multilingual students does not have baseline assessment data beyond PreLAS and EBRS and missing data is heavily concentrated among non-Spanish speaking, low-English proficiency students. Because missing baseline assessment data was not randomly distributed across the sample, we conducted a sensitivity check in which we did not match on baseline assessments aside from the English proficiency and literacy assessments.

Control variables. In addition to these primary matching variables, we also included a host of other student, teacher, class, and school covariates as control variables in our regression model. Regarding student level covariates, we included: age, special education status, whether the student repeated kindergarten, whether the student was chronically absent, and whether the student changed teachers during kindergarten. For classroom and teacher level variables, we included: whether the kindergarten class was full or half day, the teacher's number of years of teaching experience, class size, class racial composition, proportion of ELs in the class, the

proportion of the class the teacher considered to be low readers at the beginning of kindergarten, and whether the teacher considered the class to be poorly behaved. For school-level variables, we included school size, average socio-economic status of the school, and the proportion of Black and Latinx students in the school. All control variables are from students' kindergarten year.

Moderator variable. Our second research question explored the moderating variable of bilingual program enrollment. As stated above, EL-classified students are most typically served in programs where English is the language of instruction, while a smaller proportion are served in programs that incorporate home language instruction. To identify students in bilingual programs we drew on questions asked of teachers in the spring of kindergarten. Specifically, teachers were asked the following question with regard to academic instruction in reading/literacy and math: "How often is a non-English language used by teachers, aides, or other adults …?" The options given were (a) never, (b) less than half the time, (c) about half the time, (d) more than half the time, and (e) all the time. Using these questions, we created a dichotomous variable indicating that the teacher or another adult in the classroom used a language other than English in math *or* in reading/literacy for "about half the time" or more. We used this definition because a bilingual instructional model should devote a considerable amount of instructional time in core content areas to instruction in the home language (Baker, 2011). Using this definition, we identified that 14% of multilingual students were participating in a bilingual program (see Table 1). We also created alternative bilingual program indicator variables for sensitivity checks including (1) at least some non-English instruction and (2) at least half non-English instruction in *both* literacy and math.

Importantly, our indicator variable for bilingual instruction did not differentiate between different types of bilingual instruction such as two-way dual immersion programs, transitional

bilingual programs, or maintenance bilingual programs (Baker, 2011). In addition, there were very few (N=20) students in the sample who were in a bilingual program and were *not* considered English learners as defined in this study. We discuss the methodological implications of this last point below.

Identification Strategy

Federal law requires that all public schools identify incoming students with a primary language other than English. They must then assess these students' English proficiency levels in order to determine whether students qualify for EL status (Every Student Succeeds Act [ESSA], 2015). By law, EL status identification procedures must be determined *exclusively* based on these two things: dominant/home language and English proficiency level.

However, states—and prior to Every Student Succeeds Act (2015) implementation, districts—are able to set their own thresholds on the English proficiency measure they use to determine EL status. Moreover, different states use different English proficiency assessments. In the school year just prior to ECLS-K:2011 kindergarten data collection, a study found 25 separate English proficiency assessments used across U.S. states (National Research Council, 2011). Comparing eight of those 25 tests, the study identified major differences between them, including different English proficiency standards, test item types, lengths, and content. They concluded that "we cannot simply assume that a student who scores at the intermediate or proficient level on one state's ELP [English language proficiency] test will score at the intermediate or proficient level on another" (National Research Council, 2011, p. 74).

Put differently, a student with a given "true" (unobserved) English proficiency level might be classified as an EL in one school in the ECLS-K:2011 sample, while another child with the exact same "true" English proficiency level may not be classified as an EL. A substantial

body of research has confirmed these conclusions (Abedi, 2004, 2008; Linquanti & Cook, 2015; Lopez, Pooler, & Linquanti, 2016; Ragan & Lesaux, 2006; Sireci & Faulkner-Bond, 2015; Solórzano, 2008). This variation in EL classification rules and levels amounts to exogenous variation in student classification assignment, once accounting for student English proficiency level. It is plausible to expect students with very high "true" English proficiency levels to score high on numerous assessments, exceed EL test thresholds, and therefore have a relatively low likelihood of being classified as an EL across different locales. Similarly, a student with a very low "true" English proficiency level might score below the EL threshold across multiple assessments and have a high likelihood of being classified as an EL across locales. However, for students with "true" English proficiency levels somewhere in the middle, one would expect significant variation across locales in EL or non-EL identification due to the variation across assessments and thresholds. Our empirical strategy homed in on precisely these students by drawing on a region of common support where ELs and non-ELs have similar characteristics. Specifically, we examined whether teacher perceptions of student ability were different for students classified as ELs compared to students who had the same measured English proficiency level (and other characteristics) but were not classified as ELs. Given that EL classification should in theory be determined exclusively by (1) multilingual status, and (2) measured English proficiency level, and because we had data on both elements, we were able to establish a plausible counterfactual group to our sample of EL-classified students.

This identification strategy, however, rested on three critical assumptions. First, critical to our analytic strategy was the assumption that there would be significant overlap in measured English proficiency levels between those multilingual students who were classified as ELs and those who were not. Testing this assumption, we found a large area of common support (see

Figure 1 in which, for ease of interpretation, we standardized, centered, and then averaged each student's PreLAS and EBRS scores). This large area of common support is likely due to a combination of reasons including variation in different states' and districts' thresholds for EL identification as well as the lack of alignment of different English proficiency assessments and in particular between the ECLS-K:2011 assessments and local proficiency assessments.

[FIGURE 1 ABOUT HERE]

Our second assumption is that students' EL statuses were determined according to the relevant state or local English proficiency thresholds rather than being based on other characteristics that are correlated with our outcomes but which we could not observe or control for. We were not able to directly test this assumption because we did not know where each student was located, nor did we have students' local English proficiency scores. However, research examining EL classification of incoming kindergartners has demonstrated high compliance with established policy. For example, Author (2016) found 89% compliance with policy in one large school district while Shin (2018) found nearly universal compliance in a different school district.

The third assumption was that EL and non-EL students' academic skill levels were equivalent, on average, conditional on our rich set of matching and control variables. Again, we could not test this assumption beyond the variables available in the dataset, but we limited our sample to those students for whom we had the full battery of ECLS-K:2011 assessments (English proficiency, reading, math, and executive functioning) so that we could match on these directly-assessed skill- and proficiency levels.

Analytic Strategy

We used both ordinary least squares (OLS) and coarsened exact matching (CEM) analytic strategies, each of which is described below. All analyses were conducted using Stata version 15. For both OLS and CEM analyses, we took into account the complex sampling design used for ECLS-K:2011 data collection. Specifically, we used student-level sampling weights to account for different sampling probabilities. We also clustered standard errors at the school level to account for student clustering within schools.

Ordinary least squares (OLS). Research question 1 asks about the impact of EL status on teachers' academic perceptions of their students. OLS does not provide causal estimates. Instead, we included it as a first step and as a point of comparison when presenting the CEM results. We used the following model:

(1)

$$\mathbf{PERCEP}_i = \beta_0 + \beta_1 \mathrm{EL}_i + \beta_2 \mathrm{PreLAS}_i + \beta_3 \mathrm{EBRS}_i + \beta_4 \mathbf{ACHIEVE}_i + \beta_5 \mathbf{X}_i + e_i$$

where **PERCEP** represents the set of teacher academic perception outcomes in grades kindergarten through second grade, EL is our proxy for EL status in kindergarten, measured by participation in an EL program, PreLAS and EBRS are our baseline measures of English proficiency, **ACHIEVE** is our set of baseline academic skill measures, and **X** is our wide array of student, family, teacher, class, and school covariates. The coefficient of interest is β_1, which represents the correlation of EL status with teacher academic perceptions, among emergent bilingual students, holding constant students' English proficiency level, achievement levels, and a host of other background characteristics.

Research question 2 asks about the role of bilingual education in moderating the effect of EL status on teacher perceptions. To answer this question, we used the following model:

(2)

$$\mathbf{PERCEP}_i = \beta_0 + \beta_1 \text{EL\_BIL}_i + \beta_2 \text{EL\_NOTBIL}_i + \beta_3 \text{PreLAS}_i + \beta_4 \text{EBRS}_i + \beta_5 \mathbf{ACHIEVE}_i + \beta_6 \mathbf{X}_i + e_i$$

where all variables are defined as in Equation 1. We removed the EL dummy variable and replaced it with two variables, one indicating whether the student is an EL and in a bilingual class (EL_BIL) and one for whether the student is an EL and not in a bilingual class (EL_NOTBIL). As mentioned above, there are only 20 non-EL students in bilingual classrooms in the sample, which meant we could not include an interaction term of EL and BIL. The coefficients of interest in this model are β_1 and β_2, which represent the estimated difference in teacher perceptions for ELs in bilingual classes and not in bilingual classes, respectively, compared to teacher perceptions of non-EL multilingual students (the reference category). We then ran contrast tests to test the differences between the three groups of students (non-EL, EL in bilingual, EL not in bilingual).

Coarsened exact matching (CEM). CEM, like all matching strategies, matches individuals in the treatment group (in this case multilingual students classified as ELs) with students who are similar to them but who are in the control group (multilingual students not classified as ELs). It then examines the differences in outcomes between the treated and control individuals within a given match or strata. The purpose is to reduce observed variable bias by removing from the sample and subsequent estimation any individuals who cannot be matched with individuals in the alternate group. This prevented us from comparing treated individuals to hypothetical counterfactual individuals that did not exist. Further, it limited the analyses to the area of common support, in which there were students with the same observed characteristics that fell into both the EL and non-EL categories (Murnane & Willett, 2010). Conducting this

matching enabled us to achieve a better balance between the treatment and control groups (Iacus, King, & Porro, 2012), thereby reducing observed variable bias (Murnane & Willett, 2010).

Compared to other matching strategies, such as propensity score matching, CEM is a useful matching strategy because the matching algorithm is directly determined by the researcher and therefore can be theory- and research-based. In addition, results of matching including the quality of matches and the sample size can be evaluated prior to moving on to statistical estimators of primary research questions (Iacus et al., 2012). Specifically, in CEM, variables that are considered to predict the likelihood of being in the treatment group and that are correlated with the outcomes of interest are coarsened into bins. Then individuals in the treatment group are matched to those in the control group, with weights assigned based on how many matches there are per individual. In this study, we matched on English proficiency level, baseline academic and executive functioning skill levels, gender, race/ethnicity, rural locale, and socio-economic status. In the matching algorithm race, gender and rural locale were set to be exact matches while we binned the continuous variables: English proficiency level, academic skills, and socio-economic status. Following Rosenbaum and Rubin (1984), we binned each of the continuous variables into quintiles; matching by quintiles has been shown to eliminate more than 90% of bias.

Table 2 shows the balance between the pre-matched sample and the post-matched sample on the matching variables. Matching achieved a good balance between the treatment and control groups, and the matched sample and region of common support covered 48% of the analytic sample. It is important to note that the characteristics of the matched sample, which reflects the region of common support, were different from the full sample. In particular, the matched sample was much more similar to the EL sample than the overall multilingual student sample in terms of its characteristics, with the exception that the matched sample had a somewhat higher

average English proficiency level than the overall sample (specifically on the PreLAS assessment). Other differences between the full and matched samples were that the matched sample had a higher proportion of Latinx students, a smaller proportion of female students, was less likely to be in a rural location, had lower baseline reading and math skills, and had a lower average family socioeconomic level compared to the full sample. These differences made the matched analytic sample more compelling, in that it more closely aligned with characteristics of the EL population in the U.S. (relatively low SES and standardized achievement scores, and predominantly Latinx; National Center for Education Statistics, 2018). However, it is important to note these differences, because the results of our matching analyses applied to the matched sample and similar samples, not to the original analytic sample. Also of note, while there were no statistically significant differences between the matched EL and non-EL samples, once matched, there were some remaining small and nonsignificant differences between the two groups that could have biased our results.

[TABLE 2 ABOUT HERE]

Once matched, we then performed the same OLS strategy described above, including the same full set of covariates described for our OLS strategy and robust standard errors. This is considered a 'doubly-robust' model, in that we matched on key covariates and then performed a regression analysis with those and additional covariates to control for any remaining observed variation between the two groups. The first set of matching variables were solely those we considered predictive of being in the treatment group *and* correlated with teacher perceptions, while the second set of covariates included any additional variables we believed may be related to our outcomes of interest—teacher perceptions of student skill. The only difference between the second stage of the CEM model and the OLS model described in Equation 1 above is that we

also included the CEM weights in the second stage CEM model, as determined by our matching algorithms. Following DuGoff, Schuler, and Stuart (2014), we incorporated both the CEM weights and ECLS-K:2011 sampling weights by creating a new weight for each observation equivalent to the product of the two weights.

To address the second research question, which examines the moderating role of bilingual environments, we simply replaced the EL dummy variable with the two variables described above (EL_BIL and EL_NOTBIL) in the second stage regression. As in the OLS model, we ran contrast tests to test for significant differences between the three groups of students (non-EL, EL in bilingual, EL not in bilingual).

Coarsened exact matching, like any matching technique, is vulnerable to omitted variable bias. Specifically, we can only interpret our estimates causally if treatment assignment is exogenous once observations are matched and observed variables are taken into account. Because we knew the process for treatment assignment and had variables that could proxy for that process, along with other control variables, we argue that our estimates should be considered plausible causal estimates.

Sensitivity analyses. As noted, we conducted an array of sensitivity checks. First, we ran our analyses using multiply imputed data so as to not lose individuals with missing individual-level data. We also fit models using the original data where we limited the baseline achievement matching and control variables to English proficiency variables in order minimize the loss of students with missing baseline math, reading, and executive functioning data. We fit models with slightly different samples, for example dropping Black multilingual students and those that fall into the "other" racial/ethnic category, due to small sample sizes in those groups. We fit models examining individual instead of composite outcomes, models using raw scores rather than

standardized scores, and models defining bilingual programming in alternative ways. We also fit nearest neighbor matching models and propensity score matching models.

Finally, we conducted sensitivity checks using alternative matching algorithms. We include in the Appendix both the balancing results (Table A) and research question results (Table B) from one of these alternative matching algorithms, one which allowed us to achieve reasonable balance but enabled us to keep 83% of the original analytic sample, increasing the external validity of our findings and mapping more closely onto the full multilingual ECLS-K:2011 sample. The alternative matching scenario matched on the two English proficiency measures, math and reading baseline skill measures, and student race and socioeconomic status, leaving out the rural location indicator and the baseline executive functioning skill variables. All continuous variables were matched by quintile, as above.

In all sensitivity checks, the results were very similar to those reported from the main analyses, with moderate fluctuation around magnitudes and statistical significance. Together, these sensitivity checks indicate the relative robustness of the main results presented next.

Results

Research Question 1: Estimated Impact of EL Status on Teachers' Perceptions of Students' Academic Skills

Table 3 presents OLS and CEM estimates of the relationship of EL status to teacher perceptions of students' academic skills among multilingual students. Our OLS models were meant as a first examination of this relationship among the full analytic sample, while our CEM results represent estimated causal estimates based on our matched sample. The OLS results demonstrate a negative association of EL status with teacher perceptions across all four academic content areas—language/literacy, math, social studies, and science—and across all three grades:

kindergarten, first grade, and second grade. In other words, teachers had lower academic perceptions of their EL-classified students compared to their non-EL students, after controlling for our rich set of covariates. Point estimates on the EL indicator variable declined slightly in absolute value across grade levels, from between -.15 and -.18 standard deviations in kindergarten to between -.10 and -.12 standard deviations in second grade across subject areas. The estimates were fairly similar across the four subject areas.

[TABLE 3 ABOUT HERE]

The bottom half of Table 3 presents CEM results from our matched sample. With few exceptions (language and math in kindergarten), the results were significant and negative in the CEM analyses and estimated effect sizes were considerably larger, often 1.5 or two times the magnitude of the OLS results. Unlike the OLS results, the effect sizes were larger in first and second grade than in kindergarten. In first and second grade, we estimated that EL status resulted in a .19 to .28 standard deviation drop in teacher perceptions of students' academic skill levels, across the four academic domains. In kindergarten, the results were not significant in language or math (-.13 in both cases) but were significant in social studies and science (-.21 in both cases).

Research Question 2: Moderator Role of Bilingual Classrooms

In order to examine whether the relationship between EL status and teachers' academic perceptions of students varied for bilingual classrooms versus monolingual English classrooms, Table 4 shows OLS (top panel) and CEM (bottom panel) estimates from our moderator model, where we removed the EL status indicator and replaced it with two alternative indicators, one for ELs in bilingual classes and one for ELs not in bilingual classes. Non-EL students (98% of whom are not in bilingual classes) remained the reference category. Point estimates on the two indicator variables represent the estimated difference between the relevant EL group and the

non-EL reference group. In addition, the table includes results from contrast tests that examined whether there were significant differences *between* the two EL groups (those in bilingual classes and those not in bilingual classes).

[TABLE 4 ABOUT HERE]

Building on the research question 1 results, the OLS results suggested that teacher perceptions of ELs who were not in bilingual classes were generally statistically significantly lower than their perceptions of non-ELs. However, with few exceptions, teacher perceptions of ELs who were in bilingual classes were *not* significantly different from teacher perceptions of non-ELs. The contrast tests were largely statistically significant, indicating that teacher perceptions of ELs in and not in bilingual classes were systematically different from each other (specifically, teacher perceptions were more negative when an EL student was not in a bilingual class). These results were consistent across academic domains and grade levels.

As with our first research question, the CEM results were largely consistent with the OLS results. Across academic domains and grades (except for kindergarten language and math), we found a negative association of EL classification with teacher perceptions of student academic skill level when students were not in bilingual classes. These point estimates were highly significant and slightly larger than the estimates that combined ELs in and not in bilingual classrooms. By contrast, there was no evidence of a significant association of EL status with teacher perceptions when EL students were in bilingual classes (with the exception of second grade math). Contrast tests between the two EL groups indicated that teachers had generally lower perceptions of EL students who were not in bilingual classes than they did of EL students who were in bilingual classes; however, most of these tests did not reach statistical significance (perhaps due the relatively small sample size of EL students in bilingual programs in the

matched sample). As noted earlier, the results from the CEM analyses pertained to the matched sample. In the discussion, we explore possible explanations for this finding.

Discussion

This study sought to explore the effects of EL classification on teacher perceptions of student skills and abilities. While EL classification is designed to ensure the rights of a potentially vulnerable group of students (Gándara, Moran, & Garcia, 2004), scholars have highlighted how this classification is oriented around deficits (English proficiency) rather than assets (multilingualism, etc.) (Martínez, 2018). As such, prior work has documented how EL classification can have a direct and negative effect on students' opportunities and outcomes in school (Carlson & Knowles, 2016; Cimpian et al., 2017). One theorized mechanism for this negative EL classification effect is systematic differences in teacher perceptions (Blanchard & Muller, 2015).

This study utilized a nationally representative dataset that included direct measurement of individual students' English proficiency and academic skill levels, an indicator of EL status in kindergarten, and teacher perceptions of academic skills across grades. Harnessing the variation in English proficiency thresholds used in different states and districts to determine EL status eligibility, we used coarsened exact matching (CEM) to examine teacher perceptions of students who had the same English proficiency and academic skill levels (and other student, class, program, and school characteristics) but different language classifications in school (EL and non-EL). The results suggest that, as theorized, EL status in kindergarten has a direct and negative effect on teachers' perceptions of students' academic skill levels across multiple academic domains and grade levels.

Specifically, our CEM analyses point to significant causal effects of EL status on perceived skills in social studies and science in kindergarten and across all four domains (language, math, social studies, and science) in both the first and second grades. The estimated effect sizes, while modest, grew between kindergarten and first grade and then remained relatively stable in the second grade. This finding is consistent with evidence that the negative academic effects of EL labeling grow across grade level (Author, 2016). Bringing these two sets of findings together (effects on academic outcomes and effects on teacher perceptions), one hypothesis is that teachers perceive larger gaps between their EL and non-EL students in the later grades compared to the earlier grades (Author & others, under review). Alternatively, EL-classified students may be afforded systematically different learning opportunities in kindergarten compared to their non-EL peers. If these different affordances result in different skill levels in higher grades, teachers' perceptions in those grades may reflect actual differences in skill levels that arise from different educational opportunities (Garrett & Hong, 2016). Future research is needed to confirm our results showing changes across grade level, and, if confirmed, to explore potential explanations for these patterns.

The results were fairly consistent across the four academic domains of language, math, social studies, and science. Across domains, having EL status in first and second grade resulted in a negative effect of a quarter of a standard deviation. Here again, more work is needed to confirm these findings and to explore any possible differences across academic domains that we were unable to ascertain. For example, there is research that suggests that teachers may have less biased views of their ELs' math skills compared to academic domains, due to a belief that math skills rely little on language proficiency (Hansen-Thomas & Cavagnetto, 2010; Whiteford, 2009). In a study by Hansen-Thomas and Cavagnetto's 2010, for example, 70% of surveyed

teachers reported a belief that math was EL students' easiest subject, and about a quarter of teachers, across states, explicitly stated that math is "universal," transcending language.

Given that prior work has also demonstrated that the extent and characteristics of teacher bias vary based on contextual features, we sought to examine whether teacher bias toward ELs is minimized or avoided in bilingual instructional settings. Previous research has found that these settings tend to, but do not always, have more positive and asset-based orientations of multilingual students (for important work on how bilingual environments may also perpetuate deficit orientations of EL-classified students, see Cervantes-Soon et al, 2017, Martínez-Roldán & Malavé, 2004 and Valdés, 1997). Consistent with our hypothesis, we found that, when in bilingual settings, teachers do not have systematically different perceptions of their EL students compared to their non-EL multilingual peers. These findings were consistent across both grade level and academic domain. In other words, in bilingual instructional settings, we did not find evidence that teachers' perceptions are biased downward by EL status. These results give preliminary evidence that bilingual instructional environments may counteract the negative effect of EL classification on teachers' perceptions of their students' academic skill levels. This finding contributes to a growing understanding of why bilingual education, on average, benefits EL students (August & Shanahan, 2006; Steele et al., 2015; Takanishi & Le Menestrel, 2017).

The findings from both of our research questions contribute to theory on and understandings of teacher perceptions and the experiences and opportunities of EL-classified students. With regard to the research on teacher perceptions, this study adds to existing work that finds that teachers are more likely to hold negatively biased assessments of the abilities of students who already face societal and educational discrimination and unequal opportunity. For example, prior work has found that teachers tend to be more biased against African-American

students (Ferguson, 2003), special education students (Bianco, 2005), and girls (in certain domains; Hinnant, O'Brien, & Ghazarian, 2009). Like these groups of students, EL students also face societal discrimination and unequal opportunity (Gándara & Hopkins, 2010; Lippi-Green, 1997).

Importantly, this study does not examine how negative teacher perceptions may alter EL-classified students' academic outcomes. This is an important area for future research especially because prior work shows that groups of students that face societal discrimination are particularly vulnerable to teacher perception and expectancy effects (Hinnant et al., 2009; Van den Bergh et al., 2010). Research in the field of EL education gives preliminary evidence of this vulnerability. For example, Callahan (2005) showed that track placement, often determined by teacher decisions and therefore subject to teacher perception bias, is a strong predictor of students' academic performance, stronger, in fact, than English proficiency level. This lends urgency to the need for future research that examines the effects of teacher expectations on EL-classified students' educational and self-perception outcomes.

With regard to the bilingual-setting moderator results, these results similarly contribute to existing work regarding how teacher perceptions, and more specifically, levels of bias in teacher perceptions, are moderated by contextual features such as teacher-student racial congruence and the average socio-economic status of students in the classroom (Oates, 2003; Ready & Wright, 2011). This study suggests that bilingual settings likely operate as one of these moderators of teacher perceptions. What this study cannot identify is what it is about bilingual settings that drives this moderating relationship. It is important to consider two possible explanations for our results regarding bilingual programs: first, that something about bilingual settings may drive this association, or second, that bilingual settings may proxy for some other possible moderator.

Regarding the first possibility, it is plausible that the added training and education that bilingual teachers receive in working with ELs may lead to less biased perceptions. In addition, teachers' linguistic skillsets may allow them to communicate with students and their families in fuller ways that offset bias. Regarding the second explanation, it is also plausible that individuals already predisposed to not be biased against their EL students disproportionately select into bilingual settings. For example, teachers who have an underlying value for multilingualism and diversity may select into bilingual settings. Likewise, bilingual teachers may be more likely to share their EL students' linguistic and cultural roots and this shared background may be associated with less bias. In reality, both sets of factors may be in effect, with less biased teachers selecting into bilingual settings and the training and setting itself further diminishing bias. It is because of these complicated relationships that we do not assert that bilingual settings are causally linked to more or less teacher bias. Instead we use correlational language with regard to our second research question. Future research should disentangle these possible mechanisms.

Due to the wide array of matching and control variables, and in particular, our ability to examine differences in teacher perceptions after matching on students' actual measured English proficiency levels and academic skill levels, we believe our results regarding our first research question, on the effect of EL status on teacher perceptions, are credible causal estimates. A limitation of this study, however, is that coarsened exact matching is vulnerable to omitted variable bias. Specifically, if matched students classified as ELs differ from those not classified as ELs in ways that we cannot observe or control for but that are related to teacher perceptions, then our estimates are likely biased. For example, while EL classification should theoretically be triggered by English proficiency assessment scores, it is possible that those scores reflect not

only English proficiency but other characteristics of students such as having more advanced test-taking skills. If these characteristics are also associated with higher teacher perceptions, then our estimates of the effect of EL status on teacher perceptions are likely inflated. On the other hand, we use treatment assignment in kindergarten, examining treatment effects through second grade. To the extent that some students will have exited EL status in the first and second grade our estimates of the effects of EL classification on teacher perceptions are likely downwardly biased.

Another limitation of this study is that it relies on the assumption that the ECLS-K:2011 basic English proficiency and academic skill assessments accurately measure students' English proficiency and skill levels. If these measures are invalid or if they are too coarse to meaningfully differentiate between students, then our causal inference may be even more uncertain. As such, the results found and presented here should not be considered definitive causal estimates. Instead, we hope that they prompt additional research that further explores the effect of EL status on teacher perceptions and the moderating effect of bilingual settings. Future work should examine these questions using alternative datasets and measures of English proficiency and alternative experimental and quasi-experimental methods.

Although these limitations need to be kept in mind, the results of this study have important implications for educators, education leaders, and education policy-makers. For example, they suggest that interventions that attempt to decrease teacher bias - such as implicit bias training - may help teachers better understand, acknowledge, and ideally avoid bias against EL-classified students in their schools and classrooms. These results also highlight the potential risk inherent in high-stakes decisions based on teachers' judgments of students' skills in the absence of established, unbiased, measures, policies or procedures. The results of this study also support current efforts to expand students' access to bilingual instructional settings. As future

research unpacks the mechanisms by which bilingual settings may counteract teacher bias, these mechanisms can hopefully be applied to non-bilingual settings as well, be they professional training in techniques to connect with students' families or policy initiatives to increase the share of teachers who share linguistic and cultural backgrounds with immigrant populations.

Citations

Author, 2016.

Author and others, under review.

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*(1).

Abedi, J. (2008). Classification system for English language learners: Issues and recommendations. *Educational Measurement: Issues and Practice, 27*(3), 17-31.

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language Minority Children and Youth*. Mahwaw, NJ: LEA Publishers.

Auwarter, A. E., & Aruguete, M. S. (2008). Effects of Student Gender and Socioeconomic Status on Teacher Perceptions. *The Journal of Educational Research*, *101*(4), 242–246.

Baker, C. (2011). *Foundations of bilingual education and bilingualism* (Vol. 79). Bristol, UK: Multilingual Matters.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., . . . Linan-Thompson, S. (2014). Teaching Academic Content and Literacy to English Learners in Elementary and Middle School. IES Practice Guide. NCEE 2014-4012. *What Works Clearinghouse*.

Bianco, M. (2005). The effects of disability labels on special education and general education teachers' referrals for gifted programs. *Learning Disability Quarterly, 28*(4), 285-293.

Blanchard, S., & Muller, C. (2015). Gatekeepers of the American Dream: How teachers' perceptions shape the academic outcomes of immigrant and language-minority students. *Social Science Research, 51*, 262-275.

Burkam, D. T., LoGerfo, L., Ready, D., & Lee, V. E. (2007). The differential effects of repeating kindergarten. *Journal of Education for Students Placed at Risk, 12*(2), 103-136.

Byrnes, D. A., Kiger, G., & Manning, M. L. (1997). Teachers' attitudes about language diversity. *Teaching and teacher education, 13*(6), 637-644.

Callahan, R. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal, 42*(2), 305-328.

Carlson, D., & Knowles, J. (2016). The effect of English language learner reclassification on student ACT scores, high school graduation, and postsecondary enrollment: Regression discontinuity evidence from Wisconsin. *Journal of Policy Analysis and Management, 35*(3), 559–586.

Cervantes-Soon, C. G., Dorner, L., Palmer, D., Heiman, D., Schwerdtfeger, R., & Choi, J. (2017). Combating inequalities in two-way language immersion programs: Toward critical consciousness in bilingual education spaces. *Review of Research in Education, 41*(1), 403-427.

Cimpian, J. R., Thompson, K. D., & Makowski, M. B. (2017). Evaluating English learner reclassification policy effects across districts. *American Educational Research Journal, 54*(1_suppl), 255S-278S.

Cook, H. G., & Linquanti, R. (2015). Strengthening Policies and Practices for the Initial Classification of English Learners: Insights from a National Working Session.

D.J. et. al. v. State of California, No. B260075 (Second District California Court of Appeal 2015).

Dabach, D. B. (2011). Teachers as agents of reception: An analysis of teacher preference for immigrant-origin second language learners. *The New Educator, 7*(1), 66-86.

Dabach, D. B., Suárez-Orozco, C., Hernandez, S. J., & Brooks, M. D. (2018). Future perfect?: Teachers' expectations and explanations of their Latino immigrant students' postsecondary futures. *Journal of Latinos and Education, 17*(1), 38-52.

Downey, D. B., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education, 77*(4), 267-282.

Dusek, J. B., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology*, *75*(3), 327–346.

DuGoff, E. H., Schuler, M., & Stuart, E. A. (2014). Generalizing observational study results: applying propensity score methods to complex surveys. *Health services research, 49*(1), 284-303.

Estrada, P. (2014). English learner curricular streams in four middle schools: Triage in the trenches. *The Urban Review, 46*(5), 535-573.

Every Student Succeeds Act, 114-95 C.F.R. (2015).

Farkas, G. (2003). Racial disparities and discrimination in education: What do we know, how do we know it, and what do we need to know? *Teachers College Record, 105*(6), 1119-1146.

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black-White test score gap. *Urban Education, 38*(4), 460-507.

Flores, N., Kleyn, T., & Menken, K. (2015). Looking holistically in a climate of partiality: Identities of students labeled Long-Term English Language Learners. *Journal of Language, Identity & Education, 14*(2), 113-132.

Gándara, P., & Hopkins, M. (Eds.). (2010). *Forbidden Language: English learners and restrictive language policies*. New York, NY: Teachers College Press.

Gándara, P., Moran, R. F., & Garcia, E. (2004). Legacy of Brown: Lau and language policy in the United States. *Review of research in education, 28*(1).

Gándara, P., Rumberger, R. W., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *education policy analysis archives, 11*(36), 1-54.

Garcia, O. (2009). Emergent Bilinguals and TESOL: What's in a Name? *TESOL Quarterly, 43*(2), 322-326.

Garrett, R., & Hong, G. (2016). Impacts of Grouping and Time on the Math Learning of Language Minority Kindergartners. *Educational Evaluation and Policy Analysis*, *38*(2), 222-244.

Gifford, B., & Valdés, G. (2006). The linguistic isolation of Hispanic students in California: The challenge of reintegration. *Yearbook of the National Society for the Study of Education, 105*(2), 125-154.

Gutiérrez, K. D., & Orellana, M. F. (2006). At last: The" problem" of English learners: Constructing genres of difference. *Research in the Teaching of English, 40*(4), 502-507.

Hansen-Thomas, H., & Cavagnetto, A. (2010). What do mainstream middle school teachers think about their English language learners? A tri-state case study. *Bilingual Research Journal, 33*(2), 249-266.

Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology, 101*(3), 662.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis, 20*(1), 1-24.

Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. In *Advances in experimental social psychology* (Vol. 28, pp. 281-388): Elsevier.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and social psychology review, 9*(2), 131-155.

Kanno, Y., & Kangas, S. (2014). "I'm not going to be, like, for the AP": English language learners' limited access to advanced college-preparatory courses in high school. *American Educational Research Journal, 51*(5), 848-878.

Katz, S. R. (1999). Teaching in tensions: Latino immigrant youth, their teachers, and the structures of schooling. *The Teachers College Record, 100*(4), 809-840.

Lau v. Nichols, No. 414 U.S. 563 (1974).

Lee, J., & Zhou, M. (2015). *The Asian American achievement paradox*. New York, NY: Russell Sage Foundation.

Link, B., & Phelan, J. (2013). Labeling and Stigma. In C. S. Aneshensel, J. C. Phelan, & A. Bierman (Eds.), *Handbook of the Sociology of Mental Health* (pp. 525-541). Dordrecht: Springer Netherlands.

Linquanti, R., & Cook, G. (2015). *Re-examining Reclassification: Guidance from a National Working Session on Policies and Practices for Exiting Students from English Learner Status*. Retrieved from Washington, DC:

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. New York, NY: Routledge.

Lopez, A. A., Pooler, E., & Linquanti, R. (2016). Key issues and opportunities in the initial identification and classification of English learners. *ETS Research Report Series, 2016*(1), 1-10.

López, N., 2003. *Hopeful girls, troubled boys: Race and gender disparity in urban education*. New York, NY: Psychology Press.

Madon, S., Jussim, L., Keiper, S., Eccles, J., Smith, A., & Palumbo, P. (1998). The Accuracy and Power of Sex, Social Class, and Ethnic Stereotypes: A Naturalistic Study in Person Perception. *Personality and Social Psychology Bulletin*, *24*(12), 1304–1318.

Martínez, R. A. (2018). Beyond the English Learner Label: Recognizing the Richness of Bi/Multilingual Students' Linguistic Repertoires. *The Reading Teacher, 71*(5), 515-522.

Martínez-Roldán, C. M., & Malavé, G. (2004). Language ideologies mediating literacy and identity in bilingual contexts. *Journal of Early Childhood Literacy, 4*(2).

Master, B., Loeb, S., Whitney, C., & Wyckoff, J. (2016). Different skills? Identifying differentially effective teachers of English language learners. *The Elementary School Journal, 117*(2), 261-284.

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of school psychology, 46*(3), 235-261.

Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal, 38*(1), 73-95.

Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, *65*, 48–60. https://doi.org/10.1016/j.tate.2017.02.021

Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*: Oxford University Press, USA.

National Center for Education Statistics. 2018. *The condition of education 2018.* Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018144.

National Center for Education Statistics. (no date). Early Childhood Longitudinal Study Spring 2011 Kindergarten Teacher Questionnaire (Child Level). In. Washington, DC: U.S. Department of Education.

National Research Council. (2011). *Allocating federal funds for state programs for English language learners*: National Academies Press.

Oakes, J. (2005). *Keeping track: How schools structure inequality*. Princeton, NJ: Yale University Press.

Oates, G. (2003). Teacher-student racial congruence, teacher perceptions, and test performance. *Social Science Quarterly, 84*(3), 508-525.

Ochoa, G. L. (2013). *Academic profiling: Latinos, Asian Americans, and the achievement gap*. Minneapolis, MN: University of Minnesota Press.

Olsen, L. (1997). *Made in America: Immigrant students in our public schools*. New York, NY: The New Press.

Page, R. (1987). Teachers' perceptions of students: A link between classrooms, school cultures, and the social order. *Anthropology & Education Quarterly, 18*(2), 77-99.

Pepinsky, T. B. (2018). A note on listwise deletion versus multiple imputation. *Political Analysis, 26*(4), 480-488.

Ragan, A., & Lesaux, N. (2006). Federal, state, and district level English language learner program entry and exit requirements: Effects on the education of language minority learners. *education policy analysis archives, 14*, 20.

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal, 48*(2), 335-360.

Reyes, M., & Hwang, N. (2019). Middle School Language Classification Effects on High School Achievement and Behavioral Outcomes. *Educational Policy*, 0895904818823747.

Robinson, J. P. (2011). Evaluating criteria for English learner reclassification: A causal-effects approach using a binding-score regression discontinuity design with instrumental variables. *Educational Evaluation and Policy Analysis, 33*(3), 267-292.

Robinson-Cimpian, J. P., & Thompson, K. D. (2016). The Effects of Changing Test-Based Policies for Reclassifying English Learners. *Journal of Policy Analysis and Management, 35*(2), 279-305.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association*, *79*(387), 516-524.

Rosenthal, R. (1994). Interpersonal expectancy effects: A 30-year perspective. *Current directions in psychological science, 3*(6), 176-179.

Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences, 1*(3), 377-386.

Rubie-Davies, C. M. (2010). Teacher expectations and perceptions of student attributes: Is there a relationship? *The British Journal of Educational Psychology*, *80*(Pt 1), 121–135.

Ruiz, R. (1984). Orientations in Language Planning. *NABE: The Journal for the National Association for Bilingual Education, 8*(2), 15-34.

Shin, N. (2018). The effects of initial English Language Learner classification on students' later academic outcomes. *Educational Evaluation and Policy Analysis, 40*(2), 175-195.

Sireci, S. G., & Faulkner-Bond, M. (2015). Promoting validity in the assessment of English learners. *Review of research in education, 39*(1), 215-252.

Solórzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research, 78*(2), 260-329.

Sorhagen, N. S. (2013). Early Teacher Expectations Disproportionately Affect Poor Children's High School Performance. *Journal of Educational Psychology*, *105*(2), 465–477.

Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2015). Effects of Dual-Language Immersion on Students' Academic Performance.

Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2017). Effects of dual-language immersion programs on student achievement: Evidence from lottery data. *American Educational Research Journal, 54*(1_suppl), 282S-306S.

Tach, L. M., & Farkas, G. (2006). Learning-related behaviors, cognitive skills, and ability grouping when schooling begins. *Social Science Research, 35*(4), 1048-1079.

Takanishi, R., & Le Menestrel, S. (Eds.). (2017). *Promoting the educational success of children and youth learning English: Promising futures*. Washington, DC: The National Academies Press.

Tenebaum, H. R., & Ruck, M. D. (2007). Are Teachers' Expectations Different for Racial Minority Than for European American Students? A Meta-Analysis. *Journal of Educational Psychology*, *99*(2), 253–273.

Tourangeau, K., Nord, C., Le, T., Sorongon, A., Hagedorn, M., & Daly, P. (2015). User's manual for the ECLS-K: 2011 Kindergarten data file and electronic codebook. Washington, DC: National Center for Education Statistics. In.

Valdés, G. (1997). Dual-language immersion programs: A cautionary note concerning the education of language-minority students. *Harvard Educational Review, 67*(3), 391-430.

Valenzuela, A. (1999). *Subtractive schooling: US-Mexican youth and the politics of caring*. Albany, NY: State University of New York Press.

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal, 47*(2), 497-527.

Walker, A., Shafer, J., & Iiams, M. (2004). "Not in my classroom": Teacher attitudes towards English language learners in the mainstream classroom. *NABE Journal of Research and Practice, 2*(1), 130-160.

Whiteford, T. (2009). Is Mathematics a Universal Language? *Teaching Children Mathematics, 16*(5), 276-283.

Wiley, T. G., & Lukes, M. (1996). English-Only and Standard English Ideologies in the U.S. *TESOL Quarterly, 30*(3), 511-535.

Yoon, B. (2008). Uninvited guests: The influence of teachers' roles and pedagogies on the positioning of English language learners in the regular classroom. *American Educational Research Journal, 45*(2), 495-522.

Youngs, C. S., & Youngs Jr, G. A. (2001). Predictors of mainstream teachers' attitudes toward ESL students. *TESOL Quarterly, 35*(1), 97-120.

Table 1: Descriptive statistics of analytic sample

	Full sample	Non-EL	EL
English proficiency measures			
PreLAS	15.89	17.62	14.55
EBRS	11.94	13.41	10.79
Student academic skill measures			
English reading assessment	-0.82	-0.50	-1.07
Math assessment	-0.80	-0.49	-1.03
Executive functioning assessment 1	13.49	14.17	12.97
Executive functioning assessment 2	423.91	432.90	416.95
Teacher academic perceptions			
Kinder. language/literacy	-0.19	0.06	-0.39
Kinder. math	-0.13	0.07	-0.29
Kinder. social studies	-0.07	0.14	-0.23
Kinder. science	-0.07	0.13	-0.23
1st gr. language/literacy	-0.12	0.14	-0.32
1st gr. math	-0.10	0.12	-0.27
1st gr. social studies	-0.06	0.14	-0.22
1st gr. science	-0.10	0.12	-0.27
2nd gr. language/literacy	-0.17	0.11	-0.37
2nd gr. math	-0.07	0.18	-0.25
2nd gr. social studies	-0.10	0.15	-0.27
2nd gr. science	-0.08	0.17	-0.27
Student & family characteristics			
Female	49.86%	50.90%	49.06%
Age (in months)	66.14	66.43	65.91
Latinx	63.90%	50.05%	74.61%
White	8.68%	15.24%	3.60%
Asian	21.65%	26.14%	18.18%
Other racial or ethnic group incl. Black	5.77%	8.57%	3.60%
Family SES	-0.48	-0.21	-0.69
Special education status	2.95%	3.60%	2.46%
Repeated kindergarten	5.82%	4.97%	6.47%
Chronically absent in kindergarten	11.08%	12.80%	9.75%
Changed teacher in kindergarten	4.52%	6.03%	3.36%
Classroom and teacher variables			
Full day program	81.26%	77.67%	84.03%
Teacher years of experience	13.55	13.48	13.60
Prop. of class - Latinx	51.71%	38.16%	62.20%
Prop. of class - White	10.97%	11.52%	10.53%
Prop. of class - African-American	13.13%	17.38%	9.84%
Prop. of class - other race/ethnicity	7.23%	8.58%	6.18%
Prop. of class - EL	43.74%	23.99%	59.03%
Prop. of class - low reading skills (teacher perception)	17.37%	19.12%	16.01%
Prop. of class - poor behavior (teacher perception)	9.97%	10.16%	9.83%
Class size	20.96	20.91	20.99
School characteristics			
Rural location	10.57%	13.76%	8.11%
School size (1-4)	2.82	2.76	2.87
Average school SES (standardized)	-0.31	-0.15	-0.44
Prop. of school - Black and Latinx	57.47%	48.76%	64.21%
Bilingual instruction			
Bilingual classroom	14.31%	2.12%	23.75%
N	2166	945	1221

Note. All teacher academic perception variables are standardized. Gr. = grade. EL = English learner.

Table 2: Descriptive statistics on matching variables, pre and post matching

	Pre-matched full sample			Post-matched sample		
	Non-EL	EL	t-test	Non-EL	EL	t-test
PreLAS	17.62	14.55	***	16.44	16.09	
EBRS	13.41	10.79	***	11.59	11.59	
Math assessment	-0.50	-1.07	***	-0.93	-0.97	
Reading assessment	-0.49	-1.03	***	-0.91	-0.91	
Executive function assessment 1	14.17	12.97	***	14.10	14.03	
Executive function assessment 2	432.90	416.95	***	417.72	417.71	
Female	50.90%	49.06%		44.71%	44.71%	
Latinx	50.05%	74.61%	***	80.59%	80.59%	
White	15.24%	3.60%	***	2.77%	2.77%	
Asian	26.14%	18.18%	***	14.56%	14.56%	
Other race or ethnic group	8.57%	3.60%	***	2.08%	2.08%	
Rural	13.76%	8.11%	***	3.81%	3.81%	
SES	-0.21	-0.69	***	-0.62	-0.68	
N	945	1221		458	577	
Multivariate L1 distance	.9983			0.9982		

Note. The matching algorithm included the two English proficiency measures, math and reading baseline skill measures, both baseline executive functioning skill variables, student race, an indicator for whether the school attended was in a rural location, and student socioeconomic status. All categorical and dichotomous variables are exact matches. All continuous variables are matched by quintile.

Table 3: OLS and CEM estimates of effect of EL status on teacher perceptions, among multilingual students

	Kindergarten				1st grade				2nd grade			
	Language	Math	Social Studies	Science	Language	Math	Social Studies	Science	Language	Math	Social Studies	Science
OLS												
EL	-0.177***	-0.160***	-0.165***	-0.151***	-0.137**	-0.093~	-0.107*	-0.116*	-0.119**	-0.100*	-0.105*	-0.123*
	(0.038)	(0.043)	(0.043)	(0.042)	(0.046)	(0.050)	(0.050)	(0.049)	(0.046)	(0.049)	(0.051)	(0.051)
N	2,166	2,166	2,166	2,166	1,747	1,747	1,745	1,747	1,704	1,704	1,704	1,704
R2	0.466	0.332	0.304	0.317	0.405	0.318	0.292	0.313	0.403	0.323	0.265	0.267
CEM												
EL	-0.125	-0.127	-0.212**	-0.213**	-0.283**	-0.245**	-0.251*	-0.288***	-0.185*	-0.283***	-0.207**	-0.255***
	(0.078)	(0.092)	(0.066)	(0.070)	(0.086)	(0.090)	(0.103)	(0.083)	(0.077)	(0.079)	(0.071)	(0.071)
N	1,035	1,035	1,035	1,035	827	827	825	827	803	803	803	803
R2	0.393	0.257	0.242	0.253	0.327	0.263	0.275	0.232	0.322	0.262	0.237	0.229

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1

Note. All models include English proficiency measures (PreLAS & EBRS), academic skill level measures (English reading, math, and two executive functioning assessments), student characteristics (gender, age, race, family socioeconomic status, special education identification, whether repeated kindergarten, whether chronically absent, and whether experienced a teacher change in kindergarten), program and teacher characteristics (whether full day kindergarten, kindergarten teacher's years of experience), class characteristics (racial composition, EL proportion, class size, and teacher's evaluation of class behavior and reading level), school characteristics (rural locale, school size, proportion Black and Latinx, and average socioeconomic status), and sampling weights.

Table 4: OLS and CEM results for the moderating role of bilingual classroom environment on teachers' perceptions of EL students

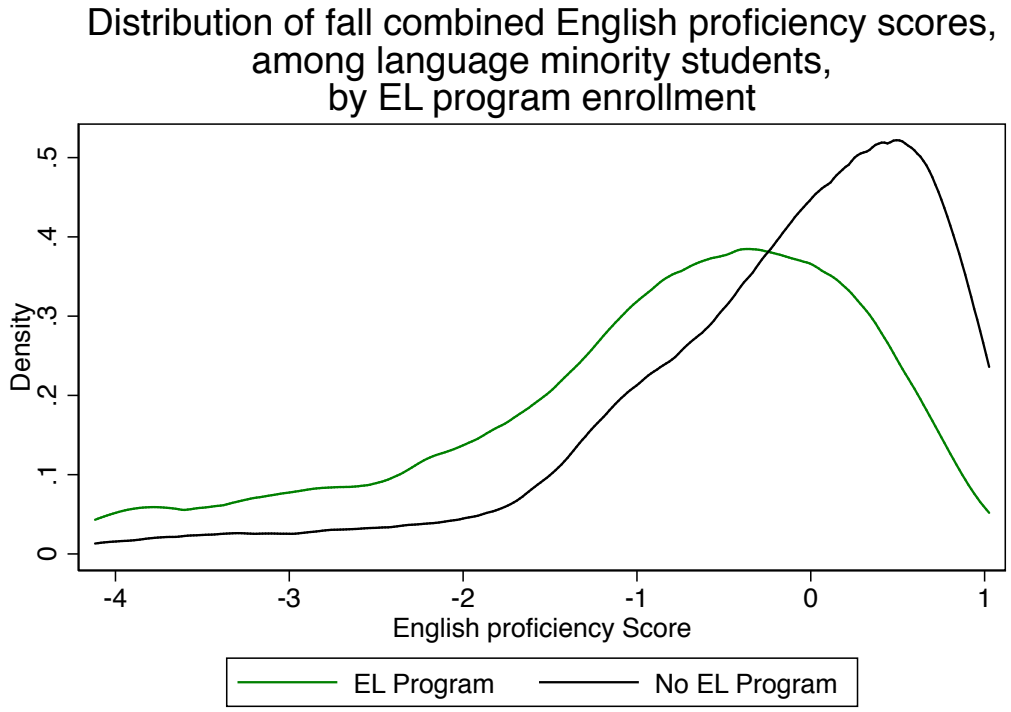
	Kindergarten				1st grade				2nd grade			
	Language	Math	Social Studies	Science	Language	Math	Social Studies	Science	Language	Math	Social Studies	Science
OLS												
EL-not bil	-0.190***	-0.162***	-0.183***	-0.166***	-0.152**	-0.106*	-0.117*	-0.141**	-0.140**	-0.100*	-0.127*	-0.141**
	(0.039)	(0.043)	(0.044)	(0.043)	(0.046)	(0.051)	(0.051)	(0.050)	(0.046)	(0.049)	(0.052)	(0.052)
EL-bil	-0.071	-0.147*	-0.021	-0.022	-0.018	0.006	-0.029	0.074	0.039	-0.105	0.058	0.015
	(0.063)	(0.070)	(0.072)	(0.070)	(0.075)	(0.082)	(0.082)	(0.081)	(0.072)	(0.078)	(0.082)	(0.081)
Contrast	-.119*	-.015	-.162*	-.145*	-.134*	-.112	-.089	-0.215**	-.179***	-.005	-0.185*	-0.156*
	(0.057)	(0.063)	(0.064)	(0.063)	(0.067)	(0.073)	(0.074)	(0.072)	(0.064)	(0.068)	(0.072)	(0.071)
N	2,166	2,166	2,166	2,166	1,747	1,747	1,745	1,747	1,704	1,704	1,704	1,704
R2	0.467	0.332	0.306	0.319	0.406	0.319	0.293	0.316	0.406	0.323	0.268	0.27
CEM												
EL-not bil	-0.118	-0.105	-0.224***	-0.227**	-0.307***	-0.286**	-0.300**	-0.315***	-0.221**	-0.282***	-0.242**	-0.295***
	(0.080)	(0.091)	(0.067)	(0.070)	(0.088)	(0.094)	(0.105)	(0.087)	(0.078)	(0.078)	(0.073)	(0.075)
EL-bil	-0.167	-0.245	-0.150	-0.136	-0.140	-0.009	0.034	-0.131	-0.017	-0.287*	-0.045	-0.069
	(0.133)	(0.168)	(0.117)	(0.121)	(0.142)	(0.136)	(0.159)	(0.123)	(0.129)	(0.134)	(0.120)	(0.113)
Contrast	0.05	.140	-.091	-.074	-.167	-.277*	-.184	-.335*	-.204~	.005	-.226*	-.197~
	(0.127)	(0.149)	(0.109)	(0.108)	(0.134)	(0.131)	(0.112)	(0.144)	(0.115)	(0.118)	(0.111)	(0.115)
N	1,035	1,035	1,035	1,035	827	827	825	827	803	803	803	803
R2	0.393	0.259	0.243	0.253	0.329	0.269	0.278	0.240	0.326	0.262	0.242	0.233

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1

Note. Bil = bilingual. All models include English proficiency measures (PreLAS & EBRS), academic skill level measures (English reading, math, and two executive functioning assessments), student characteristics (gender, age, race, family socioeconomic status, special education identification, whether repeated kindergarten, whether chronically absent, and whether experienced a teacher change in kindergarten), program and teacher characteristics (whether full day kindergarten, kindergarten teacher's years of experience), class characteristics (racial composition, EL proportion, class size, and teacher's evaluation of class behavior and reading level), and school characteristics (rural locale, school size, proportion Black and Latinx, and average socioeconomic status).

Figure 1


Distribution of fall combined English proficiency scores,
among language minority students,
by EL program enrollment
Density
0
.1
.2
.3
.4
.5
-4
-3
-2
-1
0
1
English proficiency Score
EL Program
No EL Program

Appendix

Table A: Descriptive statistics on matching variables, pre and post matching, for alternative matching algorithm with larger sample

	Pre-matched full sample			Post-matched sample		
	Non-EL	EL	t-test	Non-EL	EL	t-test
PreLAS	17.62	14.55	***	15.56	15.29	
EBRS	13.41	10.79	***	11.26	11.22	
Math assessment	-0.50	-1.07	***	-0.96	-1.00	
Reading assessment	-0.49	-1.03	***	-0.97	-0.98	
Executive function assessment 1	14.17	12.97	***	13.64	13.14	**
Executive function assessment 2	432.90	416.95	***	421.15	418.07	*
Female	50.90%	49.06%		46.25%	48.99%	
Latinx	50.05%	74.61%	***	77.53%	77.53%	
White	15.24%	3.60%	***	3.18%	3.18%	
Asian	26.14%	18.18%	***	16.49%	16.49%	
Other race or ethnic group	8.57%	3.60%	***	2.80%	2.80%	
Rural	13.76%	8.11%	***	12.05%	8.10%	**
SES	-0.21	-0.69	***	-0.67	-0.72	~
Age (in months)	66.43	65.91	***	66.37	65.93	*
N	945	1221		770	1037	
Multivariate L1 distance	0.998			0.972		

Note. This alternative matching algorithm matches on the two English proficiency measures, math and reading baseline skill measures, and student race and socioeconomic status. All categorical and dichotomous variables are exact matches. All continuous variables are matched by quintile.

Table B: Coarsened exact matching (CEM) results from alternative matching algorithm with larger analytic sample

	Kindergarten				1st grade				2nd grade			
	Language	Math	Social Studies	Science	Language	Math	Social Studies	Science	Language	Math	Social Studies	Science
Research Question 1												
EL	-0.084	-0.052	-0.102~	-0.095	-0.144*	-0.099	-0.123	-0.175*	-0.026	-0.118~	-0.097~	-0.113*
	(0.063)	(0.076)	(0.061)	(0.065)	(0.073)	(0.067)	(0.083)	(0.073)	(0.066)	(0.062)	(0.056)	(0.055)
N	1,807	1,807	1,807	1,807	1,451	1,451	1,451	1,449	1,420	1,420	1,420	1,420
R2	0.418	0.279	0.267	0.271	0.345	0.292	0.257	0.254	0.326	0.265	0.209	0.219
Research Question 2												
EL-not bil	-0.084	-0.043	-0.126*	-0.121~	-0.157*	-0.116~	-0.154~	-0.200**	-0.049	-0.115~	-0.125*	-0.149**
	(0.063)	(0.077)	(0.062)	(0.066)	(0.073)	(0.068)	(0.082)	(0.075)	(0.067)	(0.063)	(0.057)	(0.057)
EL-bil	-0.087	-0.097	0.021	0.036	-0.083	-0.016	0.031	-0.053	0.075	-0.129	0.030	0.045
	(0.112)	(0.134)	(0.095)	(0.100)	(0.115)	(0.110)	(0.145)	(0.115)	(0.114)	(0.107)	(0.099)	(0.093)
Contrast	0.003	.054	-.157~	-.147~	-.074	-.100	-.147	-.186	-.124	.014	-.193*	-.156~
	(0.104)	(0.123)	(0.083)	(0.082)	(0.099)	(0.101)	(0.099)	(0.125)	(0.103)	(0.098)	(0.088)	(0.093)
N	1,807	1,807	1,807	1,807	1,451	1,451	1,451	1,449	1,420	1,420	1,420	1,420
R2	0.418	0.279	0.269	0.273	0.346	0.292	0.260	0.255	0.328	0.265	0.211	0.223

Robust standard errors in parentheses

*** p<0.001, ** p<0.01, * p<0.05, ~ p<0.1

Note. Bil = bilingual. This alternative matching algorithm matches on the two English proficiency measures, math and reading baseline skill measures, and student race and socioeconomic status. All categorical and dichotomous variables are exact matches. All continuous variables are matched by quintile. All models include English proficiency measures (PreLAS & EBRS), academic skill level measures (English reading, math, and two executive functioning assessments), student characteristics (gender, age, race, family socioeconomic status, special education identification, whether repeated kindergarten, whether chronically absent, and whether experienced a teacher change in kindergarten), program and teacher characteristics (whether full day kindergarten, kindergarten teacher's years of experience), class characteristics (racial composition, EL proportion, class size, and teacher's evaluation of class behavior and reading level), school characteristics (rural locale, school size, proportion Black and Latinx, and average socioeconomic status), and sampling weights.