

# English Learner Labeling: How English Learner Classification in Kindergarten Shapes Teacher Perceptions of Student Skills and the Moderating Role of Bilingual Instructional Settings

Ilana M. Umansky 

University of Oregon

Hanna Dumont

DIPF | Leibniz Institute for Research and Information in Education

*Prior research has shown that English learner (EL) classification is consequential for students; however, less is known about how EL classification affects student outcomes. In this study, we examine one hypothesized mechanism: teacher perceptions. Using a national data set (Early Childhood Longitudinal Study—Kindergarten Cohort of 2010–2011 or ECLS-K:2011), we use coarsened exact matching to estimate the effect of kindergarten EL status on teachers' perceptions of students' academic skills. We further explore whether that impact is moderated by instructional setting (bilingual vs. English immersion). We find evidence that EL classification results in lower teacher perceptions. This impact is, however, moderated by bilingual environments. In bilingual classrooms, we do not find evidence that EL classification results in diminished perceptions. This study adds to research on teacher perceptions and the effects of EL classification.*

---

ILANA M. UMANSKY is an assistant professor of educational methodology, policy and leadership at the University of Oregon, 102Q Lokey Education Building, Eugene, OR 94703; e-mail: [ilanau@uoregon.edu](mailto:ilanau@uoregon.edu). Her work explores how education policy affects the educational opportunities and outcomes of immigrant, multilingual and English learner-classified students using largescale data, and longitudinal and quasi-experimental methods. She holds a PhD from Stanford University in sociology of education and is particularly interested in topics such as labeling and tracking as she focuses on how to create equitable school systems for immigrant and multilingual students.

HANNA DUMONT is a senior researcher at the DIPF | Leibniz Institute for Research and Information in Education. She holds a PhD in educational psychology, and her research focuses on the psychological mechanisms underlying social inequalities in education, including parental involvement, ability grouping, and compositional effects. In a new line of work, she investigates whether and how educational inequalities can be reduced through the practice of adaptive teaching.

**KEYWORDS:** educational inequality, English language learners, labeling, matching, teacher perceptions

With large achievement and attainment gaps between students classified as English learners (ELs) and those who are not, scholarly and practitioner attention has turned to consider the extent to which these gaps may, in part, be driven by the very services and treatments apportioned to EL students. Quasi-experimental studies examining the effects of kindergarten EL classification on later academic achievement have come to varied conclusions: Some show positive effects (Shin, 2018), while others show negative ones (Umansky, 2016). Likewise, studies measuring the effects of remaining an EL rather than exiting EL status have demonstrated a range of effects on later achievement, course placement, behavioral outcomes, graduation, and postsecondary enrollment. These include neutral effects (Reyes & Hwang, 2019; Robinson, 2011), mixed effects (Cimpian et al., 2017; Robinson-Cimpian & Thompson, 2016), and negative effects (Carlson & Knowles, 2016; Johnson, 2019). 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 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 instruction toward English language development (Baker et al., 2014), content instruction in students' home languages (Steele et al., 2017), and specially trained teachers (Master et al., 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 et al., 2003). Another, albeit infrequently examined, potential mechanism of negative EL classification effects relates to teacher perceptions of student ability and their expectations for students' future outcomes.

Drawing on labeling theory (Link & Phelan, 2013), scholars have highlighted how "English learner" 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 et al., 2015; Martínez, 2018). Research has identified how some teachers hold downwardly biased academic perceptions of their EL students and interpret students' lack of English proficiency as a lack of academic skill or potential (Blanchard & Muller, 2015; E. B. Garcia et al., 2019; García & Guerra, 2004; Katz, 1999; Olsen, 1997; Pettit, 2011; Valenzuela, 1999). The large body of teacher perception and expectancy research from the past 50 years (see Jussim & Harber, 2005) indicates that downwardly biased perceptions and/or expectations could negatively affect EL student outcomes.

However, differences in teachers' perceptions of EL and non-EL students are not necessarily the result of EL classification. Teachers may have lower academic perceptions of their EL students that accurately reflect differences in the average skill levels of their EL, compared with their non-EL, students. As such, while teacher perceptions could drive differences in student outcomes between EL and non-EL students, it could also be the case that real differences in student skill levels drive observed differences in teacher perceptions.

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 comparatively little attention as far as large-scale and quantitative research is concerned (for two exceptions, see Blanchard & Muller, 2015, and E. B. Garcia et al., 2019). This study addresses that gap by drawing on the Early Childhood Longitudinal Study—Kindergarten Cohort of 2010–2011 (ECLS-K:2011), a nationally representative data set that asks teachers a series of questions about their perceptions of individual student skill levels across a range of academic content areas and grades. In this study we were able to take advantage of a unique policy characteristic that creates what we will argue is a natural experiment to test the hypothesis that EL classification affects teachers' perceptions of their students. 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 EL 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. Because the ECLS-K:2011 data include information about both EL identification and a universally administered measure of English proficiency, we are able to identify this group of students where EL classification is as good as random. Using coarsened exact matching analysis to match students with the same English proficiency levels (and other characteristics) but different language classifications, we then estimated the causal effects of EL classification in kindergarten on teachers' perceptions of student academic skill levels.

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 EL students. Most are served in English instructional programs, that is, 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 students' home languages. It is plausible that the type of instructional program a teacher

works in moderates the impact of EL classification on teacher perceptions. Specifically, a large body of research has found that bilingual instruction is beneficial for EL students (National Academies of Sciences, Engineering and Medicine [NASEM], 2017; Steele et al., 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 (Baker, 2011; Ruiz, 1984). As such, we tested whether teacher perceptions of EL students' academic skill levels 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 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 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 affect student outcomes such as test scores or measures of intelligence (Dusek & Joseph, 1983; Hinnant et al., 2009; Jussim et al., 1996; Jussim & Harber, 2005; Sorhagen, 2013; F. A. López, 2017). These effects of teacher perceptions on student outcomes have been explained via mechanisms including grade retention (Burkam et al., 2007), track placement (Oakes, 2005), within-class ability grouping (Tach & Farkas, 2006), and instructional quality and characteristics (Page, 1987).

Importantly, teacher perceptions and expectations have been found to be systematically lower for minoritized and/or stigmatized groups of students, including African American, Latinx, and low-income students (Auwarter, & Aruguete, 2008; Ferguson, 2003; Meissel et al., 2017; McKown & Weinstein, 2008; Ready & Wright, 2011; Rubie-Davies, 2010; Tenenbaum & Ruck, 2007). A core question has been whether and to what extent teachers' differential perceptions reflect real differences in skill level versus being the result of stereotypes or bias. Taken together, the results of various studies examining this question show that teachers' perceptions of students' skills tend to be relatively accurate (Jussim et al., 1996; Jussim & Harber, 2005; Llosa, 2008; Madon et al., 1998; Meisels et al., 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 schooling system have been shown to affect teacher perceptions. In particular, research has shown that

special education labels negatively affect 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 et al., 2010).

### **Teacher Perceptions of EL-Classified Students**

The research on teacher perceptions of EL-classified students is nascent (Pettit, 2011, provides a review of this literature). Findings suggest that teachers have different perception and 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 to type of perception, such as perceptions of appropriate curricula or instructional methods, or students' personal attributes, academic knowledge, or future prospects. For example, using a nationally representative sample, Blanchard and Muller (2015) found that teachers were more likely to perceive immigrant students as hard working compared to nonimmigrant students whose home language is not English. At the same time, teachers tended to believe that these nonimmigrant students were less likely to complete college than students speaking English at home, a finding that is also reflected in qualitative research (Dabach et al., 2018).

With regard to scholastic outcomes, research has found that teachers have relatively accurate assessments of EL students' English proficiency levels (Llosa, 2008) but that they underestimate multilingual students' academic skills (Ready & Wright; 2011), with underestimation varying by grade level and student ethnicity. Perceptions of EL-classified students, the vast majority of whom are Latinx or Asian, are tied to student characteristics (Llosa, 2008), including race and ethnicity, with research demonstrating that teachers often hold stereotypes of Asian students as "model minorities" while holding stereotypes of Latinx students as "underachieving" (Lee & Zhou, 2015, N. López, 2003; Ochoa, 2013). In addition to research pointing to teachers having lower academic perceptions of their multilingual students, there is evidence that these perceptions are linked to both teachers' instructional choices and student outcomes. Murphy and Torff (2019) found that teachers believed that rigorous instructional methods involving critical thinking skills were less appropriate and beneficial for EL compared to non-EL students, while F. A. López (2017) found that teachers' expectations and beliefs about Latinx students were associated with their instructional practices.

Research that has looked specifically at EL-classified students is sparse but indicates that, on average, teachers have comparatively low perceptions of EL-classified students (E. B. García et al., 2019; García & Guerra, 2004; Katz, 1999; Pettit, 2011; Valenzuela, 1999; Walker et al., 2004). This important research,

largely ethnographic and qualitative, has rarely, however, accounted for measures of student skill level and therefore is not able to analyze the effects of EL classification on teacher perceptions. An exception is E. B. Garcia (2019), who found that teachers hold downwardly biased perceptions of EL students' executive function skills, accounting for direct measures of those skills.

There is limited research that suggests that teachers' perceptions of EL students may vary by academic domain. Specifically, teachers may have more positive views of their EL students' math skills compared to other 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 (2010), for example, 70% of surveyed teachers reported a belief that math was EL students' easiest subject, and about a quarter of teachers explicitly stated that math knowledge was "universal," transcending language. By contrast, teachers' perceptions of the academic skills of EL students may be more negative in the area of language arts. Several studies have documented how students' use of their home language and their use of code-switching practices are inaccurately interpreted by teachers as weaknesses in language arts and literacy skills (Escamilla, 2006; Salerno et al., 2019).

### **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 socioeconomic students and those in classes with lower average achievement are more likely to underestimate students' skills (Ready & Wright, 2011). In addition, 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 moderates teacher perception effects (Fox, 2015; 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 EL students vary according to school characteristics such as grade span (Gallo et al., 2014) and teacher characteristics, including how teachers understand their role, teachers' education level, their training and level of experience working with EL students (Byrnes et al., 1997; Dabach, 2011; Pettit, 2011; Walker et al., 2004; 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; NASEM, 2017; Steele et al., 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 educational resources (Baker, 2011; Matthews & López, 2019; Ruiz, 1984). Teachers' asset orientation in bilingual settings, in turn, has been attributed to teacher education and preparation to successfully work with multilingual students; teachers' critical awareness regarding educational equity and opportunity; and student-teacher congruence, familiarity, and closeness (Fránquiz et al., 2011; Hopkins, 2013; F. A. López, 2017). 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 the effects of EL status on teacher perceptions may be systematically different in bilingual versus monolingual English instructional environments.

### **Conceptual Framework and Hypotheses**

This study sought to fill two important gaps in the literature: First, it estimated the causal effect of kindergarten EL classification on teacher perceptions across the early elementary grades. Second, it looked at whether teachers' perceptions of students classified as ELs in kindergarten differed systematically in bilingual versus English instructional environments.

We posit that kindergarten EL status could affect teacher perceptions in two distinct ways. First, teachers may hold downwardly biased perceptions of their students' abilities as a direct result of the EL label (E. B. Garcia et al., 2019). We refer to this as a direct effect of EL classification on teacher perceptions. Second, EL status may result in diminished instructional access or opportunity to learn in ways that negatively affect student academic outcomes (Callahan, 2005; Johnson, 2019). EL students, for example, might be placed in lower level instructional groups, or might be called on less frequently than their peers (Harklau, 1999). In this case, EL status could affect teacher perceptions of academic skills indirectly, through real changes in student skill level. We refer to this as an indirect effect of EL classification on teacher perceptions. Of note, while direct effects indicate teacher bias, indirect effects do not indicate bias, per se, because teachers' lower perceptions accurately reflect students' skills. Both mechanisms may occur in tandem, and may, in fact, be intertwined. For example, if teachers have biased perceptions of EL students, and as a result alter their instruction to those students in ways that diminish student learning, then subsequent measures of teacher perceptions might reflect a combination of direct and indirect effects.

This article opened by theorizing that EL status might affect student outcomes via teacher perceptions. Of note, if we find evidence that EL status

affects teacher perceptions exclusively through student outcomes (i.e., an indirect effect) this would suggest that teacher perceptions are not the *cause* of lower student outcomes, but instead are a *result* of lower student outcomes. Meanwhile if we find evidence of direct effects, or a combination of direct and indirect effects, that would support the hypothesis of teacher perceptions as a potential mechanism for EL status effects on student outcomes.

Our research questions were as follows:

*Research Question 1:* Does kindergarten EL status affect teachers' perceptions of students' academic skills among multilingual students across the early elementary grades? If so, what descriptive evidence do we have as to the mechanisms of that effect (i.e., direct vs. indirect)?

*Research Question 2:* Do bilingual classrooms operate as a moderator of the impact of kindergarten EL classification on teachers' perceptions of students' academic skills?

## Method

### Data and Analytic Sample

This study drew on the ECLS-K:2011 data set, a federally collected, nationally representative sample of students who entered kindergarten in the 2010–2011 school year. The data set contains longitudinal information on this cohort of students through the fifth grade. For the purposes of this study, we included data from kindergarten, first grade, and second grade. We stopped at the second grade due to relatively high attrition of students from the EL category by the third grade (Saunders & Marcelletti, 2012).

Our sample of interest included students who spoke a language other than English at home based on either or both teacher and parent reports in kindergarten (Garrett & Hong, 2016). Throughout this article we will refer to these students as *multilingual* students. While many of these students were in the process of developing 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 EL students is mandated. This subsample of ECLS-K:2011 included 3,885 students. We omitted students who were missing one or more kindergarten variables of interest, leaving an analytic sample of 2,155 students (Pepinsky, 2018). Descriptive statistics of the analytic sample are shown in Table 1.

There were large differences between those in the sample of multilingual students who were classified as EL in kindergarten and those who were not. Non-EL multilingual students had, on average, higher kindergarten English proficiency levels, higher family socioeconomic status, and were less likely to be in bilingual classrooms. Most relevant to this study, non-EL students had



*Teacher Perceptions of EL-Classified Students*

**Table 1**  
**Descriptive Statistics of the Analytic Sample**

	Full Sample	Non-EL	EL
EL program participation			
Kindergarten	56.33%	0.00%	100.00%
1st Grade	46.26%	20.62%	66.14%
2nd Grade	39.44%	16.26%	57.41%
English proficiency measures			
PreLAS (0–20)	15.90	17.61	14.56
EBRS (0–20)	11.93	13.41	10.78
Student academic skill measures			
English reading assessment (theta score)	−0.82	−0.50	−1.07
Math assessment (theta score)	−0.80	−0.49	−1.04
Executive functioning assessment 1 (0–18)	13.49	14.16	12.97
Executive functioning assessment 2 (393–603)	423.93	432.87	417.01
Teacher academic perceptions (standardized)			
Kinder. composite	−0.18	0.06	−0.36
Kinder. language/literacy	−0.20	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	0.14	−0.23
1st Gr. composite	−0.11	0.14	−0.30
1st Gr. language/literacy	−0.11	0.15	−0.32
1st Gr. Math	−0.09	0.12	−0.26
1st Gr. Social studies	−0.06	0.14	−0.21
1st Gr. science	−0.10	0.12	−0.27
2nd Gr. composite	−0.14	0.15	−0.35
2nd Gr. language/literacy	−0.16	0.11	−0.37
2nd Gr. math	−0.06	0.18	−0.24
2nd Gr. social studies	−0.09	0.15	−0.27
2nd Gr. science	−0.08	0.17	−0.27
Student and family characteristics			
Female	49.84%	50.90%	49.01%
Age (in months)	66.14	66.44	65.90
Latinx	63.99%	50.27%	74.63%
Asian	21.72%	26.25%	18.20%
White	8.49%	14.88%	3.54%
Combined other racial or ethnic group	5.80%	8.61%	3.62%
Family SES (standardized)	−0.48	−0.21	−0.69
Special education status	2.97%	3.61%	2.47%
Repeated kindergarten	5.80%	4.99%	6.43%
Chronically absent in kindergarten	11.09%	12.75%	9.80%
Changed teacher in kindergarten	4.55%	6.06%	3.38%

*(continued)*

Table 1 (continued)

	Full Sample	Non-EL	EL
Bilingual instruction			
Enrolled in a bilingual classroom	14.39%	2.13%	23.89%
Classroom and teacher variables			
Full-day program	81.35%	77.90%	84.02%
Teacher years of experience	13.66	13.58	13.73
Teacher holds a master's degree	47.94%	50.80%	45.72%
Teacher holds a degree in education	85.29%	88.84%	82.54%
Prop. of class—Latinx	51.77%	38.30%	62.20%
Prop. of class—White	11.00%	11.57%	10.56%
Prop. of class—African American	13.17%	17.41%	9.88%
Prop. of class—other race/ethnicity	7.21%	8.61%	6.13%
Prop. of class—EL	43.75%	24.08%	59.00%
Prop. of class—low reading skills (teacher perception)	17.25%	18.86%	16.00%
Prop. of class—poor behavior (teacher perception)	10.02%	10.20%	9.88%
Class size	20.95	20.91	20.98
School characteristics			
Rural location	10.63%	13.82%	8.15%
School size (1–4) <sup>a</sup>	2.82	2.76	2.87
Average school SES (standardized)	−0.32	−0.16	−0.44
Prop. of school—Black and Latinx	57.55%	48.93%	64.24%
<i>N</i>	2,155	941	1,214

*Note.* All teacher academic perception variables are standardized. Kinder. = kindergarten; Gr. = grade; EL = English learner; Prop. = proportion; SES = socioeconomic status; PreLAS = Preschool Language Assessment Scale; EBRS = English Basic Reading Skill.

*Source.* U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Studies, Kindergarten Class of 2010–11, 2010–2014.

<sup>a</sup>School size: 1 = 0–299 students, 2 = 300–499 students, 3 = 500–749 students, 4 = 750 or more students.

considerably higher academic skill levels in kindergarten. Therefore, it is not immediately evident whether differences in teachers' perceptions of students' skills (see Table 1) reflected real skill differences across the two groups, or if they were, by contrast, caused in part by EL classification. Because of these differences in the baseline measures of EL and non-EL multilingual students, it was critical to identify a counterfactual group of kindergarten non-EL students with similar characteristics to the EL students, as we did in the present study.

## 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 over time.

We used these teacher perception variables from the spring of kindergarten, after teachers had been working with their students for approximately one academic year, and then again at the end of first grade and the end of second grade. We draw on teachers' perceptions of students' skills in the areas of (1) language and literacy, (2) math, (3) social studies, and (4) science. 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 [NCES], n.d.). For each grade, we also constructed a composite measure by averaging all four academic domains because a principal component analysis indicated that there was one latent construct underlying a given teacher's assessment of a student across the four academic domains, with similar weights across the domains.

Teacher perception measures differed by grade level. 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-type scale ranging from 1, which represented *not yet proficient* to 5, *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: eight items,  $\alpha = .89$ ; language/literacy: nine items,  $\alpha = .94$ ). 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 5-point Likert-type scale for this item ranged from 1, for *far below average*, to 5, *far above average*. For our across-domain kindergarten composite score, we averaged teachers' perceptions across the four domains of math, language/literacy, social studies, and science ( $\alpha = .94$ ).

In first grade, math, language/literacy, and science were measured via multiple items, which were each answered on the same 5-point Likert-type scale as the multiple items in kindergarten. Again, we created an average score for each domain and an overall composite across the four domains. Reliabilities for the overall composite ( $\alpha = .98$ ), and for each domain were high (math: eight items,  $\alpha = .96$ ; language/literacy, nine items,  $\alpha = .97$ ; science: eight items,  $\alpha = .97$ ). Teacher perceptions in social studies were measured on a 5-point Likert-type scale via a single question, as in kindergarten.

In second grade, teacher perceptions were assessed via one question for math, science, and social studies and three questions for language/literacy, each of which were answered on a 3-point Likert-type scale, which ranged from 1, for below grade level, to 3, for above grade level. For language/literacy, we took the average score of the three questions ( $\alpha = .87$ ); for the overall composite, we again took the average score across the four domains ( $\alpha = .88$ ).

Because the scale for the second-grade perception variables was different from the kindergarten and first grade scales, we standardized all outcome variables in kindergarten, first, and second grade. This allowed us to compare effect sizes across grade levels. It also facilitated effect size interpretation by translating unique scales into standard measures of effect size. We standardized all outcome variables using their mean and standard deviation within the full ECLS-K:2011 data set.

### *Predictor Variables of Interest*

The primary predictor variable of interest was EL status. EL status was derived from a single question posed to teachers in a questionnaire in the spring of kindergarten.<sup>2</sup> Teachers were asked about each multilingual student: “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 suggest 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, 1,214 out of 2,155 multilingual students (56%) were considered EL. The remaining 941 students were multilingual students who 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 and they were therefore not eligible for EL services. In other cases, schools may have been failing to provide EL services to eligible students, parents may have opted out of EL supports, or teachers may not have understood or correctly answered the question. We argue that teacher report for this variable may actually benefit this study. This is because our research questions surround teachers’ responses to students’ kindergarten language classification and, as such, the sample should only include those students whose language classification was known by their teachers.

We used the kindergarten measure of EL status because, as described above, we are interested in the effects of EL status over time. While students take an average of 5 to 7 years to reach English proficiency (NASEM, 2017), some of the EL-classified students in our sample exited EL status by the time they reached the first or second grade. As such, our estimates in first and second grade should be interpreted as the effects of kindergarten EL classification on later grade teacher perceptions, where some of the treatment group retained their classification and others lost it. Specifically, Table 1 shows that 66% of the treatment group retained their EL status in first grade, and 57% retained their status in second grade. Table 1 also illustrates that a nonnegligible proportion of multilingual students who were not reported

as being in an EL program in kindergarten were reported to be in an EL program in first (21%) and/or second (16%) grade. While this runs counter to federal education policy (which only allows for students to be classified as ELs when they first enter a school district), it may reflect later classifications, students that moved between districts (a given student would only remain in the ECLS-K data set if they happened to move to another school sampled in ECLS-K), or data errors. Because the end result was that some proportion of the control group (non-EL students in kindergarten) likely did, indeed, receive EL classification in later grades, this likely biased our estimates downward in those later grades. While the shifting nature of EL status is a limitation of our study, we conducted a sensitivity check that accounted for later EL status.

### *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 using state or local assessments. In the ECLS-K:2011 data set, 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 ( $\alpha = .91$ ), was used as a screener to assess each student's oral (speaking and listening) English proficiency and determine whether they 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 equal to or more than 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 less than 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 ( $\alpha = .87$ ). 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 (1.7% of the sample scored the full 20 points).

Additional matching variables included student race/ethnicity, gender, and socioeconomic status along with an indicator variable for whether the

student's district was in a rural setting. As we describe below, we found that once we matched students on these variables, there were no meaningful or significant differences in students' skill levels, as directly measured through ECLS-K: 2011 assessments. The one exception was that in some models there were small but significant differences between EL and non-EL groups on one of the two executive functioning assessments (an oral number reversing activity). As such we included that measure as a matching variable.

### *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. In our main model, all control variables are from students' kindergarten year. Regarding student-level covariates, we included scores on kindergarten reading (reliability = .95) and math (reliability = .92) assessments (both item response theory-based theta scores), two kindergarten executive functioning assessments,<sup>3</sup> age, special education status, whether the student repeated kindergarten, whether the student was chronically absent, and whether the student changed teachers midyear. For classroom and teacher-level variables, we included: whether the class was full or half day, the teacher's number of years of teaching experience, whether the teacher held a master's degree, and whether the teacher held a degree in an education-related field (Wayne & Youngs, 2003). We also controlled for class size and racial composition, proportion of EL students in the class, the proportion of the class the teacher considered to be lower-skill readers, and whether the teacher considered the class to be poorly behaved. For school-level variables, we included school size, average socioeconomic status, and the proportion of Black and Latinx students. As described below, our analyses that explore the mechanisms by which EL status might affect teacher perceptions added students' later academic achievement variables (English proficiency, reading, math, and executive functioning) to our models.

### *Moderator Variable*

Our second research question explored the moderating variable of bilingual program enrollment. 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?" There were five options given, ranging from 1 = *never*, to 5 = *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 in kindergarten (see Table 1).

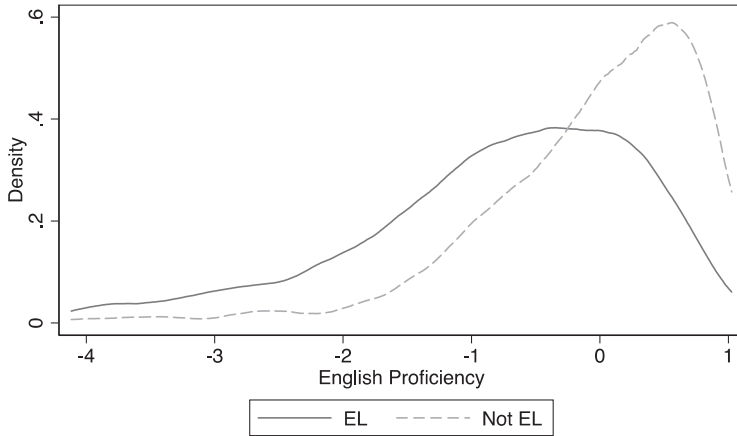
Because we focused our analyses on the effects of kindergarten EL status over time, we used an indicator of bilingual instruction from the kindergarten year. Most students who were in bilingual settings in kindergarten remained in bilingual settings in the first (63%) and second (55%) grades. Almost no students not in bilingual settings in kindergarten moved into bilingual settings in the first or second grades (<2%). There were very few students ( $N = 20$ ) in the sample who were in a bilingual program and were *not* considered ELs 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 home or primary language other than English (i.e., multilingual students, as defined in this study). Schools 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: multilingual status and English proficiency level.

However, states—and prior to implementation of ESSA (2015), districts—are able to set their own thresholds on the English proficiency measures 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 8 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). This context creates a natural experiment (Murnane & Willett, 2010) that we exploit in this study.

Because of the variation in tests and thresholds, 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 et al., 2016; Mavrogordato & White, 2017; Ragan & Lesaux, 2006; Sireci & Faulkner-Bond, 2015; Solórzano, 2008). This variation in EL classification rules and implementation amounts to exogenous variation in student classification assignment, once accounting for student



**Figure 1. Distribution of fall combined English proficiency scores (PreLAS and EBRS), among multilingual students, by EL program enrollment.**

*Note.* PreLAS = Preschool Language Assessment Scale; EBRS = English Basic Reading Skill; EL = English learner.

*Source.* U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Studies, Kindergarten Class of 2010–11 (ECLS-K:2011), 2010–2014.

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, students with very low true English proficiency levels might score below the EL threshold across multiple assessments and have a high likelihood of being classified as EL across locales. However, for students with true English proficiency levels in the middle, one would expect significant variation across locales in EL or non-EL identification due to the variation across assessments and thresholds.

In order to exploit this natural experiment, we needed a universally administered English proficiency assessment separate from those administered and used by schools to determine EL classification. ECLS-K:2011 provides just such an assessment (the PreLAS and EBRS). Our empirical strategy therefore homed in on a region of common support where EL and non-EL students had the same measured English proficiency levels (and were similar regarding other characteristics). Specifically, we examined whether teacher perceptions of student ability were different for students classified in kindergarten as ELs compared with students who had the same measured English proficiency level (and other characteristics) but were not classified as ELs. Figure 1 shows our region of common support (for ease of interpretation, we standardized, centered, and then averaged each student's PreLAS and EBRS scores).



As a matching analysis, we can interpret our estimates causally only if, conditional on observed matching and control variables, classification as an EL (treatment assignment) is exogenous, or as good as random. If, for example, EL classification is assigned, in part, on teachers' or administrators' sense of student academic need (i.e., an omitted variable), then differences in teacher perception of student skill (our outcome) could be due to systematic differences in student academic need rather than EL classification assignment.

We argue that the circumstances of this study are a near ideal use of matching and that, as such, our estimates of the effect of EL classification on teacher perception (Research Question 1) can be interpreted causally.<sup>4</sup> First, as stated above, by law, kindergarten EL classification must be assigned only based on (1) multilingual status and (2) measured English proficiency level, both of which we can account for in our models. Second, prior research has demonstrated very high compliance with EL classification law. For example, Umansky (2016) found 89% compliance in kindergarten EL classification in one large school district, while Shin (2018) reported nearly universal compliance in a different district. These first two points are important because prior research has shown that causal estimates using observational data (and methods such as matching) align with those from experiments in cases where the selection process into the intervention is known and can be effectively modeled or proxied (T. D. Cook et al., 2008). Finally, once we account for multilingual status and measured English proficiency (along with key demographic characteristics described above) there are no remaining observable differences in the measured academic skill levels of EL and non-EL students (see Table 2; also described further below). This provides evidence that our matching and control variables fully account for treatment assignment and any remaining variation is as good as random.

### **Analytic Strategy**

Coarsened exact matching (CEM), like all matching strategies, matches individuals in the treatment group (multilingual kindergartners classified as ELs) with students who are similar to them but who are in the control group (multilingual kindergartners not classified as ELs). It then examines the differences in outcomes between the matched sample of treated and control individuals. The purpose of matching 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 limited our 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 et al., 2012), thereby reducing observed variable bias (Murnane & Willett, 2010).

*Table 2*  
**Descriptive Statistics on Key Matching and Control Variables, Pre- and Postmatching**

	Prematched Full Sample			Postmatched Sample		
	Non-EL	EL	<i>t</i> Test	Non-EL	EL	<i>t</i> Test
PreLAS (0–20)	17.61	14.56	***	14.48	14.40	<i>ns</i>
EBRS (0–20)	13.41	10.78	***	10.63	10.56	<i>ns</i>
Math assessment (theta score)	−0.50	−1.07	***	−1.17	−1.13	<i>ns</i>
Reading assessment (theta score)	−0.49	−1.04	***	−1.03	−1.06	<i>ns</i>
Executive function 1 (0–18)	14.16	12.97	***	12.99	12.83	<i>ns</i>
Executive function 2 (393–603)	432.87	417.01	***	415.72	415.02	<i>ns</i>
Female	50.90%	49.01%	<i>ns</i>	49.31%	49.31%	<i>ns</i>
Latinx	50.27%	74.63%	***	81.77%	81.77%	<i>ns</i>
White	14.88%	3.54%	***	1.93%	1.93%	<i>ns</i>
Asian	26.25%	18.20%	***	15.19%	15.19%	<i>ns</i>
Other race or ethnic group	8.61%	3.62%	***	1.10%	1.10%	<i>ns</i>
Rural	13.82%	8.15%	***	2.49%	2.49%	<i>ns</i>
SES (standardized)	−0.21	−0.69	***	−0.70	−0.71	
<i>N</i>	941	1,214		538	724	
Multivariate L1 distance		0.9382			0.8533	

*Note.* The matching algorithm included the two English proficiency measures, one baseline executive functioning skill variable, student race/ethnicity, an indicator for whether the school attended was in a rural location, and student socioeconomic status. All categorical and dichotomous variables are exact matches. The English proficiency variables were matched by sample distribution quintile while the executive functioning measure was matched by sample distribution halves. All variables measured in kindergarten. EL = English learner; SES = socioeconomic status; PreLAS = Preschool Language Assessment Scale; EBRS = English Basic Reading Skill. *Source.* U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Studies, Kindergarten Class of 2010–11 (ECLS-K:2011), 2010–2014.

\*\*\* $p < .001$ .

Compared with 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 research questions (Iacus et al., 2012). Specifically, CEM allows researchers to dictate the features of the matching algorithm in substantively meaningful ways both with regard to which variables to include in the matching process and with regard to the rules for how close the matches should be for each variable. In addition, researchers evaluate the quality of the matched sample before attempting to answer research questions. 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 selected for matching. For each variable, one can require either exact matches or one can coarsen the variable into a selected number of bins and match within each bin. CEM then assigns weights based on how many matches there are per individual, and these weights are used in subsequent analytic models. In this study, we matched on kindergarten English proficiency level, executive function skill level, gender, race/ethnicity, rural locale, and socioeconomic status. All analyses were conducted using Stata version 15. In the matching algorithm race, gender and rural locale were set to be exact matches while we binned the continuous variables: English proficiency level, executive functioning skill, and socioeconomic status. Following Rosenbaum and Rubin (1984), we binned each of the continuous variables into quintiles (based on the sample distribution)<sup>5</sup>; matching by quintiles has been shown to eliminate more than 90% of bias.

As described, we were then able to evaluate the quality of our matched sample. Table 2 shows the balance between the prematched sample and the postmatched sample on matching and other key variables. The region of common support covered 59% of the analytic sample ( $N = 1,262$ ). The selected algorithm achieved a good balance between the treatment and control groups such that there were only very small, statistically insignificant, differences between the groups on all key matching and control variables, including kindergarten assessments in reading and math (see Table 2).

The characteristics of the matched sample, which reflect the region of common support, were different from the full sample of multilingual students. This may be because ECLS-K:2011 oversamples specific subgroups such as Asian and Pacific Islander students (Tourangeau et al., 2015). 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 with the full multilingual student sample. Compared with the full sample, the matched analytic sample more closely aligned with characteristics of the EL population in the United States (NCES, 2018).

We then analyzed our matched data in a regression framework. 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. Research Question 1 asks about the impact of kindergarten EL status on teachers’ academic perceptions of their students across grades. To answer this, we used the following model:

$$\text{PERCEP}_i = \beta_0 + \beta_1 \text{EL}_i + \beta_2 \text{PreLAS}_i + \beta_3 \text{EBRS}_i + \beta_4 \text{ACHIEVE}_i + \beta_5 \mathbf{X}_i + e_i \quad (1)$$

where **PERCEP** represents the set of teacher academic perception outcomes in grades kindergarten through second grade for student  $i$ , **EL** is our proxy for EL

status in kindergarten, 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 additional student, family, teacher, class, and school covariates. Standard errors were clustered at the school level to account for students within schools. To account for the complex sampling design used for ECLS-K:2011 data collection, as well as our matching results, we followed DuGoff et al. (2014), creating a new weight for each observation equivalent to the product of the CEM weights and the ECLS-K:2011 sampling weights. The coefficient of interest is  $\beta_1$ , which represents the estimated effect of kindergarten EL status on teacher academic perceptions, among multilingual students, holding constant students' English proficiency level, achievement levels, and a host of other characteristics.

After conducting our main analyses, we sought to descriptively explore the mechanisms through which EL status affects teacher perceptions. As described earlier, we argue these effects could be (1) direct effects of EL classification in the form of teacher bias toward students carrying the EL label, (2) indirect effects of EL classification through the mechanism of altered educational experiences resulting in altered educational outcomes (that teachers then accurately perceive), or (3) some combination of both. In order to descriptively test these mechanisms, we ran analyses that added measures of students' later achievement to Equation 1. Our rationale was that direct effects of EL classification would remain once controlling for later student achievement. Indirect effects of EL classification, via effects of classification on student achievement, would not be picked up in a model that controlled for later achievement. Specifically, our first-grade models included spring of kindergarten achievement measures, and our second-grade models included spring of first-grade achievement measures.

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:

$$\begin{aligned} \text{PERCEP}_i = & \beta_0 + \beta_1 \text{EL\_BIL}_i + \beta_2 \text{EL\_NOTBIL}_i + \beta_3 \text{PreLAS}_i \\ & + \beta_4 \text{EBRS}_i + \beta_5 \text{ACHIEVE}_i + \beta_6 \text{X}_i + e_i \end{aligned} \quad (2)$$

where all variables are defined as in Equation 1. We removed the kindergarten EL variable from Equation 1 and replaced it with two variables, one indicating whether the kindergartner was an EL and in a bilingual class (EL\_BIL) and one indicating whether the kindergartner was an EL and not in a bilingual class (EL\_NOTBIL). As mentioned above, there were only 20 non-EL kindergartners in bilingual classrooms in the sample, which meant we could not include an interaction term of EL and BIL. Instead, this model allowed us to estimate teacher perceptions for three groups of students: multilingual non-EL students (the reference category), multilingual EL students in bilingual classes, and multilingual EL students not in bilingual classes. The coefficients of interest

in this model are  $\beta_1$  and  $\beta_2$ , which represent the estimated difference in teacher perceptions for kindergarten ELs in bilingual classes and not in bilingual classes, respectively, compared with teacher perceptions of non-EL multilingual students. We then ran contrast tests to test the differences between the three groups of students. Of specific interest, we report results of contrast tests which test whether the relationship of EL status to teacher perceptions differed for EL kindergartners in and not in bilingual classes. Importantly, results from these analyses are not causal estimates. Prior research suggests that students are not randomly distributed across bilingual and English only settings. Instead, bilingual program enrollment is associated with characteristics such as parental value for biliteracy and multiculturalism (Parkes, 2008). We cannot fully account for such differences, nor differences between the characteristics of bilingual and English only instructional settings. As such, in contrast to Research Question 1, the results for Research Question 2 are correlational.

### *Sensitivity Analyses*

We conducted an array of sensitivity checks. First, we conducted our regression analyses without any matching (Sensitivity Check 1). Ordinary least squares analysis, in the absence of matching, does not provide causal estimates. Instead, we included it as a first step and as a point of comparison to our matching results. The remainder of our sensitivity analyses all included matching.

In Sensitivity Check 2, we used an alternate matching method, propensity score matching, instead of CEM. We included all of the matching variables used in the main model, plus reading and math assessment scores, and the second executive functioning assessment score. Propensity score matching allowed us to keep the full analytic sample, but the resulting matched sample was not as compelling because the EL treatment group had slightly but significantly lower English proficiency levels than the non-EL control group.

The remaining sensitivity checks all used CEM. Sensitivity Check 3 used the same matching variables as in the main model but added in the three additional direct assessments (reading, math, and the second executive functioning assessment). While scores on these three assessments were balanced across treatment and control groups without their inclusion, direct measures of academic skill are theoretically and empirically critical predictors of teacher perceptions of student academic skills and therefore merited inclusion as matching variables in a sensitivity check. The models used the same control variables. The resulting sample was somewhat smaller than the main model (48% of the analytic sample).

As described, our sample of multilingual students was made up of students identified by either their teacher or their parent as having a home language other than English. For the fourth sensitivity check, we removed students who were only identified as multilingual by their parent and then

proceeded with our main model matching and regression analyses. We removed these students because teachers might hold biased perceptions of multilingual students more broadly. In cases where they do, we only wanted to include in our treatment and control group students that teachers knew to be multilingual. While teachers knew their EL students were multilingual, by definition, they may not have known about their non-EL students' home languages. As such, our main model might have biased our estimates by including in our control group students that teachers did not consider multilingual.

As noted earlier, there was a significant amount of movement both out of, and into, the EL status category across grades. Specifically, 20% of the control group were reported as being in an EL program in either or both first and second grades. This could have biased our results since the control group included EL students. As such, we conducted a fifth sensitivity check where the treatment group was defined as "ever-EL" students, that is, students who were reported as being in an EL program for at least one of the three grade levels. Control group students, by contrast, were defined as "never-EL" students.

Finally, we addressed the movement in and out of the EL category through models that shifted the treatment variable to a time-varying variable indicating EL status in the current grade (Sensitivity Check 6). In these models we also shifted the classroom, teacher, and school variables to reflect the current grade. Of note, these models answer a somewhat different research question; they estimate the effect of EL status on teacher perceptions within a given grade level.

Supplemental Appendix Table A in the online version of the journal presents descriptive statistics of the matched samples for the treatment and control groups for the sensitivity checks that involve matching (2–6). Results from all sensitivity checks are presented in Supplemental Appendix Table B in the online version of the journal and described at the end of the results section.

## Results

### Research Question 1: Estimated Impact of Kindergarten EL Status on Teacher Perceptions

Table 3 presents CEM estimates of the effects of kindergarten EL status on teacher perceptions of students' academic skills among multilingual students. Results were negative across all four academic content areas—language/literacy, math, social studies, and science—and across all three grades—kindergarten, first grade, and second grade. Results were statistically significant in all domains in first grade but were not statistically significant in kindergarten. In second grade, results were significant in math, and marginally significant in the composite outcome. These results suggest that EL classification in kindergarten had a negative effect on teachers' perceptions of student academic skill level in first grade, and in math in second grade. Negative effect sizes ranged from a tenth to a third of

*Table 3*  
**Coarsened Exact Matching Estimates of Effect of EL Status on Teacher Perceptions  
of Students' Academic Skills, Among Multilingual Students**

	Kindergarten				1st Grade				2nd Grade						
	Composite	Language	Math	Social Studies	Science	Composite	Language	Math	Social Studies	Science	Composite	Language	Math	Social Studies	Science
EL	-0.100 (0.083)	-0.111 (0.080)	-0.088 (0.094)	-0.071 (0.083)	-0.115 (0.075)	-0.261** (0.083)	-0.318*** (0.082)	-0.194* (0.083)	-0.365*** (0.095)	-0.189* (0.089)	-0.136 <sup>-</sup> (0.079)	-0.083 (0.083)	-0.269** (0.084)	-0.092 (0.076)	-0.104 (0.077)
N	1,262	1,262	1,262	1,262	1,262	1,002	1,002	1,002	1,002	1,002	994	994	994	994	994
R <sup>2</sup>	0.420	0.446	0.326	0.288	0.284	0.359	0.368	0.318	0.255	0.286	0.373	0.358	0.286	0.259	0.255

*Note.* Robust standard errors in parentheses. All models include English proficiency measures (Preschool Language Assessment Scale and English Basic Reading Skill), 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, education level, and education degree), class characteristics (racial composition, EL proportion, class size, and teachers' evaluation of class behavior and reading level), and school characteristics (rural locale, school size, proportion Black and Latinx, and average socioeconomic status). EL = English learner. *Source.* U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Studies, Kindergarten Class of 2010–11 (ECLS-K:2011), 2010–2014.

<sup>-</sup>*p* < .1. \**p* < .05. \*\**p* < .01. \*\*\**p* < .001.

a standard deviation. On the composite outcomes, EL status resulted in lower teacher perceptions of approximately a quarter of a standard deviation in first grade and a seventh of a standard deviation in second grade (as noted, the later estimate was only marginally significant). Effects of EL status on teacher perceptions accounted for a considerable proportion of the average difference between teacher perceptions of multilingual EL and non-EL students (see Table 1). For example, EL status effects accounted for over half (59%) of the average differences in teacher perceptions in first grade. Across grades, there was no clear evidence supporting our hypothesis that EL status effects were larger in language arts than in math.

### *Mechanism Analyses*

Results from our models that added later grade achievement measures provide preliminary evidence that teacher perception effects in first and second grades were driven by both direct and indirect effects (see Table 4). Point estimates from this set of models represent the estimated direct effects of kindergarten EL status on teacher perceptions that remain once removing indirect effects. These point estimates remained negative, but they were smaller in magnitude than the point estimates from our main models (first- and second-grade composite outcome point estimates were 46% and 62% smaller, respectively). Four out of the six estimates that were significant or marginally significant in the main models remained so in these mechanism models. This suggests that a portion—but not all—of the effect of kindergarten EL status on teacher perceptions was explained by differences in student skill levels that emerged in the first and second grades between students who had had equivalent achievement levels in kindergarten.

### **Research Question 2: Moderator Role of Bilingual Classrooms**

Table 5 shows CEM estimates from our moderator models, where we removed the EL status indicator and replaced it with two alternative indicators, one for EL students in bilingual classes and one for EL students not in bilingual classes. Non-EL kindergartners (98% of whom were not in bilingual classes) remained the reference category. Point estimates on the two indicator variables represent the estimated correlational difference between the relevant kindergarten EL group and the non-EL reference group. The table also includes results from contrast tests that examined whether there were significant differences *between* the two EL groups.

In first and second grades, we found a negative association of kindergarten EL classification with teacher perceptions of student academic skill level among students who were not in bilingual classes. These point estimates were uniformly negative and were generally of larger magnitude than estimates in Table 3 that combined EL students in and not in bilingual classrooms. Estimates in first grade were, as in the main model, statistically significant, and



*Table 4*  
**Coarsened Exact Matching Mechanism Analyses Results Incorporating Later Student Achievement as Control Variables**

	1st Grade					2nd Grade				
	Composite	Language	Math	Social Studies	Science	Composite	Language	Math	Social Studies	Science
EL	-0.140 <sup>~</sup> (0.085)	-0.179* (0.079)	-0.089 (0.087)	-0.236** (0.072)	-0.101 (0.093)	-0.051 (0.073)	-0.001 (0.077)	-0.201** (0.077)	-0.017 (0.074)	-0.030 (0.075)
N	997	997	997	996	997	985	985	985	985	985
R <sup>2</sup>	0.477	0.501	0.415	0.376	0.362	0.486	0.469	0.377	0.321	0.325

*Note.* Robust standard errors in parentheses. These models match on kindergarten English proficiency measures (Preschool Language Assessment Scale and English Basic Reading Skill), one executive functioning assessment, and student characteristics (gender, age, race, family socioeconomic status, special education identification). They control for academic achievement measures (math, reading and two executive functioning measures) in the spring of the grade prior to the teacher perception outcomes (i.e., second-grade teacher perception models control for spring of first grade student achievement). Models also control for 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, education level, and education degree), class characteristics (racial composition, EL proportion, class size, and teachers' evaluation of class behavior and reading level), and school characteristics (rural locale, school size, proportion Black and Latinx, and average socioeconomic status). EL = English learner.

*Source.* U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Studies, Kindergarten Class of 2010–11 (ECLS-K:2011), 2010–2014.

<sup>~</sup> $p < .1$ . \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

*Table 5*  
**Coarsened Exact Matching Estimates of the Moderating Role of Bilingual Classroom Environment on Teachers' Perceptions of Students' Academic Skills, Among Multilingual Students**

	Kindergarten										1st Grade										2nd Grade									
	Composite			Language			Math			Social Studies			Science			Composite			Language			Math			Social Studies			Science		
	Estimate	SE	95% CI	Estimate	SE	95% CI	Estimate	SE	95% CI	Estimate	SE	95% CI	Estimate	SE	95% CI	Estimate	SE	95% CI	Estimate	SE	95% CI	Estimate	SE	95% CI	Estimate	SE	95% CI			
EL-not bil	-0.103 (0.083)	-0.117 (0.080)	-0.083 (0.094)	-0.111 (0.082)	-0.153* (0.075)	-0.296*** (0.083)	-0.348*** (0.082)	-0.229*** (0.084)	-0.386*** (0.099)	-0.221* (0.089)	-0.172* (0.079)	-0.114 (0.083)	-0.291*** (0.083)	-0.132* (0.076)	-0.142* (0.076)															
EL-bil	-0.088 (0.142)	-0.084 (0.131)	-0.107 (0.161)	0.090 (0.125)	0.037 (0.117)	-0.117 (0.137)	-0.196 (0.130)	-0.054 (0.128)	-0.280* (0.119)	-0.057 (0.157)	-0.012 (0.129)	0.026 (0.134)	-0.190 (0.128)	0.050 (0.126)	0.032 (0.128)															
Contrast	-0.015 (0.129)	-0.033 (0.115)	0.024 (0.146)	-0.201* (0.100)	-0.190- (0.099)	-0.179 (0.120)	-0.152 (0.112)	-0.175 (0.113)	-0.106 (0.095)	-0.164 (0.141)	-0.16 (0.112)	-0.14 (0.119)	-0.101 (0.104)	-0.181 (0.112)	-0.175 (0.112)															
N	1,262	1,262	1,262	1,262	1,262	1,002	1,002	1,002	1,001	1,002	994	994	994	994																
R <sup>2</sup>	0.420	0.446	0.326	0.292	0.287	0.362	0.370	0.320	0.256	0.288	0.376	0.360	0.287	0.262	0.258															

*Note.* Robust standard errors in parentheses. All models include English proficiency measures (Preschool Language Assessment Scale and English Basic Reading Skill), 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, education level, and education degree), 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). EL = English learner.

*Source.* U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Studies, Kindergarten Class of 2010–11 (ECLS-K:2011), 2010–2014.  $^*p < .1$ .  $^{**}p < .05$ .  $^{***}p < .001$ .

those in the second grade were also statistically significant or marginally significant across all outcomes except for language. One outcome (science) was also statistically significant in kindergarten. By contrast, there was no evidence of a significant association of kindergarten EL status with teacher perceptions in any grade or academic domain when EL students were in bilingual classes (with the exception of first grade social studies). Unlike for EL students not in bilingual classes, teachers had comparatively higher perceptions of academic skill level for their EL-bilingual students compared with their non-EL, nonbilingual students, on average, in certain academic domains in kindergarten and second grade. Focusing on the composite outcomes, point estimates of the negative association of kindergarten EL status with teacher perceptions were magnitudes larger for EL students not in bilingual classes compared with those in bilingual classes in first and second grade. Contrast tests between the two kindergarten 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, by and large these tests did not reach statistical significance.

### **Results From Sensitivity Analyses**

Supplemental Appendix Table B in the online version of the journal presents results from our sensitivity analyses as described in the methods section. In all cases results paralleled those from our main analyses indicating negative effects of kindergarten EL classification on teacher perceptions across grades and academic domains, with minor differences in magnitude and statistical significance. Sensitivity Analysis 1, which involved ordinary least squares regression analyses without matching, was meant as a first examination among the full analytic sample. These results show a consistent, negative, and significant (or in a few cases marginally significant) relationship between EL classification and teacher perceptions across academic domains and grade levels. The remainder of the checks involved matching and are thus alternative causal estimates.

Sensitivity Checks 2 and 3 both altered the matching algorithms but not the analytic samples. Results from Sensitivity Check 2, which employed propensity score matching, suggest slightly larger (and significant) negative effects in kindergarten compared with first and second grades, and first- and second-grade estimated effects were smaller than in the main model. Results from Sensitivity Check 3, which matched on the full battery of ECLS-K: 2011 assessments, suggest the opposite: larger and more significant results in first and second grades, compared with kindergarten, with results in kindergarten and first grade similar to the main model, but larger in second grade.

Sensitivity Checks 4 and 5 altered the analytic samples. In both cases, effect sizes were larger and mostly significant in first and second grades, compared to kindergarten. When compared to the main model, point estimates

were slightly larger in the latter two grades in the check that included only teacher-identified multilingual students (Sensitivity Check 4), and the check that used ever-EL students as the treatment group (Sensitivity Check 5).

Finally, Sensitivity Check 6 examined within-grade effects of EL classification rather than looking at longitudinal effects of kindergarten EL classification. Results from these analyses were smaller than the main model (and some estimates were positive rather than negative), and did not reach statistical significance, adding to evidence that indirect effects of EL classification on teacher perceptions play an important role.

## Discussion

This study sought to analyze the effects of EL classification in kindergarten on teacher perceptions of student skills and abilities in kindergarten, first, and second grade. While EL classification is designed to ensure the rights of a potentially vulnerable group of students (Gándara et al., 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).

Harnessing the variation in English proficiency thresholds used in different states and districts to determine EL status eligibility (National Research Council, 2011) as a natural experiment (T. D. Cook et al., 2008; Murnane & Willett, 2010), we used ECLS-K:2011 data and CEM to examine teacher perceptions over time of students who entered school with the same English proficiency and academic skill levels (as well as other student, class, program, and school characteristics) but different language classifications (EL and non-EL). The results suggest that, as theorized, EL status in kindergarten has a negative effect on teachers' perceptions of students' academic skills across multiple academic domains and grade levels.

In our main models, results are weaker in kindergarten and second grade, and stronger in first grade. Effect sizes on composite measures range from a tenth of a standard deviation (kindergarten—not statistically significant) to a quarter of a standard deviation (first grade—statistically significant). Results from a host of sensitivity checks, including alternate methods (propensity score matching), algorithms, and analytic samples, converge on these findings of negative effects of EL classification on teacher perceptions, although effect sizes and significance levels vary somewhat across models. Effect sizes are, by and large, meaningful, accounting for a quarter to a half of the overall differences in teacher perceptions of EL and non-EL multilingual students. They also parallel those found in prior research on teacher perceptions. For example, Ready and Wright (2011) find that teacher perceptions of

the literacy skills of Latinx students who speak a non-English language at home are underestimated by between a quarter and a third of a standard deviation, once accounting for direct measures of literacy skills. Results from our mechanism analyses, where we account for students' later skill levels, provide preliminary evidence that EL status affects teacher perceptions both directly, due to biases associated with the EL label, and indirectly through diminished opportunity to learn that results in lower student academic growth that is then accurately represented in later grade teacher perceptions (Garrett & Hong, 2016).

We examine estimated effects both across content areas (in composite perception measures) and within content areas (in language arts, math, social studies, and science). While we hypothesized that kindergarten EL status might affect teachers' language arts perceptions more than math or other content areas, our results did not support this hypothesis. Results are fairly consistent across the four academic domains. The only grade level where point estimates differ meaningfully across domains is in second grade, where effects are considerably larger in math than in the other domains. But this difference is not reflected across the sensitivity checks and we therefore conclude that more work is needed to explore any differences in EL classification effects on domain-specific perceptions.

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 negative effects of kindergarten EL status on teacher perceptions are 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; Valdés, 1997). Consistent with our hypothesis, we found that, when in bilingual settings, teachers do not have systematically different perceptions of their kindergarten EL students compared to their non-EL multilingual peers. 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.

The findings from this study contribute to theory on and understanding of teacher perceptions and the experiences and opportunities of EL-classified students. With regard to research on teacher perceptions, this study confirms and adds to existing work that finds that teachers are more likely to underestimate 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 et al., 2009). Like these groups of students, EL students also face societal discrimination and unequal opportunity (Gándara & Hopkins, 2010; Lippin-Green, 1997).

Because we find suggestive evidence that EL status may influence teacher perceptions directly and indirectly via student outcomes, we come to mixed conclusions regarding the question of whether teacher perceptions *account for* negative effects of EL status on students' outcomes reported by previous studies (Carlson & Knowles, 2016; Umansky, 2016). While we find evidence that teachers have lower perceptions of EL students even after controlling for past and current student skill level, we also find suggestive evidence that teachers are accurately picking up on emerging differences in the skill levels of their EL and non-EL students over time (Jussim et al., 1996; Jussim & Harber, 2005; F. A. López, 2017). Our findings, therefore, paint a more complex and nuanced picture of how EL status may affect students' educational outcomes. Namely, lower teacher perceptions of EL compared with non-EL students appear to reflect both biases as well as real differences in academic trajectories that may be caused by unequal access to content (Estrada, 2014; Kanno & Kangas, 2014) and other mechanisms.

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 perceptions, 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 perceptions 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 are moderated by contextual features such as teacher-student racial congruence and the average socioeconomic 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: 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 specialized training and education that bilingual teachers receive toward working with EL students may lead to less biased perceptions of EL-classified students and/or instructional choices that do not impart an academic penalty on these students (Fránquiz et al., 2011; García & Guerra, 2004; F. A. López, 2017; Moll et al., 1992). In addition, teachers' linguistic skillsets may allow them to communicate with students and their families in fuller ways that offset bias and/or increase opportunity to learn (Loeb et

al., 2014; Matthews & López, 2019). 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 and/or more beneficial instructional choices. In reality, both sets of factors may be in effect, with both teacher selection into bilingual settings and teacher preparation and training minimizing effects of EL status on teacher perceptions. Future research should disentangle these possible mechanisms. Either way, however, this study adds to a robust body of work on the benefits of bilingual instruction for multilingual students (Callahan & Gándara, 2014; Fránquiz et al., 2011; Steele et al, 2017).

While matching is vulnerable to omitted variable bias, we believe the context of a natural experiment across locales, paired with a data set providing independent and directly measured multilingual student English proficiency level (along with a rich array of other variables) warrant causal interpretation of our results. However, if matched students classified as EL in kindergarten differ from those not classified as EL in ways that we cannot observe or control for but that are related to teacher perceptions, then our estimates may be biased. Future research should explore these questions using alternate quasi-experimental methods and data sets.

Related, a second limitation of this study is that it relies on the assumption that the ECLS-K:2011 basic English proficiency assessments accurately measure students' English proficiency levels. If these measures are invalid or if they are too coarse to meaningfully differentiate between students, then our causal inference may be uncertain. This said, the fact that matching on English proficiency scores resulted in a treatment and control group that did not differ on measured reading or math scores provides at least preliminary evidence of the validity of the ECLS-K:2011 English proficiency assessments.

Although these limitations need to be kept in mind, the results of this study have important implications for educators, education leaders, and policy-makers. Because our results lend support to our hypothesis that EL status can affect teacher perceptions through both biases based on the label, and through altered instructional choices that negatively affect EL students' opportunities to learn, policy and practice implications should address both possible causal mechanisms. For example, 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 (Polat et al., 2019). Similarly, instructional policies and practices that ensure that EL-classified students have equal access to content and instruction may avoid indirect effects of EL status on teacher perceptions that operate through students' affected learning trajectories. Our 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. Finally, 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 negative teacher perception effects, these mechanisms can hopefully be applied to nonbilingual 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 multilingual populations.

### ORCID iD

Ilana M. Umansky  <https://orcid.org/0000-0001-9907-834X>

### Notes

The authors extend thanks to the following individuals: Richard Murnane, Doug Ready, John Willett, Sean Reardon, David Liebowitz, Joe Cimpian, Lauren Lanahan, and our anonymous reviewers. We are also grateful to the Jacobs Foundation for their financial support of this work, and to the National Center for Education Statistics, for data collection and sharing.

<sup>1</sup>We use the term *true English proficiency level* with hesitation and caveats, recognizing that the concept of English proficiency is contested and socially and historically situated (Cumming, 2008; Kibler & Valdés, 2016).

<sup>2</sup>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 their English proficiency level and determine EL identification (ESSA, 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.

<sup>3</sup>NCES does not provide reliabilities for the two executive functioning measures.

<sup>4</sup>As we explain in the analytic strategy section, results from our second research question, pertaining to bilingual program enrollment, should not be interpreted causally.

<sup>5</sup>The one exception is the executive functioning assessment which we only binned into two groups. Binning into two groups removed the slight difference between the treatment and control groups in executive functioning skills present when this variable was omitted as a matching variable and allowed us to keep our sample size considerably larger than would have been possible if binning the variable by quintile.

### References

- Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33(1), 4–14. <https://doi.org/10.3102/0013189X033001004>
- Abedi, J. (2008). Classification system for English language learners: Issues and recommendations. *Educational Measurement*, 27(3), 17–31. <https://doi.org/10.1111/j.1745-3992.2008.00125.x>
- August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language Minority Children and Youth*. LEA.



- Auwarter, A. E., & Aruguete, M. S. (2008). Effects of student gender and socioeconomic status on teacher perceptions. *Journal of Educational Research, 101*(4), 242-246. <https://doi.org/10.3200/JOER.101.4.243-246>
- Baker, C. (2011). *Foundations of bilingual education and bilingualism* (Vol. 79). Multilingual Matters.
- Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., Russell, R., & 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*. [https://ies.ed.gov/ncee/wwc/Docs/practice-guide/english\\_learners\\_pg\\_040114.pdf](https://ies.ed.gov/ncee/wwc/Docs/practice-guide/english_learners_pg_040114.pdf)
- 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. <https://doi.org/10.2307/4126967>
- 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. <https://doi.org/10.1016/j.ssresearch.2014.10.003>
- 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. <https://doi.org/10.1080/10824660701261052>
- Byrnes, D. A., Kiger, G., & Manning, M. L. (1997). Teachers' attitudes about language diversity. *Teaching and Teacher Education, 13*(6), 637-644. [https://doi.org/10.1016/S0742-051X\(97\)80006-6](https://doi.org/10.1016/S0742-051X(97)80006-6)
- Callahan, R. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal, 42*(2), 305-328. <https://doi.org/10.3102/00028312042002305>
- Callahan, R., & Gándara, P. (2014). *Bilingual advantage: Language, literacy, and the labor market* (R. Callahan & P. Gándara Eds.). Multilingual Matters. <https://doi.org/10.21832/9781783092437>
- 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. <https://doi.org/10.1002/pam.21908>
- 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. <https://doi.org/10.3102/0091732X17690120>
- 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. <https://doi.org/10.3102/0002831216635796>
- Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management, 27*(4), 724-750. <https://doi.org/10.1002/pam.20375>
- Cumming, A. (2008). Assessing oral and literate abilities. In E. G. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2 ed., Vol. 7, pp. 3-18). Springer.
- 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. <https://doi.org/10.1080/1547688X.2011.551736>

- 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. <https://doi.org/10.1080/15348431.2017.1281809>
- Downey, D. B., & Pribesh, S. (2004). When race matters: Teachers' evaluations of students' classroom behavior. *Sociology of Education, 77*(4), 267–282. <https://doi.org/10.1177/003804070407700401>
- 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. <https://doi.org/10.1111/1475-6773.12090>
- Dusek, J. B., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology, 75*(3), 327–346. <https://doi.org/10.1037/0022-0663.75.3.327>
- Escamilla, K. (2006). Semilingualism applied to the literacy behaviors of Spanish-speaking emerging bilinguals: Bi-illiteracy or emerging biliteracy? *Teachers College Record, 108*(11), 2329–2353. <https://doi.org/10.1111/j.1467-9620.2006.00784.x>
- Estrada, P. (2014). English learner curricular streams in four middle schools: Triage in the trenches. *Urban Review, 46*(5), 535–573. <https://doi.org/10.1007/s11256-014-0276-7>
- 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. <https://doi.org/10.1111/1467-9620.00279>
- Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black-White test score gap. *Urban Education, 38*(4), 460–507. <https://doi.org/10.1177/0042085903038004006>
- 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 and Education, 14*(2), 113–132. <https://doi.org/10.1080/15348458.2015.1019787>
- Fox, L. (2015). Seeing potential: The effects of student–teacher demographic congruence on teacher expectations and recommendations. *AERA Open, 2*(1). <https://doi.org/10.1177/2332858415623758>
- Fránquiz, M. E., Salazar, M. d. C., & DeNicolo, C. P. (2011). Challenging majoritarian tales: Portraits of bilingual teachers deconstructing deficit views of bilingual learners. *Bilingual Research Journal, 34*(3), 279–300. <https://doi.org/10.1080/15235882.2011.625884>
- Gallo, S., Link, H., Allard, E., Wortham, S., & Mortimer, K. (2014). Conflicting ideologies of Mexican immigrant English across levels of schooling. *International Multilingual Research Journal, 8*(2), 124–140. <https://doi.org/10.1080/19313152.2013.825563>
- Gándara, P., & Hopkins, M. (Eds.). (2010). *Forbidden language: English learners and restrictive language policies*. 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), 27–46. <https://doi.org/10.3102/0091732X028001027>
- 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. <https://doi.org/10.14507/epaa.v11n36.2003>
- Garcia, E. B., Sulik, M. J., & Obradovic, J. (2019). Teachers' perceptions of students' executive functions: Disparities by gender, ethnicity, and ELL status. *Journal of Educational Psychology, 111*(5), 918–931. <https://doi.org/10.1037/edu0000308>

- García, S. B., & Guerra, P. L. (2004). Deconstructing deficit thinking: Working with educators to create more equitable learning environments. *Education and Urban Society, 36*(2), 150–168. <https://doi.org/10.1177/0013124503261322>
- 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. <https://doi.org/10.3102/0162373715611484>
- 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. <https://doi.org/10.1111/j.1744-7984.2006.00079.x>
- 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. <https://doi.org/10.1080/15235882.2010.502803>
- Harklau, L. (1999). The ESL learning environment in secondary school. In C. Faltis & P. Wolfe (Eds.), *So much to say: Adolescents, bilingualism, and ESL in the secondary school* (pp. 42–60). Teachers College Press.
- 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), 662–670. <https://doi.org/10.1037/a0014306>
- Hopkins, M. (2013). Building on our teaching assets: The unique pedagogical contributions of bilingual educators. *Bilingual Research Journal, 36*(3), 350–370. <https://doi.org/10.1080/15235882.2013.845116>
- Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis, 20*(1), 1–24. <https://doi.org/10.1093/pan/mpr013>
- Johnson, A. (2019). The effects of English learner classification on high school graduation and college attendance. *AERA Open, 5*(2). <https://doi.org/10.1177/2332858419850801>
- Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in Experimental Social Psychology, 28*, 281–388. Elsevier. [https://doi.org/10.1016/S0065-2601\(08\)60240-3](https://doi.org/10.1016/S0065-2601(08)60240-3)
- 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. [https://doi.org/10.1207/s15327957pspr0902\\_3](https://doi.org/10.1207/s15327957pspr0902_3)
- 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. <https://doi.org/10.3102/0002831214544716>
- Katz, S. R. (1999). Teaching in tensions: Latino immigrant youth, their teachers, and the structures of schooling. *Teachers College Record, 100*(4), 809–840. <https://doi.org/10.1111/0161-4681.00017>
- Kibler, A. K., & Valdés, G. (2016). Conceptualizing language learners: Socioinstitutional mechanisms and their consequences. *Modern Language Journal, 100*(Suppl. 1), 96–116. <https://doi.org/10.1111/modl.12310>
- Lau v. Nichols, No. 414 U.S. 563 (1974). <https://doi.org/10.2307/1550335>

- Lee, J., & Zhou, M. (2015). *The Asian American achievement paradox*. 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). Springer. [https://doi.org/10.1007/978-94-007-4276-5\\_25](https://doi.org/10.1007/978-94-007-4276-5_25)
- 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*. Council of Chief State School Officers.
- Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Routledge.
- Llosa, L. (2008). Building and supporting a validity argument for a standards-based classroom assessment of English proficiency based on teacher judgments. *Educational Measurement, 27*(3), 32–42. <https://doi.org/10.1111/j.1745-3992.2008.00126.x>
- Loeb, S., Soland, J., & Fox, L. (2014). Is a good teacher a good teacher for all? Comparing value-added of teachers with their English learners and non-English learners. *Educational Evaluation and Policy Analysis, 36*(4), 457–475. <https://doi.org/10.3102/0162373714527788>
- 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. <https://doi.org/10.1002/ets2.12090>
- López, F. A. (2017). Altering the trajectory of the self-fulfilling prophecy: Asset-based pedagogy and classroom dynamics. *Journal of Teacher Education, 68*(2), 193–212. <https://doi.org/10.1177/0022487116685751>
- López, N. (2003). *Hopeful girls, troubled boys: Race and gender disparity in urban education*. 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. <https://doi.org/10.1177/01461672982412005>
- 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. <https://doi.org/10.1002/trtr.1679>
- Martínez-Roldán, C. M., & Malavé, G. (2004). Language ideologies mediating literacy and identity in bilingual contexts. *Journal of Early Childhood Literacy, 4*(2), 155–180. <https://doi.org/10.1177/1468798404044514>
- Master, B., Loeb, S., Whitney, C., & Wyckoff, J. (2016). Different skills? Identifying differentially effective teachers of English language learners. *Elementary School Journal, 117*(2), 261–284. <https://doi.org/10.1086/688871>
- Matthews, J. S., & López, F. (2019). Speaking their language: The role of cultural content integration and heritage language for academic achievement among Latino children. *Contemporary Educational Psychology, 57*(April), 72–86. <https://doi.org/10.1016/j.cedpsych.2018.01.005>
- Mavrogordato, M., & White, R. S. (2017). Reclassification variation: How policy implementation guides the process of exiting students from English learner status. *Educational Evaluation and Policy Analysis, 39*(2), 281–310. <https://doi.org/10.3102/0162373716687075>
- McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology, 46*(3), 235–261. <https://doi.org/10.1016/j.jsp.2007.05.001>
- 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. <https://doi.org/10.3102/00028312038001073>
- 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(July), 48–60. <https://doi.org/10.1016/j.tate.2017.02.021>
- Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory Into Practice*, 31(2), 132–141. <https://doi.org/10.1080/00405849209543534>
- Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.
- Murphy, A. F., & Torff, B. (2019). Teachers' beliefs about rigor of curriculum for English language learners. *Educational Forum*, 83(1), 90–101. <https://doi.org/10.1080/00131725.2018.1505991>
- National Academies of Sciences, Engineering and Medicine. (Eds.). (2017). *Promoting the educational success of children and youth learning English: Promising futures*. National Academies Press.
- National Center for Education Statistics. (n.d.). *Early childhood longitudinal study spring 2011 kindergarten teacher questionnaire (child level)*. 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*. Yale University Press.
- Oates, G. (2003). Teacher-student racial congruence, teacher perceptions, and test performance. *Social Science Quarterly*, 84(3), 508–525. <https://doi.org/10.1111/1540-6237.8403002>
- Ochoa, G. L. (2013). *Academic profiling: Latinos, Asian Americans, and the achievement gap*. University of Minnesota Press. <https://doi.org/10.5749/minnesota/9780816687398.001.0001>
- Olsen, L. (1997). *Made in America: Immigrant students in our public schools*. 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. <https://doi.org/10.1525/aeq.1987.18.2.04x0667q>
- Parkes, J. (2008). Who chooses dual language education for their children and why. *International Journal of Bilingual Education and Bilingualism*, 11(6), 635–660. <https://doi.org/10.1080/13670050802149267>
- Pepinsky, T. B. (2018). A note on listwise deletion versus multiple imputation. *Political Analysis*, 26(4), 480–488. <https://doi.org/10.1017/pan.2018.18>
- Pettit, S. K. (2011). Teachers' beliefs about English language learners in the mainstream classroom: A review of the literature. *International Multilingual Research Journal*, 5(2), 123–147. <https://doi.org/10.1080/19313152.2011.594357>
- Polat, N., Mahalingappa, L., Hughes, E., & Karayigit, C. (2019). Change in preservice teacher beliefs about inclusion, responsibility, and culturally responsive pedagogy for English learners. *International Multilingual Research Journal*, 13(4), 222–238. <https://doi.org/10.1080/19313152.2019.1597607>
- 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). <https://doi.org/10.14507/epaa.v14n20.2006>

- 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. <https://doi.org/10.3102/0002831210374874>
- Reyes, M., & Hwang, N. (2019). Middle school language classification effects on high school achievement and behavioral outcomes. *Educational Policy*. Advance online publication. <https://doi.org/10.1177/0895904818823747>
- 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. <https://doi.org/10.3102/0162373711407912>
- 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. <https://doi.org/10.1002/pam.21882>
- 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. <https://doi.org/10.1080/01621459.1984.10478078>
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. Holt, Rinehart & Winston
- Rubie-Davies, C. M. (2010). Teacher expectations and perceptions of student attributes: Is there a relationship? *British Journal of Educational Psychology*, 80(Pt. 1), 121–135. <https://doi.org/10.1348/000709909X466334>
- Ruiz, R. (1984). Orientations in language planning. *NABE: The Journal for the National Association for Bilingual Education*, 8(2), 15–34. <https://doi.org/10.1080/08855072.1984.10668464>
- Salerno, A. S., Andrei, E., & Kibler, A. K. (2019). Teachers' misunderstandings about hybrid language use: Insights into teacher education. *TESOL Journal*, 10(3), e00455. <https://doi.org/10.1002/tesj.455>
- Saunders, W., & Marcelletti, D. (2012). The gap that can't go away: The catch-22 of reclassification in monitoring the progress of English learners. *Educational Evaluation and Policy Analysis*, 35(2), 139–156. <https://doi.org/10.3102/0162373712461849>
- 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. <https://doi.org/10.3102/0162373717737378>
- Sireci, S. G., & Faulkner-Bond, M. (2015). Promoting validity in the assessment of English learners. *Review of Research in Education*, 39(1), 215–252. <https://doi.org/10.3102/0091732X14557003>
- Solórzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English language learners. *Review of Educational Research*, 78(2), 260–329. <https://doi.org/10.3102/0034654308317845>
- Sorhagen, N. S. (2013). Early teacher expectations disproportionately affect poor children's high school performance. *Journal of Educational Psychology*, 105(2), 465–477. <https://doi.org/10.1037/a0031754>
- 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. <https://doi.org/10.3102/0002831216634463>
- Tach, L. M., & Farkas, G. (2006). Learning-related behaviors, cognitive skills, and ability grouping when schooling begins. *Social Science Research*, 35(4), 1048–1079. <https://doi.org/10.1016/j.ssresearch.2005.08.001>

*Teacher Perceptions of EL-Classified Students*

- Tenenbaum, 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. <https://doi.org/10.1037/0022-0663.99.2.253>
- 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*. National Center for Education Statistics, U.S. Department of Education.
- Umansky, I. M. (2016). To be or not to be EL: An examination of the impact of classifying students as English learners. *Educational Evaluation and Policy Analysis, 38*(4), 714–737. <https://doi.org/10.3102/0162373716664802>
- Valdés, G. (1997). Dual-language immersion programs: A cautionary note concerning the education of language-minority students. *Harvard Educational Review, 67*(3), 391–430. <https://doi.org/10.17763/haer.67.3.n5q175qp86120948>
- Valenzuela, A. (1999). *Subtractive schooling: US-Mexican youth and the politics of caring*. 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. <https://doi.org/10.3102/0002831209353594>
- 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.
- Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89–122. <https://doi.org/10.3102/00346543073001089>
- Whiteford, T. (2009). Is mathematics a universal language? *Teaching Children Mathematics, 16*(5), 276–283. <https://doi.org/10.5951/TCM.16.5.0276>
- Wiley, T. G., & Lukes, M. (1996). English-only and standard English ideologies in the U.S. *TESOL Quarterly, 30*(3), 511–535. <https://doi.org/10.2307/3587696>
- 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. <https://doi.org/10.3102/0002831208316200>
- Youngs, C. S., & Youngs Jr, G. A. (2001). Predictors of mainstream teachers' attitudes toward ESL students. *TESOL Quarterly, 35*(1), 97–12. <https://doi.org/10.2307/3587861>

Manuscript received May 7, 2019

Final revision received January 9, 2021

Accepted January 19, 2021