Emergent bilinguals are one of the fastest growing and one of the most underserved student populations in the United States (Capps et al., 2005; Genesee et al., 2005; National Education Association, 2008). Since 2000, the emergent bilingual (EB) population in American public schools has grown by 28% (National Center for Education Statistics, 2019) and now accounts for 9.6% of all public school students in the United States. Of all EBs in the United States, 78% are Hispanic, and the majority qualify for free or reduced lunch. Emergent bilinguals are “triply segregated”—more likely to attend schools with other EBs, minorities, and low-income students (Orfield, 2001; Orfield & Lee, 2006; Rios-Aguilar & Gándara, 2012). Results from the 2019 National Assessment of Educational Progress show that EB-non EB achievement gaps are as large as gaps for students who do and do not receive free or reduced-price lunch.

To improve the educational outcomes of EBs, federal court cases have required that programs for EBs be effective. *Castañeda v. Pickard* (1981) mandated that educational programs for EBs be based on sound educational theory, implemented effectively, and tested for program efficacy. As research has shown that textbooks influence the content and methods that teachers use in the classroom (Alajmi, 2009, 2012; Robitaille & Travers, 1992), English language development (ELD) textbooks can play an integral role in program efficacy. We primarily use the term “textbooks” in this paper to refer to the traditional paper volume that contains a comprehensive, year-long, subject-specific curriculum. We sometimes also refer to “curricula” or “curriculum materials,” which include both textbooks and also other resources teachers call on to implement their lessons. We draw these definitions from Polikoff et al. (2018). This article is the first of which we are aware to address the third mandate in the *Castañeda v. Pickard* [1981] ruling by measuring the relationship between ELD curriculum materials and EBs’ English proficiency.

The association between textbook quality and student achievement has important implications for local education agencies. As the costs of textbooks across publishers are quite similar, districts have the potential to increase student achievement at minimal marginal cost (Chingos & Whitehurst, 2012). Furthermore, efforts to improve the achievement of EBs by making better curriculum choices may face fewer challenges compared to other alternatives (e.g., increasing the number of bilingually certified teachers), which may be less feasible or politically palatable. Although a few studies have looked at the relationship between mathematics textbooks and student achievement (Agodini et al., 2010; Bhatt et al., 2013; Bhatt & Koedel,
have examined the effect of ELD textbooks on the English proficiency of EBs. This study seeks to fill this important gap in the research by answering two research questions:

(a) To what extent do English proficiency outcomes differ between schools purchase any ELD curriculum materials and schools that purchase state-adopted curriculum materials?

(b) To what extent do English proficiency outcomes differ between schools that use the two most popular ELD curriculum materials (Rigby and National Geographic).

Although our work is centered in one U.S. state—Texas—the findings have potentially broad implications for the education of EB students in the United States and abroad. We discuss these in the conclusion of the manuscript.

A note about terminology. Following Burke et al. (2016), we use the term emergent bilingual in lieu of the following: English language learner (ELL), English as a second language (ESL), and limited English proficient (LEP). Scholars have objected to the terms “limited English proficient” and “English language learner” (García, 2009) due to its deficit-framing of the linguistic abilities of emergent bilinguals. The terms also suggest that English proficiency is the only desired outcome for this population of students (Menken & Solorza, 2014). Referring to this group of students by an additive name that does not focus on its limitations is an important step toward the decentering of English monolingual norms. Similar to Burke et al. (2016), we use the term emergent bilingual with two exceptions. First, the terms ELL, ESL, and LEP may be used when referencing federal and state laws. Second, the terms ELL, ESL, and LEP may be used in the literature review if the authors of the studies use the terms.

**Background Information**

**Research on Textbook Efficacy**

The choice of textbooks can significantly affect student achievement. Five recent rigorous, empirical studies examine the relationship between textbooks and student academic outcomes. We briefly review these studies and their methods and findings here, though we note that none of the studies specifically focused on ELD materials as we do.

First, Agodini et al. (2010) use a randomized control trial across 12 districts to compare the effects of four elementary mathematics on student achievement in Grades 1 and 2. The authors find that math curricula correspond to large differences in achievement. For example, by the end of Grade 2, certain curricula were associated with an increase of 0.17 standard deviations in student test scores.

Second, three quasi-experimental studies use matching techniques in Indiana (Bhatt & Koedel, 2012), Florida (Bhatt et al., 2013), and California (Koedel et al., 2017) and find positive effects of one of the studied elementary mathematics textbooks in each state (on the order of .05 to .10 standard deviations per year, and larger for low-income students). The largest and most recent study (Blazar et al., 2019), in contrast, used a value-added methodology and found no impacts of elementary mathematics textbooks across six states.

Together, these findings imply that the choice of textbook, at least in elementary mathematics, may have direct impacts on student achievement that do not depend on measures of teacher implementation. In contrast to this evidence from mathematics, to our knowledge no studies have examined the effect of ELD textbooks on EBs’ English language proficiency. This study is the first to use quasi-experimental methods to assess the effectiveness of ELD instructional materials, an important issue given large achievement gaps between EB and non-EB students, the large numbers of EB students across the nation, and federal mandates to provide a quality education to these students.

**State Textbook Adoptions**

In many countries, there are national curricula and national textbooks, such that schools use only one of a few state-approved textbooks or other curriculum materials. The situation is different in the United States. In the United States, each state sets its own content standards (what are often referred to as “curriculum” in other nations) and its own rules with regard to textbooks and other curriculum materials. Some U.S. states conduct state-level textbook adoptions and some states allow for local control of textbook purchases (Zeringue et al., 2010). Texas is one of the two largest U.S. states, and it conducts a state adoption. However, the adoption is merely advisory, meaning that the state puts out a list of approved materials but districts can purchase materials on or off the list. Historically, publishers were thought to tailor their books to the needs and preferences of the Texas State Board of Education and concentrate their selling efforts in Texas (Bowler, 1978; Stein et al., 2001). However, Texas’ national influence on textbook adoption has likely waned since the adoption of Common Core State Standards (CCSS). While 42 states and the District of Columbia have adopted the CCSS, Texas has not.

**District Textbook Adoptions**

In all 50 U.S. states, district leaders are the ultimate decision-makers with respect to which textbooks and other curriculum materials they want to adopt for their particular district. Textbook adoptions in Texas districts are uniform adoptions at the district level, meaning that all schools in the district use the same materials (Texas Education Agency [TEA], 2014). One exception to this general rule in Texas is for charter schools, which we exclude from our analyses for this reason. Unlike in other subjects, Texas districts
are not mandated to purchase any ELD textbooks at all, and they can also purchase on or off the state-approved list (TEA, 2011). Given the great flexibility that districts have over ELD textbook adoption, it is important to understand how this autonomy has impacted the English proficiency of EBs.

Prior research has shown that district leaders consult a variety of sources to make textbook adoption decisions. Using data from over 150 interviews with curriculum decision-makers in eight states, Zeringue et al. (2010) found that district leaders had limited information on textbook efficacy and, therefore, tended to make adoption decisions based on other factors such as teacher acceptance and advice from neighboring districts. In addition, Zeringue et al. (2010) found that 80% of curriculum leaders sought out information about textbook use, effectiveness, and acceptance of the materials. This finding was supported by recent research in California, which found that school district leaders were eager to learn from available data and from the experiences of other districts when making mathematics textbook adoptions (Polikoff et al., 2020). Thus, if more research was available regarding textbook efficacy, curriculum leaders may use this information to make better choices.

**ELD Textbooks in Texas**

The 2011–2012 school year is the most recent ELD textbook adoption cycle in Texas. During this adoption cycle, the state released a list of seven state-approved ELD curricula. Of the seven, two are online programs and five are traditional textbooks. The textbooks vary along several dimensions including number of lessons on forms of English, opportunities to learn new vocabulary words, and the ratio of listening/speaking activities to reading/writing activities. Of the five traditional curricula, Rigby On Our Way to English and National Geographic Reach are the most popular, comprising 69% of adoptions in Texas (see Table 1).

With 515 schools across 48 districts adopting National Geographic Reach, it is the most frequently adopted ELD curricula and represents 44% of adoptions. The author of this book also authored the widely-used Sheltered Instruction Observation Protocol, which is designed to improve instruction for EBs. National Geographic Reach was first published in 2011. A student anthology in the National Geographic Reach series is usually divided into eight units. Each unit includes multiple opportunities to use language frames, share aloud with a partner, learn conversational and academic vocabulary, and practice language-learning strategies (e.g., predicting, rereading and retelling). The anthology includes a mix of fiction and nonfiction texts.

With 313 schools across 37 districts adopting Rigby On Our Way to English, it is the second most popular ELD curriculum in Texas. Rigby is a registered trademark of Houghton Mifflin Harcourt. On Our Way to English is marketed as a comprehensive ELD program that focuses on academic language, vocabulary development, content-based instruction, and oral language. The text is typically divided into eight units with each unit providing opportunities to engage with fiction and nonfiction texts. On Our Way to English was first published in 2011 and is available in a Texas-specific version and a national version. In contrast, a Texas-specific version of National Geographic Reach does not exist.

| Type of ELD Textbook Adoption       | Number of schools | Schools |
|------------------------------------|-------------------|---------|
| State-adopted Texts               |                   |         |
| National Geographic Reach         | 551               | 44      |
| Rigby On Our Way to English       | 310               | 25      |
| McGraw-Hill Texas Treasure Chest  | 83                | 7       |
| Pearson Longman Cornerstone       | 20                | 2       |
| Santillana Spotlight on English   | 0                 | 0       |
| Achieve3000 (online)              | 0                 | 0       |
| ESL Reading Smart                 | 0                 | 0       |
| Non-State-Adopted Book            | 16                | 1       |
| No ELD textbook adoption          | 253               | 20      |

Note. ESL = English as a second language; ELD = English language development.

**Data**

We describe our quasi-experimental design below, which was approved by the institutional review board of our university. The data for this study come from the state of Texas. Texas serves roughly 5,400,000 students across approximately 1,250 public school districts. The racial composition of students in Texas’ public schools is as follows: 53% Hispanic, 27% White, 13% Black, 4% Asian, and 2% multi-racial (National Center for Education Statistics, 2019). Roughly 17% of students in Texas’ public schools are identified as English language learners.

Texas is an ideal state for the study of ELD curriculum materials on EBs’ English proficiency for several reasons. First, Texas is home to approximately 900,000 EBs, the
second highest number of EBs in the United States after California. Second, the textbook adoption process is centralized in Texas and the state tracks textbook adoption data, which allows for a more manageable analysis. Third, Texas has a large number of school districts, which allows for adequate statistical power for student achievement analyses. Fourth, as Texas grants great autonomy to districts over ELD textbook adoption, there is variation in the kinds of ELD instructional materials that are adopted. In the years under study, districts in Texas report using 7 different state-approved ELD curricula and 21 different non-state-approved ELD curricula. There are no other U.S. states that meet all these criteria, so it is unlikely that this research could have been carried out in any other state.

Textbook information was collected from districts using the TEA (TEA) online data portal. The TEA collects information, which is publicly available through Freedom of Information Law request, on instructional materials allotment (IMAT). We obtained ELD textbook purchases for the years 2011–2015. The following textbook information is available for each district: publisher, textbook title, subject, grades used, unit price, quantity of units purchased, and adoption year. We found 418 districts on the IMAT list for years 2011–2015, which left 810 districts with missing information on ELD textbook adoption.

As the IMAT report only captures purchases made using state funds, it does not capture instructional materials that were purchased using local funds nor instructional materials that were purchased prior to 2011. To obtain information on ELD instructional materials for the remaining 810 districts, we submitted Public Information Requests to each district with missing ELD textbook information and received responses from 737 districts. Overall, we obtained data on ELD textbooks from 94% of Texas school districts.

**Analytic Sample**

We focus our analysis on elementary schools in Texas that serve EBs. As the study focuses on elementary ELD books, we drop middle and high schools from the sample. In addition, elementary schools that do not have any EBs ($n = 82$ schools) are excluded from the study as they do not have a need for ELD curricular materials.

Table 2 provides details about the construction of the analytic sample starting with the universe of 3,085 elementary schools in Texas that are eligible for the study. After four exclusion criteria were applied following the criteria used by Koedel et al. (2017), 1,253 elementary schools (41%) remain in the sample. First, schools that adopted multiple ELD curricula in grades K-3 were dropped from the sample. Roughly 3% of schools were excluded for this reason. Although it is possible to examine mixed-treatment effects, there are too few schools for an effective analysis of this question.

Second, elementary schools that did not report English language proficiency scores for spring 2015 were eliminated from the study. There are two scenarios in which the state would not report Texas English Language Proficiency Assessment System (TELPAS) scores: (a) when the number of students tested is so small that it may be possible to identify particular students or 2.) when all the students in a group have the same result (TEA, 2015b). Because the number of EB students in particular schools is often quite small, 48% of Texas elementary schools did not report TELPAS scores for spring 2015 and were, thus, dropped from the study.

Third, less than 1% of schools were eliminated from the study due to missing information regarding ELD adoption. Specifically, 21 schools or 0.7% of the sample were dropped due to missing ELD textbook information. Although public information requests were submitted to all districts in Texas with missing ELD textbook information, three districts out of the remaining districts in the study did not respond. These three districts represent 21 schools. Given these data restrictions, the results from this study are only generalizable to elementary schools in Texas that adopted the same ELD curriculum in grades K-3, reported TELPAS scores in spring 2015, and reported ELD textbook adoption.

Fourth, Houston Independent School District (HISD) and Dallas Independent School District (DISD) were dropped because both districts are much larger than all other districts in the state which made it difficult to find comparable

| Reason for Exclusion                                      | Schools | Total (%) | Districts | Total (%) |
|-----------------------------------------------------------|---------|-----------|-----------|-----------|
| Initial Universe                                          | 3,085   |           | 651       |           |
| Non-uniform adopted, grades k-3                           | -95     | 3.1       | 47        | 7.2       |
| Missing outcome variable (TELPAS 2015)                    | -1486   | 48.2      | 410       | 63.0      |
| Missing textbook information                              | -21     | 0.7       | 3         | 0.5       |
| HISP or DISD                                              | -230    | 7.5       | 2         | 0.3       |
| Missing covariates                                        | 0       | 0         | 0         | 0         |
| Final Analytic Sample                                     | 1,253   | 40.6      | 189       | 29.3      |

Note. TELPAS = Texas English Language Proficiency Assessment System; HISD = Houston Independent School District; DISD = Dallas Independent School District.
matches in our propensity score models. After dropping HISD and DISD from the sample, 1,253 schools and 191 districts remain. The final analytic sample retains 41% of the universe of schools and 29% of the universe of districts.

**ELD Textbook Adoption Patterns**

In the final analytic sample, we find that the majority of elementary schools in Texas that serve EBs purchase state-adopted ELD materials (see Table 1). While roughly 1% of the final analytic sample purchased materials that were not on the state-adopted list, 79% of schools purchased materials on the state-adopted list. Approximately 20% of schools in the final sample did not purchase any ELD materials even though they did have sufficient numbers of EB students in their school for their TELPAS scores to be reported. Of the ELD materials on the state-adopted list, *National Geographic Reach* and *Rigby On Our Way to English* were the most popular with 69% of schools in the final analytic sample purchasing one of these two materials.

**School-Level Variables**

Although ELD textbook adoption typically occurs at the district level, we match at the school level for two reasons. First, matching is a “data hungry” process that requires a large sample size (Heckman et al., 1998). Second, schools within the same district can vary greatly in their demographics, test scores, teacher experience, and so on. Conducting the analysis at the school level allows us to adjust for both school- and district-level characteristics (Bhatt et al., 2013; Bhatt & Koedel, 2012).

School-level data are drawn from the TEA online data portal. Information about students and teachers are available at the school level. With regard to teachers, the following information is available: experience, gender, race, and credential type. In terms of students, the following information is available at the school level: race, gender, free/reduced price lunch eligibility, special education eligibility, English language learner status, district average math and reading score on the State of Texas Assessments of Academic Readiness (STAAR), and district average English language proficiency scores on the TELPAS (Texas English Language Proficiency Assessment System). Since the last year of ELD textbook adoption in Texas was the 2011–2012 school year, we use achievement and demographic data from 2010-2011 to match schools.

**Outcome Variables**

School-average composite scores on the spring 2015 TELPAS assessment serve as the outcome variable. The TELPAS assessment is delivered on a yearly basis to monitor EBs’ progress in attaining English proficiency. As there are four language modalities—listening, speaking, writing, and reading—the TELPAS exam is divided into four sections. Students are given a rating of 1–4 on each subtest where 1 represents Beginning, 2 represents Intermediate, 3 is Advanced, and 4 represents Advanced High (TEA, 2016). To arrive at the composite score, TEA weights each section as follows: listening (0.10), speaking (0.10), reading (0.50), and writing (0.30). Using student-level TELPAS scores, we create school-average composite TELPAS scores which are approximately normally distributed (see Figure 1).

As the STAAR reading assessment may appear to be very similar to the TELPAS exam, we briefly clarify the differences between the two assessments and justify using TELPAS scores as the outcome variable. While the reading portion of the STAAR assesses students’ mastery of content standards for English language arts, the TELPAS assessment focuses on language standards—not content standards. The reading portion of the STAAR assessment only measures reading abilities and does not include a listening and speaking portion. As the TELPAS exam covers all four language modalities (listening, speaking, reading, and writing), the TELPAS is better positioned to assess English proficiency. Increased English proficiency may lead to greater access to content standards; therefore, it is possible that increases in TELPAS scores lead to increases in STAAR reading scores. However, as TELPAS scores are a more direct measure of English proficiency than STAAR reading scores, we use TELPAS scores as the outcome variable.

**Research Design**

To model the association between ELD curriculum materials and EBs’ English proficiency, we use local linear matching. Bhatt and Koedel (2012) used a similar matching approach to evaluate relative curriculum effectiveness in the state of Indiana. As districts can self-select into the treatment (i.e., the ELD textbook), selection bias may affect our coefficient...
estimates. Matching on school characteristics reduces group differences across all the observed variables used in the propensity score equation (Shadish et al., 2002). Matching can allow for causal inferences if the conditional independence assumption is met (Heckman et al., 1997; Rosenbaum & Rubin, 1983).

The conditional independence assumption requires that potential outcomes be independent of textbook choice conditional on the observables. The conditional independence assumption will not be satisfied if unobserved characteristics influence both the treatment (choice of textbook) and the outcome (TELPAS scores). By using an estimated propensity score, we can match schools based on their observable characteristics to minimize the likelihood that unobserved characteristics may influence the outcome. Matching has two main advantages over a simple regression analysis. First, matching has fewer restrictions on functional form (Black & Smith, 2004). Second, matching can more accurately predict the relationship between covariates and the outcome variable by restricting the influence of noncomparable treatment and control units (Black & Smith, 2004).

Of schools that adopt ELD materials in the final analytic sample, 92% adopted new ELD curricula in 2011–2012. Thus, third graders in 2014–2015 will have potentially been exposed to the ELD curricula for 4 years. If an ELD curriculum is effective, then districts using this curriculum should have higher TELPAS scores for their EB students compared to districts that did not adopt this particular curriculum.

Following Bhatt and Koedel (2012), we use the term “total treatment effects” to account for the possibility of the estimates capturing systematic differences between treatment and control schools. For example, if one curriculum has a teacher’s guide that is easier to navigate, the estimates will reflect this difference. Based on the series of falsification tests we conducted for the study, we find little evidence to suggest that there are systematic differences between treatment and control schools. However, it should be noted that the estimates from this study will capture “total treatment effects.” We cannot measure teachers’ implementation of the adopted curriculum materials to estimate the impact of teacher use on student outcomes. No data on teacher implementation of curriculum materials are collected by the state or any other agency.

**Propensity Score Estimation**

In order for propensity score matching to yield unbiased estimates of the effectiveness of ELD textbooks, two types of covariates should be included: (a) pretest measures of the outcome variable and (b) direct measures of the selection process (Cook & Steiner, 2010; Glazerman et al., 2003; Steiner & Cook, 2013). The first type of covariate is easily obtainable through TEA. We include school average TELPAS composite scores from spring 2011 (pre-adoptions) in the selection model. The higher the correlation between pretest and posttest, the greater the bias reduction (Steiner & Cook, 2013). The correlation between school-average TELPAS scores in 2011 and school-average TELPAS scores in 2015 is 0.43. The second type of covariate is slightly harder to obtain. Steiner and Cook (2013) suggest investigating the actual selection process before starting the study because the omission of important variables can increase bias in the estimates (Heckman et al., 1997). Shadish and Steiner (2010) recommend interviewing relevant individuals to identify factors that influence selection into treatment and studying the pertinent research literature. In addition, economic theory and previous empirical findings should guide the choice of variables to include in the propensity score estimation (Caliendo & Kopeinig, 2008).

Heckman et al. (1997) find that selection models that include variables related to treatment assignment are better able to predict treatment assignment than models that only include demographic characteristics like education and race. The covariates that are used to estimate the propensity score should include all of the key factors affecting both the outcome variable and the treatment assignment (Smith & Todd, 2005). When creating a list of covariates for the propensity score estimation, there is a tradeoff between bias and variance. Including an irrelevant variable can lead to increased variance, while excluding a potentially important confounder can lead to increased bias (Heckman et al., 1997; Stuart, 2010).

We investigate the selection mechanism using two methods. First, we conduct in-depth interviews with 28 district leaders in Texas, which we describe in the following paragraphs. Second, we review previous research on textbook effectiveness to identify potential covariates (Rubin, 2001).

**District Leader Interviews**

To understand the factors influencing ELD textbook adoption, we conducted 28 in-depth interviews with district leaders using structured open-ended interviewing protocols. We focus on district leaders because textbook adoption typically occurs at the district level in the state of Texas. To ensure the representativeness of the interview findings, the districts selected for interviews represented variation along two dimensions: percentage of EBs (e.g., low and high) and ELD textbook adoption patterns (four possible patterns)—yielding eight total cells. Using proportional stratified sampling, we selected two to eight districts from each cell, giving a total of 28 districts. Interviews lasted approximately 45 min and were audio recorded and transcribed. Interview data were analyzed using NVivo software.

In the initial cycle of interview data review, which required at least three readings of the full transcript, we used structural codes, which entail little interpretation. Structural coding “acts as a labeling and indexing device, allowing researchers to quickly access data likely to be relevant to a particular analysis from a larger data set” (Namey et al.,
Although interviews covered a range of topics, including program model for EBs, we subjected the entire transcript to this initial coding, as district leaders spoke about instructional materials in response to questions not directly about instructional materials. The first cycle of coding revealed 27 structural codes related to ELD textbook adoption. These structural codes were further categorized and analyzed in the second stage of review.

In the second cycle of coding, the primary goal is to create categorical, thematic, and conceptual organization from the array of codes that emerged from first cycle coding (Saldaña, 2013). We used pattern coding in the later stages of data analysis to group codes into a smaller number of themes. Five main factors emerge from the interview data as shaping district leaders’ decisions around ELD textbook adoption: number of EBs, academic achievement of EBs, district leaders’ beliefs about second language acquisition, district budget, and program model for EBs. We briefly discuss each of these factors in the paragraphs that follow.

One factor that district leaders take into consideration when deciding which ELD curriculum to adopt is the number of EBs in their district. Six out of 28 district leaders cited EB enrollment as a factor for ELD textbook adoption. For example, a curriculum facilitator in district 15 states that “the number of students that we have identified [as ESL]” is a major determinant in whether or not ELD materials are purchased. This district leader elaborates,

If there’s a high number of students then, of course, the need [for ELD instructional materials] would be greater. But if there is a low number of students then our expert teachers in the bilingual department will be able to modify their instruction to meet the needs of their ESL or ELL students.

For this particular district, bilingual teachers are expected to meet the linguistic needs of EBs in the event that ELD materials are not purchased. To model the treatment assignment, we include the number of EBs—at both the school and district level—in the propensity score.

Second, district leaders consider EBs’ academic achievement when making curricular decisions. Eight out of 28 district leaders report EBs’ test scores (e.g., scores on the STAAR test, scores on language proficiency tests) as a factor that influences ELD textbook adoption. District Leader 18 shares,

We’re a pretty successful district when you look at state scores. What we have found is that a lot of the times the basalas don’t get you where you need to get. We pull from a lot of different resources.

As this district leader seems to be satisfied by her district’s test scores, she does not see a need to purchase ELD materials. Instead, she believes that the district’s current curriculum, which seems to be a mixture of various resources, is sufficient to meet the needs of the students. Furthermore, District Leader 12 looks at “proficiency projections as well as grade-level means” to determine whether or not ELD materials need to be purchased. We include pre-treatment school and district average reading achievement on the Texas Assessment of Knowledge and Skills (TAKS) and TELPAS in the estimation of the propensity score.

A third factor that influences ELD materials purchasing is district leaders’ beliefs about second language acquisition. Thirteen out of 28 district leaders use their beliefs about second language acquisition to make decisions about ELD materials. Some leaders believe that second language learning is not much different from first language learning; thus, EBs do not need additional materials. For instance, District Leader 10 states, “We don’t have a particular program that we offer or provide for [EBs] . . . We just rely on good instruction.” This district did not purchase any ELD curricular materials because school administration believes that the instruction that all students receive should be effective for EBs as well. As it is difficult to directly assess district leaders’ beliefs about second language acquisition for every district in Texas, we adjust for the following Census characteristics at the district level: median income and the percentage of non-English speakers in the district. As beliefs about language learning may vary based on professional status (Horwitz, 1999), we include median income to proxy for professional status. As classroom and social peers influence learning (Arnold, 1999), we include the percentage of non-English speakers in the district as a means of capturing the variation in beliefs about second language acquisition.

Fourth, cost is another factor that districts weigh in their decision to purchase ELD texts. Sixteen out of 28 district leaders report that cost is an important consideration in ELD materials purchasing. An elementary school principal from District 21 states, “We are a small, rural district. So we also have to think about using our money wisely too . . . the main thing we have to make sure of is that it’s something we can afford. Price would be the big decision-maker.” Along the same lines, District Leader 15 states that “the budget is always a factor.” We include district per pupil revenue along with district instructional expenditures to capture each district’s financial ability to purchase ELD materials.

Fifth, a district’s program model for EBs emerges as another factor influencing ELD materials purchasing. Eight out of 28 district leaders state that the district’s program model for EBs impacts whether or not ELD materials are adopted. Districts that have bilingual programs are less likely to purchase ELD texts because they are already purchasing texts in other languages, usually Spanish, for their EBs which reduces the amount of money that is available for ELD texts. In addition, districts that have bilingual programs also have bilingual teachers and some districts use bilingual teachers’ knowledge of second language acquisition to supplant ELD instructional materials. In the propensity score estimation, we include a dichotomous variable indicating the presence of a bilingual program.
In addition to district leader interviews, prior research on textbook adoption can help to model the selection mechanism (Caliendo & Kopeinig, 2008; Heckman et al., 1997). For instance, Bhatt and Koedel (2012), in modeling the effect of mathematics textbooks on achievement, use the following covariates in their propensity score model: enrollment, attendance rates, demographics (race, language status, free lunch status), achievement outcomes from prior years, per pupil expenditures, and Census measures (e.g., median household income, share of adults with a high school diploma). Table 3 contains the final list of covariates used to estimate the propensity score.

| Covariate                             | Mean for treated | Mean for control | Standardized difference | t-test | p value | Variance(T)/Variance(C) |
|---------------------------------------|-----------------|-----------------|--------------------------|--------|---------|-------------------------|
| School characteristics                |                 |                 |                          |        |         |                          |
| Enrollment                           | 619.59          | 632.34          | −0.072                   | −0.85  | .397    | 0.77                     |
| Share ELL                            | 0.41            | 0.41            | −0.017                   | −0.2   | .841    | 1.49                     |
| Share Hispanic                       | 0.68            | 0.72            | −0.172                   | −1.96  | .05     | 1.15                     |
| Share SPED                           | 0.07            | 0.07            | 0.09                     | 0.96   | .335    | 1.04                     |
| Average Teacher Years of Experience  | 10.69           | 10.99           | −0.112                   | −1.13  | .257    | 0.86                     |
| Number of Bilingual Teachers         | 24.24           | 28.67           | −0.255                   | −2.5   | .013*   | 0.78                     |
| Bilingual Program Indicator          | 0.9             | 0.94            | −0.107                   | −1.47  | .142    | —                       |
| School Outcomes                      |                 |                 |                          |        |         |                          |
| School Average Math Score            | 590.62          | 592.07          | −0.05                    | −0.56  | .578    | 1.08                     |
| School Average ELL Math Score        | 586.74          | 587.76          | −0.028                   | −0.3   | .764    | 1.02                     |
| School Average ELA Score             | 601.15          | 599.41          | 0.054                    | 0.61   | .539    | 1.26                     |
| School Average ELL ELA Score         | 587.26          | 586.19          | 0.025                    | 0.26   | .793    | 0.87                     |
| District Characteristics             |                 |                 |                          |        |         |                          |
| District Enrollment                  | 31,360.03       | 29,779.48       | 0.039                    | 0.69   | .488    | 0.74                     |
| Share ELL                            | 22.44           | 21.03           | 0.112                    | 1.26   | .21     | 1.76                     |
| Per Pupil Expenditure                | 10,924.73       | 11,211.18       | −0.118                   | −1.33  | .183    | 3.77                     |
| Total Revenue Per Pupil              | 1,348.79        | 1399.63         | −0.093                   | −0.92  | .36     | 1.4                      |
| Urbanicity                           | 4.08            | 3.93            | 0.051                    | 0.5    | .618    | 0.58                     |
| District Outcomes                    |                 |                 |                          |        |         |                          |
| District Avg TELPAS Score            | 3.08            | 3.08            | −0.004                   | −0.04  | .966    | 0.93                     |
| District Average Math Score          | 596.02          | 595.68          | 0.018                    | 0.2    | .844    | 1.03                     |
| District Avg Math Score for ELLs     | 587             | 587.13          | −0.006                   | −0.07  | .945    | 1.61                     |
| District Average ELA Score           | 609.87          | 608.22          | 0.077                    | 0.87   | .384    | 1.07                     |
| District Avg ELA Score for ELLs      | 587.98          | 585.54          | 0.104                    | 1.12   | .262    | 1.46                     |
| District-Area Characteristics (Census)|               |                 |                          |        |         |                          |
| Median Household Income              | 47,976.89       | 46,785.11       | 0.08                     | 0.82   | .413    | 1.09                     |
| Share Non-English Speakers           | 42.85           | 43.21           | −0.014                   | −0.15  | .879    | 1.13                     |
| Propensity score                     | 0.34            | 0.34            | 0                        | 0.00   | .997    | 1                        |

Note. ELD = English language development; ELL = English language learner; TELPAS = Texas English Language Proficiency Assessment System.

*p < .05.

In addition to district leader interviews, prior research on textbook adoption can help to model the selection mechanism (Caliendo & Kopeinig, 2008; Heckman et al., 1997). For instance, Bhatt and Koedel (2012), in modeling the effect of mathematics textbooks on achievement, use the following covariates in their propensity score model: enrollment, attendance rates, demographics (race, language status, free lunch status), achievement outcomes from prior years, per pupil expenditures, and Census measures (e.g., median household income, share of adults with a high school diploma). Table 3 contains the final list of covariates used to estimate the propensity score.

**Balancing Tests**

We test for covariate balance using three different tests. Multiple tests are used because there is no clearly preferred test for balance, and different tests may yield different results (Smith & Todd, 2005). As there are two research questions, the propensity score must be calculated separately. While the first propensity score measures the probability of not choosing an ELD text, the second propensity score measure the likelihood of adopting Rigby *On Our Way to English*. For this reason, covariate balancing is assessed twice (Tables 3 and 4).

Following Rubin (2001), we calculate the standardized difference in the mean propensity score in the two groups. To minimize bias, the standardized difference should be zero or close to zero. Although there are no formal criteria for evaluating standardized differences, Rosenbaum and Rubin (2001) suggest that values greater than 0.20 should be interpreted as large. The standardized difference for the propensity score is 0.001 for both research questions.

Second, we calculate the variance ratios of the propensity score distribution in the treatment and control groups (Rubin, 2001). To minimize bias, the variance ratios should ideally lie between 0.8 and 1.25. Variance ratios smaller than 0.5 or greater than 2.0 are considered extreme (Rubin, 2001).
The variance ratio of the propensity scores are 1.00 for both research questions.

We extend the first two balancing tests to include individual covariates. Although Rubin (2001) recommends balancing on propensity score alone, we also calculate the standardized difference and variance ratio for each covariate in the propensity score model. By doing so, we can gain a more robust understanding of the comparability of the control and treatment groups. For the first research question, we find that the variance ratio for all covariates is in the range of acceptability [0.5, 2.0] except for district per pupil expenditure which has a variance ratio of 3.77. State-adopters have a wider range of per pupil expenditures than nonadopters. For the second research question, the variance ratio for share of EBs is 0.45 which is slightly outside the range of acceptability. Schools that adopt National Geographic Reach have a larger range of percentage of EBs compared to school that adopt Rigby On Our Way to English. The absolute value of the standardized difference is less than 0.20 for all covariates except for the number of bilingual teachers which has a standardized difference of −0.26. For the first research question, nonadopters, on average, have 29 bilingual teachers while state-adopters average 24 bilingual teachers per campus.

Third, we present results from hypothesis tests. Although hypothesis tests and \( p \) values should not be used as measures of balance (Ho et al., 2007; Stuart, 2010), they are still common in balancing tests. Results from hypothesis testing may be misleading as small differences in means may be statistically significant in a large sample. Although results from hypothesis tests are not given much consideration, we present them to create a more robust understanding of covariate balance. For the first research question, treatment and control schools differ significantly on one variable—number of bilingual teachers. For the second research question, treatment and control schools differ significantly on one variable—number of bilingual teachers. For the second research question, treatment and control schools differ significantly on one variable—number of bilingual teachers.

### Table 4. Covariate-by-Covariate Balancing Details for the Comparison Between Rigby On Our Way to English and National Geographic Reach.

| Covariate                          | Mean for treated | Mean for control | Standardized difference | t-test | \( p \) value | Variance(T)/Variance(C) |
|------------------------------------|------------------|------------------|-------------------------|--------|--------------|-------------------------|
| **School Characteristics**         |                  |                  |                         |        |              |                         |
| Enrollment                         | 670.99           | 695.17           | -0.126                  | -1.53  | .127         | 1.07                    |
| Share ELL                          | 39.2             | 38.01            | 0.067                   | 0.91   | .365         | 1.16                    |
| Share Hispanic                     | 0.71             | 0.69             | 0.103                   | 1.21   | .225         | 0.86                    |
| Share SPED                         | 7.4              | 7.78             | -0.015                  | -1.94  | .053         | 1.45                    |
| Average Teacher Years of Experience| 10.92            | 11.09            | -0.066                  | -0.77  | .439         | 0.93                    |
| Number of Bilingual Teachers       | 21.85            | 18.61            | 0.188                   | 2.14   | .033*        | 1.1                     |
| Bilingual Program Indicator        | 0.88             | 0.91             | -0.097                  | -1.26  | .209         | —                       |
| **School Outcomes**                |                  |                  |                         |        |              |                         |
| School Average Math Score          | 585.29           | 585.61           | -0.011                  | -0.13  | .895         | 0.84                    |
| School Average ELL Math score      | 583.44           | 580.96           | 0.068                   | 0.84   | .404         | 0.84                    |
| School Average ELA Score           | 594.97           | 596.91           | -0.06                   | -0.76  | .446         | 1.08                    |
| School Average ELL ELA Score       | 583.58           | 583.74           | -0.004                  | -0.05  | .962         | 0.93                    |
| **District Characteristics**       |                  |                  |                         |        |              |                         |
| District Enrollment                | 37,628.28        | 36,155.02        | 0.032                   | 0.81   | .418         | 0.97                    |
| Share ELL                          | 20.75            | 19.81            | 0.098                   | 1.25   | .213         | 0.45                    |
| Per Pupil Expenditure              | 11,195.11        | 11,306.95        | -0.067                  | -0.69  | .49          | 0.86                    |
| Total Revenue Per Pupil            | 1,305.36         | 1,222.28         | 0.167                   | 1.84   | .067         | 1.03                    |
| Urbanicity                         | 2.46             | 2.68             | -0.083                  | -0.91  | .361         | 1.05                    |
| **District Outcomes**              |                  |                  |                         |        |              |                         |
| District Avg TELPAS Score          | 3.15             | 3.15             | -0.02                   | -0.24  | .81          | 0.9                     |
| District Average Math Score        | 592.09           | 593.71           | -0.008                  | -1.08  | .282         | 0.8                     |
| District Average Math Score for ELLs| 582.21           | 584.67           | -0.081                  | -0.95  | .344         | 0.73                    |
| District Average ELA Score         | 605.4            | 606.88           | -0.074                  | -0.91  | .362         | 0.95                    |
| District Avg ELA Score for ELLs    | 584.19           | 585.35           | -0.058                  | -0.71  | .48          | 0.87                    |
| **District-Area Characteristics (Census)** |          |                  |                         |        |              |                         |
| Median Household Income            | 47,996.23        | 50,695.91        | -0.189                  | -2     | .046*        | 0.81                    |
| Share Non-English Speakers         | 47.95            | 46.59            | 0.06                    | 0.67   | .505         | 0.97                    |
| **Propensity score**               | 0.52             | 0.52             | 0.001                   | 0.01   | .99          | 1                       |

*Note:* ELL = English language learner; TELPAS = Texas English Language Proficiency Assessment System.

*\( p \) < .05.
significantly on two variables—number of bilingual teachers and median household income.

Overall, these tests suggest that the individual covariates and the propensity scores are well balanced between treated and control units. The standardized difference for the propensity score is 0.001 the variance ratio is 1 for both balancing tests.

Matching Algorithm

Matching algorithms assume that the outcome variable is independent of treatment assignment conditional on a set of observable characteristics (Smith & Todd, 2005). In addition, it is assumed that for all conditioning variables, there is a positive probability of assignment to treatment or control conditions. This assumption implies that a match can be found for all treated cases. Matching estimators perform well in replicating the results of a randomized control trial when the following criteria are met: (a) the same data sources are used for both control and treated units, (b) control and treated units reside in the same labor market, and (c) the data contain a rich set of covariates that impact both treatment assignment and the outcome variable (Heckman et al., 1998, 1997; Smith & Todd, 2005). There is strong evidence suggesting that all three conditions are met in this study. First, covariates and outcome variables are the same for both treatment and control schools and is also measured in the same way for both groups. Second, all treatment and control schools reside in the same geographic area—Texas. Third, conditioning variables adjust for a wide array of characteristics—reading and math achievement, English proficiency, teacher quality, demographic characteristics, and median household income. As these foundational requirements have been met, it is appropriate to proceed with matching.

We choose local linear matching (LLR) over other matching methods for three main reasons. First, local linear matching can reduce variance relative to other matching techniques because it relies on more information. While common matching techniques like nearest neighbor matching only use a few observations from the comparison group to construct the counterfactual, local linear matching uses a weighted average of comparison units near the treatment unit. As more information is used, there is lower variance which leads to more precise estimates of the treatment effect. Second, local linear matching includes a linear term in the propensity while kernel matching does not. This added intercept is an advantage whenever there are gaps in the propensity score. Finally, local linear matching relies on weaker assumptions compared to other matching approaches. Rosenbaum and Rubin’s (1983) framework for propensity score matching requires strongly ignorable treatment assignment and common support; kernel-based matching methods such as local linear matching replace these two assumptions with weaker ones (Heckman et al., 1998). Instead of assuming strongly ignorable treatment assignment, Heckman and colleagues only require that the outcome under the control condition for the treated cases to be independent of the treatment assignment, conditional on observed covariates (Guo & Fraser, 2010). Instead of assuming full independence, Heckman and colleagues impose mean independence. Although local linear matching is our method of choice, we also include results from kernel matching, inverse probability weighting with regression adjustment (IPW-RA), and nearest neighbor matching with 1, 2, and 3 neighbors (which will be abbreviated K1, K2, and K3, respectively).

Following Dehejia and Wahba (2002), we trim the sample as needed to determine the support region. There are no control units below the minimum propensity score for treated units for either research question. We also assess for common support by visually inspecting a probability density distribution of the propensity score in both groups. Lechner (2002) argues that common support can be assessed visually, and there is no need to implement more complicated, formal processes. Estimated density functions of the propensity scores for treatment and control schools (available upon request) show strong evidence of common support.

Results

Research Question 1: Nonadopters vs. State-Adopters

Using local linear matching, we compare the effect of not purchasing an ELD text to the effect of purchasing a state-adopted ELD text. We find that elementary schools that do not purchase any elementary ELD texts score 0.30 standard deviations lower on school-average TELPAS composite scores (see Table 5). Following Bhatt et al. (2013), standard errors are estimated by bootstrapping using 250 repetitions and clustered at the district level.

To address the robustness of the findings, we compare results across all matching models. The results from the kernel matching estimator are similar to the estimates from the local linear matching model with nonadopters scoring 0.31 standard deviations lower than state-adopters. Results from the IPWRA and nearest neighbor matching models are similar in magnitude with nonadopters scoring 0.28–0.35 standard deviations lower than state-adopters. The standard errors in these models are roughly half the size of the standard errors in the local linear matching model. One reason for the large standard errors for the LLR and kernel matching models may be due to the use of bootstrapping methods. The nearest neighbor matching and IPW-RA models rely on Abadie–Imbens standard errors (Abadie & Imbens, 2008) instead of bootstrapped standard errors.

Regardless of the type of matching estimator used, a common pattern emerges across all six models; schools that do not adopt any ELD curricula have TELPAS composites scores that are roughly 0.30 standard deviations lower than schools that purchase state-adopted ELD curricula. As
previously mentioned, these results are only generalizable to schools in Texas whose TELPAS scores are not masked, use the same ELD curricula in grades K-3, and who did not have missing textbook data.

**Research Question 2: Rigby On Our Way to English vs. National Geographic Reach**

When comparing the Rigby On Our Way to English curricula to National Geographic (NG) Reach, we find that schools that use Rigby have slightly higher TELPAS composite scores in all outcome models but these differences are not statistically significant (see Table 6). The standard errors for these estimates are slightly larger than the standard errors for the previous research question due to the smaller sample size. It is possible that there is a treatment effect in the population; however, this study may not be able to identify the treatment effect given the limited sample size. Relatedly, this study was not able to assess the effectiveness of any other ELD curricular adopters that captured by the covariates in the model. If these differences are correlated with curriculum adoption and student achievement, then the estimates may be biased. For example, if there are differences in commitment to improving achievement for EBs across the various curriculum adopters, this could bias the estimates. Although it is impossible to rule out all possible sources of bias, we provide evidence about the general reliability of the findings using two falsification tests.

For the first falsification test, we estimate ELD curricular effects on math achievement for third-grade EBs (see Table 7). If there are unobserved differences across adopting districts that are not captured by the covariates in the model, then we might see significant differences between treated and control units on math scores for third-grade EBs in 2015. The confounding factor (e.g., commitment to improving educational outcomes for EBs) will be present in the form of nonzero curriculum effects. If curriculum effects are not significantly different from zero for unexposed cohorts, then it is the curricula themselves, not other unobserved differences, that are driving the estimates. As English language proficiency correlates with math achievement (Grant et al., 2011), it is possible that nonadopters have lower math achievement than state-adopters. At most, these spillover effects will be small and no larger than the effects from the main findings. We find that schools that do not adopt any ELD curricula score 0.04 standard deviations lower on the state math test than scores that do adopt ELD curricula; however, this difference is not significant. Indeed, these spillover effects are smaller than the main treatment effect of 0.30 standard deviations and are relatively small and nonsignificant. In short, this test provides evidence that the primary findings are not driven by unobserved differences between nonadopters and state adopters.

In the second falsification test, we use third-grade TELPAS scores from spring 2012 as the outcome variable. By the end of the 2011–2012 school year, third graders will have been exposed to the ELD curriculum for no more than 1 year. As ELD materials were adopted during the 2011–2012 school year, it is unlikely that the materials had a large impact on English proficiency less than a year after implementation. If there are significant curricular impacts, then the impact should be smaller in magnitude than the impact on TELPAS spring 2015 scores. Indeed, there are no significant differences between third graders in treatment and control schools during the 2012 TELPAS test (see Table 8).

### Table 5. Average Treatment Effects of Nonadoption Relative to State-Adopted ELD Instructional Materials on TELPAS Spring 2015 Composite Scores.

| Covariate | LLR     | Kernel | IPWRA | K1     | K2     | K3     |
|-----------|---------|--------|-------|--------|--------|--------|
| Treatment effect | −0.30* | −0.31* | −0.35**** | −0.28*** | −0.33*** | −0.35*** |
| N (schools) | 1,024  | 1,024  | 1,024 | 1,024  | 1,024  | 1,024  |

*Note. Standard errors are in parentheses. Standard errors are estimated by bootstrapping using 250 repetitions and clustered at the district level for the local linear and kernel matching models. Robust standard errors are calculated for K1 matching and IPWRA models. Note. ELD = English language development; TELPAS = Texas English Language Proficiency Assessment System; LLR = local linear matching.

*p < .05. ***p < .001.
In addition, the estimates are all negative which is consistent with the finding from the main analysis that nonadoption is associated with lower English language proficiency. Less than 1 year after implementation, we find that nonadopters score 0.09 standard deviations lower than state-adopters on the state English language proficiency test; however, this difference is not significant.

We also modeled the data using an alternative approach, employing a comparative interrupted time series (CITS) model to estimate the impact of ELD adoption on English language proficiency. In the CITS regression, we adjust for the same school and district characteristics that are used in the propensity score estimation. The results from the CITS model mirrors the results from the local linear matching model with nonadopters scoring 0.30 standard deviations lower than state-adopters ($p < .01$). With regard to the second research question, the CITS model reveals no significant differences in spring 2015 English language proficiency score between schools that adopt Rigby and schools that adopt National Geographic. Given the robustness of results across various quasi-experimental methods, there is strong evidence to suggest that state-adopted ELD textbooks have a positive impact on EBs’ English language proficiency.

In sum, both falsification tests point toward the validity of the findings. None of the tests suggests that there are systematic differences between treatment and control schools. In addition, we employed a different type of quasi-experimental method—comparative interrupted time series—and recovered similar treatment effects for both research questions.

### Table 6. Average Treatment Effects of Rigby On Our Way to English relative to National Geographic Reach on TELPAS Spring 2015 Composite Scores.

| Covariate | LLR | Kernel | IPWRA | K1 | K2 | K3 |
|-----------|-----|--------|-------|----|----|----|
| Treatment effect | 0.05 | 0.05 | 0.14 | 0.13 | 0.16* | 0.17* |
| (0.17) | (0.16) | (0.08) | (0.08) | (0.07) | (0.07) |
| N (schools) | 706 | 706 | 706 | 706 | 706 | 706 |

Note. Standard errors are in parentheses. Standard errors are estimated by bootstrapping using 250 repetitions and clustered at the district level for the local linear and kernel matching models. Robust standard errors are calculated for K1 matching and IPWRA models. TELPAS = Texas English Language Proficiency Assessment System; LLR = local linear matching.

### Table 7. Falsification Tests: Effects of Nonadoption Versus State-Adoption on STAAR 2015 Math Achievement.

| Covariate | LLR | Kernel | IPWRA | K1 | K2 | K3 |
|-----------|-----|--------|-------|----|----|----|
| Treatment effect | −0.04 | −0.06 | 0.01 | 0.10 | 0.10 | 0.09 |
| (0.13) | (0.13) | (0.07) | (0.08) | (0.08) | (0.08) |
| N (schools) | 1,087 | 1,087 | 1,087 | 1,087 | 1,087 | 1,087 |

Note. Standard errors are in parentheses. Standard errors are estimated by bootstrapping using 250 repetitions and clustered at the district level for the local linear and kernel matching models. Robust standard errors are calculated for K1 matching and IPWRA models. STAAR = State of Texas Assessments of Academic Readiness; LLR = local linear matching.

### Table 8. Falsification Tests: Effects of Nonadoption Versus State-Adoption on 2011-2012 3rd-Grade TELPAS Composite Scores.

| Covariate | LLR | Kernel | IPWRA | K1 | K2 | K3 |
|-----------|-----|--------|-------|----|----|----|
| Treatment effect | −0.09 | −0.06 | −0.02 | −0.05 | −0.06 | −0.07 |
| (0.16) | (0.16) | (0.08) | (0.09) | (0.08) | (0.08) |
| N (schools) | 1,037 | 1,037 | 1,037 | 1,037 | 1,037 | 1,037 |

Note. Standard errors are in parentheses. Standard errors are estimated by bootstrapping using 250 repetitions and clustered at the district level for the local linear and kernel matching models. Robust standard errors are calculated for K1 matching and IPWRA models. TELPAS = Texas English Language Proficiency Assessment System; LLR = local linear matching.
It is also important to discuss what this study is not suggesting. This study does not suggest that off-list ELD materials, or ELD materials that are not on the state-adopted list, are ineffective for ELD instruction. Off-list ELD curricula such as EL Achieve and Carousel of IDEAS may also be effective but due to small sample size we were not able to evaluate the effectiveness of off-list texts.

Discussion

This study makes two contributions to the research literature on curricular effectiveness. This study is the first to assess the effectiveness of ELD curricular materials at the state level. We find moderate to large, statistically significant differences in student achievement based on whether or not schools and districts have adopted a state-approved ELD material. In contrast, we find small-to-zero differences in effects between the two most commonly adopted books, and what differences we find are only sometimes statistically significant. Although the findings only apply directly to Texas, the implication that adopting materials can affect student achievement may be more broadly applicable in other states and beyond. And while we cannot say why adopting ELD materials leads to improvements in student achievement, the most likely and simplest explanation is that these materials lead teachers to implement better instruction (e.g., more coherent, more aligned with standards, more developmentally appropriate) for EB students.

Second, the study contributes to the larger research literature on English language development which is sparse. Saunders et al. (2013) conduct the most comprehensive review of ELD instructional practices and find that little research examines the effects of ELD instruction on EBs’ academic achievement. In fact, as ELD research is so scant, the authors had to rely on second language research in other fields to inform their ELD guidelines. For example, some of the studies included in their synthesis focused on English learners in Quebec, Canada. This study contributes directly to the research literature on ELD because it focuses on English language learners in the United States.

The current study is not able to capture the variation across and within districts with regard to ELD curriculum implementation. It is likely that the impact of ELD curriculum on EBs’ achievement is mediated by the degree of implementation, but no U.S. state tracks the implementation of curriculum materials in ELD or any other subject. If there is a significant relationship between degree of implementation and EBs’ academic outcomes, then it is important to understand the nature of that relationship, and this is an important area for future research. Furthermore, if implementation fidelity is mediated by certain factors (e.g., practitioners’ beliefs about second language acquisition), then it is of import to identify these mediating factors so that barriers to implementation can be addressed.

The qualitative data we collected from interviews with district leaders reveal that there is confusion about the purpose of ELD instruction and how it differs from instruction for monolingual English-speakers. For instance, some district leaders believe that instruction for EBs is “just good teaching” (JGT), and, therefore, there is no need to purchase materials specifically for EBs. In other words, whatever works for monolingual English-speakers is sufficient to address the academic and linguistic needs of EBs. This finding is not new as previous research has found that a “just good teaching” approach is not adequate to meet the academic and linguistic needs of EBs. Fifteen years after De Jong and Harper coined the term JGT to describe this phenomenon, we find that the JGT view of teaching EBs is still pervasive. Given our finding that the adoption of ELD curriculum materials substantially affects student achievement, we reiterate De Jong and Harper’s (2005) call to explicitly include EBs at all levels of educational policy and practice. If there are expectations that content standards be explicitly and rigorously addressed, the same expectation should exist for English language proficiency standards. Otherwise, a double standard exists based on language which is effectively a proxy for national origin (Lau v. Nichols, 1974).

Castañeda v. Pickard (1981) mandated that educational programs for EBs be tested for program efficacy, which is what we attempt to do in this study. Using local linear matching, we find that students in nonadopting districts have lower English proficiency relative to those in state-adopting districts. Schools that do not have curricula that foreground English language proficiency should consider obtaining instructional materials that explicitly address forms of English, provide ample opportunities to develop all four language modalities, and integrate conversational and academic language. States should also consider mandating the purchase and use of ELD curriculum materials for teaching EB students and supporting the costs associated with such a mandate. Though there is evidence to suggest that ELD instructional materials can improve English language proficiency, much more work needs to be done to adequately meet the academic and linguistic needs of EBs (e.g., tracking of EBs, resegregation efforts that concentrate EBs in low-income schools). Implementing high-quality ELD curricula is only one part of an educational agenda that seeks to redress educational inequities faced by emergent bilinguals.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: This work is
supported by the National Science Foundation (grant no. 1445654) and the William T Grant Foundation (grant no. 183913). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funders.

ORCID ID
Morgan S. Polikoff https://orcid.org/0000-0002-6135-6292

References
Abadie, A., & Imbens, G. W. (2008). On the failure of the bootstrap for matching estimators. *Econometrica, 76*(6), 1537–1557.
Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., & Murphy, R. (2010). *Achievement effects of four early elementary school math curricula: Findings from first graders in 39 schools (NCEE 2011-4001)*. National Center for Education Evaluation and Regional Assistance, U.S. Department of Education, Institute of Education Sciences.
Alajmi, A. H. (2009). Addressing computational estimation in the Kuwaiti curriculum: Teachers’ views. *Journal of Mathematics Teacher Education, 12*(4), 263–283.
Alajmi, A. H. (2012). How do elementary textbooks address fractions? A review of mathematics textbooks in the USA, Japan, and Kuwait. *Educational Studies in Mathematics, 79*(2), 239–261.
Arnold, J. (1999). *Affect in language learning*. Cambridge University Press.
Bhatt, R., & Koedel, C. (2012). Large-scale evaluations of curricular effectiveness: The case of elementary mathematics in Indiana. *Educational Evaluation and Policy Analysis, 34*(4), 391–412.
Bhatt, R., Koedel, C., & Lehmann, D. (2013). Is curriculum quality uniform? Evidence from Florida. *Economics of Education Review, 34*, 107–121.
Black, D. A., & Smith, J. A. (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics, 121*(1), 99–124.
Blazar, D., Heller, B., Kane, T. J., Polikoff, M., Staiger, D., Carrell, S., Goldhaber, D., Harris, D., Hitch, R., Holden, K. L., & Kurlaender, M. (2019). *Learning by the book: Comparing math achievement growth by textbook in six Common Core states*. Center for Education Policy Research, Harvard University.
Bowler, M. (1978). Textbook publishers try to please all, but first they woo the heart of Texas. *The Reading Teacher, 31*(5), 514–518.
Burke, A. M., Morita-Mullaney, T., & Singh, M. (2016). Indiana emergent bilingual student time to reclassification: A survival analysis. *American Educational Research Journal, 53*(5), 1310–1342.
Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31–72.
Capps, R., Fix, M., Murray, J., Ost, J., Passel, J. S., & Hervantoro, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Urban Institute.
Castañeda v. Pickard. 1981. 648 F.2d 989 (5th Cir.)
Chingos, M. M., & Whitehurst, G. J. (2012). *Choosing blindly: Instructional materials, teacher effectiveness, and the Common Core*. Brookings Institution.
Cook, T. D., & Steiner, P. M. (2010). Case matching and the reduction of selection bias in quasi-experiments: The relative importance of pretest measures of outcome, of unreliable measurement, and of mode of data analysis. *Psychological Methods, 15*(1), 56–68.
Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics, 84*(1), 151–161.
De Jong, E. J., & Harper, C. A. (2005). Preparing mainstream teachers for English-language learners: Is being a good teacher good enough? *Teacher Education Quarterly, 32*(2), 101–124.
Garcia, O. (2009). Emergent bilinguals and TESOL: What’s in a name? *Tesol Quarterly, 43*(2), 322–326.
Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English language learners in U.S. schools: An overview of research findings. *Journal of Education for Students Placed at Risk, 10*(4), 363–385.
Glazerman, S., Levy, D. M., & Myers, D. (2003). Nonexperimental versus experimental estimates of earnings impacts. *The ANNALS of the American Academy of Political and Social Science, 589*(1), 63–93.
Grant, R., Cook, H. G., & Phakiti, A. (2011). *Relationships between language proficiency and mathematics achievement*. WIDA Consortium.
Guo, S., & Fraser, M. W. (2010). Propensity score analysis: *Statistical methods and analysis*. SAGE.
Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies, 64*(4), 605–654.
Heckman, J. J., Ichimura, H., & Todd, P. E. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies, 65*(2), 261–294.
Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*(3), 199–236.
Horwitz, E. K. (1999). Cultural and situational influences on foreign language learners’ beliefs about language learning: A review of BALLI studies. *System, 27*(4), 557–576.
Koedel, C., Li, D., Polikoff, M. S., Hardaway, T., & Wrabel, S. (2017). Mathematics curriculum effects on student achievement in California. *AERA Open, 3*(1), 1–22.
Lau v. Nichols, 414 US 563 (1974).
Lechner, M. (2002). Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *Review of Economics and Statistics, 84*(2), 205–220.
Menken, K., & Solorza, C. (2014). No child left bilingual: Accountability and the elimination of bilingual education programs in New York City schools. *Educational Policy, 28*(1), 96–125.
Namey, E., Guest, G., Thairu, L., & Johnson, L. (2008). Data reduction techniques for large qualitative data sets. In G. Guest & K. M. MacQueen (Eds.), *Handbook for team-based qualitative research* (pp. 137–162). Rowman & Littlefield.
National Center for Education Statistics. (2019). *Digest of education statistics*. 
National Education Association. (2008). English Language Learners face unique challenges.
Orfield, G. (2001). Schools more separate: Consequences of a decade of resegregation. Harvard Civil Rights Project.
Orfield, G., & Lee, C. (2006). Racial transformation and the changing nature of segregation. Harvard Civil Rights Project.
Polikoff, M. S., Campbell, S. E., & Korn, S. A. (2018). Using quantitative and qualitative methods to study the content and effects of curriculum materials. In C. R. Lochmiller (Ed.), Complementary research methods for educational leadership (pp. 193–212). Palgrave MacMillan.
Polikoff, M. S., Campbell, S. E., Rabovsky, S., Koedel, C., Le, Q. T., Hardaway, T., & Gasparian, H. (2020). The formalized processes districts use to evaluate mathematics textbooks. Journal of Curriculum Studies, 52(4), 451–477.
Rios-Aguilar, C., & Gándara, P. (2012). Horne v. Flores and the future of language policy. Teachers College Record, 114(9), 1–13.
Robitaille, D. F., & Travers, K. J. (1992). International studies of achievement in mathematics. In D. A. Grouws (Ed.), Handbook of research in mathematics teaching and learning (pp. 687–709). Macmillan.
Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41–55.
Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. Health Services and Outcomes Research Methodology, 2(3–4), 169–188.
Saldaña, J. (2013). The coding manual for qualitative researchers. SAGE.
Saunders, W., Goldenberg, C., & Marcelletti, D. (2013). English language development: Guidelines for instruction. American Educator, 37(2), 13–39.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Wadsworth Cengage Learning.
Shadish, W. R., & Steiner, P. M. (2010). A primer on propensity score analysis. Newborn and Infant Nursing Reviews, 10(1), 19–26.
Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde’s critique of nonexperimental estimators? Journal of Econometrics, 125(1), 305–353.
Stein, M., Stuen, C., Carnine, D., & Long, R. M. (2001). Textbook evaluation and adoption. Reading & Writing Quarterly, 17(1), 5–23.
Steiner, P. M., & Cook, D. (2013). Matching and propensity scores. In T. D. Little (Ed.), The Oxford handbook of quantitative methods (pp. 237–259). Oxford University Press.
Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical Science: A Review Journal of the Institute of Mathematical Statistics, 25(1), 1–21.
Texas Education Agency. (2011). Textbook funding and adoption process gets overhauled.
Texas Education Agency. (2014). Reports and data. http://tea.texas.gov/Reports_and_Data/
Texas Education Agency. (2015a). A brief overview of the adoption process. Accessed October, 1, 2015. http://tea.texas.gov/interiorpage.aspx?id=2147485614
Texas Education Agency. (2015b). Interpreting assessment reports. http://tea.texas.gov/student.assessment/interpguide/
Texas Education Agency. (2016). Texas English Language Proficiency Assessment System. Author.
Zeringue, J. K., Spencer, D., Mark, J., Schwinden, K., & Newton, M. A. (2010, April). Influences on mathematics textbook selection: What really matters [Paper presentation]. The NCTM Research Pre-session, San Diego, CA, United States.