AERA Open
July-September 2019, Vol. 5, No. 3, pp. 1–14
DOI: 10.1177/2332858419870488

Addressing Literacy Needs of Struggling Spanish-Speaking First Graders: First-Year Results From a National Randomized Controlled Trial of Descubriendo la Lectura

Trisha H. Borman
American Institutes for Research
Geoffrey D. Borman
University of Wisconsin–Madison
Scott Houghton
So Jung Park
Bo Zhu
Alejandra Martin
Sidney Wilkinson-Flicker
American Institutes for Research

Given the growing number of Latino English learners and the lack of evidence-based educational opportunities they are provided, we investigated the impact of one potentially effective literacy intervention that targets struggling first-grade Spanish-speaking students: Descubriendo La Lectura (DLL). DLL provides first-grade Spanish-speaking students one-on-one literacy instruction in their native language and is implemented at an individualized pace for approximately 12 to 20 weeks by trained bilingual teachers. Using a multisite, multicohort, student-level randomized controlled trial, we examined the impact of DLL on both Spanish and English literacy skills. In this article, we report findings from the first of three cohorts of students to participate in the study. Analyses of outcomes indicate that treated students outperformed control students on all 11 Spanish literacy assessments with statistically significant effect sizes ranging from 0.34 to 1.06. Analyses of outcomes on four English literacy assessments yielded positive effect sizes, though none were statistically significant.

Keywords: *English language learners, Spanish, randomized controlled trial, literacy*

The rich cultural and language diversity of our nation continues to grow. As of fall 2015, English language learners (ELLs) made up nearly 10% (4.8 million) of public school students in the United States (Office of English Language Acquisition, 2015), with recent projections suggesting that ELLs will comprise 25% of the nation's students by 2030 (National Clearinghouse for English Language Acquisition, 2011). Spanish speakers represent the majority of ELL students (77%) and the overall Latino population is expected to double to 28.6% (or 119 million) of the U.S. population by 2060.

Even as Latino ELL students represent a larger and larger share of the U.S. student population, interventions and policies to support their educational opportunities and outcomes have been insufficient. For instance, Latino ELL students have fewer opportunities to practice literacy, particularly in English, before formal schooling begins (Schneider, Martinez, & Ownes, 2006) and generally attend schools with less favorable learning environments relative to White students, including less experienced teachers (Galindo & Reardon, 2006; Reardon & Galindo, 2009). The progress of ELL students may be further constrained by statewide English-only programming policies that are associated with higher rates of special education referrals and dropout for ELLs (Gandara & Hopkins, 2010). Finally, the assessments themselves can contribute to inaccurate and biased estimates of ELLs' achievement when the test items confound students' cognitive skills in the measured content area with unrelated cultural and linguistic knowledge (Abedi, Hofstetter, & Lord, 2004).

Addressing the Literacy Needs of Spanish-Language ELLs: Best Practices and the Descubriendo La Lectura (DLL) Theory of Action

As existing policies and practices have largely failed to capitalize on the strengths of Latino and ELL students, compromising students' educational and labor market

opportunities, a critical policy question comes into focus: *How do policy makers and practitioners advance the literacy skills of struggling Spanish-language ELLs?* This article highlights one promising and replicable approach: Descubriendo La Lectura (DLL), which is the Spanish-language reconstruction of the widely used Reading Recovery program. DLL is an intensive, one-to-one literacy intervention that helps struggling first-grade ELLs, while also drawing on Spanish-speaking students' cultural and native language assets through the unique potential of bilingualism—an approach with cognitive, social–emotional, and employment benefits (Bialystock, 2001; Galambos & Hakuta, 1988; Nagy, Berninger, Abbott, Vaughan, & Vermeulen, 2003; Portes & Hao, 2002; Vom Dorp, 2000).

First, we ground the current study within the context of prior approaches to advancing the literacy skills of Spanish-language ELLs by exploring the research literature concerning best practices. Next, we discuss both how the DLL program addresses these research-based best practices and the theory of action through which the program may achieve the hypothesized positive Spanish- and English-language achievement impacts.

Language of Instruction. The research literature suggests that extensive use of native language instruction supports ELLs' language and literacy development (Cummins, 2005; Lindholm-Leary & Block, 2010; Ramírez, Yuen, Ramey, & Pasta, 1991; Thomas & Collier, 2002). Moreover, learning basic literacy skills in one's native language tends to transfer to development of second-language skills (Durgunoglu, Nagy, & Hancin-Bhatt, 1993; Quiroga, Lemos-Britton, Mostafapour, Abbott, & Berninger, 2001; Roberts, 1994). This latter point is supported by evidence suggesting that early native language reading ability is predictive of later English proficiency (August & Shanahan, 2008; Genesee, Lindholm-Leary, Saunders, & Christian, 2006).

Indeed, relative to English-only programs, recent research suggests that bilingual programs produce stronger long-term English proficiency and English/language arts achievement (National Academies of Sciences, Engineering, and Medicine, 2017; Umansky & Reardon, 2014; Valentino & Reardon, 2015). A recent meta-analysis similarly concluded that bilingual programs boosted academic achievement, especially reading proficiency, for language minority children across countries and languages (McField & McField, 2014). Several research syntheses also suggest that ELLs experience better English reading outcomes in bilingual programs than in English-only programs (Francis, Lesaux, & August, 2006; Rolstad, Mahoney, & Glass, 2005; Slavin & Cheung, 2004).

Timing of Intervention. Beyond the language of instruction, the research literature highlights the importance of early intervention and points to enduring challenges experienced by students who do not learn to read in the early grades (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Juel, 1988). Given the optimal neurological and developmental stage of early childhood, preschool, and early-elementary interventions offer potential for substantial growth and a particularly strong return on investment (Knudsen, Heckman, Cameron, & Shonkoff, 2006). A research synthesis by Wanzek and Vaughn (2007) specifically found larger effects for reading interventions that begin in first grade relative to interventions that begin in second and third grade, as reading challenges become more complex in the later grades.

Mode of Instruction. While research supports the efficacy of a handful of whole class and whole school reforms (e.g., Success for All, cooperative learning, and Direct Instruction) for struggling ELL readers, small group and one-to-one tutoring have proven to be effective for those students who require the greatest level of support (Cheung & Slavin, 2012). An extensive research synthesis of effective programs for struggling readers by Slavin, Lake, Davis, and Madden (2011) concluded that one-on-one tutoring is a particularly effective approach to improving reading achievement (effect size: $d = 0.38$). Those tutoring programs that focus on phonics and are delivered by teachers ($d = 0.69$) appear particularly effective relative to those that employ phonetic models delivered in small groups ($d = 0.31$) or by paraprofessionals ($d = 0.38$). A research synthesis of effective instructional approaches for Spanish-dominant ELLs in the elementary grades confirms the efficacy of one-on-one tutoring and small-group instruction for struggling ELL readers (Cheung & Slavin, 2012).

A meta-analysis on early reading interventions by Waznek et al. (2018) suggested that one-on-one tutoring is generally more effective than group instruction of 2 to 8 students, though the authors reported that too few studies have examined small-group instruction (2–4 students) to understand its relative effectiveness. While highlighting the superiority of phonetic tutoring, Slavin et al. (2011) concluded that small-group instruction may be appropriate for students with "mild reading problems," given the cost of tutoring, while individual tutoring is optimal for students with "the most serious difficulties."

While one-on-one tutoring is associated with large effect sizes, the evidence suggests that tutoring ought to be coupled with improvements in classroom instruction through elementary school to sustain effects over time. For example, Success for All is associated with longer term gains than tutoring alone. It combines classroom interventions focused on reading with tutoring for first-grade students who continue to struggle, as well as sustained classroom interventions in later elementary grades. Slavin et al. (2011) thus conclude that "with a continuing focus on effective classroom instructional models, most children who receive

effective tutoring interventions in first grade can be kept on track in reading."

Another research review (U.S. Department of Education, 2001) substantiates the efficacy of tutoring programs and specifies that programs with research-based components lead to reading improvements, as well as increases in students' self-confidence and motivation to read. This synthesis identified six elements of tutoring programs that produce the best academic outcomes: (1) close coordination with the classroom or reading teacher, (2) intensive and ongoing training for tutors, (3) well-structured tutoring sessions in which the content and delivery of instruction is carefully scripted, (4) careful monitoring and reinforcement of progress, (5) frequent and regular tutoring sessions, with each session between 10 and 60 minutes daily, and (6) specially designed interventions for the 17% to 20% of children with severe reading difficulties.

Implementation Approach. Well-trained teachers are key to the success of any educational program or intervention. In addition to the benefit of tutoring models that employ certified teachers (Slavin et al., 2011), the research literature suggests that high-quality professional development (PD) is a key ingredient of effective tutoring programs (U.S. Department of Education, 2001), programs providing dual-language instruction (Boyle, August, Tabaku, Cole, & Simpson-Baird, 2015), and reading programs for struggling readers (Slavin et al., 2011), ELLs (National Academies of Sciences, Engineering, and Medicine, 2017), and Spanish-dominant ELLs (Cheung & Slavin, 2012). In fact, a synthesis of evidence on effective reading programs for Spanish-dominant ELLs by Slavin and Cheung (2004) suggested that, more important than the language of instruction, the most effective interventions "provide extensive PD and coaching to help teachers effectively implement promising models." Indeed, many successful education interventions have failed to maintain their effectiveness once they are brought to scale, and implementation research suggests that successful innovations are supported and sustained by change agents that facilitate learning and reform (Fullan, 1993; May et al., 2013) and networks that support continuous learning and development within and across sites (Bryk, Gomez, & Grunow, 2010; Englebart, 2003).

Key Literacy Components. Beyond the language of instruction, timing of intervention, instructional grouping, instructional support, and support for implementation, the research literature highlights the importance of evidence-based literacy content. Though the National Reading Panel report did not specifically focus on ELLs nor students learning in non-English languages, DLL—like its sister English-language program, Reading Recovery—has adopted the five essential elements of reading instruction identified by the panel's report: phonemic awareness, phonics, fluency, vocabulary, and text comprehension (National Institute of Child Health and Human Development, 2000). It has been 18 years since the publication of this report, which relied on the review of over 100,000 studies of literacy instruction, yet the findings remain relevant and influential today as both Reading Recovery and DLL practices continue to be guided by the "five essential elements" identified by the report's authors.

In 2006, another expert panel produced the Report of the National Literacy Panel on Language-Minority Children and Youth, titled, *Developing Literacy in Second Language Learners* (August & Shanahan, 2006). This report has been criticized for privileging monolingualism, and language-minority students are, by default, defined as "behind" from the outset (Escamilla, 2009). In addition, the report drew on a limited evidence base relative to the original National Reading Panel study and, despite these limitations, presumed that many of the National Reading Panel findings applied to ELLs as well as monolingual students. Nonetheless, most fundamentally, the 2006 report agreed with several other research reviews mentioned earlier by concluding that instructional programs that invest in the development of students' first language are more effective than those that are English medium or English only (Escamilla, 2009).

Alignment of DLL With the Research Evidence

The DLL model was designed to deliver research-based practices to ELLs who are struggling with reading. As displayed in Table 1, the research literature supports the various components of the DLL model—an early, one-on-one Spanish-language tutoring intervention composed of evidence-based literacy and tutoring components, delivered by well-trained teachers.

DLL is taught to first-grade students in their native Spanish language, reflecting the evidence that early interventions that build literacy proficiency in students' first language support long-term language and literacy achievement in both Spanish and English. It aligns with research-based, one-on-one tutoring practices by providing daily 30-minute, one-on-one lessons for the lowest achieving Spanish-language ELLs (i.e., the lowest performing 20% to 25%). The tutoring is highly structured with a set of prescribed activities for each lesson, including a portion devoted to phonics instruction. Teacher procedures include a high degree of progress monitoring through multiple record-keeping procedures, such as the running record form—a detailed accounting of student progress completed during each lesson to inform the student's next lesson. In addition, the tutors regularly communicate with the classroom teachers to discuss student progress and coordinate reading strategies.

Along with a rigorous teacher selection process, a defining characteristic of the DLL model is a highly structured,

TABLE 1
Alignment of Descubriendo La Lectura (DLL) and the Research Literature

Research domain	DLL component	Research reviews supporting DLL components
Language of instruction	Native language instruction	• August and Shanahan (2008) • Genesee et al. (2006) • McField and McField (2014) • Francis et al. (2006) • Rolstad et al. (2005) • Slavin and Cheung (2004)
Timing of intervention	First-grade intervention	• Wanzek and Vaughn (2007)
Mode of instruction	One-on-one tutoring	• Slavin et al. (2011) • Cheung and Slavin (2012) • Wanzek et al. (2018) • U.S. Department of Education (2001)
Implementation approach	Professional development and systemic implementation	• U.S. Department of Education (2001) • Cheung and Slavin (2012)
Key literacy components	Evidence-based reading instruction	• National Institute of Child Health and Human Development (2000)

ongoing PD system. At the first level, DLL teacher leaders train to become "expert literacy coaches" through postgraduate training and ongoing mentorship from university faculty. These trained teacher leaders then deliver a year-long, graduate-level DLL course to beginning DLL teachers, who teach DLL children while receiving at least four school visits from their teacher leader. Teachers "bridging" from Reading Recovery to DLL participate in a second year of graduate coursework and training in Spanish (May et al., 2013; Reading Recovery Council of North America [RRCNA], 2017a). DLL teachers and teacher leaders then maintain an ongoing professional learning relationship in which the teacher participates in a minimum of six PD sessions each year. They regularly communicate with the teacher leader about student progress, share student data, and receive at least one school visit from their teacher leader annually. These prescribed PD processes aim to highlight the importance of teacher quality, foster highly skilled teachers, and sustain high teaching standards throughout a teacher's career (May et al., 2013; RRCNA, 2017b).

DLL is also designed to provide systemic support for implementation fidelity across sites, as envisioned by its founder and reflected in the program's implementation Standards and Guidelines (Clay, 1987; May et al., 2013). Teacher leaders serve as change agents by working with district administrators to oversee program implementation, maintain PD standards, and exchange information on student outcomes with university trainers, district-level site coordinators, principals, and DLL teachers. District-level site coordinators communicate with teacher leaders to make policies, design structures, recruit teachers, and secure resources that support and sustain implementation. Principals ensure the structures, personnel, and resources to facilitate school-level implementation, and DLL teachers coordinate with classroom teachers and principals to align instruction. This structured network is further supported by resources from the RRCNA, training that begins with the North American Trainers Group, and a central data management entity, the International Data Evaluation Center (IDEC). DLL site and personnel expectations are clearly defined through the Standards and Guidelines to support systemic implementation (May et al., 2013; RRCNA, 2017b), while the PD and data from the IDEC system help maintain an approach that emphasizes a model of ongoing and data-driven improvement.

Similar to the Reading Recovery approach, DLL lessons consistently include each of the essential reading elements as identified by the National Reading Panel (Forbes & Doyle, 2004). For example, DLL teachers facilitate text comprehension by helping students anticipate meaning through reading continuous text, rereading familiar books, and discussing the content (Forbes & Doyle, 2004). DLL also promotes self-monitoring and sustained literacy by emphasizing self-directed and active learning through independent reading and writing opportunities (DeFord, 2007; May et al., 2013). Finally, lessons focus on improving students' vocabulary skills and knowledge of comprehension strategies, which can improve reading comprehension outcomes for ELLs (Jimenez, 1997; Jimenez, Garcia, & Pearson, 1996; Proctor, Dalton, Grisham, 2007).

Prior Research on Reading Recovery and DLL

In addition to the theoretical and practical evidence supporting the key components of DLL, there is a long-standing

history of rigorous empirical support for the English-language version of DLL, Reading Recovery (D'Agostino & Murphy, 2004; May et al., 2015; Sirinides, Gray, & May, 2018). Furthermore, quasi-experimental studies of the DLL program have demonstrated substantial reading gains by students, wherein they meet the reading level of their first-grade peers (Escamilla, 1994; Escamilla, Loera, Ruiz, & Rodríguez, 1998; Neal & Kelly, 1999), and descriptive studies indicate that DLL students maintain these gains through elementary school (Escamilla et al., 1998; Neal & Kelly, 1999). With limited quasi-experimental and descriptive evidence to assess the impacts of DLL, this study adds considerable rigor to the literature on evidence-based practices for advancing the literacy skills of struggling Spanish-speaking ELLs in general, and to our understanding of the DLL intervention's efficacy in particular. Specifically, with this experimental study, we respond to the question: What is the impact of assignment to DLL on the literacy achievement of Spanish-speaking, first-grade ELLs, as measured by two Spanish-language assessments (*Instrumento de Observación* [IdO] and *Logramos*) and an English-language assessment (*Iowa Test of Basic Skills* [ITBS])?

Method

Study Design

To estimate the impact of DLL on literacy achievement, we designed this study as a multisite, student-level randomized controlled trial in which first-grade Spanish-speaking students, who were determined to be struggling readers, were assigned to receive DLL services either at the start of the 2016–2017 school year (immediate group) or later in the 2016–2017 school year (delayed group). The delayed group served as the control group. This article reports outcomes for the first cohort from this study.

Sample

Schools were recruited to participate in this study if they had been implementing DLL for at least one year prior to the start of the study. Students were eligible for DLL services if their home language was Spanish and if they were among the lowest scoring 25% of students in their schools as determined by the IdO (see below for a description of all measures).

The Cohort 1 sample included 152 first-grade struggling readers, whose home language was Spanish (78 treatment and 74 control) nested within 22 schools and seven school districts across three states: Texas, Illinois, and Arizona. Of these students, approximately 45% are female, 99% are Hispanic, and 82% are economically disadvantaged, as indicated by students' participation in free or reduced-priced meal programs.

Random Assignment Procedure

We conducted student-level random assignment within each school-level block. Assignment to experimental conditions was accomplished via use of a random number generator. Each DLL-eligible student was assigned a random number and those with the lowest numbers were served first. DLL teachers served half of the students in the fall semester (immediate/treatment group) and the other half in the spring (delayed/control group).

Testing Procedure

All eligible students (immediate and delayed) were tested in the fall on three measures: IdO, Logramos, and ITBS (see below for a description of all measures). Students served first progressed through the program at an individualized pace, lasting between 12 and 20 weeks. As each "immediate" student exited the program, she or he was posttested on the three measures. At the same time as a student exited the program, the next student from the delayed group with the lowest randomly generated number was also tested, and then received DLL services. In this way, both groups of students were pre- and posttested at the same time, but at that time, only the immediately served students had received DLL services. Figure 1 depicts the DLL services and testing timeline for both experimental groups.

Measures

To evaluate the impacts of DLL on literacy achievement, we used a battery of two Spanish assessments and one English assessment including IdO, Logramos, and the ITBS. These tests were administered prior to randomization and as each immediate student exited the program and each delayed student entered. DLL teacher leaders, who were not direct instructors of the students in our study, administered these assessments to students. Each test is described below.

IdO is a routinely used DLL assessment that serves as a diagnostic tool to assess students' literacy skills prior to, and on exiting, DLL services (Escamilla, Andrade, Basurto, & Ruiz, 1996). It is administered as a part of routine practice for all implementations of DLL for Spanish-language students and includes six subtests: Letter Identification, Word Test, Concepts about Print, Writing Vocabulary, Hearing and Recording Sounds in Words, and Text Reading.

IdO is designed for the systematic observation of young children's early literacy competencies. The National Center on Response to Intervention reviewed the Observation Survey, giving it its highest rating of "convincing evidence" in each category (see http://www.rti4success.org/observation-survey-early-literacy-achievement-reading). Escamilla et al. (1996) reported evidence of construct validity for the IdO using data from approximately 500 first-grade students,

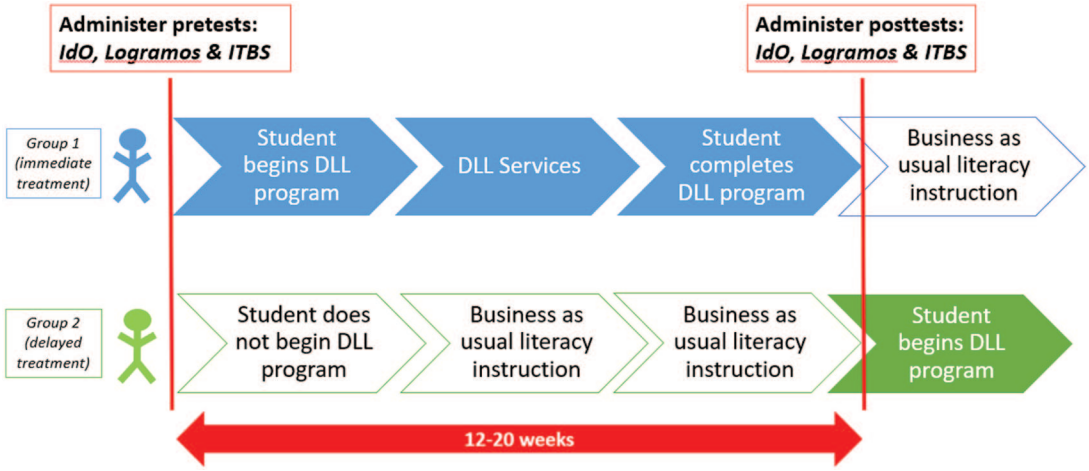

FIGURE 1. *Timeline of Descubriendo la Lectura (DLL) program delivery and student assessments.*

yielding Cronbach's alphas ranging from $\alpha = .51$ to $\alpha = .82$. In addition, concurrent validity was established by comparing the six subtests of IdO with a norm-referenced test, *Aprenda: La Prueba de Logros en Español* (Pearson, 3rd ed.; see http://www.pearsonassessments.com/learningassessments/products/100000585/aprenda-3-aprenda-la-prueba-de-logros-en-espanol-tercera-edicion.html). Content validity was established through a series of translations and back translations (Escamilla et al., 1996). Furthermore, Escamilla and her colleagues (1996) reported that the IdO performed equally well across first-grade students from three different cultural backgrounds: Mexican American, Puerto Rican, and Cuban American. The results suggested no differential understanding of the survey questions across the three groups.

Logramos (2006) is the Spanish-language version of the ITBS achievement test battery and is available from Houghton Mifflin Harcourt. The items in Logramos follow the scope and sequence of Form E of the English-language Iowa Assessments and have been translated, when appropriate, from the original English items in Form E of the Iowa Assessments. Though the vast majority of items were direct translations of the English test into Spanish, some items required adaptation and replacement of English items in the Spanish version in order to target the same skills and maintain the underlying psychometrics of the test items.

Finally, to assess students' English-literacy skills, we administered the ITBS. We concede that content-based assessments can conflate students' cognitive skills in the measured content area with unrelated cultural and linguistic knowledge and are, thus, not always valid and reliable assessments of ELLs' content knowledge (Abedi et al., 2004). One might argue that as such an English-language assessment is not a valid and reliable measure of ELLs' literacy skills, as the test conflates an ELLs' knowledge of English with general literacy skills. We administered two Spanish-language literacy assessments, the Logramos and IdO, which *do not* conflate English-language knowledge with general literacy skills. We use the ITBS test not to measure general literacy skills, but rather to measure the extent to which the students may have realized improvements in their *English* literacy skills. We do, however, assume that measurement error is likely to be a factor, due to ELLs' limited opportunities to learn English.

For both the ITBS and Logramos, we administered three subtests during each assessment, Vocabulary/Vocabulario, Reading/Lectura, and Language/Lenguaje, which are combined to create an overall total score. Vocabulary/Vocabulario includes 26 items. Reading/Lectura, includes 35 items and assesses literacy in the following domains: Literary Text/Texto literario, Informational Text/Texto informative, Explicit Meaning/Significado explícito, Implicit Meaning/Significado implícito, Key Ideas/Ideas principals. The Language (Lenguaje) section of the assessment includes 34 items assessing: Spelling/Ortografía, Capitalization/Uso de mayúsculas, Punctuation/Puntuación, and Written Expression/Expresión escrita. These tests are standardized, norm-referenced, and not overaligned with the intervention and, therefore, provide fair assessments of achievement change for both control- and treatment-group students.

Beyond these test scores, we collected demographic data for each student through the IDEC (2012) at the Ohio State University, which gathers data on all Reading Recovery and DLL sites and students each year.

Analytic Method

We fit a series of hierarchical linear models to estimate the student-level "intent to treat" impact, or impact of being *assigned* to receive DLL services, on each of the outcome measures. The unit of analysis is at the student-level because students were randomly assigned within schools in this study. To account for the nesting structure of data (where students were nested within schools), the model includes a random error term at the school level. The model adjusts outcomes by using a student-level pretest measure, thus improving model fit and statistically adjusting for any potential chance pretest differences between treatment and control groups. The model used for the impact analysis is conceived as follows:

$$Y_{ij} = \alpha + \beta(DLL_{ij}) + \gamma(PRETEST_{ij}) + u_j + \varepsilon_{ij}$$

where Y_{ij} represents the test score of student *i* in school *j*, DLL_{ij} is an indicator variable equal to 1 for DLL (immediately treated group) and 0 for the control group (delayed treatment group) and β is the coefficient representing the impact of DLL for student *i* in school *j*. $PRETEST_{ij}$ is each student's standardized pretest measure. In this formulation, the model intercept, α, represents the grand mean for the control group members with average pretest scores (reference group), γ is a coefficient representing the association between the pretest measure and the outcome, u_j represents the school-specific error, and ε_{ij} represents the student-specific error.

All pre- and posttest scores were standardized to have a mean of 0 and standard deviation of 1. Therefore, the coefficient for the treatment variable represents the standardized mean difference—that is, the effect size—between treatment and control students. To reduce the false discovery rate, the Benjamini–Hochberg correction for multiple comparisons was used.

Results

Attrition Analysis

We estimated potential bias due to sample attrition following standards developed by the What Works Clearinghouse (WWC; Institute of Education Sciences, 2017). Sample attrition occurs when the final analytical sample differs from the randomly assigned sample due to missing data. In our study, students missing pre- or posttest data were excluded from the analytic samples. WWC describes whether the combination of overall attrition (i.e., the rate of attrition for the entire sample) and differential attrition (i.e., the difference in the rates of attrition for the treatment and control groups) is high or low, then determines whether the expected bias due to attrition rate is tolerable or unacceptable.

For IdO, with an overall attrition rate of 6.58% and a differential attrition rate of 0.35 percentage points, the IdO sample falls in the low attrition category, as defined by WWC standards. For ITBS, with an overall attrition rate of 8.55% to 26.32% and a differential attrition rate of 0.31 to 1.25 percentage points, the ITBS sample also falls within the WWC's low attrition category. Finally, for the Logramos assessment, with an overall attrition rate of 7.24% to 19.08% and a differential attrition rate of 1.70 to 3.60 percentage points, the Logramos sample is also within the WWC's low attrition category. Therefore, overall, all subtests meet the criteria for "low" attrition, resulting in tolerable levels of potential bias. Given that our study used a randomized controlled design and sample attrition falls into the low attrition category, our study is eligible to "Meet WWC Group Design Standards Without Reservations" (Institute of Education Sciences, 2017). Table 2 shows the combination of overall and differential attrition for each outcome measure of our study.

Baseline Equivalence

Baseline equivalence testing between immediate and delayed treatment students was completed for each analytic sample. As preintervention baseline measures, seven subtests of IdO as well as four subtests of both the ITBS and Logramos were included. The mean differences between treatment and control group and the standardized mean differences (Hedges' *g*) for the IdO, ITBS, and Logramos are shown in Tables 3, 4, and 5, respectively. For IdO, the differences between the treatment and control groups are greater than 0.25 standardized units for Hearing and Recording Sounds in Words ($g = 0.27$), Text Reading Level ($g = 0.26$), and fall Observation Survey Total Score ($g = 0.27$; see Table 3). For ITBS, the differences between treatment and control groups are less than 0.25 standardized units for all pretest measures except the Reading pretest ($g = 0.30$ for the Reading sample; $g = 0.26$ for the Language sample; and $g = 0.30$ for the ELA [English language arts] total sample; see Table 4). Finally, for the Logramos assessment, the differences between treatment and control groups are less than 0.25 standardized units for all pretest measures except the Language pretest ($g = -0.30$ to -0.34 across analytic samples; see Table 5).

Because we established that there were low overall and differential attrition rates, we can be confident that any observed baseline differences between the treatment and control samples are not due to systematic sources of bias. Though some baseline pretest differences were greater than 0.25 *SD*s, these differences occurred by chance through sampling error alone. To improve the precision of our analytical models, and to account for these baseline differences,

TABLE 2
Student Sample Attrition by Outcome Measure

Outcomes	Subtest	Percent missing				WWC attrition category
		Delayed (C)	Immediate (T)	Difference	Overall	
IdO	Letter Identification	6.76	6.41	−0.35	6.58	Low
	Sounds in Words	6.76	6.41	−0.35	6.58	Low
	Writing Vocabulary	6.76	6.41	−0.35	6.58	Low
	Concepts about Print	6.76	6.41	−0.35	6.58	Low
	Ohio Word Test	6.76	6.41	−0.35	6.58	Low
	Text Reading	6.76	6.41	−0.35	6.58	Low
	Language Total	6.76	6.41	−0.35	6.58	Low
ITBS	Reading	25.68	26.92	1.25	26.32	Low
	Language	18.92	19.23	0.31	19.08	Low
	Vocabulary	8.11	8.97	0.87	8.55	Low
	ELA total	25.68	26.92	1.25	26.32	Low
Logramos	Reading	20.27	16.67	−3.60	18.42	Low
	Language	10.81	7.69	−3.12	9.21	Low
	Vocabulary	8.11	6.41	−1.70	7.24	Low
	ELA total	20.27	17.95	−2.32	19.08	Low

Note. WWC = What Works Clearinghouse; IdO = Instrumento de Observación; ITBS = Iowa Test of Basic Skills; ELA = English language arts.

TABLE 3
Summary of Instrumento de Observación Pretest Baseline Equivalence

Pretest measure	Immediate			Delayed			Mean difference	SMD
	M	*SD*	*N*	*M*	*SD*	*N*		
Letter Identification	46.6	15.45	73	44.49	15.48	69	2.11	0.14
Hearing and Recording Sounds in Words	27.21	10.51	73	24.39	10.41	69	2.82	0.27
Writing Vocabulary	11.9	7.46	73	12.22	7.81	69	−0.32	−0.04
Concepts about Print	11.71	3.11	73	11.14	3.05	69	0.57	0.19
Ohio Word test	9.71	5.94	73	8.71	6.01	69	1.00	0.17
Text Reading Level	1.76	1.61	73	1.35	1.52	69	0.41	0.26
Fall OS Total Score	472.03	39.58	73	460.9	41.9	69	11.13	0.27

Note. SD = standard deviation; SMD = standardized mean difference; OS = Observation Survey.

we statistically control for the pretests in all of our analytical models.

Impact Analyses

The results are presented separately for each literacy measure. As noted earlier, treatment coefficients can be interpreted as the standardized mean differences (or effect sizes, *g*) between treatment and control students on the three literacy assessments.

Instrumento de Observacion. We found statistically significant student-level impacts of assignment to DLL services on all IdO posttest outcomes, ranging from an effect size of 0.34 for the Sound in Words outcome to $g = 1.01$ for the Text Reading subtest. Table 6 shows the average DLL effects on all six IdO outcomes: Letter Identification, Ohio Word Test, Concepts about Print, Writing Vocabulary, Hearing and Recording Sounds and Text Reading Sounds, and Text Reading Level.

Iowa Test of Basic Skills. Students assigned to DLL scored higher on all ITBS subtests than those assigned to the control (delayed) group as demonstrated by positive effect sizes; however, the effect sizes were not statistically significant. Table 7 shows the average impacts of assignment to DLL on each of the ITBS outcomes including the following: Reading, Language, Vocabulary, and ELA total.

TABLE 4
Summary of Iowa Test of Basic Skills Pretest Baseline Equivalence

Analytic sample	Pretest subject	Immediate			Delayed			Mean difference	SMD
		M	*SD*	*N*	*M*	*SD*	*N*		
Reading sample	Reading	133.58	5.41	61	131.93	5.75	59	1.65	0.30
	Language	123.48	9.15	61	122.54	10.33	59	0.94	0.10
	Vocabulary	124.57	12.7	61	124.22	14.48	59	0.35	0.03
	ELA total	126.89	6.28	61	126.14	7.09	59	0.75	0.11
Language sample	Reading	133.46	5.36	64	132.05	5.66	63	1.41	0.26
	Language	123.56	9.17	64	122.74	10.15	63	0.82	0.08
	Vocabulary	125.44	13.06	64	124.68	14.3	63	0.76	0.06
	ELA total	127.11	6.36	64	126.32	6.98	63	0.79	0.12
Vocabulary sample	Reading	132.48	7.25	71	132.05	5.66	70	0.43	0.07
	Language	122.65	10.18	71	122.84	10.33	70	−0.19	−0.02
	Vocabulary	124.68	14.34	71	124.4	14.02	70	0.28	0.02
	ELA total	127.13	6.31	71	126.32	6.98	70	0.81	0.12
ELA total sample	Reading	133.58	5.41	60	131.93	5.75	57	1.65	0.30
	Language	123.34	9.16	60	122.98	10.14	57	0.36	0.04
	Vocabulary	124.68	12.77	60	124.77	14.42	57	−0.09	−0.01
	ELA total	126.89	6.28	60	126.14	7.09	57	0.75	0.11

Note. SD = standard deviation; SMD = standardized mean difference; ELA = English language arts.

TABLE 5
Summary of Logramos Pretest Baseline Equivalence

Analytic sample	Pretest measure	Immediate			Delayed			Mean difference	SMD
		M	*SD*	*N*	*M*	*SD*	*N*		
Reading sample	Reading	153.91	15.23	71	151.69	12.96	63	2.22	0.16
	Language	166.58	16.55	71	171.41	12.29	63	−4.83	−0.33
	Vocabulary	159.36	10.89	71	160.29	8.94	63	−0.93	−0.09
	ELA total	162.5	8.47	71	162.98	9.39	63	−0.48	−0.05
Language sample	Reading	153.91	15.23	72	151.23	13.34	66	2.68	0.19
	Language	166.21	16.44	72	171.06	11.96	66	−4.85	−0.34
	Vocabulary	159.03	10.55	72	160.2	8.8	66	−1.17	−0.12
	ELA total	162.5	8.47	72	162.87	9.35	66	−0.37	−0.04
Vocabulary sample	Reading	153.91	15.23	74	151.69	12.96	68	2.22	0.16
	Language	166.46	16.32	74	170.84	12.26	68	−4.38	−0.30
	Vocabulary	159.34	10.75	74	160.28	8.71	68	−0.94	−0.10
	ELA total	162.5	8.47	74	162.98	9.39	68	−0.48	−0.05
ELA total sample	Reading	153.91	15.23	70	151.69	12.96	61	2.22	0.16
	Language	166.36	16.57	70	171.23	12.36	61	−4.87	−0.33
	Vocabulary	159.03	10.62	70	160.25	9.07	61	−1.22	−0.12
	ELA total	162.5	8.47	70	162.98	9.39	61	−0.48	−0.05

Note. SD = standard deviation; SMD = standardized mean difference; ELA = English language arts.

Logramos. Students who received DLL services outperformed those who had not yet received services on all Logramos subtests and the associated effect sizes were statistically significant ranging from $d = 0.41$ to $d = 0.55$. Table 8 shows the average DLL effects for the four Logramos outcomes: Reading, Language, Vocabulary, and ELA total.

Because we tested the statistical significance of multiple outcomes within one literacy domain, we applied the

TABLE 6
Hierarchical Linear Model Results for Instrumento de Observación Outcomes

	Letter Identification	Sound in Words	Writing Vocabulary	Concepts about Print	Ohio Word Test	Text Reading Level	Total score
(Intercept)	–0.316** (0.095)	–0.177 (0.101)	–0.505 *** (0.118)	–0.439 *** (0.112)	–0.271** (0.097)	–0.546*** (0.085)	–0.474*** (0.091)
Treatment	0.602*** (0.116)	0.337* (0.132)	0.995*** (0.116)	0.883*** (0.126)	0.528*** (0.133)	1.062*** (0.119)	0.939*** (0.099)
Letter Identification Pretest	0.637*** (0.065)						
Sound in Words Pretest		0.572*** (0.069)					
Writing Vocabulary Pretest			0.377*** (0.069)				
Concepts about Print Pretest				0.353*** (0.068)			
Ohio Word Test Pretest					0.534*** (0.067)		
Text Reading Level Pretest						0.423*** (0.060)	
Total Score Pretest							0.565*** (0.057)
No. of observations	142	142	142	142	142	142	142
No. of groups	22	22	22	22	22	22	22

*$p < .05$. **$p < .01$. ***$p < .001$.

Benjamini–Hochberg method to adjust for multiple comparisons to control for a false-discovery rate.[1] The results suggest that even after adjusting for multiple comparisons across seven IdO and four Logramos tests, these statistically significant results held. In other words, our analyses indicated that the treated students served by DLL outperformed the control students on all Spanish assessments (i.e., all IdO and Logramos measures).

Discussion

The results from the first cohort of our national randomized controlled trial of DLL indicate that treatment students served by DLL outperformed control students on all IdO and Logramos measures. In addition, although the effect sizes on the English-language assessment, ITBS, did not meet conventional levels of statistical significance, all treatment effect estimates were positive and ranged from a standardized effect size of 0.12 to 0.29. Indeed, according to the U.S. Department of Education's WWC conventions, the 0.29 effect size for the ITBS Vocabulary subtest was greater than the 0.25 criterion—suggesting that the impact may be substantial enough to recognize as practically or *educationally significant*, even in the absence of achieving a traditional criterion of *statistical significance* (Institute of Education Sciences, 2017).

This study provides valuable experimental estimates of the literacy impacts of the DLL program. Similar to the recent findings of Sirinides et al. (2018) for DLL's sister English-language program, Reading Recovery, DLL demonstrates a clear benefit to students across many dimensions of literacy in reliable ways that are replicated from student to student and teacher to teacher. Moreover, these observed effects are, in general, of considerable magnitude. Relative to the findings of Slavin et al. (2011), whose synthesis of the evidence concluded that one-on-one tutoring programs, in general, improve the reading achievement of struggling readers by an effect size of 0.38 relative to the performance of control groups, our findings suggest that the impact of DLL on students' Spanish-language test outcomes was nearly twice that magnitude, with an average effect size of approximately 0.66. As another potentially informative reference, Hill, Bloom, Black, and Lipsey (2008), concluded that the average amount of achievement growth experienced by the typical U.S. first grader, after receiving a full year of instruction (as measured by typical U.S. achievement tests), is equivalent to an effect size of 0.97. After exposure to only 12 to 20 weeks of DLL, students' Spanish-language achievement grew by approximately two thirds that amount, and actually exceeded a *full year* of first-grade achievement growth for two of the eleven outcomes that we measured.

Previous studies of DLL have not included assessments of English-language acquisition and achievement and, thus,

TABLE 7
Hierarchical Linear Model Results for Iowa Test of Basic Skills Outcomes

	Reading	Language	Vocabulary	ELA total
(Intercept)	−0.120 (0.141)	−0.035 (0.138)	−0.099 (0.144)	−0.075 (0.140)
Treatment	0.189 (0.178)	0.122 (0.176)	0.292 (0.160)	0.167 (0.178)
Reading Pretest	0.152 (0.096)			
Language Pretest		0.118 (0.091)		
Vocabulary Pretest			0.062 (0.089)	
ELA total Pretest				0.204* (0.094)
No. of observations	112	123	137	112
No. of groups	18	20	22	18

Note. ELA = English language arts.
*$p < .05$. **$p < .01$. ***$p < .001$.

TABLE 8
Hierarchical Linear Model Results for Logramos Outcomes

	Reading	Language	Vocabulary	ELA total
(Intercept)	−0.278* (0.115)	−0.204 (0.132)	−0.240 (0.122)	−0.292* (0.123)
Treatment	0.542*** (0.159)	0.407* (0.163)	0.473** (0.157)	0.548*** (0.159)
Reading Pretest	0.417*** (0.080)			
Language Pretest		0.153 (0.083)		
Vocabulary Pretest			0.239** (0.082)	
ELA Total Pretest				0.425*** (0.081)
No. of observations	124	138	141	123
No. of groups	20	21	22	20

Note. ELA = English language arts.
*$p < .05$. **$p < .01$. ***$p < .001$.

have not explored the idea that learning in one's native language can transfer to learning a second language. The DLL program serves schools that are dedicated to the idea of Spanish-language instruction, and bilingualism. School-based programs for ELLs of the type sometimes referred to as *maintenance programs* are designed to foster the minority language of the child's ethnic group and promote bilingualism and biculturalism in the pupils (Ferguson, Houghton, & Wells, 1977). Indeed, all participating schools in this study target cultivating ELL students' native language skills in Spanish with the goal that students will, ultimately, acquire both English and Spanish literacy skills as they progress through elementary school. As such, the results may more readily generalize to schools with these characteristics. ELL students in the participating schools receive approximately 80% to 90% of their instruction in Spanish. Therefore, opportunities to acquire English-language literacy skills through classroom instruction were, by design, quite limited, and expected gains in English-language skills may rely more strongly on Spanish–English language transference, than direct English-language instruction. Nevertheless, a number of studies on a variety of languages support the idea that learning literacy in one's first language augments literacy in the second (Richards-Tutor et al., 2016; Roberts, 1994). So, there is some evidence to support the idea that promoting students' literacy learning in their native language can transfer to learning in a second language.

Though these results suggest that the DLL program provided an impact that was substantially greater than the typical one-on-one tutoring program, our ongoing work will replicate this experimental study of DLL impacts across two additional student cohorts and will help determine the extent to which positive effects are reliably produced. In addition, we are hopeful that additional ongoing data collection activities will provide new insights to evaluate the quality and variability of implementation across sites and teachers. Clear, evidence-based strengths of the DLL program are its well-specified guidelines and practices, and ongoing oversight, support, and training by teacher leaders and university-based trainers. However, much may be learned by measuring such things as variability in the contexts in which DLL is implemented, the quality and quantity of DLL instruction and how it may vary, and how such contextual and programmatic variation may be associated with differences in student outcomes. Future analyses of the

longitudinal outcomes for the participating students also will help inform how these initial achievement effects are sustained over time. Finally, through planned cost analyses, we hope to inform additional questions related to cost-effectiveness, such as how the program may prevent the need for other potentially more costly interventions, including special education, continuing specialized services for English learners, and retention in grade.

Acknowledgment

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305A160060 to American Institutes for Research. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Note

1. Specifically, we rank ordered the statistically significant findings within the domain in ascending order of the p values, such that $p1 \leq p2 \leq p3 \cdots \leq p6 \leq pM$. For each p value, $p1 - M$, we computed: $p_i = \alpha_i/M$, where i is the rank for p_i, with $i = 1, 2, \ldots M$; M is the total number of findings within the domain; and α is the target level of statistical significance.

References

Abedi, J., Hofstetter, C., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, *74*, 1–28.

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the national literacy panel on language minority children and youth*. Mahwah, NJ: Lawrence Erlbaum.

August, D., & Shanahan, T. (Eds.). (2008). *Developing reading and writing in second-language learners: Lessons from the report of the National Literacy Panel on Language-Minority Children and Youth*. New York, NY: Routledge.

Bialystock, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. New York, NY: Cambridge University Press.

Boyle, A., August, D., Tabaku, L., Cole, S., & Simpson-Baird, A. (2015). *Dual language education programs: Current state policies and practices*. Washington, DC: U.S. Department of Education. Retrieved from https://ncela.ed.gov/files/rcd/TO20_DualLanguageRpt_508.pdf

Bryk, A. S., Gomez, L., & Grunow, A. (2010). *Getting ideas into action: Building networked improvement communities in education*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching. Retrieved from: http://www.carnegiefoundation.org/spotlight/webinar-bryk-gomez-building-networked-improvement-communities-in-education

Cheung, A. C., & Slavin, R. E. (2012). Effective reading programs for Spanish-dominant English language learners (ELLs) in the elementary grades: A synthesis of research. *Review of Educational Research*, *82*, 351–395.

Clay, M. (1987). Implementing Reading Recovery: Systemic adaptations to an education innovation. *New Zealand Journal of Educational Studies*, *22*(1), 35–57.

Cummins, J. (2005, September 23). *Teaching for cross-language transfer in dual language education: Possibilities and pitfalls*. Paper presented at the TESOL Symposium on Dual Language Education: Teaching and Learning Two Languages in the EFL Setting, Bogazici University, Istanbul, Turkey. Retrieved from http://www.wakeed.org/wp-content/uploads/2014/04/tesol.turkey.pdf

D'Agostino, J. V., & Murphy, J. A. (2004). A meta-analysis of Reading Recovery in United States schools. *Educational Evaluation and Policy Analysis*, *26*, 23–38.

DeFord, D. (2007). Marie Clay: An honored mentor, colleague, and friend. *Journal of Reading Recovery*, *7*(1), 101–103. Retrieved from http://readingrecovery.org/images/pdfs/Journals/JRR/Vol7_No1_Fall-2007/JRR_7.1-DeFord.pdf

Durgunoglu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology*, *85*, 453–465.

Englebart, D. C. (2003, September). Improving our ability to improve: A call for investment in a new future. In *IBM Co-Evolution Symposium*. Retrieved from http://worrydream.com/refs/Engelbart%20-%20Improving%20Our%20Ability%20to%20Improve.pdf

Escamilla, K. (1994). Descubriendo la Lectura: An early intervention literacy program in Spanish. *Literacy Teaching and Learning: An International Journal of Early Literacy*, *1*(1), 57–70.

Escamilla, K. (2009). English language learners: Developing literacy in second-language learners—Report of the National Literacy Panel on Language-Minority Children and Youth. [Review of the book Developing literacy in second language learners—Report of the National Literacy Panel on Language-Minority Children and Youth, edited by D. August & T. Shanahan]. *Journal of Literacy Research*, *41*, 432–452.

Escamilla, K., Andrade, A., Basurto, A., & Ruiz, O. (1996). *Instrumento de observacion de los logros de la lecto-escritura inicial: Spanish reconstruction of an observation survey a bilingual text*. Portsmouth, NH: Heinemann.

Escamilla, K., Loera, M., Ruiz, O., & Rodríguez, Y. (1998). An examination of sustaining effects in Descubriendo la Lectura programs. *Literacy Teaching and Learning: An International Journal of Early Reading and Writing*, *3*(2), 59–81.

Ferguson, C. A., Houghton, C., & Wells, M. H. (1977). Bilingual education: An international perspective. In B. Spolsky & R. Cooper (Eds.), *Frontiers of bilingual education: An international perspective* (pp. 159–194). Rowley, MA: Newbury House.

Forbes, S., & Doyle, M. A. (2004). How Reading Recovery teaches the five essential elements of reading instruction and more: Second of a two-part series. *Journal of Reading Recovery*, *3*(2), 16–30.

Francis, D., Lesaux, N. K., & August, D. (2006). Language of instruction. In D. L. August & T. Shanahan (Eds.), *Developing literacy in a second language: Report of the National Literacy Panel* (pp. 365–410). Mahwah, NJ: Erlbaum.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth

curves analysis. *Journal of Educational Psychology*, *88*, 3–17. doi:10.1037/0022-0663.88.1.3

Fullan, M. (1993). *Change forces: Probing the depths of educational reform*. New York, NY: Routledge.

Galambos, S. J., & Hakuta, K. (1988). Subject-specific and task-specific characteristics of metalinguistic awareness in bilingual children. *Applied Psycholinguistics*, *9*, 141–162.

Galindo, C., & Reardon, S. F. (2006). *Latino students' educational experiences and opportunities during kindergarten*. Tempe, AZ: National Task Force on Early Childhood Education for Hispanics.

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

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

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*, 172–177.

Institute of Education Sciences. (2017, October). *Standards handbook: Version 4.0*. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Jimenez, R. T. (1997). The strategic reading abilities and potential of five low-literacy Latina/o readers in middle school. *Reading Research Quarterly*, *32*, 224–232.

Jimenez, R. T., Garcia, G. E., & Pearson, P. D. (1996). The reading strategies of Latina/o students who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly*, *31*, 90–112.

Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grade. *Journal of Educational Psychology*, *80*, 437–447.

Knudsen, E., Heckman, J., Cameron, J., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 10155–10162.

Lindholm-Leary, K., & Block, N. (2010). Achievement in predominantly low SES/Hispanic dual language schools. *International Journal of Bilingual Education and Bilingualism*, *13*, 43–60.

May, H., Gray, A., Gillespie, J., Sirinides, P., Sam, C., Goldsworthy, H., . . . Tognatta, N. (2013). *Evaluation of the i3 scale-up of Reading Recovery: Year one report (2011–2012)*. Philadelphia, PA: Consortium for Policy Research in Education. Retrieved from http://www.cpre.org/sites/default/files/researchreport/1488_readingrecoveryreport.pdf

May, H., Gray, A., Sirinides, P., Goldsworthy, H., Armijo, M., Sam, C., . . . Tognatta, N. (2015). Year one results from the multisite randomized evaluation of the i3 scale-up of Reading Recovery. *American Educational Research Journal*, *52*, 547–581.

McField, G., & McField, D. (2014). The consistent outcome of bilingual education programs: A meta-analysis of meta-analyses. In G. McField (Ed.), *The miseducation of English learners: A tale of three states and lessons to be learned* (pp. 267–299). Charlotte, NC: Information Age.

Nagy, W., Berninger, V., Abbott, R., Vaughan, K., & Vermeulen, K. (2003). Relationship of morphology and other language skills to literacy skills in at-risk second graders and at-risk fourth grade writers. *Journal of Educational Psychology*, *95*, 730–742.

National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the educational success of children and youth learning English: Promising futures*. Washington, DC: National Academies Press. doi:10.17226/24677

National Clearinghouse for English Language Acquisition. (2011). *The growing numbers of English learner students, 1998/99–2008/09*. Washington, DC: Author.

National Institute of Child Health and Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (Reports of the subgroups). Bethesda, MD: Author. Retrieved from http://www.nichd.nih.gov/publications/pubs/nrp/documents/report.pdf

Neal, J., & Kelly, P. (1999). The success of Reading Recovery for English language learners and Descubriendo la Lectura for bilingual students in California. *Literacy Teaching and Learning*, *4*(2), 81–108.

Office of English Language Acquisition. (2015). *English Learners (ELs) and NAEP: Fast facts*. Retrieved from https://www2.ed.gov/about/offices/list/oela/fast-facts/elnaep.pdf

Portes, A., & Hao, L. (2002). The price of uniformity: Language, family and personality adjustment in the immigrant second generation. *Ethnic and Racial Studies*, *25*, 889–912.

Proctor, C. P., Dalton, B., & Grisham, D. L. (2007). Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research*, *39*, 71-93.

Quiroga, T., Lemos-Britton, Z., Mostafapour, E., Abbott, R. D., & Berninger, V. W. (2001). Phonological awareness and beginning reading in Spanish-speaking ESL first graders: Research into practice. *Journal of School Psychology*, *40*, 85–109.

Ramírez, J., Yuen, S., Ramey, D., & Pasta, D. (1991). *Final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit bilingual education programs for language minority children* (Vol. I). San Mateo, CA: Aguirre International.

Reading Recovery Council of North America. (2017a). *Descubriendo la Lectura* [Web page]. Retrieved from http://readingrecovery.org/descubriendo-la-lectura

Reading Recovery Council of North America. (2017b). *Standards and guidelines of reading recovery in the United States* (8th ed.). Worthington, OH: Author. Retrieved from https://readingrecovery.org/wp-content/uploads/2018/06/RR_Standards_and_Guidelines_8ED_2017.pdf

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*, 853–891. doi:10.3102/0002831209333184

Richards-Tutor, C., Baker, D. L., Gersten, R., Baker, S. K., & Smith, J. M. (2016). The effectiveness of reading interventions for English learners: A research synthesis. *Exceptional Children*, *82*, 144–169.

Roberts, C. A. (1994). Transferring literacy skills from L1 to L2: From theory to practice. *Journal of Educational Issues of Language Minority Students*, *13*, 209–221.

Rolstad, K., Mahoney, K., & Glass, G. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy*, *19*, 572–594.

Schneider, B., Martinez, S., & Ownes, A. (2006). Barriers to educational opportunities for Hispanics in the United States. In National Research Council (US) Panel on Hispanics in the United States, M. Tienda, & F. Mitchell (Eds.), *Hispanics and the future of America*. Washington, DC: National Academies Press. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK19909/?report=printable

Sirinides, P., Gray, A., & May, H. (2018). The Impacts of Reading Recovery at scale: Results from the 4-year i3 external evaluation. *Educational Evaluation and Policy Analysis*, *40*, 316–335.

Slavin, R., & Cheung, A. (2004). How do English language learners learn to read? *Educational Leadership*, *61*(6), 52–57.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, *6*, 1–26.

Thomas, W., & Collier, V. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Santa Cruz, CA: Center for Research on Education, Diversity and Excellence. Retrieved from http://files.eric.ed.gov/fulltext/ED475048.pdf

Umansky, I., & Reardon, S. F. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal*, *51*, 879–912.

U.S. Department of Education. (2001). *Evidence that tutoring works*. Retrieved from https://www.gpo.gov/fdsys/pkg/ERIC-ED464343/pdf/ERIC-ED464343.pdf

Valentino, R. A., & Reardon, S. F. (2015). Effectiveness of four instructional programs designed to serve English learners: Variation by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis*, *37*, 612–637. doi:10.3102/0162373715573310

Vom Dorp, I. E. (2000). *Biliteracy, monoliteracy and self-concept in native Spanish-dominant and native English-dominant fifth graders* (Unpublished doctoral dissertation). University of New Mexico, Albuquerque.

Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities*, *51*, 612–624. doi:10.1177/0022219418775110

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, *36*, 541–561.

Authors

TRISHA H. BORMAN is a managing director at the American Institutes for Research. Her research interests are most concentrated in quantitative methodology, English Language Learners, and issues of educational equity.

GEOFFREY D. BORMAN is a Vilas Distinguished Achievement Professor at the University of Wisconsin–Madison. His research interests are education policy and research methodology.

SCOTT HOUGHTON is a researcher at the American Institutes for Research. His research interests are education policy and early childhood education.

SO JUNG PARK is a researcher at the American Institutes for Research. Her research interests are the development and application of advanced statistical methods for education policy research.

BO ZHU is a senior researcher at the American Institutes for Research. Her research interests are international education policy and educator effectiveness.

ALEJANDRA MARTIN is a research associate at the American Institutes for Research. Her research interests are early childhood development and language acquisition/learning.

SIDNEY WILKINSON-FLICKER is a researcher at the American Institutes for Research. Her main research interest is in early childhood education.