Routledge
Taylor & Francis Group

Journal of Research on Educational Effectiveness

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uree20

The Spanish- and English-Language Literacy Impacts of *Descubriendo la Lectura* across Three Experimental Replications

Trisha H. Borman, Geoffrey D. Borman, So Jung Park, Bo Zhu & Scott Houghton

To cite this article: Trisha H. Borman, Geoffrey D. Borman, So Jung Park, Bo Zhu & Scott Houghton (09 Nov 2023): The Spanish- and English-Language Literacy Impacts of *Descubriendo la Lectura* across Three Experimental Replications, Journal of Research on Educational Effectiveness, DOI: 10.1080/19345747.2023.2269936

To link to this article: https://doi.org/10.1080/19345747.2023.2269936

Published online: 09 Nov 2023.

JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
https://doi.org/10.1080/19345747.2023.2269936

Routledge
Taylor & Francis Group

INTERVENTION, EVALUATION, AND POLICY STUDIES

Check for updates

The Spanish- and English-Language Literacy Impacts of *Descubriendo la Lectura* across Three Experimental Replications

Trisha H. Borman[a], Geoffrey D. Borman[b], So Jung Park[c], Bo Zhu[a] and Scott Houghton[a]

[a]Human Services Division, American Institutes for Research, Arlington, Virginia, USA; [b]Division of Leadership and Innovation, Arizona State University, Tempe, Arizona, USA; [c]Department of Education, Korea University, Seongbuk-gu, Korea (The Republic of)

ABSTRACT

We present findings across three randomized trials of *Descubriendo la Lectura* (DLL), an intervention designed to improve literacy skills of Spanish-speaking first graders struggling with reading. DLL is a one-to-one Spanish-language literacy program lasting 12–20 weeks offered to a school's lowest performing emerging bilingual first graders. Directly replicating procedures across three cohorts, we examined Spanish and English literacy outcomes for approximately 400 students across 30 schools, 10 districts, and five states. Analyses revealed statistically significant effects of student-level assignment to DLL on all 11 Spanish-language outcomes, with an average impact of $d = 0.57$. Impacts across the four English-language outcomes were positive but not statistically significant. Descriptive results suggest that the Spanish-language impacts are reliably replicated, most likely due to the intervention's strong implementation supports.

ARTICLE HISTORY
Received 22 March 2023
Revised 28 July 2023
Accepted 29 August 2023

KEYWORDS
Early literacy; English learners; high-intensity tutoring; randomized controlled trial; replication

Prevailing educational policies and practices have failed to capitalize on the strengths of Latinx English learners (ELs), including the unique potential of bilingualism—a versatile asset with cognitive, social-emotional learning, and future employment benefits (Bialystock, 2001; Chen & Padilla, 2019; Nagy et al., 2003; Portes & Hao, 2002; Vom Dorp, 2000). A consensus of evidence indicates that the most effective strategy for cultivating the early literacy skills of emerging bilingual students is a strengths-based approach, which draws on the Spanish-language skills that students possess as they begin formal schooling. Research suggests that early elementary bilingual programs produce stronger long-term English proficiency and English language arts (ELA) achievement compared with English-only programs while simultaneously developing students' Spanish-language assets and promoting bilingualism (Francis et al., 2006; National Academies of Sciences Engineering & Medicine, 2017; Rolstad et al., 2005; Slavin & Cheung, 2004; Umansky & Reardon, 2014; Valentino & Reardon, 2015). To take

CONTACT Trisha H. Borman tborman@air.org Human Services Division, American Institutes for Research, Arlington, Virginia, USA.

advantage of the strengths associated with bilingual education, educators need proven programs and strategies that enrich opportunities for ELs.

As of 2019, ELs accounted for more than five million (10%) of the total K–12 student population, with a majority being Spanish speakers (77%) and approximately 15% of ELs concentrated in first grade (de Brey et al., 2022; National Center for Education Statistics, 2022). As the fastest growing population of students, a recent projection suggested that ELs will comprise one of every four students in the United States by 2025 (National Education Association, 2020). As the Spanish-speaking EL population in the United States continues to grow (National Center for Education Statistics, 2022), schools have increasingly turned to bilingual programming, particularly dual-language approaches, to support students' language needs (Arteagoitia & Yen, 2020; Escamilla, 2018). With the significant historical and current barriers to school opportunities and success that Latinx ELs have faced, such as poorly funded schools, suboptimal learning environments, less experienced teachers, and English-only programming (Galindo & Reardon, 2006; Gandara & Hopkins, 2010; Reardon & Galindo, 2009), effective Spanish-language literacy supports will be more important than ever in the coming years.

Indeed, all students, including ELs, face literacy challenges during their first years of formal schooling. Intervening early, when many ELs confront reading and language difficulties simultaneously, can mitigate longer term future literacy challenges that are more difficult to resolve (Cheung & Slavin, 2005; Francis et al., 1996; Knudsen et al., 2006). Providing literacy supports in first grade can be particularly beneficial, before reading complexity increases (Wanzek et al., 2018; Wanzek & Vaughn, 2007) and before bilingual programs transition to increasing the use of English instruction (Calderón & Slakk, 2019). What scalable evidence-based solutions are available to support the beginning-literacy skill development and school success of those Spanish-language ELs who are struggling?

Our review of the What Works Clearinghouse (WWC) website,[1] multiple meta-analyses (Ludwig et al., 2019; Richards-Tutor et al., 2016), and research reviews (August et al., 2014; August & Shanahan, 2006; Baker et al., 2014; Cheung & Slavin, 2005, 2012; Gersten et al., 2007; Snyder et al., 2017) revealed very few proven programs designed to support early elementary ELs who are struggling with reading. One review, in fact, prominently noted that there were "dramatically" fewer studies of quality reading interventions for ELs relative to those for native English speakers (Richards-Tutor et al., 2016, p. 161). Indeed, our review identified only three programs with any evidence of impact on emerging bilingual students' Spanish-language outcomes: *Enhanced Proactive Reading* (currently distributed by McGraw Hill Education as *Intervenciones tempranas de la lectura*; see Vaughn, Cirino, et al., 2006; Vaughn, Linan-Thompson, et al., 2006); the former Spanish-language adaptation of Success for All, *Exito Para Todos* (see Dianda & Flaherty, 1995; Nunnery et al., 1997); and the intervention that is the subject of our study here, *Descubriendo la Lectura* (DLL).

[1]See https://ies.ed.gov/ncee/wwc.

The DLL Program

DLL is the Spanish-language reconstruction of Reading Recovery (RR), providing one-to-one literacy tutoring in Spanish to first-grade ELs struggling to learn to read. The DLL program comprises material and personnel resources, advanced training and ongoing professional development (PD), and high-quality student instruction designed to improve literacy skills and comprehension for emerging bilingual students for whom Spanish is their first language. The overall DLL theory of change is summarized in Figure 1.

Resources

At the top left side of Figure 1, which is the heart of DLL, are several evidence-based literacy strategies, including early intervention, native language instruction, essential research-based reading components, tutoring, and self-directed learning. Also, as highlighted among the other resources summarized within the left panel, DLL calls for rigorous teacher selection and training standards, along with a personnel structure designed to include coordination across a district, a school, and the DLL teacher. Finally, a variety of proprietary material resources support the delivery of DLL as a replicable program.

Early Intervention

DLL reaches students in first grade before reading complexity increases and many bilingual programs transition to more English instruction (Calderón et al., 2011; Wanzek & Vaughn, 2007). Research suggests that students face substantial difficulties catching up if they fall behind in reading in first grade (Juel, 1988; Lyon & Fletcher, 2001; Shaywitz et al., 1999). Investing in early literacy instruction has profound, positive effects on

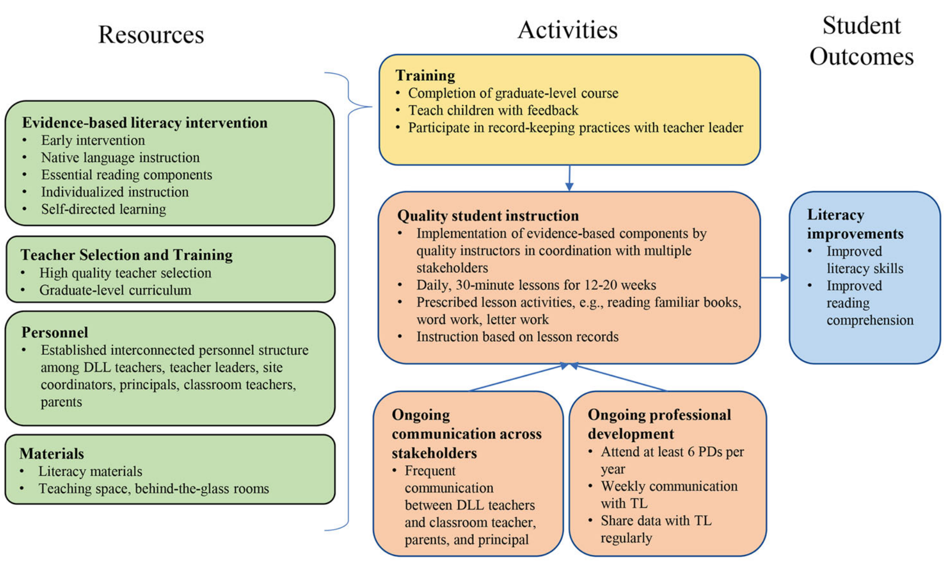

Figure 1. Theory of change for the Descubriendo la Lectura program.

students' neurological and skill development and long-term academic trajectories (Cheung & Slavin, 2005; Francis et al., 1996; Juel, 1988; Knudsen et al., 2006).

Native Language Learning

Aligned to early instruction in most bilingual programs (Calderón et al., 2011), DLL is provided in a student's native language. Prior research (Francis et al., 2006; Greene, 1997; McField & McField, 2014; Rolstad et al., 2005; Slavin & Cheung, 2004; Willig, 1985) indicates that teaching students in both English and their native language produces stronger literacy results than English-only programs. Research also demonstrates that fostering children's first language development has positive impacts on acquiring the second language (August & Hakuta, 1997; Cummins, 2005; Durgunoğlu et al., 1993; Gottardo, 2002). Differing amounts of English instruction can support initial learning, but literacy programs using native language instruction promote long-term academic gains (Ashdown & Simic, 2000).

Essential Components of Reading Instruction

DLL lessons are grounded in Marie Clay's scholarship on literacy development (Clay 1993, 2005). Clay (2005) argued that a child challenged by literacy learning requires a skilled teacher who is professionally trained and who embraces the complexity of literacy learning required for supporting the range of diversity among literacy learners. DLL incorporates the National Reading Panel's essential elements—phonemic awareness, phonics, fluency, vocabulary, and text comprehension (National Institute of Child Health & Human Development, 2000)—which are emphasized across student lessons.

Tutoring

Research supports DLL's use of one-to-one instruction (i.e., tutoring) for students struggling with reading, including ELs (Cheung & Slavin, 2012; Slavin et al., 2011). Several DLL components align to evidence-based tutoring practices: coordination with the classroom teacher, intensive tutor training, well-structured sessions, progress monitoring, and a daily cadence (Planning and Evaluation Service, Office of the Deputy Secretary, 2001). DLL instruction is tailored to the individual needs of each student. For example, teachers use running records that identify the source(s) of students' challenges (e.g., meaning, syntax), serve as the basis for coaching conversations with their teacher leader, and inform the next day's lesson. This individualized approach highlights a child's ever-changing processing and the teacher's observation, adaptation, and deep understanding of that child's strengths (Reading Recovery Council of North America, 2017).

Self-Directed Learning

Aligned to Clay's emphasis on independent learning (Clay, 2001), DLL fosters active learning through opportunities to problem-solve, self-monitor, read books, and write messages with substantial independence and teachers serving as a guide (DeFord, 2007). DLL aims to support sustainable literacy gains by providing opportunities to employ writing, motivation, and independence (May et al., 2013).

Teacher Selection and Training

Rather than a program that can be purchase, DLL is an "investment in the development of teacher expertise" (Reading Recovery Council of North America, 2017, p. 5). High teaching standards begin with selection criteria for teachers: they must hold state-required certification; have at least 3 years of overall teaching experience; and have demonstrated, through teacher evaluation, successful bilingual teaching (Reading Recovery Council of North America, 2017, p. 8). When entering the program, teachers receive training from teacher leaders in a yearlong master's level RR class and then "bridge" to DLL in the second year, during which they must teach lessons and implement all aspects of DLL. Alternatively, if enough teachers are training in DLL at once, a teacher leader may choose to train teachers directly in DLL.

Personnel

DLL devotes substantial attention to ensuring that the program is implemented in coordination with school systems (May et al., 2013). It calls for a district-level site coordinator who supports and sustains implementation of DLL within a district and facilitates communication with teacher leaders who support DLL teachers and work with school-level administrators to ensure that DLL is integrated within schools, rather than being an "add-on" program. Within schools, the program calls for coordination among DLL teachers and principals, classroom teachers, and parents.

Materials

In addition to using the validated *Instrumento de Observación* (*IdO*), a Spanish-language diagnostic assessment used by teachers to measure student skill levels and progress, teachers use several other forms to assess and monitor student progress, including a running record form that allows them to calculate error ratios and accuracy rates and to determine which sources of information (e.g., meaning, syntax, visual) children are using when they make errors or self-corrections. Additional monitoring forms allow teachers to measure change across time in text level and known reading and writing vocabulary. A daily lesson form provides a platform for teachers to plan daily lessons, take notes on their systematic observations during the lesson, and make instructional decisions both in the moment and for future lessons. Teachers also have an extensive collection of books so that students can read a new book during each lesson, and they also employ engaging manipulatives such as word cutouts for story creation and assembly.

Activities

With these resources as a foundation, DLL's implementation relies on rigorous teacher training that supports quality student instruction through the evidence-based intervention. This instruction is reinforced by ongoing PD and communication across key stakeholders. These central DLL program activities are in the middle panel of Figure 1.

Training

The research on tutoring supports interventions delivered by trained teachers (Slavin et al., 2011) and emphasizes intensive training (Planning and Evaluation Service, Office of the Deputy Secretary, 2001). DLL training requires the completion of yearlong graduate-level coursework. As they train, DLL teachers also teach students and ultimately "demonstrate effective teaching" (Reading Recovery Council of North America, 2017, p. 11). The graduate-level courses focus on theory, an analysis of children's reading and writing behaviors, and the design of individualized lessons tailored to the specific needs of individual children. In addition, DLL teachers receive direct training from DLL teacher leaders who observe student lessons through a one-way glass.

Quality Student Instruction

With qualified teachers trained to implement an evidence-based intervention, DLL provides quality student instruction. Trained literacy teachers provide daily 30-minute, one-to-one instruction to DLL students for 12–20 weeks, with the goal of meeting the literacy skills of their peers (Reading Recovery Council of North America, 2021). Prescribed lesson activities include rereading familiar books; reading a recently assigned book while teachers take a running record; working with letters or words using magnetic letters; writing a story; assembling a child's cutup story; and reading a new, strategically selected book (Reading Recovery Council of North America, 2021). Lessons also include instruction on letter and sound relationships, spelling patterns, comprehension, producing oral language, and increasing vocabularies (Clay, 2001; Reading Recovery Council of North America, 2021). Teachers are required to keep records of their student's progress as a basis for future lessons and submit data to their assigned teacher leader to inform coaching discussions.

Ongoing Professional Development

Ongoing PD and coaching is another cornerstone of DLL often associated with student success (Slavin & Cheung, 2004; Joyce & Showers, 2002; Planning and Evaluation Service, Office of the Deputy Secretary, 2001). Throughout their involvement with DLL, DLL teachers must participate in at least six PD sessions each year, and teachers are expected to participate in behind-the-glass sessions, in which DLL teacher leaders observe their instruction and provide feedback. DLL teacher leaders coach DLL teachers during the school year by discussing strategy, observing instruction, and/or modeling instructional approaches.

Ongoing Professional Development and Communication across Stakeholders

In addition to an interconnected personnel structure between the district and school, DLL requires that DLL teachers work closely with building administrators, classroom teachers, and the parents of individual students (Reading Recovery Council of North America, 2017). Rather than an isolated intervention within a school, the program aims to foster communication across stakeholders to support student learning. Implementation research supports this coordinated approach, including the importance of school-level administration support (Bryk et al., 2010), collaboration (Patterson et al., 2018), and coordination with classroom teachers (Planning and Evaluation Service, Office of the Deputy Secretary, 2001). Along with ongoing PD, these communications across stakeholders help support quality student instruction.

Student Outcomes

When these activities (training, quality instruction, ongoing PD, and communication) are implemented as intended, they support the student literacy outcomes, which appear on the right-hand side of Figure 1. That is, DLL teachers implement an evidence-based one-on-one tutoring intervention in coordination with the school community, which supports students' Spanish-language literacy skill development, including skills such as decoding, vocabulary building, and fluency, to help promote overall reading, writing, and oral proficiency. This focus on heritage language development supports emerging bilingual students and facilitates their later development of English-language skills.

Prior Evidence of DLL Impacts on Student Outcomes

Two prior randomized controlled trials (RCTs) demonstrated consistently large and positive impacts of DLL on the Spanish-language literacy outcomes of ELs (Borman et al., 2019, 2020), with both studies receiving the WWC's highest evidence rating—*meets standards without reservations*—in recently published individual study reviews.[2] The first RCT, conducted in 2011–12 (Borman et al., 2020), comprised 187 (98 treatment and 89 control) students who were served by 27 previously trained DLL teachers in 19 schools in two districts from two states. The researchers used two student assessments to measure academic progress: the *IdO* (Escamilla et al., 1996)[3] and the *Logramos* (2006).[4] Treatment students served by DLL outperformed wait-list control students on all *IdO* and *Logramos* measures. Effect sizes (d) for *IdO* outcomes ranged from $d = 0.48$ to $d = 1.24$, and effect sizes for the broad-based, standardized *Logramos* outcomes ranged from $d = 0.36$ to $d = 0.48$.

The more recent study conducted in 2016–17 (Borman et al., 2019) included approximately 150 students from 22 schools across three states. In addition to the *IdO* and *Logramos* assessments, this study included the broad-based English-language ITBS. The findings revealed results similar to those in the prior study, with statistically significant student-level impacts of assignment to DLL services on all *IdO* posttest outcomes, ranging from $d = 0.36$ to $d = 0.97$ across the six literacy subtests, and for the four *Logramos* subtests, ranging from $d = 0.18$ to $d = 0.34$. The impacts on Spanish-language outcomes did not translate into statistically significant English literacy impacts as measured by the ITBS, but this result is not surprising given DLL's focus on students' Spanish-language literacy development.

[2]The WWC reviews of these studies are available at https://ies.ed.gov/ncee/wwc/study/89881 and https://ies.ed.gov/ncee/wwc/Study/90683.

[3]*IdO* is the Spanish version of RR's Observation Survey (Clay, 2019), which tests six literacy tasks (letter identification, word test [i.e., reading vocabulary], concepts about print, writing vocabulary, phonemic awareness, and text reading) that, taken together, can measure emerging literacy abilities for early readers.

[4]The *Logramos* is an evidence-based Spanish-language literacy assessment that parallels the English-language Iowa Test of Basic Skills (ITBS) and is used to monitor growth and achievement in Spanish literacy skills.

Method

In this section, we report the final results of the replicated findings across three independent yearly cohorts from 30 schools across 10 districts in five states: Arizona, California, Illinois, Texas, and Wisconsin. With direct replications of the intervention and research methods across three independent first-grade cohorts, this study offers the largest body of evidence to date about the overall Spanish-language and English-language impacts of DLL and the relative consistency of those impacts across replications. In addition to describing the overall impacts, we describe the extent to which those impacts varied across the 30 participating schools.

Sample

Schools were recruited to participate in this study if they had been implementing DLL for at least 1 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 students in their schools as determined by the *IdO*. (See the Measures section for a detailed description of all measures.) The combined sample across the three cohorts—2016–17, 2017–18, and 2018–19—included nearly 400 first-graders who were struggling readers, and whose home language was Spanish, nested within the 30 schools. The Institutional Review Board (IRB) of the American Institutes for Research® (FWA00003952, IRB00000436) reviewed the project for compliance with the Common Rule and approved it under determination 85649/04410. The IRB also determined that there were no risks greater than those ordinarily encountered in daily life or during routine testing or instructional activities. Because the conditions stipulated in 45 CFR §46.117(c) were met, a waiver of documentation of consent was granted.

All sampled schools offered transitional bilingual programs, which provided Spanish-language regular-classroom reading and language arts instruction to their first-grade ELs. DLL teachers provided supplemental one-to-one Spanish-language literacy tutoring to approximately four students each semester, and each child progressed through the program at an individualized pace, generally lasting 12–20 weeks. All students were native Spanish speakers.

Study Design

For this study, we employed an RCT design in which the lowest performing first-grade native Spanish-speaking students were randomly assigned to immediate treatment or a delayed, poststudy treatment condition. Our study focused on the fall semester outcomes for students and contrasted the pre-post outcomes for the immediately treated students to the pre-post outcomes of those students whose treatment was delayed until after the study was completed. Following typical DLL practices, during the fall of each year, DLL teachers administered the *IdO*, a Spanish literacy assessment, to all Spanish-speaking students who were identified by their classroom teachers as performing below grade level in literacy in their school. Based on the *IdO* results, the lowest performing students at each campus were identified and randomized to an immediate or a poststudy delayed DLL treatment condition. Those students randomized to the

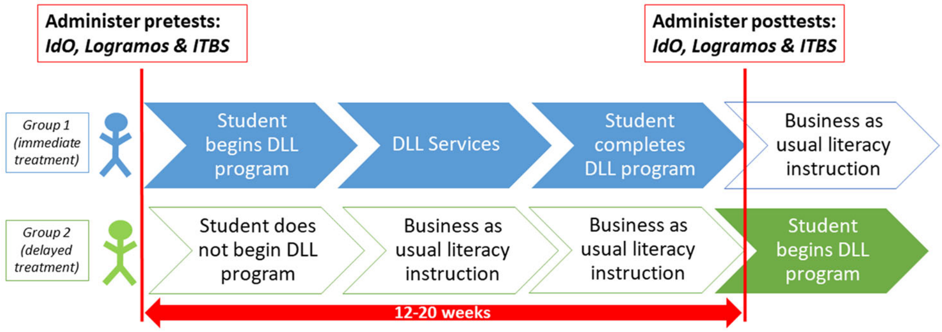

Figure 2. Timeline of DLL program delivery and student assessments.

immediately treated condition received DLL services at the beginning of the academic year, and those in the delayed condition received services 12–20 weeks later—as the immediately served students exited the program.

The researchers achieved this randomized design by generating a random number for each eligible child (i.e., the lowest performing ELs) at the beginning of the school year. Teachers served those with lower randomly assigned numbers first, and those students with higher numbers remained on a waiting list and were served later, at the conclusion of the study period. As each immediately treated student completed DLL, the teachers served the next child with the lowest randomly assigned number on the waiting list. As students assigned to immediate service during the fall completed DLL, they were tested using three instruments: the *IdO*, *Logramos,* and the ITBS. At the same time, prior to entry into DLL, the next child on the randomly determined waiting list also was administered the three instruments. All students—both immediate and delayed treatment—were administered the *IdO*, *Logramos*, and ITBS assessments at baseline in August or September, prior to being assigned to DLL, and all students were again tested with all three instruments after 12–20 weeks—between November and February. The timeline for program delivery and testing for the immediate- and delayed-treatment students appears in Figure 2.

Measures

The three student assessments were the *IdO* (Escamilla et al., 1996), the *Logramos* reading achievement test, and the ITBS. The *IdO* is the Spanish-language version of the *Observation Survey of Early Literacy Achievement* (*OS*; Clay, 1993). The *OS* and the *IdO* measure student performance on the following six subtests: Letter Identification, Ohio Word Test, Concepts About Print, Writing Vocabulary, Hearing and Recording Sounds in Words, and Text Reading Level (Clay, 2005). Both *OS*, for English-language students, and *IdO*, for Spanish-language students, are administered as a regular component of the Reading Recovery and DLL programs, respectively. The *OS* and the *IdO* tests are administered at the beginning of the first-grade school year as a screener to determine eligibility for RR and DLL services and after students complete services to measure academic progress. The *IdO* also has been used to examine student outcomes in prior research studies, including Escamilla et al. (1996).

IdO is not a literal translation of the English *OS* but rather a conceptual recreation which measures literacy skill acquisition for young native Spanish speakers. Both *OS* and *IdO* are tools designed for the systematic observation of young children's early literacy competencies. The National Center on Intensive Intervention reviewed *OS*, giving it its highest rating of "convincing evidence."[5] Construct validity for the *IdO* was evaluated, using data from approximately 500 first-grade students, yielding Cronbach's alphas from $\alpha = 0.51$ to $\alpha = 0.82$ (Escamilla et al., 1996). Concurrent validity was assessed by comparing the six subtests of *IdO* with a norm-referenced test, *Aprenda: La prueba de logros en Español* (Pearson, 3rd ed.).[6] Content validity was supported through a series of translations and back translations (Escamilla et al., 1996).

Logramos is the Spanish-language version of the ITBS achievement test battery and is available from Riverside Insights.[7] The items in *Logramos* follow the scope and sequence of Form E of the ITBS and have been translated, when appropriate, from the original English items in Form E of the ITBS. Although 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 to target the same skills and maintain the underlying psychometrics of the test items.

Finally, to assess students' English-language skills, we administered the ITBS. For both the ITBS and *Logramos*, we administered three subtests during each assessment—Vocabulary/Vocabulario, Reading/Lectura, and Language/Lenguaje—which were 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 idea/ideas principals. The Language section of the assessment includes 34 items assessing Spelling/Ortografía, Capitalization/Uso de mayúsculas, Punctuation/Puntuación, and Written Expression/Expresión escrita.

The *Logramos* and *ITBS* tests are standardized, norm-referenced, and not overaligned with the intervention and, therefore, provide fair assessments of achievement change for both treatment- and control-group students. Both norm-referenced tests were administered as additional assessments for the purposes of the study and are not typically given to DLL students under routine, nonstudy conditions. All three assessments (*IdO*, *Logramos*, and ITBS posttests) were proctored by a DLL teacher leader or another teacher who did not provide DLL or classroom instruction to the tested students.

Beyond these test scores, we collected demographic data for each student through the International Data Evaluation Center (2012) at The Ohio State University, which gathers data on all RR and DLL sites and students each year.

[5]See https://charts.intensiveintervention.org/screening/tool/?id=77c5c64492268897.
[6]See https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Academic-Learning/Brief/Aprenda%3A-La-Prueba-de-Logros-en-Espa%C3%B1ol-%7C-Tercera-edici%C3%B3n/p/100000585.html.
[7]See https://riversideinsights.com/log_tercera.

Program Implementation

For this study, the program was implemented under routine conditions that would support any typical DLL implementation. The services provided one-on-one tutoring in Spanish for first-grade Spanish-speaking students receiving regular-classroom literacy instruction in Spanish and experiencing substantial challenges with reading and writing. The program offered daily 30-minute instruction for 12–20 weeks for students performing in the bottom 20–25% of their class, as specified by the DLL protocol (Reading Recovery Council of North America, 2017). DLL-trained literacy teachers provided intensive, individualized support as a complement to students' regular-classroom instruction, with the goal of raising students' literacy skills to the level of their peers (Reading Recovery Council of North America, 2017). DLL teachers worked with classroom teachers to avoid providing this supplemental literacy instruction during the regular classroom literacy block. Typically, DLL services were offered at times that did not interrupt students' core academic instruction.

As the delayed-treatment students awaited DLL services, they experienced business as usual literacy instruction, which comprised Spanish-language regular classroom literacy instruction and, in some cases, small-group supplemental reading instruction or, in a smaller number of cases, special education services. The DLL teachers monitored the academic progress of all their assigned immediate- and delayed-treatment students throughout the school year, occasionally providing supplemental small-group instruction to their delayed-treatment students. However, only immediate-treatment students received the one-to-one DLL programming during the first 12–20 weeks of the school year.

Features of the DLL Daily Lessons

Teachers had flexibility to adapt lessons based on individual student needs, but the program required that several activities occur during each student's daily 30-minute lesson. These activities included rereading familiar books; reading a recently assigned book while teachers took a running record; working with letters or words using magnetic letters; writing a story; assembling a child's cutup story; and reading a new, strategically selected book (Reading Recovery Council of North America, 2017). Lessons provided teaching on letter and sound relationships, spelling patterns, comprehension, producing oral language, and increasing vocabularies (Clay, 2001; Reading Recovery Council of North America, 2017). Rather than a prescribed curriculum, however, teachers had the flexibility to adjust lessons based on each child's strengths and needs, offering support and strategic guidance as needed. Further, teachers provided extensive opportunities for student problem solving and self-correction with limited interruption (Clay, 2001).

Individually tailored instruction is a key defining feature of the program. With typical implementations of DLL, the teaching decisions depend on careful observation and an understanding of student progress; thus, program implementation relies on and demands extensive record-keeping. Teachers maintain records of each student—using data to inform their practice—by employing a tool that captures literacy progress before entering the program, when leaving the program, and at the end of the school year (the *IdO*; Reading Recovery Council of North America, 2017). Teachers also use a variety of tools to routinely assess student progress, including a planning report, daily monitoring tools, and daily assessment forms. The running records form is an example of a crucial assessment tool that allows for responsive instruction. Using this form, DLL teachers calculate error ratios, calculate accuracy rates and self-correction ratios, and determine which sources of information (meaning, syntax, visual) children used when they made errors or self-corrections. They also made notes about fluency, observed strategic processing, and the use of high frequency words. DLL teachers use the running record analysis to determine the focus of the next day's lesson.

Additional Teacher and Teacher Leader Activities

To provide high-quality, individualized instruction, the program includes rigorous teacher training and standards. The present study evaluated the program as run by already-trained DLL teachers, who had to keep complete records on each student as a basis for instruction, use the *IdO*, communicate with parents and teachers, submit data to the DLL teacher leader, and monitor the progress of students who exited the program (Reading Recovery Council of North America, 2017).

DLL teachers must regularly consult with DLL teacher leaders about those students who are not making satisfactory progress, and DLL teachers receive at least one school visit from a teacher leader annually (Reading Recovery Council of North America, 2017). In addition to their training role, teacher leaders provide program implementation oversight in schools by regularly collecting program data on student achievement, providing PD to teachers (at least four sessions per year), serving as specialists to support teachers facing struggling readers, supporting the development of school teams, and creating annual site reports for the university training center (Reading Recovery Council of North America, 2017). Teacher leaders themselves receive continuous support through PD activities, attending three sessions throughout the year: once in the fall and once in the spring at their respective university training centers and once in the summer at an annual program-sponsored institute.

Ultimately, DLL extends the RR approach (Sirinides et al., 2018) to ELs struggling with English literacy by first addressing literacy in their native Spanish language. The program aims to improve early reading and writing outcomes for the lowest achieving ELs through a focused approach that offers the needed boost to catch up to their classmates, which, ultimately, aims to improve long-term academic achievement (Reading Recovery Council of North America, 2017).

Analytical Methods

Intent-to-Treat Impact Analyses

We fit two-level hierarchical linear models (HLMs) to estimate the student-level intent-to-treat impact, or the impact of being assigned to receive DLL services, on each outcome measures. The primary analysis of interest is at the student-level because students were randomly assigned within schools in this study. To account for the nested structure of data—students nested within schools—the model includes random effects for the school-level intercept and the treatment slope. The model adjusts outcomes by using a student-level pretest measure, thus improving model fit and statistically controlling for any potential chance pretest differences between treatment and control groups. Finally, we account for the blocked random assignment by yearly cohort by including cohort fixed effects. The two-level model used for the overall impact analysis is specified as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(DLL_{ij}) + \beta_{2j}(PRETEST_{ij}) + \beta_{3j}(COHORT2_{ij}) + \beta_{4j}(COHORT3_{ij}) + r_{ij} \tag{1}$$

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{2}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{3}$$

$$\beta_{2j} = \gamma_{20} \tag{4}$$

$$\beta_{3j} = \gamma_{30} \tag{5}$$

$$\beta_{4j} = \gamma_{40} \tag{6}$$

Beginning with the student-level model (Equation (1)), Y_{ij} represents the test score of student i in school j; the model intercept, β_{0j}, represents the grand mean for control group members with average pretest scores from Cohort 1 (reference group); β_{1j} is a coefficient representing the impact of assignment to DLL, with DLL_{ij} representing an indicator variable equal to 1 for DLL and 0 for the control group. The coefficient β_2 is the association between the pretest measure $(PRETEST_{ij})$ and the outcome, which we include to reduce unexplained variance in the outcome measures and improve the precision of our treatment-effect estimates. The indicator variables $(COHORT2_{ij})$ and $(COHORT3_{ij})$, which specify the year in which a student received treatment, account for the fixed effects for each cohort, with Cohort 1 as the referent category. Finally, r_{ij} is the student-specific error term.

Equations (2)–(5) are the Level 2 school-level models. With Equation (2), we estimate the extent to which the mean achievement intercept varies across schools. With Equation (3), we estimate how the impact of assignment to DLL, β_{1j}, may vary across DLL sites. Because we have no compelling theoretical or practical reason to believe that the cross-level relationships between the student-level premeasure and outcomes and student-level cohort indicators and outcomes should vary across schools, the Level 2 equations estimate the fixed effects of these covariates. Equation (4) estimates the fixed effect of the pretest covariate, and Equations (5) and (6) estimate the fixed effects for the two cohort indicators. In these equations, the mean effects of the pretest and the cohort indicators are predicted solely by their respective school-level intercepts, γ.

Most importantly, this two-level model estimates both the overall student-level impact of receiving DLL across schools (β_{1j}) and the random variation in DLL's impact across schools, which is captured by the variance term u_{1j} in Equation (3). The student-level impact speaks to the overall effectiveness of DLL, and the key random site-level component of our model addresses the extent to which DLL impacts may vary across schools. In this respect, the model provides evidence about the overall DLL effect and the extent to which that effect may depend on differences across implementation sites.

In addition to this main impact model, we specified an exploratory model with cohort-by-treatment interaction terms to evaluate potential differences in treatment effects by cohort. The Level 1 model is as follows:

$$\begin{aligned} Y_{ij} = {} & \alpha + \beta_1(DLL_{ij}) + \beta_2(PRETEST_{ij}) + \beta_3(COHORT2_{ij}) + \beta_4(COHORT3_{ij}) \\ & + \beta_5(COHORT2{*}DLL_{ij}) + \beta_6(COHORT3{*}DLL_{ij}) + r_{ij} \end{aligned} \tag{7}$$

We specified the Level 2 equations for this exploratory model in the same way as our primary analytic model discussed previously. We treated the additional student-level cohort-by-DLL interaction terms as fixed effects at Level 2 of the model. In all models, the pre-post scores were standardized ($M = 1$; $SD = 0$). Therefore, the coefficient for the treatment variable can be interpreted directly as the covariate-adjusted standardized mean difference between treatment and control.

Cross-Site Distribution of Site Mean DLL Impacts

After estimating the overall impacts of DLL, we further explored the cross-site variation of DLL impacts across the 30 participating schools. Our previously specified multilevel impact analyses employed empirical Bayes (EB) estimates (or shrinkage estimators), which, as noted by Raudenbush and Bryk (2002), have the smallest mean squared error for predicting a specific parameter value but are biased toward the overall mean of those parameter values. As a result, the cross-site variance of EB estimates ($Var(\hat{B}^2{}_j)$) from our previous two-level HLMs with a randomly varying DLL treatment coefficient tended to understate the cross-site variance of true mean program effects (τ_B^2).

To address this overshrinkage, we employed methods outlined by Bloom et al. (2017), which, for our study, involved multiplying each EB estimate by the ratio of the random effect covariance estimate for the school-specific DLL treatment effect to the variance of the raw EB estimates. Multiplying the deviation of each EB estimate from its grand mean "stretches" these deviations to produce a sample variance that equals the estimated variance of true program effects. Because of the relatively large number of subtests, we restricted these exploratory estimates of the cross-site distribution of DLL intent-to-treat impacts to the overall total score outcomes from each literacy assessment. For *IdO*, we used the total score and for the *Logramos* and ITBS assessments, we report results for each test's ELA total score. We present descriptive statistics, including the mean and standard deviation of each outcome, and produce histograms, which display the distributions of impacts across the 30 school-based sites.

Results

Attrition

Attrition in an RCT occurs when the final analytical sample is different from the initial sample that was randomly assigned at baseline. In our case, attrition occurred only when the participating students failed to complete the assigned pretests and posttests. Both overall attrition and differential attrition, between treatment and control groups, can affect the statistical equivalence of the initial randomized sample and can potentially create bias in outcome estimates (What Works Clearinghouse, 2022). Table 1 summarizes the number of students with complete data included in the final analytical samples, the treatment and control attrition rates, and the overall and differential attrition rates. Relative to WWC's attrition standards, which describe guidelines for determining low, moderate, and high attrition, based on a combination of overall and differential attrition rates, each measure meets the criteria for "low" attrition according to the WWC

Table 1. Data attrition from randomly assigned treatment ($n = 206$) and control ($n = 195$) samples by assessment.

	Analytic sample sizes		Data attrition rates			
	Treatment	Control	Treatment (%)	Control (%)	Overall (%)	Differential (%)
IdO	197	178	4	9	6	4
Logramos	177	160	14	18	16	4
ITBS	167	156	19	20	19	1

Note. IDO: *Instrumento de Observación*; ITBS: Iowa Tests of Basic Skills.

Table 2. Baseline treatment-control equivalence for *Instrumento de Observación* (*IdO*), *Logramos*, and Iowa test of basic skills (ITBS) pretest measures.

	Treatment		Control		
IdO pretest measures	*M*	*SD*	*M*	*SD*	Standardized mean difference
Letter identification	44.61	14.54	45.20	13.34	−0.04
Sounds in words	24.19	11.26	22.90	10.94	0.12
Writing vocabulary	10.72	7.45	10.82	7.92	−0.01
Concepts about print	11.13	3.40	10.69	3.54	0.13
Ohio Word Test	7.86	5.74	7.52	5.56	0.06
Text Reading Level	1.41	1.52	1.10	1.34	0.22
Total score	463.6	39.15	457.39	41.11	0.15
	($n = 197$)		($n = 178$)		
Logramos pretest measures					
Reading	145.86	35.92	141.38	40.14	0.12
Language	169.25	14.34	169.31	17.98	0.00
Vocabulary	159.08	15.58	159.84	10.25	−0.06
ELA total	154.43	37.11	151.21	42.14	0.08
	($n = 177$)		($n = 160$)		
ITBS pretest measures					
Reading	120.41	38.29	116.57	42.54	0.09
Language	123.09	13.88	123.04	10.15	0.00
Vocabulary	123.75	16.37	122.74	16.60	0.06
ELA total	115.44	36.95	111.58	40.96	0.10
	($n = 167$)		($n = 156$)		

Version 5.0 standards (What Works Clearinghouse, 2022). Therefore, the small amount of data attrition from our analytical sample is unlikely to bias our results.

Baseline Equivalence

We also evaluated baseline equivalence between treatment and control students on all pretest measures from the *IdO*, *Logramos*, and ITBS assessments. This check for baseline equivalence helps ensure that no substantial pretreatment differences exist between the treatment and control students that may compromise analyses of their posttest outcomes. Table 2 shows the treatment and control means and standard deviations and the standardized treatment-control mean differences for each subtest. Following WWC procedures (What Works Clearinghouse, 2022s), when the difference between treatment and control groups on observable baseline characteristics is greater than 0.25 *SDs*, the two groups are judged to be not equivalent. When the difference in baseline characteristics is greater than 0.05 but less than 0.25 *SDs*, the analysis should include a statistical adjustment for the baseline measure. Differences of less than or equal to 0.05 *SDs* require no statistical adjustment. As shown in Table 2, none of the 15 pretests had a standardized mean difference greater than 0.25. To adjust for baseline differences less than 0.25 *SDs*, though, and to improve the power and precision of our impact analyses, we included the applicable pretest covariate for each analysis of the 15 achievement outcomes.

Impact Analyses

The results are presented separately for each literacy outcome. For each outcome, we first estimated a main impact model and then estimated a second model, which assesses any potential cohort-by-treatment interaction effects. As noted earlier, treatment coefficients can be interpreted as the standardized mean differences (or effect sizes) between treatment and control students on the three literacy assessments.

Instrumento de Observación

Table 3 shows the estimated DLL effects on the total score and all six *IdO* subtest outcomes: Letter Identification, Ohio Word Test, Concepts About Print, Writing Vocabulary, Hearing and Recording Sounds and Text Reading Sounds, and Text Reading Level. The results for the main impact analyses, labeled as Model 1 for each *IdO* outcome shown in Table 3, reveal statistically significant student-level impacts of assignment to DLL services on all posttest outcomes, ranging from $d = 0.36$ for the Sounds in Words outcome to $d = 0.97$ for the Text Reading Level subtest. The results for each outcome labeled Model 2 evaluate the cohort-by-treatment interaction effects. As suggested by the nonsignificant cohort-by-treatment interactions for each outcome, these exploratory analyses revealed no evidence to suggest that there were differences for the DLL impacts across the three independent yearly student cohorts.

Finally, in the bottom panel of Table 3, we summarize the random effects, including the variance of the school-level intercept, τ^2; the school-level variance of the DLL treatment effect, ζ^2; the correlation, ρ, between the school intercept and treatment effect; and the unexplained student-level residual variance, σ^2. The results for the school-level variance of the DLL impacts reveal statistically significant treatment-effect variability across the 30 schools for all *IdO* outcomes, except for the Concepts About Print subtest. For instance, the results for the *IdO* total score at the far right-side of Table 3 show an average $d = 0.91$, with a random effect covariance estimate for the school-level impacts of 0.19. Taking the square-root of this covariance estimate yields a standard deviation of 0.44.

To further explore this variability in DLL impacts across the 30 participating schools, we used methods outlined by Bloom et al. (2017) to describe the estimated variance of true DLL program effects. We restricted our estimates of the cross-site distribution of DLL intent-to-treat impacts to the total score for the *IdO* assessment, which is equivalent to $d = 0.91$ with an adjusted standard deviation of the school-specific DLL impacts of 0.58 *SDs*. Figure 3 is a histogram of the adjusted EB estimates of school-specific impact estimates for the 30 schools included in the overall analysis of *IdO* outcomes. Of the 30 schools, 29 schools (97%) have positive impact estimates, and 12 schools (40%) have impact estimates greater than 1 *SD*. The figure shows only two relative outliers,

Table 3. Multilevel estimates of Intention-to-Treat impacts on *Instrumento de Observación* outcomes.

	Letter identification		Sounds in words		Writing vocabulary		Concepts about print		Ohio Word Test		Text Reading Level		Total score	
Fixed effects	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	−0.40***	−0.38**	−0.21	−0.25	−0.52***	−0.59***	−0.49***	−0.49***	−0.26*	−0.30*	−0.49***	−0.57***	−0.50***	−0.54***
	(0.10)	(0.12)	(0.12)	(0.13)	(0.10)	(0.11)	(0.09)	(0.11)	(0.12)	(0.13)	(0.08)	(0.10)	(0.09)	(0.10)
Treatment	0.72***	0.67***	0.36**	0.42*	0.87***	0.99***	0.87***	0.87***	0.49***	0.57***	0.97***	1.13***	0.91***	0.98***
	(0.12)	(0.16)	(0.12)	(0.17)	(0.11)	(0.15)	(0.08)	(0.12)	(0.12)	(0.16)	(0.11)	(0.15)	(0.10)	(0.14)
Pretest	0.52***	0.52***	0.50***	0.50***	0.40***	0.40***	0.38***	0.38***	0.44***	0.44***	0.39***	0.39***	0.54***	0.54***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Cohort 2	0.02	−0.03	0.04	0.11	0.07	0.15	0.03	0.03	0.02	0.03	−0.02	0.09	0.03	0.09
	(0.09)	(0.13)	(0.09)	(0.14)	(0.09)	(0.12)	(0.09)	(0.12)	(0.09)	(0.14)	(0.09)	(0.13)	(0.07)	(0.11)
Cohort 3	−0.02	−0.03	−0.02	0.02	0.06	0.22	0.02	−0.01	−0.03	0.15	−0.04	0.16	−0.00	0.07
	(0.11)	(0.17)	(0.12)	(0.19)	(0.11)	(0.16)	(0.11)	(0.16)	(0.12)	(0.18)	(0.11)	(0.16)	(0.10)	(0.14)
Treatment * Cohort 2		0.11		−0.13		−0.16		−0.01		−0.03		−0.20		−0.11
		(0.18)		(0.19)		(0.17)		(0.17)		(0.19)		(0.18)		(0.15)
Treatment * Cohort 3		0.02		−0.06		−0.31		0.04		−0.31		−0.39		−0.13
		(0.23)		(0.24)		(0.21)		(0.20)		(0.24)		(0.22)		(0.19)
Random effects														
School intercept variance	0.14***	0.14***	0.22***	0.23***	0.14***	0.15***	0.10***	0.10***	0.20***	0.20***	0.04	0.04	0.11***	0.11***
School treatment impact variance	0.24***	0.23***	0.23***	0.26***	0.19***	0.18***	0.01	0.01	0.25***	0.23***	0.15**	0.16	0.19***	0.19***
School intercept/impact correlation	−0.17	−0.16	−0.22*	−0.24*	−0.06	−0.06	0.02	0.02	−0.21	−0.20	−0.06	−0.07	−0.10	−0.10
Student-level residual variance	0.52***	0.52***	0.60***	0.60***	0.47***	0.47***	0.50***	0.50***	0.60***	0.60***	0.54***	0.54***	0.36***	0.37***

Note. Student $n = 375$; school $n = 30$. Standard errors presented in parentheses.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. (1) Refers to Model 1. (2) Refers to Model 2.

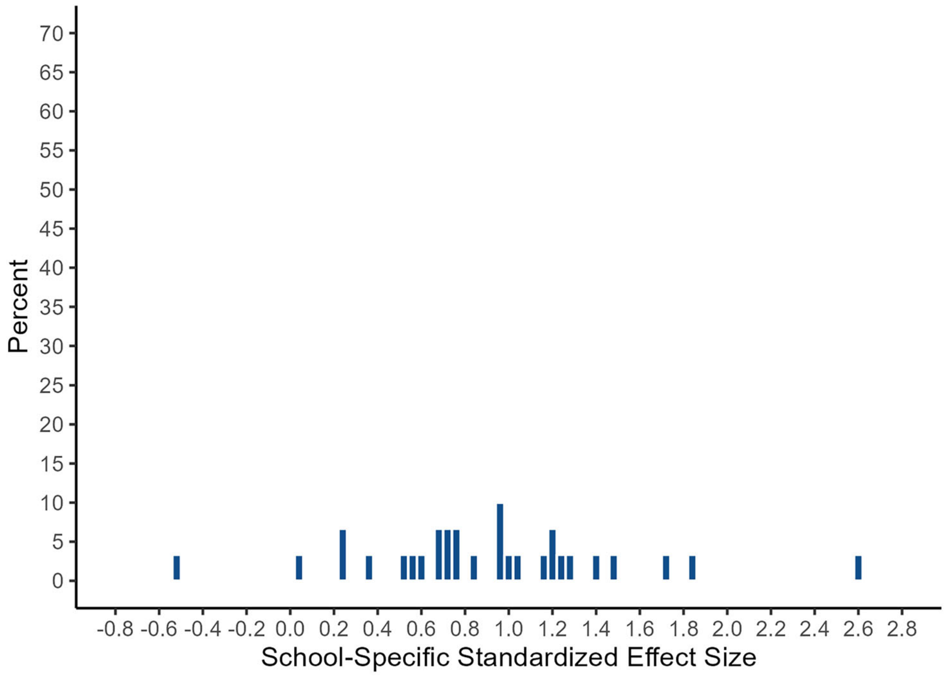

Figure 3. Adjusted empirical Bayes estimates of School-Specific impacts of DLL on mean *IdO* outcomes.

one school with an estimated positive impact of approximately $d = 2.60$ and one with a negative impact of $d = -0.50$.

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 $d = 0.26$ to $d = 0.57$. Model 1 results shown in Table 4 indicate the average DLL effects for the four *Logramos* outcomes: Reading, Language, Vocabulary, and ELA total. As was the case for the *IdO* outcomes, the exploratory analyses of each *Logramos* outcome labeled as Model 2 suggest no evidence of statistically significant differences for the DLL impacts across the three independent yearly student cohorts.

The random effects, including the school-level variance of the DLL treatment effect, ζ^2, are shown in the bottom panel of Table 4. These outcomes reveal statistically significant treatment-effect variability across the 30 schools for two of the four *Logramos* outcomes: Reading and Vocabulary. For instance, the results for the *Logramos* Vocabulary subtest yielded an estimated average effect size of $d = 0.30$, with a random effect covariance estimate for the school-level impacts of 0.36, which is equivalent to 0.60 *SDs*.

We calculated adjusted EB estimates of the school-specific impacts on the composite *Logramos* total score for the 30 schools. Figure 4 provides a graphical representation of this school-level variability in DLL impacts on the *Logramos* total score, which was equivalent to $d = 0.25$ with an adjusted standard deviation of the school-specific DLL impacts of 0.23. As with the *IdO* outcome, 29 of the 30 schools (97%) had positive

Table 4. Multilevel estimates of Intention-to-Treat impact on *Logramos* outcomes.

	Reading		Language		Vocabulary		English/Language Arts total	
Fixed effects	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	−0.23	−0.25	−0.23	−0.22	−0.20	−0.30	−0.20	−0.19
	(0.17)	(0.18)	(0.19)	(0.19)	(0.19)	(0.21)	(0.11)	(0.12)
Treatment	0.34**	0.36*	0.20***	0.18	0.30*	0.48*	0.25***	0.24*
	(0.11)	(0.16)	(0.06)	(0.10)	(0.14)	(0.19)	(0.07)	(0.11)
Pretest	0.39***	0.39***	0.07*	0.08*	0.05	0.05	0.63***	0.63***
	(0.06)	(0.06)	(0.03)	(0.03)	(0.05)	(0.05)	(0.04)	(0.04)
Cohort 2	0.00	0.01	0.10	0.07	0.01	0.19	0.04	−0.00
	(0.11)	(0.15)	(0.08)	(0.10)	(0.11)	(0.16)	(0.08)	(0.11)
Cohort 3	−0.02	0.09	0.01	0.09	−0.12	−0.02	−0.07	0.04
	(0.15)	(0.23)	(0.11)	(0.15)	(0.15)	(0.24)	(0.11)	(0.16)
Treatment* Cohort 2		−0.00		0.07		−0.31		0.07
		(0.19)		(0.13)		(0.20)		(0.14)
Treatment* Cohort 3		−0.17		−0.13		−0.17		−0.16
		(0.27)		(0.18)		(0.28)		(0.19)
Random effects								
School intercept variance	0.58***	0.58***	0.97***	0.95***	0.85***	0.89***	0.23***	0.21***
School treatment impact variance	0.13**	0.13**	0.00	0.00	0.36**	0.41	0.04	0.03
School intercept/ impact correlation	−0.27**	−0.27**	0.00	0.01	−0.56***	−0.60	−0.09*	−0.08*
Student-level residual variance	0.61***	0.62***	0.29***	0.29***	0.65***	0.65***	0.32***	0.32***

Note. Student $n = 337$; school $n = 30$. Standard errors presented in parentheses.
$*p < 0.05$; $**p < 0.01$; $***p < 0.001$. (1) Refers to Model 1. (2) Refers to Model 2.

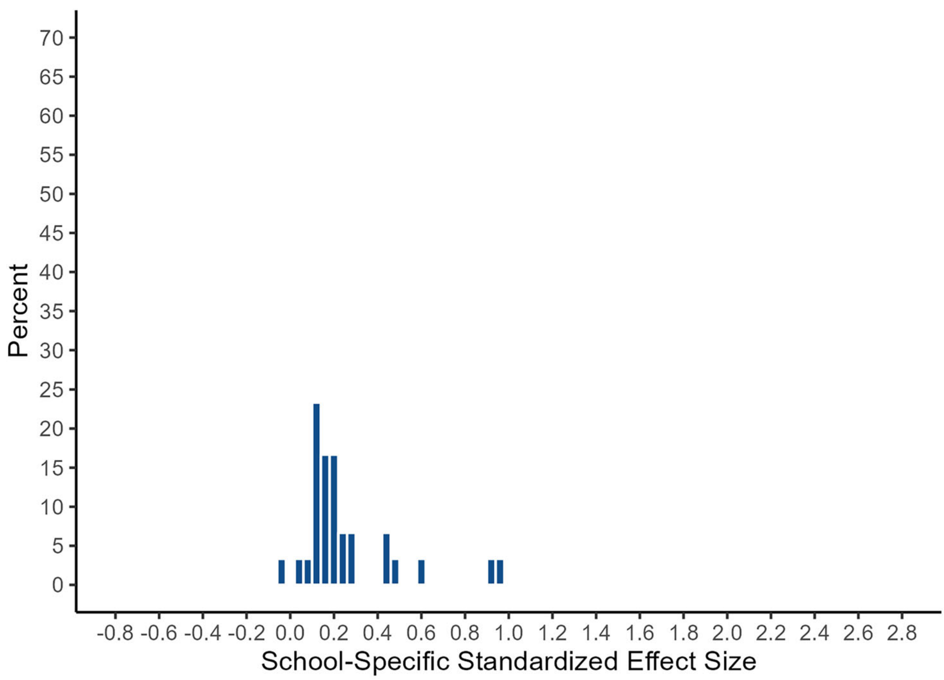

Figure 4. Adjusted empirical Bayes estimates of School-Specific impacts of DLL on mean *Logramos* outcomes.

Table 5. Multilevel estimates of Intention-to-Treat impact on *Iowa test of Basic skills* outcomes.

	Reading		Language		Vocabulary		English/Language Arts total	
Fixed Effects	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Intercept	0.10	0.11	−0.07	−0.06	0.08	0.07	0.12	0.12
	(0.15)	(0.16)	(0.20)	(0.21)	(0.19)	(0.21)	(0.13)	(0.14)
Treatment	0.06	0.03	0.06	0.06	0.07	0.08	0.02	0.02
	(0.08)	(0.13)	(0.06)	(0.10)	(0.13)	(0.19)	(0.07)	(0.12)
Pretest	0.35***	0.35***	0.02	0.02	−0.01	−0.01	0.37***	0.37***
	(0.06)	(0.06)	(0.03)	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)
Cohort 2	−0.27**	−0.29*	−0.08	−0.09	−0.19	−0.12	−0.30**	−0.29*
	(0.10)	(0.14)	(0.08)	(0.10)	(0.11)	(0.17)	(0.09)	(0.13)
Cohort 3	−0.24	−0.28	−0.03	−0.02	−0.05	−0.22	−0.23	−0.24
	(0.14)	(0.19)	(0.11)	(0.15)	(0.16)	(0.24)	(0.13)	(0.17)
Treatment* Cohort 2		0.02		0.01		−0.13		−0.00
		(0.18)		(0.13)		(0.21)		(0.16)
Treatment* Cohort 3		0.08		−0.02		0.25		0.02
		(0.23)		(0.17)		(0.28)		(0.20)
Random effects								
School intercept variance	0.43***	0.43***	1.12***	1.12***	0.81***	0.84***	0.33***	0.33***
School treatment impact variance	0.00	0.00	0.01	0.01	0.28*	0.32*	0.00	0.00
School intercept/impact correlation	0.01	0.01	−0.12***	−0.12***	−0.47***	−0.50***	0.02	0.02
Student-level residual variance	0.48***	0.49***	0.26***	0.26***	0.64***	0.64***	0.40***	0.40***

Note. Student $n = 323$; school $n = 30$. Standard errors presented in parentheses.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. (1) Refers to Model 1. (2) Refers to Model 2.

impacts on the *Logramos* Total Language Arts outcome, ranging from $d = -0.04$ to $d = 0.94$.

ITBS

Students assigned to DLL scored somewhat higher on all ITBS subtests than those assigned to the control group as demonstrated by positive effect sizes; however, the effect sizes were not statistically significant. Model 1 in Table 5 shows the average impacts of assignment to DLL on each ITBS outcome: Reading, Language, Vocabulary, and ELA total. Similar to the Spanish-language outcomes, the results from Model 2 reveal no statistically significant cohort-by-treatment interaction effects. The random effects shown at the bottom of Table 5 do not reveal substantial treatment-effect variability across the 30 schools. The results suggested no statistically significant treatment effect heterogeneity for the ITBS outcomes.

Although the ITBS outcomes revealed no statistically significant treatment effects or cross-site treatment effect variability, we calculated adjusted EB estimates of the school-specific impacts for the 30 schools to examine their variability descriptively. Figure 5 shows this school-level variability in DLL impacts on the ITBS total score composite outcome, which was equivalent to $d = 0.02$ with an adjusted standard deviation of the school-specific DLL impacts of 0.04 *SDs*. Of the 30 schools, 24 schools (80%) had positive impact estimates. Overall, the estimated impacts ranged from $d = -0.08$ to $d = 0.08$. Thus, no schools had impacts greater than 0.10 *SDs*, and no schools had negative impacts less than 0.10 *SDs*.

Multiple Comparisons Corrections

Because we tested the statistical significance of multiple outcomes within one literacy domain, we applied the Benjamini-Hochberg method to adjust for multiple comparisons

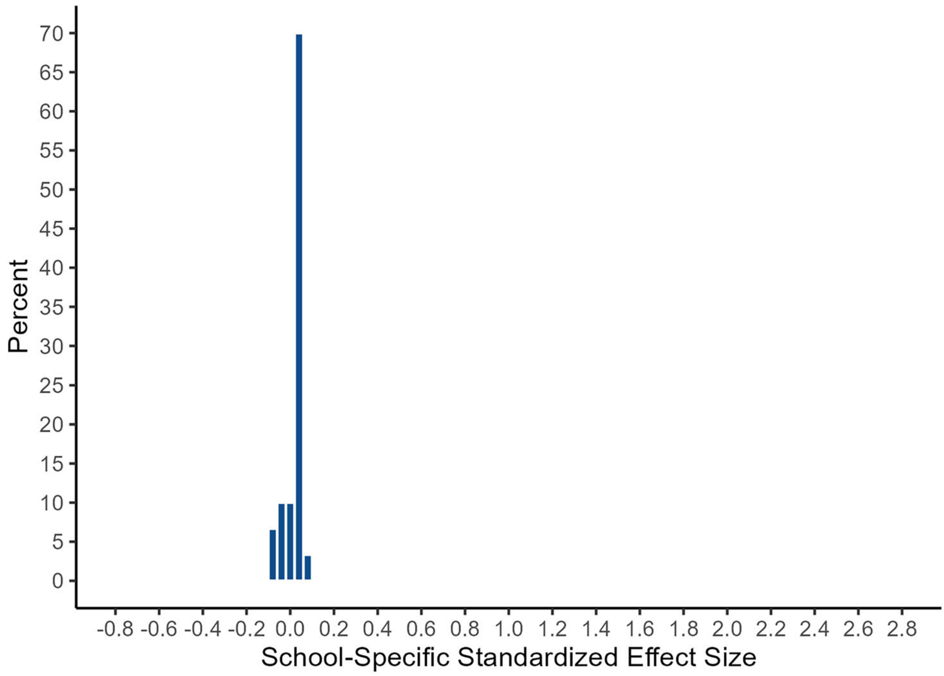

Figure 5. Adjusted empirical Bayes estimates of School-Specific impacts of DLL on mean ITBS outcomes.

to control for a false-discovery rate.[8] The results suggest that after adjusting for multiple comparisons across the seven *IdO* and four *Logramos* tests, the 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). Of course, because there were no statistically significant impacts across the four ITBS measures, there was no need to make the corresponding adjustments to these outcomes.

Discussion

Our analyses indicate that treatment students served by DLL outperformed control students on all Spanish-language *IdO* and *Logramos* measures. We found statistically significant student-level impacts of assignment to DLL services on all *IdO* posttest outcomes, ranging from $d = 0.36$ for the Sounds in Words outcome to $d = 0.97$ for the Text Reading subtest. The DLL effect sizes for the *Logramos* posttests ranged from $d = 0.20$ to $d = 0.34$. No analyses of interaction effects suggest that the impacts of DLL depend on the year of instructional delivery: 2016–17, 2017–18, or 2018–19. Further, our analyses of the school-to-school differences in DLL impacts consistently revealed

[8]Specifically, we rank ordered the statistically significant findings within the domain in ascending order of the *p*-values, such that $p_1 \leq p_2 \leq p_3 \leq p_4 \leq p_5 \leq p_6 \leq p_M$. For each *p*-value, $p_{1\text{-}M}$, 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.

that the overall Spanish-language impacts are larger than they are variable, with an *IdO* total score outcome equivalent to $d = 0.91$ and a standard deviation of 0.58 and an average *Logramos* outcome equivalent to $d = 0.25$ with a standard deviation of 0.23. Although 80% of the schools also realized positive impacts on the ITBS assessment, the highly positive Spanish-language outcomes did not appear to translate immediately into statistically significant English-language impacts.

Interpreting the Magnitudes of the DLL Impacts

Similar to the sister English-language version (RR), DLL demonstrates strong impacts for the sampled students across many dimensions of literacy in reliable ways that are replicated across years and from teacher to teacher. Comparing our results to those from the recent large-scale, multicohort evaluation of RR reported by Sirinides et al. (2018), the magnitudes of the impacts are quite similar. The RR effect size for the overall score from the *OS* was nearly one full standard deviation and similarly DLL's impact on the *IdO* total score was equivalent to 0.91 standard deviations. RR impacts on the English-language ITBS, which were somewhat greater than one-third of a standard deviation, were comparable to those found for the Spanish-language *Logramos* assessment, which ranged from 0.20 to 0.34. Based on the first-year results from the national study of RR, May et al. (2013) indicated that 81% of the 147 participating RR schools had positive English-language ITBS impacts on the participating treatment students. Although our sample was considerably smaller, with only 30 schools, all but one school (97%) in the current study of DLL had positive impacts on the Spanish-language *Logramos* assessment.

Other benchmarks may provide additional context to understand the impacts of DLL. For instance, Kraft's (2020) review of nearly 250 recent RCTs of education interventions suggested that a "large" effect size is 0.20 or greater. More specifically, by examining the distribution of impacts across those 242 RCTs, the impacts of DLL on the *Logramos* and *IdO* total scores are, respectively, above the 75th percentile and close to the 99th percentile of that distribution. When comparing the largest impact of 0.97 standard deviation units to other more intuitive benchmarks provided by Hill et al. (2008), the DLL effect is equivalent to a learning advantage relative to controls that is approximately equal to the amount of reading achievement growth experienced by the typical first-grade student during a full school year. With effect sizes ranging from $d = 0.20$ to $d = 0.34$ on the *Logramos* assessment, the impacts of DLL were equivalent to approximately one fifth to more than one third of the achievement growth found for the typical first-grade student across a full year of school. Finally, compared with typical supplemental Title I programs, which were reviewed in a meta-analysis by Borman and D'Agostino (1996), which frequently serve as the funding source for DLL, the DLL impacts on the *Logramos* assessment are more than two to three times larger, and the *IdO* impacts are greater than nine times larger than those observed in supplemental Title I programs. Indeed, as a supplemental intervention spanning only 12–20 weeks, DLL appears to produce impressive impacts of a magnitude not typically seen for educational programs of any type.

Interpreting the Variability of the DLL Impacts

Relatively few benchmarks are available to assess the consistency of impacts found across the 30 participating DLL sites. Recently, there has been growing interest in investigating treatment-effect heterogeneity in multisite RCTs of education interventions (Bloom et al., 2017; Bloom & Spybrook, 2017; Weiss et al., 2017). In reviewing results from 16 multisite RCTs, Weiss et al. (2017) concluded that findings of treatment impacts and site-to-site variability of the impacts follow several patterns. Specifically, the authors found that studies of education interventions fall into one of four categories: (1) yielding consistent near-zero impacts across sites; (2) yielding near-zero average impacts with substantial cross-site variation; (3) yielding consistent positive impacts across sites; and (4) yielding large average positive impacts with substantial cross-site variation. Our results would seem to fall into the latter two categories. Concluding which of these two characterizations best fits our results requires more careful consideration of the magnitudes of our cross-site variation.

Weiss et al. (2017) reported that about one third of the 16 multisite education RCTs that they reviewed yielded little cross-site impact variation, ranging from 0 to 0.05 standard deviations (*SD*s). Another third of the estimates ranged from 0.05 *SDs* to 0.15 *SDs*, indicating moderate cross-site variation, and the final third ranged from 0.15 *SDs* to 0.35 *SDs*, which suggested considerable cross-site variation. On the surface, our estimates of cross-site variation for the Spanish-language outcomes (0.25 *SDs* to more than 0.5 *SDs*) seem to align with the category of findings characterized by Weiss et al. as "large average positive impact[s] with substantial cross-site variation" (p. 13). Yet our cross-site impact estimates also can be described by the authors' other description of "consistently positive impacts across sites" (p. 13). Specifically, only one of the 30 sites in our sample had nonpositive impacts for the *Logramos* and *IdO* assessments.

When examining the reasons behind greater consistency in cross-site impacts, Weiss et al. (2017) identified three explanations: (1) highly specific and consistent interventions that tend to deliver similar services across sites; (2) low-intensity programs, which are characterized by small treatment-control service contrasts and limited opportunity for cross-site differences; and (3) interventions delivered to treatment and control students in the same school building yielding smaller treatment-control service contrasts than when treatment and control students are served in separate and potentially more variable school contexts.

Because DLL is a highly specific intervention (explanation #1) and we studied both treatment and control students in the same building (explanation #3), we would expect consistent impacts across sites. However, because DLL is not a low-intensity program, (explanation #2), this may induce greater opportunity for impact variation across sites. Although the sample of 16 multisite studies reviewed by Weiss et al. (2017) revealed no evidence that cross-site impact variation increases as the magnitude of cross-site mean impacts increases, none of the 16 studies had effect sizes approaching the magnitudes of the DLL intervention. Perhaps it is this combination of factors that explain the consistently positive but relatively variable impacts across the 30 participating schools in our study.

Conclusion

DLL is predicated on the notion that responsive, high-quality Spanish-language literacy programs require coordinated supports for students who are struggling with early literacy learning. DLL provides supplemental resources needed for Spanish-speaking students to participate—and potentially thrive—in their school's Spanish-language literacy program. Also, furthering the development of learning in one's native language can transfer to learning a second language and since DLL is taught in Spanish, it may contribute to students acquiring not only Spanish but also English literacy skills as they progress through elementary school. Indeed, much research indicates that teaching reading in a student's native language may be the optimal approach to improving reading in English. Research from multiple studies on a variety of languages supports the notion that learning literacy in one's first language augments literacy in the second (Richards-Tutor et al., 2016; Roberts, 1994). In the initial months of intervention, though, our results do not suggest that such transference occurred. As past research has suggested, this process of improved Spanish-language literacy helping students acquire additional English-language skills may take several years (Umansky & Reardon, 2014). Given the range of Spanish literacy skills that we measured, more nuanced analyses of which particular skills may predict students' future Spanish and English development could help further elucidate this process and suggest important points of focus for early intervention.

Our sample was relatively small, therefore more research is necessary to document the generalizability of these favorable results across other districts, schools, teachers, and students. However, these experimental outcomes across 30 schools are very promising. The literacy outcomes evaluated revealed little evidence of variability across years, suggesting the reliability and replicability of DLL's impacts. Although impacts varied across sites, the impacts were positive in all but one site. Future research should investigate the extent to which these effects are sustained across multiple years, and whether these early advantages in students' native language of Spanish translate, as past evidence suggests, into later advantages in English literacy.

In addition, future work should systematically link data concerning the quality and variability of implementation across sites and teachers with impact data. Clearly, key 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, future work might provide greater understandings of how 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 variations may be associated with differences in student outcomes. Although there was some variability in DLL teachers' experience with DLL, on average, teachers had been delivering DLL for 8 years. Furthermore, all school sites had been implementing DLL when our study began. Our cost analysis suggests DLL requires significant financial commitment (Borman et al., 2022), so it is possible that newly implementing sites may not yield the large and positive impacts observed in this study or the variation in those impacts may be larger or smaller. Thus, understanding impact and variation in the context of early implementers is also important.

Finally, future longitudinal studies could help inform how the program may prevent the need for other potentially more costly interventions, including special education, continuing specialized services for ELs, and retentions in grade. Prior quasi-experimental evidence suggests that DLL students maintain average to high literacy performance through elementary school (Escamilla et al., 1998; Neal & Kelly, 1999), and studies of early literacy programs emphasizing native language instruction suggest similar long-term benefits (Ashdown & Simic, 2000; Richards-Tutor et al., 2016). Therefore, even though DLL targets intensive resources for first-grade struggling readers, it is quite possible that these early investments have long-term benefits that may, ultimately, save schools and districts valuable resources and prevent future remediation efforts.

Open Research Statements

Study and Analysis Plan Registration

There is no study and analysis plan registration associated with this manuscript.

Data, Code, and Materials Transparency

The data and code underlying the results reported in this manuscript are available in openICPSR: https://www.openicpsr.org/openicpsr/project/193586/

Design and Analysis Reporting Guidelines

Not applicable.

Transparency Declaration

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Replication Statement

This manuscript reports an original study.

Open Scholarship

This article has earned the Center for Open Science badge for Open Data. The data are openly accessible at https://www.openicpsr.org/openicpsr/project/193586/version/V1/view.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through [Grant #R305A160060] to the 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.

ORCID

Trisha H. Borman http://orcid.org/0000-0003-4176-4202
Geoffrey D. Borman http://orcid.org/0000-0002-7039-8208

References

Arteagoitia, I., & Yen, S. J. (2020). Equity in representing literacy growth in dual language bilingual education for emerging bilingual learners. *TESOL Quarterly*, *54*(3), 719–742. https://doi.org/10.1002/tesq.588

Ashdown, J., & Simic, O. (2000). Is early literacy intervention effective for English language learners? Evidence from reading recovery. *Literacy Teaching and Learning*, *5*(1), 27–42. https://readingrecovery.org/wp-content/uploads/2017/03/LTL_5.1-Ashdown-Simic.pdf.

August, D. and Hakuta, K., Eds. (1997). *Improving schooling for language minority children: A research agenda*. National Academy Press.

August, D., McCardle, P., & Shanahan, T. (2014). Developing literacy in English language learners: Findings from a review of the experimental research. *School Psychology Review*, *43*(4), 490–498. https://doi.org/10.1080/02796015.2014.12087417

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

Baker, S., Geva, E., Kieffer, M. J., Lesaux, N., Linan-Thompson, S., Morris, J., Proctor, C. P., Russell, R., Gersten, R., Dimino, J., Jayanthi, M., Haymond, K., & Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE 2014-4012). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/english_learners_pg_040114.pdf

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

Bloom, H. S., Raudenbush, S. W., Weiss, M. J., & Porter, K. (2017). Using multisite experiments to study cross-site variation in treatment effects: A hybrid approach with fixed intercepts and a random treatment coefficient. *Journal of Research on Educational Effectiveness*, *10*(4), 817–842. https://doi.org/10.1080/19345747.2016.1264518

Bloom, H. S., & Spybrook, J. (2017). Assessing the precision of multisite trials for estimating the parameters of a cross-site population distribution of program effects. *Journal of Research on Educational Effectiveness*, *10*(4), 877–902. https://doi.org/10.1080/19345747.2016.1271069

Borman, G. D., & D'Agostino, J. V. (1996). Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis*, *18*(4), 309–326. https://eric.ed.gov/?id=EJ538585 https://doi.org/10.3102/01623737018004309

Borman, G. D., Brodziak de los Reyes, I., Borman, T. H., Carbuccia-Abbott, M., Houghton, S., Park, S. J., Zhu, B., & Martin, A. (2022). *The costs and impacts of Descubriendo la Lectura: Evidence from a multisite experimental study*. Division of Leadership and Innovation, Arizona State University.

Borman, G. D., Borman, T. H., Park, S. J., & Houghton, S. (2020). A multisite randomized controlled trial of the effectiveness of *Descubriendo la Lectura*. *American Educational Research Journal*, *57*(5), 1995–2020. https://doi.org/10.3102/0002831219890612

Borman, T. H., Borman, G. D., Houghton, S., Park, S. J., Zhu, B., Martin, A., & Wilkinson-Flicker, S. (2019). Addressing literacy needs of struggling Spanish-speaking first graders: First-year results from a national randomized controlled trial of Descubriendo la Lectura. *AERA Open*, *5*(3), 233285841987048. https://doi.org/10.1177/2332858419870488
Bryk, A. S., Gomez, L., & Grunow, A. (2010). *Getting ideas into action: Building networked improvement communities in education*. Carnegie Foundation for the Advancement of Teaching. https://www.carnegiefoundation.org/wp-content/uploads/2014/09/bryk-gomez_building-nics-education.pdf.
Calderón, M., Slavin, R., & Sanchez, M. (2011). Effective instruction for English learners. *The Future of Children*, *21*(1), 103–127. https://doi.org/10.1353/foc.2011.0007
Calderón, M. E., & Slakk, S. (2019). *Success with multicultural newcomers & English learners: Proven practices for school leadership teams*. ASCD.
Chen, X., & Padilla, A. M. (2019). Role of bilingualism and biculturalism as assets in positive psychology: Conceptual dynamic GEAR model. *Frontiers in Psychology*, *10*, 2122. https://doi.org/10.3389/fpsyg.2019.02122
Cheung, A., & Slavin, R. E. (2005). Effective reading programs for English language learners and other language minority students. *Bilingual Research Journal*, *29*(2), 241–267. https://doi.org/10.1080/15235882.2005.10162835
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*(4), 351–395. https://doi.org/10.3102/0034654312465472
Clay, M. M. (1993). *An observation survey of early literacy achievement*. Heinemann.
Clay, M. M. (2001). *Change over time in children's literacy development*. Heinemann.
Clay, M. M. (2005). *Literacy lessons designed for individuals*. Heinemann.
Clay, M. M. (2019). *Observation survey of early literacy achievement* (4th ed.). Heinemann.
Cummins, J. (2005, September 23). *Teaching for cross-language transfer in dual language education: Possibilities and pitfalls* [Paper presentation]. TESOL Symposium on Dual Language Education: Teaching and Learning Two Languages in the EFL Setting, Bogazici University.
de Brey, C., Zhang, A., & Duffy, S. (2022). *Digest of education statistics*. U.S. Department of Education, National Center for Education Statistics. https://nces.ed.gov/programs/digest/d20/index.asp
DeFord, D. (2007). Marie Clay: An honored mentor, colleague, and friend. *Journal of Reading Recovery*, *7*(1), 101–103. http://readingrecovery.org/images/pdfs/Journals/JRR/Vol7_No1_Fall-2007/JRR_7.1-DeFord.pdf.
Dianda, M., & Flaherty, J. (1995, April). *Effects of Success for All on the reading achievement of first graders in California bilingual programs* [Paper presentation]. Annual Meeting of the American Educational Research Association, San Francisco, CA, United States.
Durgunoğlu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology*, *85*(3), 453–465. https://doi.org/10.1037/0022-0663.85.3.453
Escamilla, K. (2018). Growing up with the Bilingual Education Act: One educator's journey. *Bilingual Research Journal*, *41*(4), 369–387. https://doi.org/10.1080/15235882.2018.1529641
Escamilla, K., Andrade, A., Basurto, A., & Ruiz, O. (1996). *Instrumento de Observación de los logros de la lecto-escritura inicial*. 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.
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). Lawrence 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*(1), 3–17. https://doi.org/10.1037/0022-0663.88.1.3

Galindo, C., & Reardon, S. F. (2006). *Hispanic students' educational experiences and opportunities during kindergarten*. National Task Force on Early Childhood Education for Hispanics. https://stacks.stanford.edu/file/druid:xz483fw4636/hispanic_experience.pdf.

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

Gersten, R., Baker, S. K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades* (NCEE 2007-4011). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/20074011.pdf.

Gottardo, A. (2002). The relationship between language and reading skills in bilingual Spanish-English speakers. *Topics in Language Disorders*, *22*(5), 46–70. https://doi.org/10.1097/00011363-200211000-00008

Greene, J. (1997). A meta-analysis of the Rossell and Baker review of bilingual education research. *Bilingual Research Journal*, *21*(2–3), 103–122. https://doi.org/10.1080/15235882.1997.10668656

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*(3), 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

International Data Evaluation Center. (2012). *2012–2013 Descubriendo la Lectura statistical abstract for the U.S.* The Ohio State University. https://readingrecovery.org/wp-content/uploads/2023/02/2021-2022-Descubriendo-la-Lectura-National-Summary-Report-for-the-United-States.pdf.

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Association for Supervision & Curriculum Development.

Juel, C. (1988). Learning to read and write: A longitudinal study 54 four children from first through fourth grades. *Journal of Educational Psychology*, *80*(4), 437–447. https://eric.ed.gov/?id=EJ389953 https://doi.org/10.1037/0022-0663.80.4.437

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*(27), 10155–10162. https://doi.org/10.1073/pnas.0600888103

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Logramos. (2006). *2005 Norms and score conversions with technical information*. Riverside. http://www.riverpub.com/scoring/pdf/.

Ludwig, C., Guo, K., & Georgiou, G. K. (2019). Are reading interventions for English language learners effective? A meta-analysis. *Journal of Learning Disabilities*, *52*(3), 220–231. https://doi.org/10.1177/0022219419825855

Lyon, G. R., & Fletcher, J. M. (2001). Early warning systems: How to prevent reading disabilities. *Education Matters*, *1*(2), 22–29.

May, H., Gray, A., Gillespie, J. N., Sirinides, P., Sam, C., Goldsworthy, H., Atmijo, M., & Tognatta, N. (2013). *Evaluation of the i3 scale-up of Reading Recovery. Year one report (2011–12)*. Consortium for Policy Research in Education. http://www.cpre.org/sites/default/files/researchreport/1488_readingrecoveryreport.pdf.

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* (pp. 267–299). 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*(4), 730–742. https://doi.org/10.1037/0022-0663.95.4.730

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

National Center for Education Statistics. (2022). *English learners in public schools. Condition of Education*. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cgf.

National Education Association. (2020). *English language learners*. https://www.nea.org/resource-library/english-language-learners.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. 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*. 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.

Nunnery, J., Slavin, R., Ross, S., Smith, L., Hunter, P., & Stubbs, J. (1997, March 24–28). *Effects of full and partial implementations of Success for All on student reading achievement in English and Spanish* [Paper presentation]. Annual meeting of the American Educational Research Association, Chicago, IL, United States.

Patterson, N., Weaver, J., Fletcher, J., Connor, B., Thomas, A., & Ross, C. (2018). Leveraging literacies through collaborative, source-based planning and teaching in social studies and language arts. *Journal of Adolescent & Adult Literacy*, *61*(5), 523–532. https://www.jstor.org/stable/26631714 https://doi.org/10.1002/jaal.712

Planning and Evaluation Service, Office of the Deputy Secretary. (2001). *Evidence that tutoring works*. U.S. Department of Education. https://www.gpo.gov/fdsys/pkg/ERIC-ED464343/pdf/ERIC-ED464343.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*(6), 889–912. https://doi.org/10.1080/0141987022000009368

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage.

Reading Recovery Council of North America. (2017). *Standards and guidelines of Reading Recovery in the United States* (8th ed.). https://readingrecovery.org/wp-content/uploads/2018/12/SG2018-Master-Full.pdf.

Reading Recovery Council of North America. (2021). *Lessons*. https://readingrecovery.org/reading-recovery/teaching-children/lessons/.

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*(3), 853–891. https://doi.org/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*(2), 144–169. https://doi.org/10.1177/0014402915585483

Roberts, C. A. (1994). Transferring literacy skills from L1 to L2: From theory to practice. *The Journal of Educational Issues of Language Minority Students*, *13*, 209–221. https://eric.ed.gov/?id=EJ489585.

Rolstad, K., Mahoney, K., & Glass, G. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy*, *19*(4), 572–594. https://doi.org/10.1177/0895904805278067

Shaywitz, S. E., Fletcher, J. M., Holahan, J. M., Shneider, A. E., Marchione, K. E., Stuebing, K. K., Francis, D. J., Pugh, K. R., & Shaywitz, B. A. (1999). Persistence of dyslexia: The Connecticut Longitudinal Study at adolescence. *Pediatrics*, *104*(6), 1351–1359. https://doi.org/10.1542/peds.104.6.1351

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*(3), 316–335. https://doi.org/10.3102/0162373718764828

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), 1–26. https://doi.org/10.1016/j.edurev.2010.07.002

Snyder, E., Witmer, S. E., & Schmitt, H. (2017). English language learners and reading instruction: A review of the literature. *Preventing School Failure: Alternative Education for Children and Youth*, *61*(2), 136–145. https://doi.org/10.1080/1045988X.2016.1219301

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*(5), 879–912. https://doi.org/10.3102/000283121454511

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*(4), 612–637. https://doi.org/10.3102/0162373715573310

Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., Pollard-Durodola, S. D., Fletcher, J. M., & Francis, D. J. (2006). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal*, *43*(3), 449–487. https://doi.org/10.3102/00028312043003449

Vaughn, S., Linan-Thompson, S., Mathes, P. G., Cirino, P. T., Carlson, C. D., Pollard-Durodola, S. D., Cardenas-Hagan, E., & Francis, D. J. (2006). Effectiveness of Spanish intervention for first-grade English language learners at risk for reading difficulties. *Journal of Learning Disabilities*, *39*(1), 56–73. https://doi.org/10.1177/00222194060390010601

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.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, *36*(4), 541–561. https://eric.ed.gov/?id=EJ788353 https://doi.org/10.1080/02796015.2007.12087917

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*(6), 612–624. https://doi.org/10.1177/0022219418775110

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook* (Version 5.0). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/Handbooks.

Weiss, M. J., Bloom, H. S., Verbitsky-Savitz, N., Gupta, H., Vigil, A. E., & Cullinan, D. N. (2017). How much do the effects of education and training programs vary across sites? Evidence from past multisite randomized trials. *Journal of Research on Educational Effectiveness*, *10*(4), 843–876. https://doi.org/10.1080/19345747.2017.1300719

Willig, A. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, *55*(3), 269–317. https://doi.org/10.3102/00346543055003269