AERA Open
January-March 2020, Vol. 6, No. 1, pp. 1–16
DOI: 10.1177/2332858419897886
Article reuse guidelines: sagepub.com/journals-permissions

Early Efficacy of Multitiered Dual-Language Instruction: Promoting Preschoolers' Spanish and English Oral Language

Trina D. Spencer
University of South Florida

Meghan Moran
Northern Arizona University

Marilyn S. Thompson
Arizona State University

Douglas B. Petersen
Brigham Young University

M. Adelaida Restrepo
Arizona State University

The purpose of this cluster randomized group study was to investigate the effect of multitiered, dual-language instruction on children's oral language skills, including vocabulary, narrative retell, receptive and expressive language, and listening comprehension. The participants were 3- to 5-year-old children (n = 81) who were learning English and whose home language was Spanish. Across the school year, classroom teachers in the treatment group delivered large-group lessons in English to the whole class twice per week. For a Tier 2 intervention, the teachers delivered small-group lessons 4 days a week, alternating the language of intervention daily (first Spanish, then English). Group posttest differences were statistically significant, with moderate to large effect sizes favoring the treatment group on all the English proximal measures and on three of the four Spanish proximal measures. Treatment group advantages were observed on Spanish and English norm-referenced standardized measures of language (except vocabulary) and a distal measure of language comprehension.

Keywords: *ANOVA/MANOVA, bilingual/bicultural, early childhood, language comprehension/development, vocabulary*

Reading comprehension and academic achievement are dependent on oral language skills (Catts, Fey, Tomblin, & Zhang, 2002; Gough & Tunmer, 1986; Griffin, Hemphill, Camp, & Wolf, 2004; Storch & Whitehurst, 2002). Although interventions to promote code-related skills have proliferated, interventions to systematically teach oral language and its components, such as vocabulary, narratives, listening comprehension, and use of complex sentences (Cain & Oakhill, 2011; Elleman, Lindo, Morphy, & Compton, 2009; Mehta, Foorman, Branum-Martin, & Taylor, 2005; Verhoeven & van Leeuwe, 2008), remain largely unavailable to early-childhood educators (Zucker, Cabell, Justice, Pentimonti, & Kaderavek, 2013). Spanish-speaking children entering English-only elementary schools are in particular need of effective interventions that are strategically and intensely designed to prepare them for the academic language demands of school (Castro, Páez, Dickinson, & Frede, 2011). The purpose of this study was to examine the effect of an innovative instructional model designed specifically for young dual-language learners on children's oral language skills preparatory to their entrance into kindergarten.

The Oral Language and Literacy Connection

Oral language is a unique and meaningful indicator of academic success (Barton-Hulsey, Sevcik, & Romski, 2017; Catts, Nielsen, Bridges, & Liu, 2016; Chaney, 1998; Clarke, Snowling, Truelove, & Hulme, 2010; Larney, 2002).

Specifically, vocabulary (Bleses, Makransky, Dale, Højen, & Ari, 2016; National Institute of Child Health and Human Development, 2000), narrative ability (Griffin, et al., 2004), listening comprehension (Catts, Adlof, & Ellis Weismer, 2006), and the use of complex sentences (Craig, Connor, & Washington, 2003) are key contributors to reading comprehension. Limited reading comprehension can be the direct result of limited academic English oral language skills (Cain, Lemmon, & Oakhill, 2004; Catts et al., 2006; Dickinson, McCabe, Anastasopoulos, Peisner-Feinberg, & Poe, 2003; Hammer, Lawrence, & Miccio, 2007). Many young children with typical language-learning ability may not produce or understand language on par with academic expectations for a variety of reasons, including economic, cultural, and linguistic diversity (National Center for Education Statistics, 2017). The idea that children with language differences must wait until their language difficulties evolve into reading difficulty and poor academic performance to receive special, individualized help is problematic because with early identification and intervention, their difficulties may be prevented (Catts, 1993; Catts et al., 2006).

With the adoption of higher language and reading standards across states, expectations of what children are to understand and produce linguistically in school have likewise increased. Young children who have typical language-learning abilities but who are far behind their peers in English language development, for whatever reason, have few options. The outdated dichotomous system of general and special education cannot fully meet the needs of children with typically developing language skills who are learning English. More research is needed to develop effective models of instruction that are strategically designed to facilitate and hasten the acquisition of English (Vaughn et al., 2006).

Multitiered Systems of Support

One model that may have utility for promoting English language acquisition before children experience academic failure, is multitiered system of supports (MTSS). The idea of providing special services to children who are not performing as expected, irrespective of ability status, is not new. In 2004, the reauthorized IDEA (Individuals with Disabilities Education Act) clearly outlined the concept of response to intervention that has been shaped into the contemporary framework of MTSS. In general, MTSS is a framework for identifying children with emerging difficulties so that timely differentiated and preventative instruction can be dispensed according to individual children's needs. As a conceptual basis for early identification and prevention (Fuchs & Deshler, 2007), MTSS is a paradigmatic model, not a formula, method, or procedure. Therefore, there are many effective ways to actualize the chief MTSS attributes, which are (a) multiple tiers of instruction and intervention, (b) students who need more support transition to more intense arrangements of intervention, (c) interventions are intensified by adjusting the duration and frequency of intervention, and the expertness of the interventionist, (d) educators other than classroom teachers assist in the delivery of targeted and intensive interventions, and (e) tiered placement is determined irrespective of special education classification (Marston, 2005).

MTSS has several advantages over the traditional general-special education dichotomy. Perhaps the greatest is that rather than focusing on what caused the delays, MTSS delivers supplemental intervention to all who need it, not just those with the appropriate diagnosis. Despite the success of MTSS for early reading intervention, language has been neglected. If the goal is to ensure all children receive what they need to succeed in school, then more systematic language intervention should be considered for children with language differences. In the traditional system, children who receive language supports experience no intermediate step such as Tier 2 intervention. There is no strategy for eliminating environmental confounds to language delays and no way to prevent language-related disabilities. Students go straight from classroom instruction to special education, and that pathway is only available to students who have a disability. Nonetheless, a multitiered approach for language, one that affords an intermediate, preventative step, is possible, especially in early childhood (Carta & Young, 2019; Durán & Wackerle-Hollman, 2019; Greenwood et al., 2013).

Dual-Language Approach to Intervention

Recent recommendations for creating powerful interventions for Spanish-speaking English learners include incorporating children's first language (L1) to facilitate development of their second language (L2; Baker, 2000; Barnett, Yarosz, Thomas, Jung, & Blanco, 2007; Castro, Garcia, & Markos, 2013; Collier & Thomas, 2017; Coltrane, 2003; MacSwan & Rolstad, 2005; Restrepo, Morgan, & Thompson, 2013). Those who receive sustained dual-language instruction tend to be two to three years ahead of those who receive English-only instruction in terms of academic performance (Mahoney, MacSwan, & Thompson, 2005; Rolstad, Mahoney, & Glass, 2005). Collier and Thomas (2017) argued that sustained L1 and L2 instruction engages sociocultural, linguistic, cognitive, and academic processes that lead to high academic achievement in children's L2. Furthermore, they posit that when schools provide strong dual-language programs, children from low–socioeconomic backgrounds overcome the negative effects of poverty. Such sentiments are echoed in the recent National Academies of Sciences, Engineering, and Medicine (2017) report on

promoting educational success of children learning English, to include recommendations for incorporating children's L1 and involving families in the promotion and retention of their home language.

The possibility of skills learned in one language transferring with minimal direct teaching to another language helps explain the facilitative effects seen in dual-language instruction research (Méndez, Crais, Castro, & Kainz, 2015; Miller, Heilmann, Nockerts, Iglesias, Fabiano, & Francis, 2006; Proctor, Carlo, August, & Snow, 2005; Restrepo et al., 2013; Rolstad et al., 2005). That is, when children receive strategic language instruction in L1, it is possible that their knowledge and skills transfer to L2, and vice versa in some cases (Marian & Kaushanskaya, 2007). It is theorized that cross-language interactions will occur across structures that have a similar, underlying cognitive schema (MacWhinney, 1999). Schemas are the mental organization of prior experiences (Anderson & Pearson, 1984), and such schemas can be expressed through narration (Stein & Glenn, 1979). Narrative organization is very similar across English and Spanish, which implies that the narrative schemas for the two languages are similar. This underlying similarity suggests that narrative structure will have linguistic reciprocity between L1 and L2 (and vice versa). For example, Petersen, Thompsen, Guiberson, and Spencer (2016) found that the effects of an L2 intervention targeting narrative and linguistic structures transferred to typically developing children's L1. In vocabulary programs, this transfer is evidenced by the faster acquisition of concepts from L1 instruction to L2 than when children receive the instruction only in L2 (English in the case of the United States; Perozzi, 1985; Perozzi & Chavez Sanchez, 1992). Moreover, Miller et al. (2006) found that sentence complexity and story structure in L1 at school entry predicted academic achievement in L2 in Spanish-English dual-language learners. These studies, correlational and causal, indicate that one language can facilitate the acquisition of another language and that the stronger the child's L1, the greater their acquisition in L2.

The Current Study

This study represents an early efficacy pilot study to determine the promise of a multitiered dual-language curriculum for a large-scale efficacy trial. As such, it was particularly important to understand the extent to which measures of narrative, vocabulary, language comprehension, and general language abilities could be affected. Therefore, we addressed the following research questions:

Research Question 1: To what extent does multitiered dual-language instruction enhance preschoolers' oral language skills when they are assessed using proximal narrative retell and targeted vocabulary measures?

Research Question 2: To what extent does multitiered dual-language instruction enhance preschoolers' oral language skills when they are assessed using distal story comprehension and general language measures?

Because the curriculum was new, the extent to which preschool teachers perceived it to be feasible in their classrooms was not known. Feasibility of an intervention can depend on how well teachers like it, its contextual fit to the school system, how well teachers understand it and how to deliver the lessons, and the extent to which teachers can make reasonable modifications. Therefore, we also examined the curriculum's feasibility in a secondary research question: *Research Question 3*: To what extent is multitiered dual-language instruction feasible?

Method

Setting and Participants

This study was conducted in Head Start preschool classrooms in a southwestern state. During the spring prior to the commencement of the study, the first author gave a presentation regarding the study to administrators of two Head Start grantees (one urban and one rural). Once the administrators volunteered to allow their centers to participate, the first and second authors visited each center to speak directly with teachers about the study. Head Start teachers who were interested in participating signed an informed consent form and completed a demographic survey. When school started at the beginning of August the next year, the research team collected parental permission for children to participate. Using the parent-completed forms at their sites, the teachers identified children from Spanish-speaking homes. All children for whom Spanish was one of the languages spoken at home were invited to participate.

Teachers/Classrooms. In total, 25 classrooms were included in this study. The classrooms were randomly assigned to treatment and control groups at the completion of the consenting and screening process, resulting in 12 classrooms in the treatment group and 13 in the control group. One lead teacher and one teaching assistant provided instruction to 18 to 20 children (3-, 4-, and 5-year-olds) in each classroom. Although efforts were made to recruit classrooms that had at least one teacher or teaching assistant who spoke Spanish fluently, given the available workforce and frequent turnover, three of the treatment classrooms and five of the control classrooms were without a Spanish-speaking teacher or teaching assistant. Children in 18 (nine in treatment and nine in control) of the classrooms attended preschool Monday through Thursday. In the remaining seven classrooms, children attended 5 days a week. All the teachers reported using Creative Curriculum (Dodge, Colker, & Heroman, 2002) as their core curriculum, which was complemented by Teaching Strategies Gold (Heroman, Tabors, & Teaching Strategies, Inc.,

TABLE 1
Teacher and Classroom Characteristics

Characteristic	Treatment Group	Control Group
Number of classrooms	12	13
Years teaching, mean (range)	10 (3 months to 20)	9 (3 months to 18)
Highest level of education, number of teachers		
High school diploma	2	2
Associate's degree	5	6
Bachelor's degree	4	5
Graduate degree	1	0
Race/ethnicity, number of teachers (percentage)		
White	6 (50)	8 (62)
Hispanic/Latino	6 (50)	4 (31)
American Indian	0 (0)	1 (7)
Language of instruction, number of classrooms		
English only	8	6
Mostly English	3	5
50/50 Bilingual	1	1
Type of classroom, number of classrooms		
Half-day	9	8
Full day	3	5
CLASS scores, mean		
Emotional support	6.03	6.07
Classroom organization	5.79	5.72
Instructional support	3.65	4.46

Note. CLASS = Classroom Assessment Scoring System.

2010). The Head Start programs completed Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008) observations of all of their teachers during September or October of the school year. These data are reported, along with additional information about the teachers and classrooms, in Table 1.

Children. During the recruitment phase, the research team went to each center during drop-off or pickup times and met with the parents or guardians of the children. The researchers explained the study to the parents in their preferred language (Spanish or English). Consent was obtained from the parents of 144 children of ages 3 to 5 years who were exposed to Spanish at home. Once signed consents were obtained, the research team administered screening measures to assess the children's language skills in English and Spanish. Screening involved the use of the Expressive Vocabulary (EV) subtest of Clinical Evaluations Language Fundamentals–Preschool (CELF-P; Semel, Wiig, & Secord, 2004; Wiig, Secord, & Semel, 2009), a norm-referenced test of language, and the Narrative Language Measures (NLM) Listening retell subtest of the CUBED Assessment (Petersen & Spencer, 2016).

The goal for participant recruitment was to identify Spanish-speaking children who did not perform according to age expectations on English measures, indicating that they may benefit from a Tier 2 oral language intervention. To select participants, we conducted a multistep process. First, we examined the children's English NLM Listening retell scores, and any child who earned a retell score of 8 or higher in English was excluded. A retell score of 8 presupposes the use of key story grammar features and places a preschool student above the 20th percentile based on normative data from 281 preschool students across the United States (Petersen & Spencer, 2016). Second, children who earned an English retell score of 0 to 7 but scored within the normal range on the English EV subtest of the CELF-P were also excluded. In other words, scores within age expectations for English on either screening measure disqualified children from being participants. Therefore, children who displayed low English skills and low, moderate, or high Spanish language skills were included as participants. The screening process resulted in 43 children recruited for 12 treatment classrooms and 40 children in 13 control classrooms. Shortly after the pretesting, 2 children from the control group moved away from the area, which resulted in 38 children in the control group.

In five of the 12 treatment classrooms, more than three children qualified to be research participants (i.e., they could potentially benefit from Tier 2 intervention). However, the

TABLE 2
Child Characteristics

Child Characteristic	Treatment Group ($n = 43$)	Control Group ($n = 38$)
Gender, number (percentage)		
Male	16 (37)	13 (34)
Female	25 (58)	21 (55)
Age in months, mean (range)	50 (39–59)	49 (37–59)
Race/ethnicity, number (percentage)		
White	1 (2)	0 (0)
Hispanic/Latino/a	39 (91)	38 (87)
Primary language, number (percentage)		
English	0 (0)	0 (0)
Spanish	31 (72)	27 (71)
Both English and Spanish	7 (16)	6 (16)
Number of children with IEP	0	0

Note. IEP = Individualized Education Program. The percentages do not add up to 100 within each variable and group due to incomplete demographic survey data.

teachers were not able to feasibly provide the Tier 2 intervention to more than one group every day. Therefore, only three children in each class were able to receive the intended multitiered instruction, and the rest of the children ($n = 10$) received only large-group instruction in English. The teachers determined which children would receive the small-group instruction, and the researchers did not guide them in making those choices. Although 10 of the 43 children who were identified as needing Tier 2 small-group support only received large-group instruction, these 10 children were included as research participants because they received part of the intended intervention and the control group received none of it.

The parents completed a brief survey to report demographic information about their children. Child characteristics are shown in Table 2. The parents also reported their highest level of education and annual family income. Only 7% of the treatment group's parents had attended college, with two of them having earned a college degree, and 8% of the control group's parents had attended college, with none having earned a college degree. Only 26% of the treatment group parents and 21% of the control group parents reported the family's annual income to be more than $22,000.

Research Assistants

Research assistants (RAs) were responsible for all screening, data collection, and supporting teachers as they implemented the intervention. The RAs visited each classroom once or twice a week to check in with the teachers and teaching assistants and to conduct fidelity observations. The first author completed rigorous training with the RAs prior to their participation in the study. Because they were all needed to observe fidelity, support teachers' delivery of the intervention, and collect data, they were not blind to assignment.

Video Manual and Training

We created a video manual to explain the rationale and teaching procedures needed to deliver the multitiered language curriculum. The video manual consisted of 13 short (5–15 minutes) modules that covered the active ingredients of the program, its materials, and guidelines for delivering lessons. During a full-day group training prior to the beginning of the school year, the modules were played one by one for the teachers, teaching assistants, and directors from the treatment group. Each teacher and teaching assistant practiced teaching a lesson to the other attendees. Question-and-answer sessions were interspersed throughout the day to address any questions or concerns. In addition to the training, the teachers were given their own flash drives with the video manual so that they could review any module at any time throughout the year. Once they began using the curriculum, the RAs spent 1 to 2 weeks coaching the Head Start teachers and teaching assistants until they felt comfortable delivering the lessons independently.

Research Design and General Procedures

Because the 81 child participants were nested within classrooms, a cluster-randomized group study design was employed to investigate the effect of multitiered dual-language instruction on the children's language skills. After the children were screened and included as research participants, the RAs completed the pretesting (September). The intervention consisted of three units of instruction (Units A–C), with each unit lasting 8 to 10 weeks. Throughout the school year, the children in both control and intervention classrooms were administered several proximal and distal

measures to examine the extent to which the multitiered curriculum affected important child outcomes. Dependent variables included narrative retells, receptive vocabulary, listening comprehension, and general oral language abilities (e.g., understanding and use of grammar). Posttesting was completed at the end of the study (April/May); however, the proximal measures (e.g., receptive vocabulary and narrative retells) were repeated four times across the year to ensure that the participants' skills were assessed both before and after each of the three units of instruction. The Head Start teachers and teaching assistants completed all of the intervention components by integrating them within the routine of their classroom, although each teacher decided how and when to implement each component.

All the research activities, including assessments and intervention, took place in Head Start classrooms. In an effort to minimize noise and distractions, the RAs conducted the assessments with individual children during scheduled activities that were moderately quiet (e.g., drop-off and pickup times, as the children finished snack time, and when the class was at circle time). Although there were a large number of assessments that were administered to the children individually and repeatedly, all of the assessments were extremely brief (most were under 5 minutes), and only one assessment was completed at a time.

Multitiered Dual-Language Narrative Curriculum

The multitiered dual-language narrative curriculum is called Puente de Cuentos ("Bridge Made of Stories"). It features 36 English stories (three units of 12 stories each), with 36 corresponding Spanish stories. Each story was written to include two target vocabulary words (e.g., *rough/áspero*). As the units progressed, coordinating and subordinating conjunctions were folded into the stories and lessons. To accompany each story, a set of five illustrations was created. These illustrations were simple line drawings with minimal color and few details. Photos of the target vocabulary words were included in the materials so teachers could show how the words could be used in contexts other than the stories. Additional information about the stories and lessons can be found in the online supplemental material.

Stories served as the basis for language instruction in small-group and large-group arrangements. Lessons were scripted for teachers and adhered to a consistent format across the three units. During each lesson, the teacher or teaching assistant read the featured story and then guided the children through a series of activities designed to help them learn the meaning of target words and to retell the stories. Some activities required children to respond together as a group to increase active responding, whereas other activities required children to respond individually. When individual children retold the featured stories, they were prompted (and supported) to use all of the story grammar elements, the target vocabulary words, and complex sentences (e.g., with coordinating and subordinating conjunctions).

The Head Start teachers and teaching assistants worked together to determine how they would deliver the components of Puente de Cuentos. All the children in the classrooms participated in the large-group activities, but the research participants received small-group lessons in addition to the large-group lessons as their Tier 2 intervention. A typical implementation consisted of two English large-group lessons, two Spanish small-group lessons, and two English small-group lessons each week. Spanish small-group lessons preceded English small-group lessons to facilitate cross-language transfer. In the three treatment classrooms that did not have a Spanish-speaking teacher or teaching assistant, the children only received English large- and small-group lessons, each twice a week. In addition to the explicit, teacher-led instruction, the teachers embedded several child-directed extension activities throughout their daily routine.

Parents of the children who qualified for the Tier 2 Puente de Cuentos intervention in the classroom received a set of family engagement activities in Spanish. Each activity featured one of the 72 stories from the Puente de Cuentos curriculum and listed questions and suggestions on how to help their children to retell the story and to use the target words in Spanish.

The control group was considered a "business as usual" condition. The center directors reported that teachers used small-group instruction to differentiate individual students but most consistently delivered instruction in large groups. Because the teachers did not have access to a Spanish curriculum or a systematic Spanish program, instruction was completed in English with occasional directions or explanations in Spanish (if the teacher spoke Spanish).

Proximal Measures and Data Collection

Narrative Language Measures Listening. The NLM Listening is a subtest of the CUBED Assessment (Petersen & Spencer, 2016). To collect retell language samples in English and Spanish using NLM Listening, the RAs read a brief story to a child, and the child retold the story. The RAs scored the children's retells in real time, giving points for each story grammar element and indicators of complex language use (e.g., subordinating conjunctions such as *because, when, after*). At each assessment time point, the children were administered three of the NLM Listening parallel forms in a single session lasting 3 to 4 minutes. However, only the retell with the highest score was used in the analysis and to identify the participants. The NLM Listening stories were unfamiliar to the children and children never heard the same story twice. Because NLM Listening stories are similar to those directly taught in Puente de Cuentos, this was considered a proximal outcome measure for this study.

Receptive Picture Vocabulary Assessment. The researcher-designed receptive picture vocabulary assessment measured the children's mastery of the Spanish and English words targeted in the Puente de Cuentos curriculum. The children were shown four different black-and-white line drawings and asked to point to the target word.

Distal Measures and Data Collection

Assessment of Story Comprehension. The Assessment of Story Comprehension (ASC; Spencer & Goldstein, 2019) is a narrative-based, criterion-referenced assessment for preschoolers. It is only in English. During its administration, the RAs read a short story to a child, then asked a series of factual and inferential questions. Examiners wrote the children's answers word for word on record forms and rated each answer for correctness and clarity on a 0 to 2 or 0 to 3 scale, yielding a total of 17 possible points. Six parallel forms were administered, three at pre-intervention (September) and three at postintervention (April/May). The highest score was used for analysis. Because the ASC stories are significantly different from the Puente de Cuentos stories and children answer factual and inferential questions instead of retelling stories, it is considered a distal measure of language comprehension.

Clinical Evaluation of Language Fundamentals: Preschool. The CELF-P in English and Spanish (Semel et al., 2004; Wiig et al., 2009) includes three language subtests that measure general oral language proficiency. The Sentence Structure (SS) subtest requires children to point to pictures corresponding to a spoken sentence. The Word Structure (WS) subtest requires an expressive response that examines children's grammatical abilities. In the EV subtest, children label pictures of objects and actions. The EV subtests of the English and Spanish versions were used for screening, but participants who qualified for Tier 2 intervention also completed the SS and WS subtests in English and Spanish as part of pretesting. Raw scores were calculated and used in the analysis.

Feasibility Measures and Data Collection

Usage Rating Profile–Intervention. At the end of the intervention phase, the classroom teachers and teaching assistants completed the Usage Rating Profile–Intervention (URP-I; Chafouleas, Briesch, Riley-Tillman, & McCoach, 2009). The URP-I consists of 35 questions, each with 6-point Likert scale responses regarding four intervention dimensions: acceptability, understanding, feasibility, and system support. Because each dimension has a different number of items, we converted scores to percent so they can be interpreted.

Fidelity Checklists. The RAs monitored the fidelity of the Puente de Cuentos lessons. During each observation, an RA completed a fidelity checklist that documented adherence (12 items), responsiveness (3 items), and quality (9 items) of the intervention (Dane & Schneider, 1998). The RAs recorded the fidelity of 21% of large-group lessons, 21% of Spanish small-group lessons, and 17% of English small-group lessons. To yield a percent fidelity, the number of items completed as intended or with high quality was divided by the total number of items on the checklist and multiplied by 100.

Intervention Logs. To capture information about the extent to which the children received the intended dose, the researchers provided intervention and attendance logs for each classroom. Dose for each type of teacher-directed lesson (i.e., large-group English, small-group Spanish, or small-group English) was recorded, as well as how many extension activities were completed and for which words and concepts.

Implementation Survey. At the end of the school year, the Head Start teachers completed a short survey. This consisted of nine researcher-generated questions that probed the teachers' perceptions about the modifications completed and needed, planned sustainment, and contextual fit of the Puente de Cuentos curriculum in Head Start settings. Responses were rated using a Likert-type scale of 1 to 5.

Results

Descriptive statistics for the focal measures are shown in Table 3. Less than 1% of the scores were missing overall (18/2754 = 0.0065, or 0.65%), and all available data were used in the multilevel model results that follow. Detailed descriptions of the analyses are available in the online supplemental material.

Proximal Child Outcomes

We evaluated baseline equivalence across the treatment and control groups on pretest measures. As shown in Table 4, tests of pretest differences on these measures were nonsignificant (gs = −.10 to .46), except for Spanish Vocabulary B, for which the treatment group had a significantly higher pretest mean (g = .53). We proceeded to test the differences in posttest scores adjusted for the respective pretest to control for any baseline differences between groups.

NLM Listening English and Spanish. On the English and Spanish NLM posttests, tests of the estimated difference between groups on the adjusted means in the random-intercept analyses of covariance showed statistically significant differences in favor of the treatment group (see Table 5). The 95% confidence intervals, although somewhat wide given the pilot study sample size, support the estimated positive effects for the treatment group. On the English NLM, the effect size was large (g = .85), and the improvement index

TABLE 3

Descriptive Statistics for Pretest, Posttest, and Adjusted Posttest Scores by Treatment Group

	Treatment Group (N_T = 43)			Control Group (N_C = 38)		
Measure	Pretest, *M* (*SD*)	Posttest, *M* (*SD*)	Posttest, M_{adj} [a]	Pretest, *M* (*SD*)	Posttest, *M* (*SD*)	Posttest, M_{adj} [a]
E NLM	1.00 (1.93)	6.86 (5.41)	6.91	1.24 (2.10)	2.92 (3.77)	2.86
S NLM	5.51 (6.15)	10.62 (6.35)	10.09	3.55 (4.86)	6.50 (6.06)	7.08
E Vocab A	9.84 (3.75)	12.72 (4.47)	12.65	9.45 (4.18)	10.54 (3.66)	10.66
E Vocab B	8.21 (2.48)	11.81 (3.73)	11.73	8.32 (3.10)	9.47 (3.26)	9.48
E Vocab C	10.07 (3.59)	12.86 (4.30)	12.53	9.18 (3.97)	10.18 (4.31)	10.51
S Vocab A	13.09 (4.02)	14.86 (4.02)	14.38	11.29 (4.31)	11.16 (4.09)	11.80
S Vocab B	11.86 (3.81)	13.95 (3.95)	13.24	10.08 (2.74)	11.39 (3.04)	12.11
S Vocab C	9.58 (4.34)	11.93 (3.84)	11.55	8.34 (2.39)	9.32 (3.66)	9.75
ASC	1.14 (1.66)	4.23 (3.39)	4.24	1.19 (1.60)	2.71 (2.95)	2.66
E SS CELF	5.74 (4.69)	9.14 (4.30)	9.23	6.19 (3.21)	7.24 (3.67)	7.01
E WS CELF	2.60 (3.40)	6.30 (4.76)	6.18	2.45 (3.03)	4.24 (3.76)	4.40
E EV CELF	3.09 (3.66)	7.49 (5.16)	7.21	2.79 (4.34)	7.29 (5.85)	7.47
S SS CELF	9.65 (4.27)	14.19 (4.84)	13.76	8.45 (3.89)	10.71 (4.23)	10.84
S WS CELF	10.28 (5.51)	14.91 (5.84)	14.39	8.37 (5.40)	11.26 (6.32)	11.89
S EV CELF	17.98 (9.56)	18.88 (10.03)	17.22	13.11 (8.34)	14.45 (9.45)	16.66

Note. E = English; S = Spanish; NLM = Narrative Language Measure; Vocab = Puente de Cuentos Picture Vocabulary Assessment; ASC = Assessment of Story Comprehension; CELF = Clinical Evaluation of Language Fundamentals—Preschool; SS = Sentence Structure; WS = Word Structure; EV = Expressive Vocabulary; *M* = mean; *SD* = standard deviation.

[a]Adjusted posttest means have been adjusted for group differences on the pretest and were used in conducting the analyses of covariance.

TABLE 4

Unconditional Pretest ICCs and Tests of Baseline Equivalence (Random-Intercept ANOVAs[a]), with Hedges' g Effect Sizes With Small-Sample Adjustment

Measure	Pretest, ICC	Estimated $M_T - M_C$, γ_{01} (95% CI)	*p* (for γ_{01})	Hedges *g* (Effect Size)
E NLM	.24	−0.19 (−1.39, 1.00)	.74	−.10
S NLM	.01	1.96 (−0.52, 4.43)	.12	.35
E Vocab A	.09	0.31 (−1.80, 2.41)	.76	.08
E Vocab B	.00	−0.12 (−1.36, 1.13)	.85	−.04
E Vocab C	.07	0.86 (−1.04, 2.75)	.36	.23
S Vocab A	.13	1.94 (−0.27, 4.15)	.08	.46
S Vocab B	.00	1.78 (0.30, 3.27)	.02	.53
S Vocab C	.07	1.26 (−0.52, 3.04)	.16	.35
ASC	.10	−0.09 (−0.97, 0.79)	.83	−.05
E SS CELF	.11	−0.51 (−2.75, 1.74)	.64	−.12
E WS CELF	.00	0.16 (−1.27, 1.59)	.83	.05
E EV CELF	.07	0.33 (−1.73, 2.38)	.74	.08
S SS CELF	.00	1.20 (−0.61, 3.02)	.19	.29
S WS CELF	.01	2.06 (−0.65, 4.77)	.13	.37
S EV CELF	.00	4.87 (0.88, 8.86)	.02	.54

Note. N_T = 43, NC = 38. ICC = intraclass correlation coefficient; T = treatment group; C = control group; E = English; S = Spanish; NLM = Narrative Language Measure; Vocab = Puente de Cuentos Picture Vocabulary Assessment; ASC = Assessment of Story Comprehension; CELF = Clinical Evaluation of Language Fundamentals–Preschool; SS = Sentence Structure; WS = Word Structure; EV = Expressive Vocabulary; ANOVA = analysis of variance; CI = confidence interval.

[a]Solutions for S NLM, E Vocab B, S Vocab B, E WS CELF, S SS CELF, and S EV CELF are equivalent to general linear model–based ANOVAs, as the between-class random-intercept variance component estimate was 0 (or near 0).

TABLE 5

Unconditional Posttest ICCs and Tests of Postintervention Differences in Adjusted Means (Random-Intercept ANCOVAs[a]), with Hedges' g Effect Sizes and Improvement Indexes

Measure	Posttest, ICC	[b]Estimated $M_{adjT} - M_{adjC}$, γ_{01} (95% CI)	p (for γ_{01})	Hedges g (effect size)	Improvement index
E NLM	.21	4.05 (2.06, 6.05)	<.01	.85	30%
S NLM	.02	3.01 (0.53, 5.50)	.02	.48	18%
E Vocab A	.13	1.99 (0.32, 3.67)	.02	.48	18%
E Vocab B	.26	2.25 (0.42, 4.09)	.02	.63	24%
E Vocab C	.25	2.02 (0.39, 3.66)	.02	.46	18%
S Vocab A	.08	2.58 (0.94, 4.22)	<.01	.63	24%
S Vocab B	.22	1.12 (−0.60, 2.85)	.19	.31	12%
S Vocab C	.26	1.80 (0.39, 3.22)	.02	.48	18%
ASC	.10	1.59 (0.18, 2.99)	.03	.49	19%
E SS CELF	.00	2.22 (0.60, 3.84)	.01	.55	21%
E WS CELF	.09	1.78 (−0.12, 3.69)	.07	.41	16%
E EV CELF	.09	−0.26 (−2.38, 1.86)	.80	−.05	−2%
S SS CELF	.25	2.91 (0.23, 5.60)	.03	.63	24%
S WS CELF	.05	2.50 (−0.27, 5.27)	.07	.41	16%
S EV CELF	.00	0.56 (−3.00, 4.11)	.75	.06	2%

Note. $N_T = 43$, $N_C = 38$. ICC = intraclass correlation coefficient; T = treatment group; C = control group; E = English; S = Spanish; NLM = Narrative Language Measure; Vocab = Puente de Cuentos Picture Vocabulary Assessment; ASC = Assessment of Story Comprehension; CELF = Clinical Evaluation of Language Fundamentals–Preschool; SS = Sentence Structure; WS = Word Structure; EV = Expressive Vocabulary; ANCOVA = analysis of covariance; CI = confidence interval.

[a]Solutions for E NLM, S NLM, and E SS CELF are equivalent to general linear model–based analyses of variance, as the between-class random-intercept variance component estimate was 0 (or near 0).

[b]Adjusted posttest means were adjusted for group differences on the pretest and used in conducting the ANCOVAs.

was 30%, indicating that an average student in the control group would be expected to score about 30% higher if receiving the intervention. The effect size for the Spanish NLM was moderately strong (g = .48), with an improvement index of 18%.

Receptive Picture Vocabulary Assessment. With the exception of the posttest for Spanish Unit B, tests of the estimated difference between groups on the adjusted posttest means for English and Spanish vocabulary were statistically significant, favoring the treatment group (see Table 5). Effect sizes for these five measures (English Vocabulary A, B, and C; Spanish Vocabulary A and C) were moderate (gs = .46 to .63). The improvement indices suggested that an average student in the control group would be expected to score from 18% to 24% higher on the vocabulary assessments if receiving the intervention. Although the vocabulary posttest for Spanish B was not statistically significant, the effect size was not trivial (g = .31), and the improvement index was 12% in favor of the treatment group.

Distal Child Outcomes

As shown in Table 4, tests of pretest differences on distal measures were not significant with small to moderate effect sizes (gs = −.12 to .37), except for Spanish EV, for which the treatment group had a significantly higher pretest mean (g = .54). We evaluated differences in posttest scores adjusted for the respective pretest to control for any baseline differences between groups.

Assessment of Story Comprehension. The random-intercept analysis of covariance on the ASC adjusted posttest means was statistically significant, with a moderate effect size (g = .49). The improvement index estimated that an average student in the control group would be expected to score 19% higher on the ASC if receiving the intervention, which would be a meaningful gain.

Clinical Evaluation of Language Fundamentals—Preschool. Results for adjusted posttest differences between the treatment and control groups differed across the CELF-P SS, WS, and EV subtests, but were very consistent for subtests across English and Spanish. The treatment group clearly outperformed the control group on SS, evidenced by statistically significant differences, moderate effect sizes (gs = .55 for English and .63 for Spanish), and improvement indices. An average student in the control group would be expected to score 21% higher on SS for English and 24% higher for Spanish.

Differences in adjusted posttest means were not statistically significant for WS in either language, but effect sizes

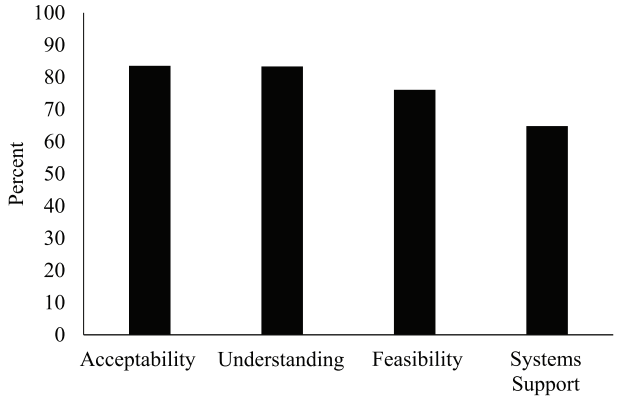

FIGURE 1 *Teachers' mean ratings on the Usage Rating Profile–Intervention.*

approached moderate (*g*s = .41), with improvement indices of 16% in support of intervention effects. The final two distal measures—English and Spanish EV—did not evidence any appreciable differences between treatment and control group adjusted means.

Feasibility

Usage Rating Profile–Intervention. Mean percentage for each dimension of the URP-I is displayed in Figure 1. Higher scores in acceptability, understanding, and feasibility suggest that the intervention was perceived as useful and doable. The teachers and teaching assistants reported Puente de Cuentos to be more acceptable than feasible, although the scores for both characteristics were moderately high. The teachers also reported having a good understanding of the curriculum. For systems support, the teachers reported lower scores than for the other dimensions, but because of the nature of the scale, higher scores were not necessarily desired.

Fidelity. After the Head Start teachers and teaching assistants felt comfortable delivering the lessons (1–2 weeks), the RAs began assessing their intervention fidelity using the fidelity checklists. The teachers and teaching assistants demonstrated consistently high fidelity to the Puente de Cuentos procedures. For the small-group lessons in Spanish, the mean fidelity scores were 97%, 96%, and 98% for Units A, B, and C, respectively. For the small-group English lessons, they were 97%, 96%, and 97%, respectively. For the large-group lessons, fidelity was slightly lower, with mean fidelity scores of 91%, 97%, and 94% for the respective units.

Intervention Logs. Based on a review of the intervention logs the teachers completed, very few teacher-directed lessons were omitted, with the exception of small-group lessons in Spanish in the three treatment classrooms without a Spanish-speaking teacher. All planned lessons had been implemented by the middle of May. The small-group intervention portion of the log revealed that all the research participants were present for at least 85% of the Tier 2 lessons intended for them. Moreover, 90% or more of the target words and concepts were addressed through extension activities in all the treatment classrooms.

Implementation Survey. Mean ratings of all the teachers and teaching assistants who completed the implementation survey are displayed in Table 6. Overall, they reported that they made few modifications during the study, but some had plans to make more. Most of the teachers planned to continue using Puente de Cuentos after the study. The mean ratings indicate that there is a reasonable contextual fit between the intervention and their values, students, and setting.

Discussion

The importance of building oral language skills is clear as there is a strong link between oral language and reading comprehension (Cain et al., 2004; Chaney, 1998; Clarke et al., 2010; Dickinson & Tabors, 2001; Larney, 2002). Vocabulary and narrative skills are particularly important areas to develop early so that children can benefit more from subsequent instruction and comprehension of what is read to them and what they read (Cain & Oakhill, 2011; Elleman et al., 2009; Mehta et al., 2005; Verhoeven & van Leeuwe, 2008). If oral language instructional efforts can incorporate children's L1 and produce meaningful improvements in English, there is an added benefit of helping cultivate a bilingual and biliterate society (Collier & Thomas, 2017). The purpose of this early-stage efficacy study was to examine the extent to which multitiered dual-language instruction improved children's Spanish and English language skills on proximal and distal measures of vocabulary, narrative retells, language comprehension, and general language abilities.

Proximal Measures of Vocabulary and Narrative Retell

Consistent with prior English, oral narrative–based language intervention studies that have focused on proximal outcomes (e.g., Spencer, Petersen, & Adams, 2015; Spencer, Petersen, Slocum, & Allen, 2015; Spencer, Weddle, Petersen, & Adams, 2017), we found statistically significant effects for narrative retells in English. Narrative retelling was the most salient instructional activity in the Puente de Cuentos instruction, with all the large-group and half of the small-group lessons based on English stories. The teachers supported the children's practice of each model story, English vocabulary, and English language complexity through retelling activities in every lesson. Only half of the small-group lessons featured Spanish story retelling, which may account for the differences in the effect sizes for English (g = .85) and Spanish (g = .48) retell outcomes. Although improvement in the proximal narrative retell outcome was expected,

TABLE 6
Implementation Survey Results

Implementation survey item		Mean rating
Modifications		1 = *not at all*, 5 = *very much*
1	To what extent was the Puente de Cuentos program implemented as it was written and designed?	4.67
2	To what extent have you made changes to the Puente de Cuentos program by shortening the lessons?	2.00
3	To what extent have you made changes to the Puente de Cuentos by incorporating new materials and activities?	3.08
Planned sustainment		1 = *definitely not*, 5 = *probably*
4	To what extent do you plan to continue using Puente de Cuentos in your classroom?	4.83
5	Do you intend to make changes to the Puente de Cuentos program?	3.36
Contextual fit		1 = *strongly disagree*, 5 = *strongly agree*
6	The Puente de Cuentos program is compatible with your values and teaching philosophy.	4.42
7	The Puente de Cuentos program is more effective than other programs that address language development.	4.20
8	The complexity of the content, activities, and structure of the Puente de Cuentos program is appropriate for preschoolers.	4.08
9	The complexity of the content, activities, and structure of the Puente de Cuentos program is appropriate for Head Start preschool classrooms.	4.83

growth in narrative language can have meaningful immediate and future consequences. Narrative language has been shown to be correlated and causally related to later academic success (Barton-Hulsey et al., 2017; Catts et al., 2016; Clarke et al., 2010). It is worth noting that the Puente de Cuentos curriculum improved the oral narrative language of one of the populations most at risk of not meeting future reading comprehension standards (National Center for Educational Statistics, 2017).

Improvements on the researcher-made receptive vocabulary assessment were statistically significant for all three units of English words, and significant for two of the three units in Spanish. All effect sizes were considered educationally meaningful ($g > .25$; Institute of Education Sciences, 2017), although the children in the treatment group made smaller gains on the Spanish Unit B vocabulary assessment than those in the control group. Across the year, the teachers explicitly taught 36 verbs and adjectives in English and 36 verbs and adjectives in Spanish. They were strategically selected to be less common, Tier 2 words (Beck, McKeown, & Kucan, 2002). The multitiered dual-language curriculum was intentionally designed to ensure that the most attention would be given to words that are most difficult to learn. Thus, the teachers were able to direct their explicit instruction and intentional practice toward these less common and more challenging verbs and adjectives.

The meaningful improvements in Spanish receptive vocabulary suggest that the combined dose of small-group Spanish lessons in the classroom and family engagement activities was sufficient to help the children learn the words in Spanish. In previous studies, we found little evidence of improvement on the Spanish receptive vocabulary assessment but adequate evidence for improved English vocabulary (Spencer, Moran, Petersen, Thompson, & Restrepo, 2018; Spencer, Petersen, Restrepo, Thompson, & Gutierrez Arvizu, 2019). In these studies, children received English instruction in large groups, in small groups, and through extensions (e.g., storybook reading and child-directed center activities); however, Spanish instruction was only delivered in small groups twice a week for 20 minutes. In the current study, the children's families received a set of family engagement activities that aligned with all the lessons, and they were only in Spanish. We speculate that by boosting the children's exposure to the Spanish vocabulary through the family engagement activities, a better English-Spanish instructional balance was created. Another difference between previous studies and the current study was that all the families viewed a video module that showed them how to use the family engagement activities to facilitate storytelling, encourage the use of the target words, and help their children answer questions about the stories. Because we did not isolate the effect of the family engagement activities, this supposition will require replication and more rigorous investigation in the future.

Distal Measures of Language Comprehension and General Language Skills

The chain of logic for building vocabulary and narrative skills is that, if truly successful, improvements will also be

detected on language-related measures that do not closely match the intervention. If children's language comprehension can be improved before they enter kindergarten, there is a chance that their future reading comprehension will also benefit. Although this was not investigated experimentally in this study, other research suggests that language outcomes mediate the effects of language intervention on reading comprehension for students in primary grades (Bowyer-Crane et al., 2008; Clarke et al., 2010; Language and Reading Research Consortium, Jiang, & Logan, 2019). It is the same logic that underpins early-childhood intervention aimed at enhancing language comprehension. The ASC is a standardized, criterion-referenced assessment tool that uses stories and comprehension questions to assess children's language comprehension skills, similar to the common reading comprehension tasks in elementary grades. The ASC stories are longer and more complex than the stories featured in the multitiered curriculum and were strategically designed to capture inferential comprehending. At pretest, the children's scores were extremely low, indicating that they were unable to answer questions about a story. This could mean that they did not understand the stories or that they understood the stories but had insufficient expressive language to respond to the questions. At posttest, the children in the treatment group showed small, but important, gains over the children in the control group. Although we can be confident that the multitiered dual-language instruction was responsible for the observed gains, there is substantially more room for growth, as the children in the treatment condition had mean posttest ASC scores of 4.24 out of a total of 17 possible points. Given that the ASC is a distal measure and the children were not directly trained in the intervention to answer factual and inferential questions, this is considered a meaningful outcome with a moderate effect size ($g = .49$), indicating the significant promise of the intervention.

As further evidence of promise, the multitiered dual-language instruction had a statistically significant impact on the treatment group's scores on the CELF-P SS subtest in English and Spanish. Although not statistically significant, group differences on the English and Spanish WS subtest were meaningful, with moderate effect sizes. This pattern of responding corresponds to developmental expectations. SS is a receptive task in which children point to the picture that corresponds to the sentence the examiner says, while WS requires children to produce a grammatically complex phrase or sentence. It is reasonable that children learn to understand a second language before they are able to speak it. It is possible that stronger effects would be seen if children received 2 years of focused dual-language instruction.

The children in the treatment group did no better than the children in the control group on the EV subtest of the CELF-P in English or in Spanish. Again, this is a consistent and expected pattern. Both groups made equivalent gains from pretest to posttest in English and no gains in Spanish. Although the intervention targeted a large number of words, none of them are on the EV subtests. The EV subtests feature words that are commonly learned in preschools, such as *crying, riding, carrot*, and *firefighter*, which are distinctly different from the types of words taught in Puente de Cuentos (e.g., *narrow, tremble*). From these results, it can be deduced that general classroom English instruction, to which both groups were exposed, was sufficient to improve the children's ability to expressively identify the items on the English EV subtest. Evidence of this can be seen in the lack of growth observed for Spanish EV. Because their general classroom instruction was primarily in English, they only learned the English words. While a goal of most vocabulary interventions is to improve children's ability to learn new words, there are distinct barriers to validly measuring this construct (Camilleri & Botting, 2013). Many have argued that because standardized measures of preschoolers' vocabulary are inappropriate to detect the effects of vocabulary-rich language interventions, this gap in the literature warrants urgent attention (Hoffman, Teale, & Paciga, 2014; National Institute of Child Health and Human Development, 2000).

Feasibility

The URP-I data suggest that the Puente de Cuentos curriculum was generally acceptable to the teachers and teaching assistants. It was easy to understand and regarded as feasible in their setting. This is further evidenced by the high fidelity of lesson delivery and completion of all of the planned lessons before the end of the year. The dimension of systems support of the URP-I cannot be interpreted as easily because high scores suggest that the teacher is unable to implement the intervention without the help of others and low scores indicate that they can easily implement the intervention on their own. A multitiered instructional system necessitates a fair amount of help and teamwork. For example, administrative support is needed so that schedules can be adapted and lesson plans modified to better fit a multitiered delivery. Moreover, the teachers often divided the lesson delivery responsibilities with the teaching assistant, which is an acceptable use of personnel and resources.

The implementation survey revealed that the teachers implemented the program as it was designed, although some of them made changes to the duration of lessons and the materials and activities used. The ability to modify a research-based practice has been associated with sustainability of that practice (Klinger, Vaughn, Hughes, & Arguelles, 1999), so midrange scores (2.00–4.00) on the implementation survey may indicate that teachers feel empowered and knowledgeable about how to adapt Puente de Cuentos for their classrooms. Multitiered instructional systems may pose paradigmatic shifts for early-childhood educators. Likewise, not all early-childhood professionals value teacher-directed instruction. We attribute the high

contextual fit scores (4.08–4.83) to the balance of short explicit instruction sessions with child-directed activities in Puente de Cuentos.

Contributions to MTSS in Early Childhood

The implementation of MTSS across early-childhood settings has been limited, and multitiered systems of language support have rarely been attempted or reported in the research literature. This is one of the first studies to report on the efficacy and feasibility of a dual-language multitiered curriculum for preschool children. The promise of MTSS transcends special education and extends services to any and all students who may need extra support. Thus, through MTSS, students who are not meeting English language expectations due to various external factors are eligible to receive the language support they need. Dual-language instruction has been shown to have an equal or stronger impact on academic performance as compared with English-only approaches (Collier & Thomas, 2017; Collins, 2014; Mahoney et al., 2005; Rolstad et al., 2005), with the added benefit of sociocultural, socioeconomic, linguistic, and cognitive gains (Collier & Thomas, 2017). The implementation of a dual-language multitiered system of support merges two powerful, evidence-based approaches. With a tiered system in place that provides special services to all students in need of additional support, and a focus on both L1 and L2, there is a real possibility of meaningful change and, for the first time, significant improvement in reading outcomes for dual-language learners.

Limitations and Future Directions

Despite the valuable contributions this study makes to the literature on dual-language interventions and to the literature on MTSS in early-childhood contexts, there are a number of limitations and points to consider for future research. First, because this was a pilot early-efficacy study, we were limited by our financial resources. These limitations reduced the number of classrooms that could realistically be managed and our ability to monitor conditions in the control classrooms. The small sample may be responsible for the lack of statistical significance found for Spanish receptive vocabulary Unit B and for the WS subtest in English and Spanish. It is possible that significance will be observed when a larger, fully powered efficacy trial can be completed. A second limitation is also related to resources. We were unable to mask the classrooms' assignment to conditions because all of the RAs were needed to collect pre- and posttest data and observe the teachers for fidelity. With greater financial resources, a second group of data collectors can remain blind to condition.

A number of limitations were related to the dose of the intervention. Several children assigned to the treatment group did not receive the full intervention for various reasons. First, we were unable to be more selective about the classrooms we recruited to participate. Three of the treatment classrooms did not have a Spanish-speaking teacher or teaching assistant. This meant that nine children in the treatment group received multitiered English language instruction instead of dual-language instruction. There were also more control classrooms without a Spanish-speaking teacher than treatment classrooms. Second, in five of the treatment classrooms, more than three children qualified to be research participants. Because the teachers did not have the time to conduct more than one small-group intervention every day, they selected three children for Tier 2 intervention, and the rest ($n = 10$) received only large-group instruction with the rest of the class. The researchers did not advise the teachers how to select the children, but it was hypothesized that they selected the three children about whom they were most concerned. The effects of these limitations are not known because the samples were too small to analyze for possible differential effects. It should be noted that most of the research participants in the treatment groups received some level of Spanish exposure through the family engagement activities, so there is a possibility that this compensated somewhat for what was missed in school.

Although not necessarily weaknesses of the current study, there are a few recommendations that future research in this area can address. The extent to which Spanish instruction benefited the children should be examined in future research. We did not attempt to isolate the effect of the Spanish components or examine cross-language transfer directly, but future researchers should plan for a systematic and rigorous analysis of the value added of using children's L1 in multitiered dual-language instruction. Likewise, the impact of the small-group instruction on top of the large-group instruction is assumed to have added benefit. However, this should be examined empirically, comparing different variations and possible configurations of the Puente de Cuentos curriculum.

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A140093 for $1,481,960. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

ORCID iD

Trina D. Spencer https://orcid.org/0000-0002-3531-8276

References

Anderson, R., & Pearson, P. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. Pearson, R. Barr, M. Kamil, P. Mosenthal, & R. Dykstra (Eds.), *Handbook of reading research* (pp 255–291). New York, NY: Longman.

Baker, C. (2000). *A parents' and teachers' guide to bilingualism* (2nd ed.). Clevedon, England: Multilingual Matters.

Barnett, W. S., Yarosz, D. J., Thomas, J., Jung, K., & Blanco, D. (2007). Two-way and monolingual English immersion in preschool education: An experimental comparison. *Early Childhood Research Quarterly*, *22*, 277–293. doi:10.1016/j.ecresq.2007.03.003

Barton-Hulsey, A., Sevcik, R. A., & Romski, M. A. (2017). Narrative language and reading comprehension in students with mild intellectual disabilities. *American Journal on Intellectual and Developmental Disabilities*, *122*, 392–408. doi:10.1352/1944-7558-122.5.392

Beck, I. L., McKeown, M., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford Press.

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*, 1461–1476. doi:10.1017/S0142716416000060

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., . . . Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *49*, 422–432. doi:10.1111/j.1469-7610.2007.01849.x

Cain, K., Lemmon, K., & Oakhill, J. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *9*, 671–681. doi:10.1037/0022-0663.96.4.671

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities*, *44*, 431–443. doi:10.1177/0022219411410042

Camilleri, B., & Botting, N. (2013). Beyond static assessment of children's receptive vocabulary: The dynamic assessment of word learning (DAWL). *International Journal of Language & Communication Disorders*, *48*, 565–581.

Carta, J. J., & Young, R. M. (2019). *Multi-tiered systems of support for young children: Driving change in early education*. Baltimore, MD: Brookes.

Castro, D. C., Garcia, E. E., & Markos, A. M. (2013). *Dual language learners: Research informing policy*. Chapel Hill: University of North Carolina, Frank Porter Graham Child Development Institute, Center for Early Care and Education Research-Dual Language Learners.

Castro, D. C., Páez, M., Dickinson, D. K., & Frede, E. (2011). Promoting language and literacy in young dual language learners: Research, practice and policy. *Child Development Perspectives*, *5*, 15–21. doi:10.1111/j.1750-8606.2010.00142.x

Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research*, *36*, 948–958. doi:10.1044/jshr.3605.948

Catts, H. W., Adlof, S. M., & Ellis Weismer, S. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*, *49*, 278–293. doi:10.1044/1092-4388(2006/023)

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research*, *45*, 1142–1157. doi:10.1044/1092-4388(2002/093)

Catts, H. W., Nielsen, D. C., Bridges, M. S., & Liu, Y. S. (2016). Early identification of reading comprehension difficulties. *Journal of Learning Disabilities*, *49*, 451–465. doi:10.1177/0022219414556121

Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., & McCoach, D. B. (2009). Moving beyond assessment of treatment acceptability: An examination of the factor structure of the Usage Rating Profile—Intervention (URP-I). *School Psychology Quarterly*, *24*, 36–47. doi:10.1037/a0015146

Chaney, C. (1998). Preschool language and metalinguistic skills are links to reading success. *Applied Psycholinguistics*, *19*, 433–446. doi:10.1017/S0142716400010250

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading comprehension difficulties: A randomized controlled trial. *Psychological Science*, *21*, 1106–1116. doi:10.1177/0956797610375449

Collier, V., & Thomas, W. (2017). Validating the power of bilingual schooling: Thirty-two years of large-scale, longitudinal research. *Annual Review of Applied Linguistics*, *37*, 203–217. doi:10.1017/S0267190517000034

Collins, B. A. (2014). Dual language development of Latino children: Effect of instructional program type and the home and school language environment. *Early Childhood Research Quarterly*, *29*, 389–397. doi:10.1016/j.ecresq.2014.04.009

Coltrane, B. (2003). *Working with young English language learners: Some considerations*. Washington, DC: ERIC Clearinghouse on Languages and Linguistics.

Craig, H. K., Connor, C. M., & Washington, J. A. (2003). Early positive predictors of later reading comprehension for African American students: A preliminary investigation. *Language, Speech, and Hearing Services in Schools*, *34*, 31–43. doi:10.1044/0161-1461(2003/004)

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*, 23–24.

Dickinson, D., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E., & Poe, M. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology*, *95*, 465–481. doi:10.1037/0022-0663.95.3.465

Dickinson, D. K., & Tabors, P. O. (Eds.). (2001). *Beginning literacy with language: Young children learning at home and school*. Baltimore, MD: Paul H. Brookes.

Dodge, D. T., Colker, L. J., & Heroman, C. (2002). *The creative curriculum for preschool* (4th ed.). Washington, DC: Teaching Strategies.

Durán, L. K., & Wackerle-Hollman, A. (2019). Meeting the needs of young dual language learners in multi-tiered systems of support. In J. J. Carta & R. M. Young (Eds.), *Multi-tiered systems of support for young children: Driving change in early education* (pp. 171–190). Baltimore, MD: Brookes.

Elleman, A., Lindo, E., Morphy, P., & Compton, D. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Educational Effectiveness*, *2*, 1–44. doi:10.1080/19345740802539200

Fuchs, D., & Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice*, *22*, 129–136. doi:10.1111/j.1540-5826.2007.00237.x

Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*, 6–10. doi:10.1177/074193258600700104

Greenwood, C. R., Carta, J. J., Atwater, J., Goldstein, H., Kaminski, R., & McConnell, S. (2013). Is a response to intervention (RTI) approach to preschool language and early literacy instruction needed? *Topics in Early Childhood Special Education*, *33*, 48–64. doi:10.1177/0271121412455438

Griffin, T. M., Hemphill, L., Camp, L., & Wolf, D. P. (2004). Oral discourse in the preschool years and later literacy skills. *First Language*, *24*(71 Pt. 2), 123–147. doi:10.1177/0142723704042369

Hammer, C. S., Lawrence, F. R., & Miccio, A. W. (2007). Bilingual children's language abilities and early reading outcomes in Head Start and kindergarten. *Language, Speech, and Hearing Services in Schools*, *38*, 237–248. doi:10.1044/0161-1461(2007/025)

Heroman, C., Tabors, P. O., & Teaching Strategies, Inc. (2010). *Teaching strategies GOLD: Birth through kindergarten: Assessment toolkit*. Washington, DC: Teaching Strategies.

Hoffman, J. L., Teale, W. H., & Paciga, K. A. (2014). Assessing vocabulary learning in early childhood. *Journal of Early Childhood Literacy*, *14*, 482–509.

Institute of Education Sciences, U.S. Department of Education. (2017). *What Works Clearinghouse™ standards handbook* (Version 4.0). Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Klinger, J. K., Vaughn, S., Hughes, M. T., & Arguelles, M. E. (1999). Sustaining research-based practices in reading: A 3 year follow-up. *Remedial and Special Education*, *20*, 263–274. doi:10.1177/074193259902000502

Language and Reading Research Consortium, Jiang, H., & Logan, J. (2019). Improving reading comprehension in the primary grades: Mediated effects of a language-focused classroom intervention. *Journal of Speech, Language, and Hearing Research*, *62*, 2812–2828. doi:10.1044/2019_JSLHR-L-19-0015

Larney, R. (2002). The relationship between early language delay and later difficulties in literacy. *Early Child Development and Care*, *172*, 183–193. doi:10.1080/03004430210890

MacSwan, J., & Rolstad, K. (2005). Modularity and the facilitation effect: Psychological mechanisms of transfer in bilingual students. *Hispanic Journal of Behavioral Sciences*, *27*, 224–243. doi:10.1177/0739986305275173

MacWhinney, B. (1999). *The emergence of language*. Mahwah, NJ: Psychology Press.

Mahoney, K., MacSwan, J., & Thompson, M. (2005). English language learners. In D. García & A. Molnar (Eds.), *The condition of pre-K–12 education in Arizona: 2005* (pp. 1–24). Tempe: Arizona State University, Education Policy Research Laboratory.

Marian, V., & Kaushanskaya, M. (2007). Language context guides memory content. *Psychonomic Bulletin & Review*, *14*, 925–933. doi:10.3758/BF03194123

Marston, D. (2005). Tiers of intervention in responsiveness to intervention: Prevention outcomes and learning disabilities identification patterns. *Journal of Learning Disabilities*, *38*(6), 539–544. doi:10.1177/00222194050380061001

Mehta, P., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in Grades 1–4. *Scientific Studies of Reading*, *9*, 85–116. doi:10.1207/s1532799xssr0902_1

Méndez, L., Crais, E. R., Castro, D., & Kainz, K. (2015). A culturally and linguistically responsive vocabulary approach for young Latino dual language learners. *Journal of Speech, Language, and Hearing Research*, *58*, 93–106. doi:10.1044/2014_JSLHR-L-12-0221

Miller, J. F., Heilmann, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research and Practice*, *21*, 30–43. doi:10.1111/j.1540-5826.2006.00205.x

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

National Center for Education Statistics, U.S. Department of Education. (2017). *National assessment of educational progress: Reading assessment*. Retrieved from https://nces.ed.gov/nationsreportcard/reading/

National Institute of Child Health and Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (Report of the National Reading Panel, NIH Publication No. 00-4769). Washington, DC: Government Printing Office.

Perozzi, J. A. (1985). A pilot study of language facilitation for bilingual, language-handicapped children: Theoretical and intervention implications. *Journal of Speech and Hearing Disorders*, *50*, 403–406. doi:10.1044/jshd.5004.403

Perozzi, J. A., & Chavez Sanchez, M. L. (1992). The effect of instruction in L1 on receptive acquisition of L2 for bilingual children with language delay. *Language, Speech, and Hearing Services in Schools*, *23*, 348–352. doi:10.1044/0161-1461.2304.348

Petersen, D. B., & Spencer, T. D. (2016). *CUBED*. Laramie, WY: Language Dynamics Group.

Petersen, D. B., Thompsen, B. E., Guiberson, M., & Spencer, T. D. (2016). Cross-linguistic interactions from second language to first language as the result of individualized narrative language intervention with children with and without language impairment. *Applied Psycholinguistics*, *37*, 703–724. doi:10.1017/S0142716415000211

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™*. Baltimore, MD: Paul H. Brookes.

Proctor, C. P., Carlo, M. S., August, D., & Snow, C. E. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology*, *97*, 246–256. doi:10.1037/0022-0663.97.2.246

Restrepo, M. A., Morgan, G. P., & Thompson, M. S. (2013). The efficacy of a vocabulary intervention for dual-language learners with language impairment. *Journal of Speech, Language, and Hearing Research*, *56*, 748–765. doi:10.1044/1092-4388(2012/11-0173)x

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

Semel, E., Wiig, E. H., & Secord, W. A. (2004). *Clinical evaluation of language fundamentals® screening test* (4th ed.). Toronto, Ontario, Canada: Psychological Corporation.

Spencer, T. D., & Goldstein, H. (2019). *Assessment of story comprehension*. Baltimore, MD: Paul H. Brookes.

Spencer, T. D., Moran, M. K., Petersen, D. B., Thompson, M. S., & Restrepo, M. A. (2018). *Feasibility of a multi-tiered dual language narrative curriculum for Spanish-speaking preschoolers: An indispensable phase of iterative development*. Manuscript submitted for publication.

Spencer, T. D., Petersen, D. B., & Adams, J. L. (2015). Tier 2 language intervention for diverse preschoolers: An early-stage randomized control group study following an analysis of response to intervention. *American Journal of Speech-Language Pathology*, *24*, 619–636. doi:10.1044/2015_AJSLP-14-0101

Spencer, T. D., Petersen, D. B., Restrepo, M. A., Thompson, M., & Gutierrez Arvizu, M. N. (2019). The effect of Spanish and English narrative intervention on the language skills of young dual language learners. *Topics in Early Childhood Special Education*, *38*, 204–219. doi:10.1177/0271121418779439

Spencer, T. D., Petersen, D. B., Slocum, T. A., & Allen, M. M. (2015). Large group narrative intervention in Head Start preschools: Implications for response to intervention. *Journal of Early Childhood Research*, *13*, 196–217. doi:10.1177/1476718X13515419

Spencer, T. D., Weddle, S. A., Petersen, D. B., & Adams, J. L. (2017). Multi-tiered narrative intervention for preschoolers: A Head Start implementation study. *Dialog*, *20*, 1–28. doi:10.1016/j.ecresq.2019.01.004

Stein, N., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Norwood, NJ: Ablex. doi:10.1353/lan.1980.0030

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal model. *Developmental Psychology*, *38*, 934–947. doi:10.1037/0012-1649.38.6.934

Vaughn, S., Cirino, P., Linan-Thompson, S., Mathes, P., Carlson, C., Hagan, E., . . . Francis, D. (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. doi:10.3102/00028312043003449

Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, *22*, 407–423. doi:10.1002/acp.1414

Wiig, E. H., Secord, W., & Semel, E. M. (2009). *Clinical evaluation of language fundamentals: Preschool 2 Spanish* (2nd ed.). San Antonio, TX: Psychological Corporation.

Zucker, T. A., Cabell, S. Q., Justice, L. M., Pentimonti, J. M., & Kaderavek, J. N. (2013). The role of frequent, interactive prekindergarten shared reading in the longitudinal development of language and literacy skills. *Developmental Psychology*, *49*, 1425–1439. doi:10.1037/a0030347

Authors

TRINA D. SPENCER is an associate professor at the Rightpath Research & Innovation Center at the University of South Florida. Her research interests include multitiered system of supports, curriculum and instructional design, and early-childhood language and literacy.

MEGHAN MORAN is an instructor in the Department of Applied Linguistics at Northern Arizona University. Her areas of interest include teaching English as a second language.

MARILYN S. THOMPSON is a professor in the School of Social and Family Dynamics at Arizona State University. Her areas of expertise include quantitative data analysis and children's academic development.

DOUGLAS B. PETERSEN is an associate professor in the Department of Communication Disorders at Brigham Young University. His areas of interest include assessment development, dynamic assessment, and dual-language learners.

M. ADELAIDA RESTREPO is a professor of speech and hearing science at Arizona State University. She researches bilingualism and bilingual language and literacy interventions.