



Literacy Skill Development in Bilingual Learners: A Quasi-Experiment of Small Group Instruction

Darrell M. Hull, Marcus A. Fagan, Ricardo González-Carriedo, Rossana Boyd, Franklin Lemmons, Patricia Parker & Princy Sebastian

To cite this article: Darrell M. Hull, Marcus A. Fagan, Ricardo González-Carriedo, Rossana Boyd, Franklin Lemmons, Patricia Parker & Princy Sebastian (2026) Literacy Skill Development in Bilingual Learners: A Quasi-Experiment of Small Group Instruction, *Journal of Latinos and Education*, 25:3, 833-846, DOI: [10.1080/15348431.2025.2569587](https://doi.org/10.1080/15348431.2025.2569587)

To link to this article: <https://doi.org/10.1080/15348431.2025.2569587>



Published online: 09 Oct 2025.



Submit your article to this journal [↗](#)



Article views: 162



View related articles [↗](#)



View Crossmark data [↗](#)



Literacy Skill Development in Bilingual Learners: A Quasi-Experiment of Small Group Instruction

Darrell M. Hull^a, Marcus A. Fagan^b, Ricardo González-Carriedo ^c, Rossana Boyd^c, Franklin Lemmons^a, Patricia Parker^d, and Princy Sebastian^a

^aDepartment of Educational Psychology, University of North Texas; ^bOffice of Research and Sponsored Programs, Texas Women's University; ^cDepartment of Teacher Education and Administration, University of North Texas; ^dEducational Services, Carrollton-Farmers Branch ISD

ABSTRACT

This study evaluated the impact of an 18-week small-group literacy intervention for third-grade bilingual learners following teacher professional development and coaching. Teachers worked with Tier 2 students in groups of four to six, emphasizing foundational reading skills, multimodal instruction, and vocabulary development. Using propensity score matching ($n = 121$; 22 treatment) and hierarchical linear modeling, the study examined effects on the Measures of Academic Progress, Fountas and Pinnell, and STAAR assessments. Findings indicate positive effects supporting continued implementation of small-group literacy instruction and provide recommendations for future professional development and research on bilingual learner interventions.

KEYWORDS

Small group instruction; second language acquisition; bilingual learners; English language development; HLM; propensity scores

Bilingual learners¹ (BLs) represent a growing contingent of students in the United States (Hussar et al., 2020). In the fall of 2021, 10.6% of public-school students in grades K-12 were classified as BLs, representing 5.3 million students (National Center for Education Statistics [NCES], 2023). In Texas, where this study took place, there were 1.2 million BLs during the 2022–2023 school year, which constituted 22% of the student population (Texas Education Agency [TEA], 2024).

Under civil rights law, public schools are obligated to ensure BLs equal access to education (U.S. Department of Education, Office for Civil Rights 2025). In this context, it is imperative that schools adapt their practices to the academic and linguistic needs of these students who are expected to master content of different subject areas at the same time as they are acquiring English as an additional language. Rather than adopting one-size-fits-all approaches for teaching BLs, there is consensus about the need to implement interventions and instructional strategies that target specific areas of need with the ultimate goal of improving the literacy skills and second language acquisition of BLs (Huebner, 2009).

For this purpose, the United States Department of Education, through the Institute of Education Sciences and What Works Clearinghouse, published in 2014 a practice guide under the title of *Teaching Academic Content and Literacy to English Learners in Elementary and Middle School*. This guide provides a series of four recommendations related to the instruction of BLs with a focus on the linguistic and literacy skills that students need to master to be successful in school, in particular listening, speaking, reading, and writing for academic purposes (Baker et al., 2014). The present study focuses on the fourth of these recommendations: “**Recommendation 4:** Provide small-group

CONTACT Ricardo González-Carriedo  Ricardo.Gonzalez@unt.edu  Department of Teacher Education and Administration, University of North Texas, 1155 Union Circle #310740, Denton, TX 76203

¹This article adopts an asset-laden perspective to refer to students who are learning multiple languages concurrently as bilingual learners. We are aware that at the federal level the terms English learner and English language learner are used (NCES, n.d.). However, the term bilingual learner provides a fairer representation of the socio-cultural context of these students.

instructional intervention to students struggling in areas of literacy and English language development” (p. 59). As the guide asserts, the goal “is to provide teachers with guidelines for (and examples of) systematically – and at times explicitly – building students’ English language and literacy, skills” (p. 3). The emphasis on small-group interventions was already present in previous editions of the guide, which stressed the importance of direct, explicit, and interactive literacy instruction (Gersten et al., 2007). The use of small-group interventions for BLs has been suggested as a key component in the curriculum of teacher education programs (Richards-Tutor et al., 2016). Small-group interventions have shown to provide positive results in Spanish oral language, print knowledge, phonological awareness, and phonics (Landry et al., 2019). They have also been used in general education classrooms with bilingual learners, where teachers have used parallel instruction (Hopkins et al., 2019). Small-group is usually considered a Tier 2 intervention, more intensive than Tier 1 whole-class instruction and more likely to result in better reading and language outcomes (Foorman et al., 2017).

The aim of this study was to ascertain the effectiveness of small-group instruction to enhance the literacy outcomes of BLs. Our research questions were formulated as follows:

- (1) Does teacher professional development in small guided-reading group instruction, supported by coaching, improve student academic performance?
- (2) Is this improvement measurable through the MAP (Measures of Academic Progress), Fountas and Pinnell Benchmark Assessment System, and STAAR (State of Texas Assessments of Academic Readiness) assessments?

Next, we will present a concise review of relevant academic literature and the theoretical framework used for this study

Literature review

Small-group instruction has shown to be an effective intervention even in the short term. In a study with 36 students in pre-kindergarten through Grade 1, Phillips (2014) found substantial potential for the intervention, which resulted in positive results on syntax assessments and listening-comprehension measures after three weeks of small-group targeted syntax interventions. In Phillips (2014), bilingual learners received instruction in groups of four or six for 30 minutes per day, four days per week.

Other studies have looked at the results of long-term Tier 2 interventions. To evaluate the efficacy of small-group instruction among bilingual learners in kindergarten, Nelson et al. (2011) designed a randomized trial ($n = 210$) focused on word vocabulary and decoding interventions. In the study, students were provided with small-group instruction for 20 minutes per day, five days per week, from October to April. The results showed large effects on learning for the treatment group, both regarding root word vocabulary ($d = 1.04$) and word reading ($d = 0.69$) compared to the comparison group. In a different long-term randomized ($n = 94$) intervention study, Vaughn et al. (2006) examined the efficacy of small group (three to five students) instruction for 50 minutes a day, 5 days a week, from October to May. The analysis of the results showed significant differences in favor of the students in the intervention group with an effect size of 0.33 to 0.81 “on the Spanish measures of letter naming fluency, phonological awareness, letter-sound identification, verbal analogies, word reading and word reading fluency, and spelling” (p. 479), with similar results in on the equivalent English measures with effect size of 0.35 to 0.42.

One of the central barriers to effective BL instruction is the confusion and exclusion BLs experience in the classroom as well as teachers frequently underestimating the value of peers in the inclusion and education of BLs (Helfrich & Bosh, 2011). In the absence of intentional attempts to serve the linguistic and cultural needs of the BL, teachers risk alienating these students. This was underscored by Meyer (2000) when he pointed out to the challenges BLs face in classrooms where instruction is solely in

English and where students face cognitive, cultural, and linguistic barriers that can impede their academic success:

The English learner sits in class, not completely understanding or fluently speaking English and possessing limited experience with cultural practices and expectations of school. What is it like for this student to go to school in English only? When the learner's language is not the school's language, how does he or she feel when confronted with academic lessons, school texts, and classroom learning activities that are partially or totally incomprehensible? (p. 228)

These questions highlight the difficulties BLs experience when instruction does not accommodate their linguistic needs. Without support, such as bilingual instruction or scaffolding strategies, these students may struggle to engage with the curriculum and develop academic proficiency.

Theoretical framework

Following the work of Vygotsky (1978), this study assumed that learning is a socially constructed process where knowledge is constructed through interactions with others in meaningful social contexts. In the case of BLs, these students benefit from collaborative learning with peers, teachers, and family members who use multiple languages to facilitate comprehension and meaning-making (García & Wei, 2014). In this context, language becomes a key mediational tool for learning (Vygotsky, 1986). In dual-language programs, students use their first language (L1) and second language (L2) to construct understanding. Using their full linguistic repertoires (García, 2009), and with the support of teachers who act as facilitators (González-Carriedo et al., 2016), students aim for the zone of proximal development, where learning occurs with the support or a more knowledgeable other (Vygotsky, 1978). This other can be a teacher, a peer or any adult providing assistance in or outside of the classroom. Teachers use a variety of instructional strategies to provide scaffolding, such as allowing and encouraging a translanguaging approach where students are free to combine languages (García, 2009) or using visual aids and other techniques to provide contextualized instruction (Wood et al., 1976). In the use of scaffolding techniques, teachers can decide to incorporate cultural elements that tap into the students' linguistic backgrounds, traditions, and lived experiences, making direct connections with the curriculum to foster deeper engagement with the concepts taught in class (Paris & Alim, 2017). It is important, then, to provide meaningful and contextualized learning where BLs have an opportunity to develop literacy and content knowledge more effectively through the use of authentic, real-world tasks, such as storytelling, project-based learning, and community-based activities (Cummins, 2000). The fundamental aspect of instruction, consequently, is to promote language learning through dialogue, discussion, and negotiation of meaning with others (Bakhtin, 1981). For this, peer collaboration, cooperative learning, and bilingual discussion groups where students practice and develop both L1 and L2 in authentic communicative settings become central components of the instruction (Mercer, 2000).

In particular, our study focused on the use of small groups to support co-construction of knowledge, a strategy that is not free of potential problems as it can be difficult to create successful small peer groups due to the potential stigmatization, the exacerbation of differences of those included in the small group, and the generation of dysfunctional interactions among those in the group (Blumenfeld et al., 1996).

Consequently, the present study takes a socio-constructivist approach to learning. We start with the assumption that small-group instruction allows for knowledge to be shared and meaning to be negotiated and co-constructed in socially mediated, collaborative learning environments (Vygotsky, 1978, 1986). Under this perspective, the social function that occurs in schools is a key component in the learning process because it is through the interactions of the learners with their social environment that they construct knowledge. Socially mediated group discourses and activities, therefore, provide the ideal milieu through which students internalize and co-construct knowledge (Hull & Saxon, 2009). True learning occurs when

students are actively involved in the classroom (Aronowitz, 1996) and collectively share their interpretations of reality. This model of learning rejects the transmission approach to teaching and learning and departs from what Freire (1970/2000) termed as the banking method, the idea that teachers *deposit* knowledge into the students. Instead, socio-constructivist approaches involve students who react and respond to others with more expertise by working together and constructing their own understanding (García et al., 2011).

One way of attaining high levels of student engagement is using small groups, which allow teachers to implement several strategies aimed at enhancing specific skills (Hull et al., 2016). However, the implementation of any strategy specifically aimed at improving the academic performance of BLs needs to be preceded by a series of professional development activities for teachers. Here, we use the term professional development to refer to “processes and practices that improve the job-related knowledge, skills, and attitudes of school employees” (Wilde, 2010, p. 5). Professional development has a direct effect on teachers’ sense of self-efficacy to teach BLs. In a study conducted by Gándara et al. (2005), it was found that “greater preparation for teaching English learners equaled greater teacher confidence in their skills for working with these students successfully” (p. 12). Yet, for professional development to be effective, it needs to include opportunities stemming from personalized coaching focusing on hands-on practice and teaching techniques directly applicable in the classroom (Calderón et al., 2011).

Methodology

Participants

Third-grade teachers ($j = 18$; 10 treatment) and their students ($n = 587$; 22 treatment) were recruited from 24 schools across a large school district in the Southern United States. Student participants ($n = 587$, 50.1% female) identified predominantly as Hispanic (62.7%) with Asian (10.9%) being the second largest group. In this group, 10.9% were classified as needing special education; 62.6% were economically disadvantaged (classified through free and reduced lunch status); 44.5% were bilingual learners; 37.4% participated in a bilingual program; 9.9% participated in an ESL program; and 53.8% spoke a language other than English at home.

Sampling procedures

Third-grade teachers from 24 primary schools in the district were provided printed materials describing the study along with personal invitations following discussions with principals at each of the schools. In order to preserve the integrity of the treatment by preventing teacher conversations from contaminating comparison group teacher instruction, half ($j = 12$) of the primary schools were identified as treatment in which teachers would be recruited to participate in professional development and were expected to implement the instructional interventions, and the other half of the primary schools ($j = 12$) were identified as schools from which comparison teachers would be recruited. Consequently, recruitment materials supplied to teachers were different depending upon whether the school was a treatment or comparison school. Teacher participation occurred through self-selection following recruitment from a member of the research team housed within the school district. All teachers provided informed consent in accordance with an approved protocol from the lead author’s institutional review board.

In general, third-grade teachers in the district have class sizes of 20–28 students, and approximately five students in any teacher’s classroom are “tier 2” students. Under the Response to Intervention (RTI) framework, these are students that are not making adequate progress in the regular classroom and require intensive instruction, specifically in English language acquisition, reading, and mathematics. These

students were identified as study participants for the dependent variable outcomes. Under the approved IRB protocol, the district supplied de-identified data on student participants to carry out an analysis.

Measures and covariates

Measures of Academic Progress

Measures of Academic Progress (MAP) is a norm-referenced tool that measures student growth over time. MAP (created by Northwest Evaluation Association) is a computer-adaptive test, meaning a computer algorithm selects the items supplied, from a large pool of items with varying difficulties and discrimination parameters, to create a unique test for each student. The difficulty of each question is determined by how well the student answered the previous question. It measures students' reading and language skills. Students are assigned to MAP tests based on grade level. Along with other data, MAP assessments provide valuable information on reading levels (low to high within grade level) for each child. MAP test scores allow teachers to customize instruction to fit each student's literacy need. The marginal reliability estimates (Samejima, 2016) for total MAP scores range from 0.92 to 0.96.

Fountas and Pinnell Benchmark Assessment System

The Fountas and Pinnell Benchmark Assessment System (Fountas & Pinnell, 2010) provides teachers with tools to observe and quantify specific reading behaviors, such as decoding, fluency, vocabulary, and comprehension skills. Each assessment determines students' independent reading levels. Students receive a letter grade based on the composite of these skills. The test-retest reliability estimate for the present sample is 0.79.

The State of Texas Assessment of Academic Readiness (STAAR)

STAAR is a standardized test that is used in Texas public primary and secondary schools to assess the students' achievement and knowledge learned in specific grade levels. STAAR is based on the state curriculum standards, the Texas Essential Knowledge and Skills (TEKS), in core subjects including reading and writing. Reading readiness standards are emphasized and tested on STAAR. It specifically measures comprehensions skills, reading strategies, fluency, and vocabulary development (National Center for Education Statistics [NCES], n.d.). The scale scores were utilized as a distal outcome for this study purposes.

Covariates

Pretest scores for MAP and Fountas and Pinnell were supplied along with end-of-year scores for all student participants on STAAR. Pretest scores associated with the dependent variables were used to improve the precision of estimates in the HLM analyses (Bloom et al., 2007).

In addition, the district supplied demographic covariates on all student participants, including sex, ethnic origin, bilingual status, and ESL status. These covariates were used for construction of propensity scores to match treatment and comparison group participants.

Research design

The study design is a nonequivalent comparison group design containing an untreated control group with pretest and posttest samples of the dependent variables in both groups (Shadish et al., 2002). In this design, students were nested within teachers, resulting in a multi-level structure in which teacher exposure to the independent variable of professional development and subsequent coaching on small group instructional strategies were expected to result in greater student language acquisition proximally, and distally in enhanced academic performance. Pretests were administered at the beginning of the academic year as a covariate to

improve the precision of parameter estimates (Bloom et al., 2007) and posttests administered at the conclusion of the academic year as the dependent variable outcomes of interest.

Intervention

Teachers in the treatment group enrolled in a five-week graduate-level face-to-face course (40 contact hours of instruction) provided by the lead author's institution and taught by a professor of child literacy. The course centered on the study of the development of literacy in young BLs through oral language, listening, comprehension, alphabetic knowledge, and print awareness. In addition, teachers were exposed to small group instructional practices that supported English language learning where teachers facilitated and encouraged informal student interaction, as these informal interactions are theoretically essential to facilitating progress with BLs in both language and literacy development. Small group interactions were created within several multimodal activities, including visual art, music, and drama (digital collaging, visual response talks), listening comprehension strategies, active comprehension activities (read-alouds, shared reading, reading workshops), literacy writing activities appropriate for second language acquisition, vocabulary development, and diverse learners in book clubs. Treatment group teachers were also supplied with a variety of academic resources to support the course (Henn-Reinke & Yang, 2017; Herrell & Jordan, 2016; Pahl & Rowsell, 2012; Souto-Manning & Martell, 2016). At the conclusion, the course required teachers to create five lesson plans meeting reading standards as specified by the Texas Essential Knowledge and Skills (TEKS) and the English Language Proficiency Standards (ELPS) utilizing instructional strategies and small group activities.

At the start of the academic year teachers met to review student assessment data and selected the BLs for their small group. They developed a coaching schedule for the 18 weeks implementing interventions. During this time, a facilitator observed teachers once per week during a 30-minute lesson sequence following the Leverage Leadership 2.0 coaching model and observation form (Bambrick-Santoyo, 2018). Following observation, debriefing meetings were held between the coach and teacher to provide feedback. The coaching model emphasized teachers seeing the intervention being modeled, naming the components of the intervention, and practicing implementation of the intervention prior to using it in small groups with students. During the middle of the school year, teachers met with their peers to discuss their experiences implementing small group instruction, and the moderator focused on how well teachers were able to facilitate student-to-student interaction, and techniques teachers employed to engage all small group participants in these non-formal interactions.

Statistical data analysis

Propensity score matching

Propensity scores (Rosenbaum & Rubin, 1983) were used to match comparison and treatment participants on covariates collected at baseline. The covariates utilized to construct propensity scores for students (Sex, Hispanic, Bilingual, ESL, and Teacher) were used to construct propensity scores with a logistic model for matching (Powell et al., 2020) with optimal matching (Austin, 2014). Treatment group participants ($n = 22$) were matched with a larger

Table 1. Group comparisons on covariates before propensity score matching.

Variables	Treatment Group		Comparison Group		<i>t</i>	bias	<i>d</i>
	<i>(n = 22)</i>		<i>(n = 99)</i>				
	Mean	<i>SD</i>	Mean	<i>SD</i>			
Sex	.27	.46	.56	.50	−2.58	−0.29	−.60
Hispanic	.68	.48	.90	.30	−2.04	−0.22	−.55
Bilingual Program	.41	.50	.81	.39	−3.54	−0.40	−.89
ESL Program	.50	.51	.11	.32	3.39	0.31	.91

Table 2. Group comparisons on covariates after propensity score matching.

Variables	Treatment Group		Comparison Group		<i>t</i>	bias	<i>d</i>
	(n = 19)		(n = 19)				
	Mean	<i>SD</i>	Mean	<i>SD</i>			
Sex	.32	.46	.26	.45	−.35	0.06	.13
Hispanic	.63	.49	.74	.45	.68	−0.11	−.23
Bilingual Program	.47	.51	.53	.51	.31	−0.06	−.11
ESL Program	.42	.51	.32	.48	.66	0.10	.20

Table 3. Percent reduction in bias after propensity score matching.

Variables	b_i	b_m	$ b_i - b_m $	Δ
Sex	-0.29	0.06	0.23	120
Hispanic	-0.22	-0.11	0.11	50
Bilingual Program	-0.40	-0.06	0.36	85
ESL Program	0.31	0.10	0.21	68

set of comparison students ($n = 99$). A caliper width of 0.10, which is 0.20 standard deviation of the propensity score logit was used as recommended by Austin (2011). This procedure was conducted using nearest neighbor matching in SPSS v26. This resulted in $n = 38$ participants remaining following matching, with $n = 19$ in each group. Table 1 demonstrates the bias present in the full sample prior to propensity score matching with Table 2 demonstrating the bias after propensity score matching. Table 3 indicates the extent to which bias reduction was produced following propensity score matching, demonstrating balance on all observed covariates.

Data analysis

Two-level hierarchical linear modeling (HLM) was conducted to assess changes in outcomes of students ($n = 38$) nested within teachers ($j = 15$). In each model, students are represented at Level 1 with Group (treatment versus control) estimated as a predictor of intercepts at Level 2 (teacher level). Additionally, pre-intervention covariates were included in the model to account for preexisting differences on the outcomes of interest (e.g., beginning of the year MAP scores were included as a covariate for end of the year MAP scores). However, STAAR scores are an exception as this exam is only taken once a year. All analyses were conducted in HLM 8 (Raudenbush et al., 2004) utilizing full information maximum likelihood estimation (FIML) with the model presented below (with OUTCOME referring to the outcome variable, i referring to individuals, j referring to the teachers that provided the intervention, β_{0j} and β_{1j} as individual level predictors, γ_{00} and γ_{01} as teacher level predictors, and all error terms being estimated):

$$\text{Level-1: } OUTCOME_{ij} = \beta_{0j} + \beta_{1j} * (COVARIATE_{ij}) + r_{ij}$$

$$\text{Level-2: } \beta_{0j} = \gamma_{00} + \gamma_{01} * (GROUP_j) + u_{0j}$$

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

$$\text{Mixed Model: } OUTCOME_{ij} = \gamma_{00} + \gamma_{01} * (GROUP_j) + u_{0j}$$

$$+ \gamma_{10} * (COVARIATE_{ij}) + u_{1j} + r_{ij}$$

Intraclass correlation coefficients (ICCs) were used to determine variance due to clustering and to determine if multilevel models were appropriate for analyses. Additionally, Hedge's g was calculated to determine mean differences between the groups.

Results

Descriptive statistics for all outcome measures are presented in Table 4. Results for all models including ICCs are presented in Table 5. ICCs ranged from .001 to .087 with only MAP vocabulary and Fountas and Pinnell demonstrating the necessary variance for conducting multilevel modeling. However, due to the theoretical appropriateness of conducting HLM analyses with clustered data, multilevel analyses were conducted for all outcomes. Additionally, effect size differences of all outcomes can be seen in Figure 1.

Table 4. Descriptive Statistics after propensity score matching.

Variables	Treatment Group		Comparison Group	
	(n = 19)		(n = 19)	
	Mean	SD	Mean	SD
MAP T1	171.39	14.00	173.88	14.17
Foundational Skills T1	170.72	14.84	172.76	13.50
Comprehension T1	170.61	15.57	175.41	17.88
Vocabulary T1	172.97	14.10	174.12	14.23
MAP T2	184.84	12.54	182.94	18.79
Foundational Skills T2	184.11	13.36	183.94	20.32
Comprehension T2	185.84	12.22	182.67	19.55
Vocabulary T2	184.89	14.66	182.39	18.22
Fountas and Pinnell T1	11.06	2.84	11.06	2.46
Fountas and Pinnell T2	14.74	2.47	12.79	3.61
STAAR Reading	16.89	6.61	16.26	5.85
	Percent	Count	Percent	Count
Student Female	31.6%	6	26.3%	5
Hispanic	63.2%	12	73.7%	14
Bilingual Program	47.4%	9	52.6%	10
ESL Program	57.9%	8	31.6%	6

Table 5. Treatment and control means, standard deviations, baseline equivalence, intraclass correlations, and treatment model effect sizes.

	TMT (n = 19)		CTL (n = 19)		Baseline Equivalence (g)	ICC	Effect size (g)	p
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)				
MAP	184.84 (4.55)	182.95 (3.65)			-.18	.001	.46	.686
Foundational Skills	184.10 (5.47)	183.95 (3.92)			-.14	.001	.03	.978
Comprehension	185.84 (4.27)	182.67 (3.45)			-.29	.001	.82	.471
Vocabulary	184.89 (5.38)	182.65 (3.89)			-.08	.083	.48	.684
Fountas and Pinnell	14.72 (0.89)	12.78 (0.63)			0	.087	2.52	.049
STAAR Reading	17.47 (2.18)	17.84 (1.54)			N/A	.001	-.20	.870

Note. Reported means are covariate adjusted. Standard deviations are calculated based off cluster standard deviations and the corresponding ICC. Hedge's *g* effect sizes were not adjusted for ICCs as the ICCs are non-significant; ICC = Intraclass Correlations.

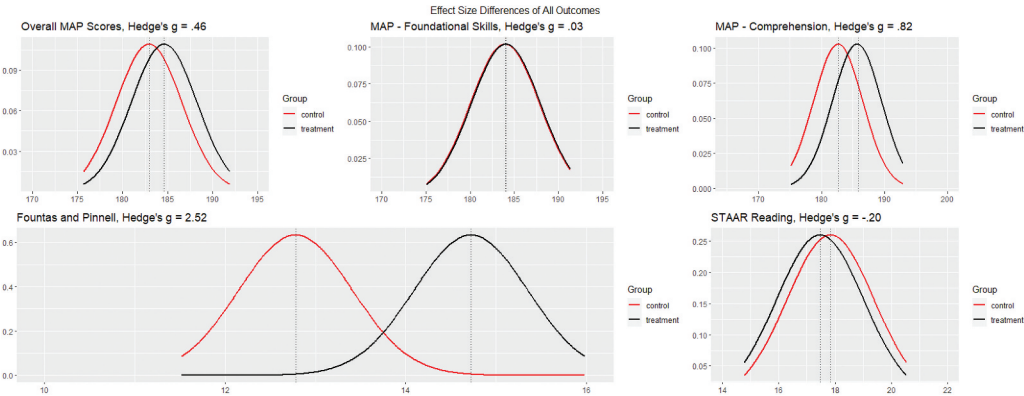


Figure 1. Effect size differences for all outcomes in Hedge's *g*.

Measures of Academic Progress

As shown in Table 5, four separate models were conducted examining MAP (overall, foundational skills, comprehension, and vocabulary). No model was statistically significant; however, comprehension ($\gamma_{01} = 3.17$, $t(13) = .742$, $p = .471$, $g = .82$), vocabulary ($\gamma_{01} = 2.24$, $t(13) = .416$, $p = .684$, $g = .48$), and overall ($\gamma_{01} = 1.88$, $t(13) = .371$, $p = .717$, $g = .46$) displayed medium to large effect size differences. Foundational skills ($\gamma_{01} = 0.15$, $t(13) = .028$, $p = .978$, $g = .03$) demonstrated almost no difference between the intervention and control group. Additionally, introducing the intervention into the model decreased the unexplained level-2 variance (teacher level variance) for all aspects of MAP: overall (30.33% reduced), foundational skills (1.11% reduced), comprehension (27.10% reduced), and vocabulary (73.39% reduced).

Fountas and Pinnell

Introducing the intervention effect into the model reduced the unexplained level-2 variance by 99.06%. Moreover, the intervention produced a statistically significant difference and a large effect size, $\gamma_{01} = 1.95$, $t(13) = 2.167$, $p = .049$, $g = 2.52$. Specifically, on average, participants in the intervention group saw an increase of two Fountas and Pinnell levels greater than the control group (Table 5).

State of Texas Assessments of Academic Readiness

The intervention created a 7.99% reduction in unexplained variance; however, it did not produce a significant change in STAAR reading scores, $\gamma_{01} = -0.37$, $t(13) = -.202$, $p = .843$, $g = -.20$. Moreover, participants in the intervention did worse on the STAAR reading exam than participants in the control group (Table 5).

Discussion

The present study represents a field trial of small, guided reading group instruction that resulted as a consequence of intensive professional development followed by in-classroom coaching for 18 weeks. The causal pathway for such an intervention is complex since there are fidelity concerns related to the extent to which professional development leads to extant changes in teacher instructional practices, and whether those instructional practices then result in positive academic outcomes for students. Given this two-step approach to improve student performance, there are multiple instances in which failure to implement the intervention as desired, or where a lack of implementation efficacy would result in attenuation of effects on the outcomes of interest. Despite these potential pitfalls that might attenuate the outcomes, we generally found positive effects, one of which was statistically significant, even though the statistical power to detect significant effects was quite small ($1 - \beta = 0.37$). Consequently, we preferred to interpret effect sizes of the intervention's causal impact on outcomes.

Using propensity score matching to construct statistically matched treatment and comparison groups for statistical analyses is rapidly gaining attention as a means for inferring causal effects of interventions in educational field trial research when random assignment is not feasible or possible (Powell et al., 2020). While causal implications from the present study are not as strong as they might be for a randomized trial where all confounding covariates are controlled for (at least in theory), we controlled for several observable covariates as evidenced by reduction in bias in constructing groups. Accordingly, we were able to infer weak causal support for the intervention in producing the effects, most notably the extraordinarily large outcome effect of more than 2.5 standard deviations on the Fountas and Pinnell Benchmark Assessment, suggesting that students exposed to small, guided reading group instruction from trained teachers engaged in comprehension conversations that went beyond retelling when compared to the matched counterfactual group. At the same time, we also observed medium to large effects on the comprehension and vocabulary subscales on the Measures of

Academic Progress. The modest effect on STAAR is to be expected (Keiffer et al., 2009), as BLs often do not benefit from accommodations on large-scale assessments, possibly bringing into question the validity of their scores.

The focal point of this study was an intervention composed of two elements. The teacher professional development course focused on language and literacy development for BLs in the context of small group instruction. This included strategies for the development of vocabulary, peer interactions, collaborative activities, and multi-modal instruction with an emphasis on second language acquisition. The second intervention was a coaching model based on weekly 30-minute meetings followed by classroom observations and debriefing meetings. These sessions lasted 18 weeks. The improvements in the reading skills of decoding, fluency, vocabulary and comprehension shown on the Fountas and Pinnell Benchmark Assessment and the progress on comprehension and vocabulary indicated by the scores on the Measures of Academic Progress support the continued use of the structured professional development sessions and instructional coaching. These results indicate that the professional development received by the teachers had a causal relationship with the improvement of specific reading skills that their third-grade students experienced. This is in alignment with findings from studies showing that professional development positively impacts the sense of self-efficacy to teach bilingual learners (Gándara et al., 2005), especially when it is accompanied by personalized coaching (Calderón et al., 2011). In turn, teacher-self-efficacy has an impact on how students learn from their teachers. Teachers with higher levels of self-efficacy tend to have higher expectations from their students (Bandura, 1981). In other words, when teachers display higher levels of self-efficacy the overall environment of the class is more confident, which positively impacts student outcomes (Guo et al., 2010; Skaalvik & Skaalvik, 2010).

The focus of the professional development was small group instruction. In this regard, the results described in the previous section are aligned with previous studies that support the use of small group instruction for BLs on account of the social and academic benefits that they yield for students (Kamps et al., 2007; Richards-Tutor et al., 2016). This is even more pronounced in the case of BLs, for whom the development of second language skills rests on the opportunity to interact with the language. Second language acquisition, in educational settings, depends largely on the opportunities afforded to students to engage in active, social, and collaborative processes (Lee, 2007). Small group instruction provides BLs with close interactions with peers who serve as linguistic role models and with whom they engage in what constructivist theorists have termed as dialogic learning (Zhang, 2008). This context provides students with an environment that is much less threatening than interactions with the whole class, lowering their affective filter (Krashen, 1982) and allowing students to experiment with language. Small groups provide teachers with an ideal setting with which to implement scaffolding techniques that are vital for language maintenance and development (Baker & Wright, 2017). Small groups allow not only the peer interactions but also opportunities for guided or targeted support and scaffolding techniques where the teacher focuses on specific skills that may be beyond the individual reach of the student. For example, the teacher may want to focus on disciplinary literacy skills, the linguistic and symbolic tools shared in a specific academic discipline. This will allow students to access the content knowledge of the discipline, negotiating meaning and expressing their understandings (Ardasheva et al., 2019).

Limitations

As mentioned at the outset of the Discussion, the present study is limited by the lack of fidelity of implementation measures at both the teacher and student levels. A great deal of variance in the outcomes may be explained by the extent to which teachers utilize small group instructional strategies following exposure to professional development, and how coaching serves to bolster the effects of the professional development over time. Future work should consider these covariate effects in estimating effects. In addition, the present study is limited in causal attribution of the intervention to the effects noted due to (a) lack of randomization that would permit more careful control over confounding

factors that might explain the observations, and (b) the lack of statistical power due to the small sample size. We recommend future studies replicate the present study with larger samples when possible. In addition, there are limitations due to generalizability of the present study since it occurs only with third-grade teachers/students in a single urban district.

Conclusion

While small group instructional strategies represent a complex intervention to implement, we recommend teachers and administrators consider the approach as a central approach to RTI, specifically for tier 2 students, as the procedure represents a framework for proactively altering student outcomes in early grades within a population of learners that supports intervention proactively, before potential failures occur. As we have suggested, small group instruction allows for teachers to modify their approach, utilizing the strategies that fit their own unique abilities and the unique needs of their students. Because the use of the students' first language as medium of instruction is not universal, small group instruction provides a mechanism that permits students to access language and content support in a way that is less threatening, and provides opportunities for them to express their limitations and obtain the assistance needed. This allows teachers to customize tasks that fit their students' needs and abilities, accommodating the full range of learning styles and background experiences of the students, all while engaging in reflection and dialogic activities with their peers (Schcolnik et al., 2006). These tasks may include the implementation of guided reading, a strategy that provides support to students by modeling reading strategies and enables students to comprehend increasingly more difficult texts.

Moreover, small groups provide students with opportunities to interact with each other using their full range of linguistic repertoires. A common strategy in dual language classrooms is the use of bilingual pairs, where the teacher teams up two students with differing linguistic backgrounds (Gómez et al., 2005). These bilingual pairs collaborate and create meaning together as they work through different activities, assisting each other. Under this small group strategy, students assist each other while they work on class activities, allowing the teacher to take a secondary position while students build their own understandings. When this happens, the learning process shifts from teachers to students and students become teachers, following a Freirean model as students make meaning of the world around them (Aronowitz, 1996). Given the diversity of schools today, the scaffolding that small group instruction allows is of paramount importance.

Disclosure statement

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

Funding

This publication was supported by the Office of English Language Acquisition (OELA) of the U.S. Department of Education PR/Award #T365Z170181. The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement by the U.S. Department of Education.

ORCID

Ricardo González-Carriedo  <http://orcid.org/0000-0003-3413-2207>

References

- Ardasheva, Y., Newcomer, S. N., Ernst-Slavit, G., Morrison, S. J., Morrison, J. A., Carbonneau, K. J., & Lightner, L. K. (2019). Decision-making practices of experienced upper elementary teachers during small group instruction to support English learners. *Elementary School Journal*, 120(1), 88–108. <https://doi.org/10.1086/704392>
- Aronowitz, S. (1996). Paulo Freire's radical democratic humanism. In P. McLaren & P. Leonard (Eds.), *Paulo Freire a critical encounter* (pp. 8–24). Routledge.

- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.568786>
- Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, 33(6), 1057–1069. <https://doi.org/10.1002/sim.6004>
- Baker, C., & Wright, W. (2017). *Foundations of bilingual education and bilingualism* (6th ed.). Multilingual Matters.
- Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., Gersten, R., Haymond, K., Kieffer, M. J., Linan-Thompson, S., & Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE 2014-4012). National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/publications_reviews.aspx
- Bakhtin, M. M. (1981). *The dialogic imagination: Four essays*. University of Texas Press.
- Bambrick-Santoyo, P. (2018). *Leverage leadership 2.0: A practical guide to building exceptional schools*. Jossey-Bass.
- Bandura, A. (1981). Self-referent thought: A developmental analysis of self-efficacy. In J. Flavell & L. Ross (Eds.), *Social cognitive development: Frontiers and possible future* (pp. 200–239). Cambridge University Press.
- Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, 29(1), 30–59. <https://doi.org/10.3102/0162373707299550>
- Blumenfeld, P. C., Marx, R. W., Soloway, E., & Krajcik, J. (1996). Learning with peers: From small group cooperation to collaborative communities. *Educational Researcher*, 25(8), 37–40. <https://doi.org/10.3102/0013189X025008037>
- Calderón, M., Slavin, R., & Sánchez, M. (2011). Effective instruction for English learners. *The Future of Children*, 21(1), 103–127. <https://doi.org/10.1353/foc.2011.0007>
- Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire*. Multilingual Matters.
- Foorman, B., Herrera, S., & Dombek, J. (2017). *Stated briefly: The relative effectiveness of two approaches to early literacy intervention in grades K-2* (REL 2017-258). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. <http://ies.ed.gov/ncee/edlabs>
- Fountas, J. C., & Pinnell, G. S. (2010). *Benchmark assessment system 1: Assessment guide* (2nd ed.). Heinemann
- Freire, P. (2000). *Pedagogy of the oppressed*. Continuum. (Original work published 1970).
- Gándara, P., Maxwell-Jolly, J., & Driscoll, A. (2005). Listening to teachers of English language learners: A survey of California teachers' challenges, experiences, and professional development needs. Policy Analysis for California Education, PACE (NJ1).
- García, G. E., Pearson, P. D., Taylor, B. M., Bauer, E. B., & Stahl, K. A. (2011). Socio-constructivist and political views on teachers' implementation of two types of reading comprehension approaches in low-income schools. *Theory into Practice*, 50(2), 149–156. <https://doi.org/10.1080/00405841.2011.558444>
- García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Wiley-Blackwell.
- García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.
- 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: A practice guide* (NCEE 2007-4011). Institute of Education Sciences, U.S. Department of Education. <http://ies.ed.gov/ncee/wwc/pdf/practiceguides/20074011.pdf>
- Gómez, L., Freeman, D., & Freeman, Y. (2005). Dual language education: A promising 50-50 model. *Bilingual Research Journal*, 29(1), 145–164. <https://doi.org/10.1080/15235882.2005.10162828>
- González-Carriedo, R., Bustos, N., & Ordóñez, J. (2016). Constructivist approaches in a dual-language classroom. *Multicultural Perspectives*, 18(2), 108–111. <https://doi.org/10.1080/15210960.2016.1159102>
- Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains. *Teaching & Teacher Education*, 26(4), 1094–1103. <https://doi.org/10.1016/j.tate.2009.11.005>
- Helfrich, S. R., & Bosh, A. J. (2011). Teaching English language learners: Strategies for overcoming barriers. *The Educational Forum*, 75(3), 260–270. <https://doi.org/10.1080/00131725.2011.578459>
- Henn-Reinke, K., & Yang, X. (2017). *The literacy club: Effective instruction and intervention for linguistically diverse learners*. Caslon.
- Herrell, A. L., & Jordan, M. (2016). *50 strategies for teaching English language learners*. Pearson.
- Hopkins, M., Gluckman, M., & Vahdani, T. (2019). Emergent change: A network analysis of elementary teachers' learning about English learner instruction. *American Educational Research Journal*, 56(6), 2295–2332. <https://doi.org/10.3102/0002831219840352>
- Huebner, T. A. (2009). What research says about small-group intervention for ELLs. *Educational Leadership*, 66(7), 90–91.
- Hull, D. M., Booker, D. D., & Näslund-Hadley, E. I. (2016). Teachers' self-efficacy in Belize and experimentation with teacher-led math inquiry. *Teaching & Teacher Education*, 56, 14–24. <https://doi.org/10.1016/j.tate.2016.01.026>
- Hull, D. M., & Saxon, T. F. (2009). Negotiation of meaning and co-construction of knowledge: An experimental analysis of asynchronous online instruction. *Computers and Education*, 52(3), 624–639. <https://doi.org/10.1016/j.compedu.2008.11.005>

- Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., Smith, M., Bullock Mann, F., Barmer, A., & Dilig, R. (2020). *The condition of education 2020 (NCES 2020-144)*. U.S. Department of Education. National Center for Education Statistics. <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020144>.
- Kamps, D., Abbott, M., Greenwood, C., Arreaga-Mayer, C., Wills, H., Longstaff, J., Culpper, M., & Walton, C. (2007). Use of evidence-based, small-group reading instruction for English language learners in elementary grades: Secondary-tier intervention. *Learning Disability Quarterly*, 30(3), 153–168. <https://doi.org/10.2307/30035561>
- Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners on large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research*, 79(3), 1168–1201. <https://doi.org/10.3102/0034654309332490>
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Landry, S. H., Assel, M. A., Carlo, M. S., Williams, J. M., Wu, W., & Montroy, J. J. (2019). The effect of the preparing pequeños small-group cognitive instruction program on academic and concurrent social and behavioral outcomes in young Spanish-speaking dual-language learners. *Journal of School Psychology*, 73, 1–20. <https://doi.org/10.1016/j.jsp.2019.01.001>
- Lee, L. (2007). Fostering second language oral communication through constructivist interaction in desktop videoconferencing. *Foreign Language Annals*, 40(4), 635–649. <https://doi.org/10.1111/j.1944-9720.2007.tb02885.x>
- Mercer, N. (2000). *Words and minds: How we use language to think together*. Routledge.
- Meyer, L. M. (2000). Barriers to meaningful instruction for English learners. *Theory into Practice*, 39(4), 228–236. https://doi.org/10.1207/s15430421tip3904_6
- National Center for Education Statistics. (2023). *English learners in public schools*. U.S. Department of Education. <https://nces.ed.gov/programs/coe/indicator/cgf/english-learners>
- National Center for Education Statistics. (n.d.). *Glossary*. The condition of education. <https://nces.ed.gov/programs/coe/glossary.asp#e>
- Nelson, J. R., Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of a tier 2 supplemental root word vocabulary and decoding intervention with kindergarten Spanish-speaking English learners. *Journal of Literacy Research*, 43(2), 184–211. <https://doi.org/10.1177/1086296X11403088>
- Pahl, K., & Rowsell, J. (2012). *Literacy and education*. SAGE Publications Ltd.
- Paris, D., & Alim, H. S. (2017). *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world*. Teachers College Press.
- Phillips, B. M. (2014). Promotion of syntactical development and oral comprehension: Development and initial evaluation of a small-group intervention. *Child Language Teaching & Therapy*, 30(1), 63–77. <https://doi.org/10.1177/0265659013487742>
- Powell, M. G., Hull, D. M., & Beaujean, A. A. (2020). Propensity score matching for education data: Worked examples. *The Journal of Experimental Education*, 88, 145–164. <https://doi.org/10.1080/00220973.2018.1541850>
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2004). *Hlm 6: Hierarchical linear and nonlinear modeling*. Scientific Software International.
- Richards-Tutor, C., Aceves, T., & Reese, L. (2016). Evidence-based practices for English learners (document no. IC-18). <http://cedar.education.ufl.edu/tools/innovation-configurations/>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Samejima, F. (2016). Graded response models. In W. J. van der Linden (Ed.), *Handbook of item response theory, volume one* (pp. 123–136). Chapman and Hall/CRC.
- Scholnik, M., Kol, S., & Abarbanel, J. (2006). Constructivism in theory and in practice. In *English Teaching Forum* (Vol. 44, No. 4, pp. 12–20). US Department of State. Bureau of Educational and Cultural Affairs, Office of English Language Programs.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage Learning.
- Skaalvik, E. M., & Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations. *Teaching & Teacher Education*, 26(4), 1059–1069. <https://doi.org/10.1016/j.tate.2009.11.001>
- Souto-Manning, M., & Martell, J. (2016). *Reading, writing, and talk: Inclusive teaching strategies for diverse learners, K-2*. Teachers College Press.
- Texas Education Agency. (2024). *Emergent bilingual students in Texas public schools*. <https://kinder.rice.edu/urbanedge/bilingualism-strength-texas-failing-nurture-students>
- U.S. Department of Education, Office for Civil Rights. (2025). Education and title VI of the civil rights act of 1964. <https://www.ed.gov/laws-and-policy/civil-rights-laws/race-color-and-national-origin-discrimination/education-and-title-vi>
- Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E. C., Francis, D. J., 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>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

- Vygotsky, L. S. (1986). *Thought and language*. MIT Press.
- Wilde, J. (2010). Guidelines for professional development: An overview. In C. J. Casteel & K. G. Ballantyne (Eds.), *Professional development in action: Improving teaching for English learners* (pp. 1–10). National Clearinghouse for English Language Acquisition. http://www.ncela.gwu.edu/files/uploads/3/PD_in_Action.pdf
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Zhang, L. J. (2008). Constructivist pedagogy in strategic reading instruction: Exploring pathways to learner development in the English as a second language (ESL) classroom. *Instructional Science*, 36(2), 89. <https://doi.org/10.1007/s11251-007-9025-6>