American Educational Research Journal
June 2018, Vol. 55, No. 3, pp. 488–531
DOI: 10.3102/0002831217741089

Early Evaluation Findings From the Instructional Conversation Study: Culturally Responsive Teaching Outcomes for Diverse Learners in Elementary School

Pedro R. Portes
University of Georgia
Manuel González Canché
University of Pennsylvania
Diego Boada
Melissa E. Whatley
University of Georgia

This study explores preliminary results from a pedagogical intervention designed to improve instruction for all students, particularly emergent bilinguals in the United States (or English language learners). The study is part of a larger efficacy randomized controlled trial (RCT) of the Instructional Conversation (IC) pedagogy for improving the school achievement of upper elementary grade students. Standardized achievement student data were

---

Pedro R. Portes, PhD, is the Goizueta Foundation distinguished chair for Latino teacher education, professor in the department of counseling and human development services, and executive director of the Center for Latino Achievement and Success in Education (CLASE), 125 Aderhold Hall, College of Education, Athens, GA 30602; e-mail: *portes@uga.edu*. His interests center on linking primary prevention practices to human development within a cultural-context perspective, the development of culturally valid methods, improving teacher and counselor education, and equity in educational policy.

Manuel González Canché, PhD, is an associate professor of higher education at the University of Pennsylvania. His research follows two interconnected paths. The first concerns issues of access, persistence, and success, with an emphasis on institution effects—such as two-year versus four-year college and distance from home—on students' outcomes. The second focuses on higher education finance, with emphases on spatial modeling and competition based on spatial proximity and spillover effects.

Diego Boada, MEd, is a doctoral candidate in learning, design, and technology at the University of Georgia. His areas of interest include teacher education, instructional design, and English language learners.

Melissa E. Whatley, MA, is a doctoral candidate in the Institute of Higher Education at the University of Georgia. Her research focuses on issues surrounding the democratization of study abroad and the stratification of education abroad opportunities.

*gathered from (*N = *74) randomized teachers' classrooms. Preliminary ordinary least squares analyses of the intervention appear promising for English language arts in general. Limitations in baseline equivalency for students after teacher randomization are discussed along with strategies to overcome them and implications concerned with the education of all students, notably those whose parents speak languages other than English at home.*

KEYWORDS: English language learners, instructional conversation, professional development, randomized trial, school achievement

Introduction

Teaching through conversation in small groups characterizes a dialogic approach that is increasingly being employed as a primary means to facilitate academic learning in schools. Current research on dialogic and enriched language approaches is being advanced that contrasts with traditional and direct instruction methods (Clarke, Resnick, & Rose, 2016; Lawrence, Crosson, Paré-Blagoev, & Snow, 2015; Lawrence, Francis, Paré-Blagoev, & Snow, 2016; Lesaux, Kieffer, Kelley & Harris, 2014). Also, there is some evidence that the language of instruction for diverse students may not be as critical as teacher practices grounded in pedagogical research (Cheung & Slavin, 2012). Evidence-based methods and strategies for reading may accelerate school learning in helping students achieve at grade level in specific or most content areas (Foorman, Francis, Fletcher, Schatshneider, & Mehta, 1998; McNamara, Burns, Griffin, Snow, & Strickland, 2002). In general, a focus on reading and English language arts (ELA) has been part of research of interventions seeking to improve student comprehension and learning across content areas and test outcomes (Alfassi, 2004; Unrau & Alvermann, 2013). Specifically, fluency, phoneme awareness, vocabulary, and comprehension must be mastered by these students early through whatever pedagogical means are necessary before they can be successful in secondary school (Shanahan & Shanahan, 2008). Interventions that improve student learning are particularly important for diverse student who are placed at risk early in their schooling and for teacher education today.

Discussions regarding the best policies and evidence-based practices for emergent bilingual students referred to as English language learners (ELLs)[1] are salient today as public education seeks to become effective for this growing population. ELLs represent 1 in every 10 students in the K–12 system (National Center for Education Statistics, 2015), with 77.2% being Spanish speakers. Emergent bilingual (EB) learners today comprise a significant portion of the United States's school-age population that is often underserved and not exposed to evidence-based teaching methods in school (Bunch, 2013; García, 2009; Lucas, Villegas, & Freedson-Gonzalez, 2008; Nichols, Glass, & Berliner, 2006). Nationally, the public educational system leaves

mostly low-income ELLs years behind the norm in most content areas long before the end of high school despite well-meaning educational and language policies (Kena et al., 2014; Portes, Salas, Baquedano-López, & Mellom, 2014; Rios-Aguilar, González Canché, & Sabetghadam, 2012; Salas & Portes, 2017).

ELLs have been and still remain treated much from a deficit perspective, grounded in part on special education (Peal & Lambert, 1962; Rueda & Windmueller, 2006). Public school bilingual or dual-language programs were strongly discouraged in many parts of the country at the turn of this century in contested political contexts (Powers, 2014). Ironically, around this same time period, Thomas and Collier (2002) and others reported important positive outcomes for emergent bilingual learners taught in bilingual education programs. Nevertheless, many districts have opted for sheltered instruction since as a more economic approach in mainstreaming emergent bilingual children with mostly English-only teachers. The majority of these educators have limited preparation for teaching the ELL population effectively through evidence-based methods. Nevertheless, it should also be noted that today many schools are again offering bilingual or dual-language programs in some states, and asset-based pedagogies that take into account learners' social capital have gained popularity (García & Kleifgen, 2010). Other promising program initiatives that require longer periods of time have also been implemented across the country, yet their results are difficult to assess and generalize relative to shorter intervention studies.

As Gersten and associates (2015) noted, randomized controlled trial (RCT) studies serve as critical precedents before determining whether larger scale implementations can sustain positive effects on school achievement in real-world settings. RCTs (e.g., Fuchs et al., 2005, 2013; Pinnell, 1989; Schwartz, 2005) have examined the efficacy of models aimed toward special populations often through content-focused curricular programs. Some of the latter involve reading-focused models in which other than emergent bilingual students are of primary interest. This study is part of a larger clustered randomized evaluation that presents early results for a promising conversational model for teaching emergent bilingual students and their peers. Next, we describe this Instructional Conversations (IC) approach that is part of the Center for Research on Excellence and Diversity in Education (CREDE) pedagogical framework. We also summarize related research concerning dialogic approaches associated with both ELL and non-ELL elementary school students.

Defining Instructional Conversations

The IC is one of five standards for effective pedagogy that define a constructivist pedagogical system initially developed by the CREDE. CREDE's five standards work together as a system in making teaching meaningful

and challenging to students through frequent assistance generated in small group discussions. This socio-cultural pedagogical approach was based originally on professional development of teachers in Hawaii for cultural and linguistic minority students (Tharp & Gallimore, 1989). The CREDE model promotes learning by building on students' first language and cultural experiences, thus it is an example of culturally responsive teaching (Gay, 2000; Ladson-Billings, 2009). The concept in general has evolved from earlier interdisciplinary foundations (see Bruner, 1966, 2009; Dewey, 2007; Vygotsky, 1978). From this socio-cultural framework (Cole, 1996; Moll, 1992; Portes, 1996), to promote learning driven by activation of linguistic tools and signs, teachers must utilize knowledge of their students' lived experiences in their cultural context for meaningful learning and instruction. Instructional practices centered on students' backgrounds and experiences promote meaningful learning, student engagement and motivation, and access to high-quality curriculum (Gay, 2000; Wlodkowski & Ginsberg, 1995).

The five CREDE pedagogical standards are:

1. joint productive activity, whose goal is to facilitate learning through collaborative and problem-based tasks between teacher and students;
2. language and literacy development, which aims to develop competence in the language(s) of instruction (English) and the academic disciplines across the curriculum (e.g., science and math);
3. contextualization, which makes meaning and connections by embedding curricular instruction in the interests, experiences, and skills of students' families and communities;
4. teaching complex thinking to consistently challenge students toward their next level of cognitive complexity or zone of proximal development (ZPD) in second language (L2) and literacy (e.g., verbal and written/symbolic comprehension, input/output, concrete/formal operations).
5. teaching through instructional conversations to develop students' cognitive and linguistic skills through facilitated discussion, social learning, and modeled academic/disciplinary conversation.

The last of these five standards, Instructional Conversations, is a regularly scheduled teacher-led event involving three to seven students, lasting about 20 minutes, with a clear instructional goal. The teacher leads through topic control, and thus the event is instructional, but the ordinary "courtesies" and characteristics of conversation apply. That is, students regulate their own speaking turns, everyone participates, and the teacher speaks less, thus allowing for close monitoring of students' comprehension and language development. This model allows for ongoing assessment and is in marked contrast to traditional "class discussions," events well known for teacher domination and participation by only the most able and verbal students (Cazden, 2001). Providing students the opportunity to dialogue with the

instructor about content in an open and responsive forum affords them sufficient language input as well as ample time to practice English. The IC provides teachers sufficient time and space to model academic language and respond collaboratively, extensively, and intensively to the student, offering the corrective feedback theorists argue is necessary for the learner to "notice" the correct forms of language (DeKeyser, 1998; Krashen, 1985) and develop implicit knowledge out of meaning-focused communication (Ellis, 2008). The topics for ICs integrate diverse academic disciplines: reading, science, math, social studies, and English language arts.

The implementation of the CREDE pedagogy in a classroom requires organizing activities to facilitate each standard before, during, and after the lesson. Before an IC, teachers must create a classroom structure that supports small group instruction and must consider management strategies, such as the implementation of rules and norms that guide students toward collaborative work that is independent from the teacher. Teachers must also develop activities for students that are collaborative in nature and encourage and even require conversational exchange. During ICs, teachers implement strategies to maintain classroom structure so that IC lessons are not interrupted and also provide instruction to students concerning how to appropriately participate in ICs. To this end, teachers may explicitly instruct students in the norms of conversation (e.g., how to disagree respectfully). Teachers facilitate ICs by keeping students focused on the goal of conversations and encouraging all students to participate. After ICs are over, teachers must reflect on the learning that happened during the IC to modify future implementations of the IC pedagogy. Additional information on teacher practices involved in this intervention is reported by Gokee (2017).

In sum, the IC is coordinated with the other four interrelated standards by teachers prepared through intensive and expert professional development. For students, IC discussions and joint problem solving center on specific academic language and academic goals as they engage in joint productive activities and contextualized instruction throughout the school year. The activities and conversations contextualize knowledge by using students' own cultural experiences and resources. Teachers scaffold student thinking by modeling logical reasoning and cognitive strategies that demonstrate how to advance complex thinking and scientific concepts (Vygotsky, 1978). The *IC* term is used as a proxy for all five standards in further describing the intervention in this study.

Review of the Literature

A growing literature is emerging with respect to pedagogical models and strategies that promote students' academic success in reading and other areas. Studies concerning English learners also vary in foci ranging from response to intervention (RTI) to best practices from interrelated disciplines

(Artiles, Rueda, Salazar, & Higareda, 2002, 2005; August et al., 2014; Harrison & Thomas, 2014; Rueda & Windmueller, 2006). This research area includes some experimental studies that address key questions regarding the most effective practices that teachers might implement to improve diverse students' learning.

Improving teacher quality for ELLs is attracting increased research attention to specific approaches that can improve the capacity of English-only teachers. Efficacy studies of teacher quality for these students are most relevant in addressing interrelated questions concerning why students whose English language competencies improve over time still fail to reach grade-level proficiency (Abedi & Linquanti, 2012). As Vaughn, Wanzek, Murray, and Roberts (2012) noted, both explicit and systemic pedagogical means should serve to increase phoneme awareness, often through discussion and visuals, modeling differently what readers are expected to do depending on their developmental characteristics (i.e., regular, special needs, or emergent bilingual status). To achieve this goal, elementary school teachers must be prepared to design and implement complex learning activities within students' multiple zones of proximal development. Using an integrated set of teaching principles centered on ELLs' affective and cognitive lived experiences can serve to maintain students' academic engagement and motivation. Short and Fitzsimmons (2007) noted in a Carnegie Corporation report that most Latino ELLs do not meet content standards due to a lack of educator capacity and evidence-based teaching. Teachers, on the other hand, require evidence-based professional development beyond that generally offered in their preparation.

Depending on a number of in-school factors, as well as family social capital and community context variables, achievement gaps created in elementary grades tend to widen through secondary school, particularly for reading and writing (Kim, 2011). These achievement gaps are of particular concern among ELLs, who are likely to remain in restrictive learning environments (Salas & Portes, 2017) that often fail to motivate learners. The overlap between Latino post–first generation students and ELLs is considerable and reflected in poverty-laden communities (Kena et al., 2014).

Transitions for Emergent Bilingual Learners

Part of the background for this study concerns the political context noted earlier regarding bilingual education issues in public schools earlier this century (Genesee, 2006; Powers, 2014; Rolstad, Mahoney, & Glass, 2005; Thomas & Collier, 2002). The political response in many school districts since leans generally toward mandated sheltered or transitional English for speakers of other languages (ESOL) models that lack evidence-based research in terms of ELLs attaining grade-level proficiency. This policy response remains problematic because academic gaps for many emergent

bilingual learners are developed prior to middle and high school and continue to amplify over time even after reclassification (Kena et al., 2014; Portes & Salas, 2015; Umansky & Reardon, 2014). A resistance to such policies has been attenuated by the reemergence of dual language instruction options in some states, such as Colorado, North Carolina, Utah, and Delaware. To date, strong evidence for sheltered instruction programs in closing learning gaps is lacking compared to that found for dual language instruction or bilingual programs as noted previously.

An important question underlying this study is whether a culturally responsive instructional model integrating best practices can prove effective in improving teaching and learning outcomes for ELLs and all students in general. In Table 1, we present a cross-section of recent literature on improving ELL academic success that complement two comprehensive meta-analyses (Adesope, Lavin, Thompson, & Ungerleider, 2010; Cheung & Slavin, 2012). Some studies reflect specific practices addressing one or more reading indicators while others reflect instructional models that approach or meet rigorous What Works Clearinghouse (WWC) design standards. Few studies, however, address how the practices are derived from a theory of action model or how they might be sustained.

In sum, we frame this study on literature examining the role of policies, laws, and their impact on ELLs' academic success and equity in education (Gándara & Orfield, 2010; Powers, 2014; Rios-Aguilar & Gándara, 2012). It explores teaching practices in relation to the many challenges in implementing professional development effectively (Desimone, 2009; Desimone, Porter, Garet, Yoon, & Birman, 2002; Garet, Porter, Desimone, Birman, & Yoon, 2001; Penuel, Fishman, Yamaguchi, & Gallagher, 2007). The extent to which the IC pedagogy improves both ELL and non-ELL academic learning in mainstreamed classrooms is explored next.

Prior Research on the IC and Theory of Change

The IC model is anchored in cognitive-developmental theory and is part of a pedagogical model that has been the subject of four decades of multimethod quasi-experimental studies (Dalton, 2008; Tharp, Estrada, Dalton, & Yamauchi, 2000). One study (Saunders, 1999) delivered the IC intervention as a key component of an elementary language arts program. A subsequent study (Saunders & Goldenberg, 1999) showed statistically significant effects of IC on achievement. Saunders's work on IC found differences between treatment and control groups in terms of English language development and reading achievement that were recognized as promising by the WWC (2007).

In this study's theory of change, it is posited that as the five standards of effective teaching are enacted, a sufficiently high level of responsive assistance for ELLs can be generated that increases student learning. This effect is hypothesized to work mainly through the interaction between cognitive,

Table 1

Recent English Language Learner (ELL) Interventions by Grade, Effect Size, and Other Characteristics

Study	Intervention Description	Design	Duration	*N*	Grade	Sample Characteristics	Posttest/ Effect Size	Overall ES
Apthorp et al. (2012)	Elements of Reading	RCT	1 year	16,471 in 44 schools	K–4	45 districts in an unnamed Southeastern state	Proximal effects: TOIW-V +.95 to +1.24 TOIW-C +.09 to +.44 No significant distal effects	N/A
Coyne et al. (2013)	Early Reading Intervention	RCT	126 daily lessons	162 in 10 schools	K	One district in Florida: 17.2% ELL (experimental), 13.3% ELL (control)	Within-study results: LNF (DIBELS) = –.08 Sound Matching (CTOPP) = +.15 WRMT-R/NU Word Attack = 0 Word Identification = –.18	N/A
Dyson, Jordan, and Glutting (2013)	Number Sense	RCT	8 weeks	121 in 5 schools	K	25% ELL	Woodcock-Johnson III Tests of Achievement Form C Brief Battery (WJ) Applied Problems and Calculation subtests Effect sizes run from +.18 (Number Combinations NSB posttest) to +.32 (Number Recognition NSB posttest)	N/A
Filippini, Gerber, and Leafstedt (2012)	Vocab+	RCT	Average 394.5 minutes per student over 8 weeks	71 in 1 school	1	Southern California school district: 61% ELL	PPVT Researcher designed target vocabulary knowledge DIBELS: NWF	

(continued)

Table 1 **(continued)**

Study	Intervention Description	Design	Duration	*N*	Grade	Sample Characteristics	Posttest/ Effect Size	Overall ES
Gersten et al. (2015)	Number Rockets	RCT	1 year	994 in 76 schools	1	4 urban districts: 46.7% Hispanic in treatment	Test of Early Mathematics Ability–Third Edition	+.34
Lawrence, Crosson, Paré-Blagoev, AND Snow (2015)	Word Generation	RCT	1–2 years	1,558 in 28 schools	6–8	2 unnamed urban districts: 5.4%/11% Latina/o, respectively	Researcher designed Math = +1.13 Science = + .47 Social studies = +.38 English language arts = +.44	+.62
May (2015)	Reading Recovery	RCT	1 year	866 in 158 schools	1	Unnamed schools: ~18% ELL	Iowa Tests of Basic Skills Reading Words = +.45 Reading Comprehension = +.44	N/A
Smith, Cobb, Farran, Cordray, and Munter (2013)	Mathematics Recovery	RCT	2 years	775 in 20 schools	2	5 districts in two states: ~17% LEP	MR Initial Assessment = +1.04 WJIII subtests: WJ AP, WJ QC, WJ MF = +.31 to +.40	+.15 to +.30

Note. CTOPP = Comprehensive Test of Phonological Processing; DIBELS = Dynamic Indicators of Basic Early Literacy Skills; ELL = English language learner; ES = effect size; K = kindergarten; LEP = limited English proficient; LNF = Letter Naming Fluency; MR = Mathematics Recovery; N/A = not available; NSB = Number Sense Brief; NWF = Nonsense Word Fluency; PPVT = Peabody Picture Vocabulary Test-Revised; RCT = randomized controlled trial; TOIW-C = Tests of Instructed Word Knowledge in Comprehension; TOIW-V = Tests of Instructed Word Knowledge in Vocabulary; WJ AP = Woodcock Johnson Applied Problems; WJIII = Woodcock Johnson III Achievement; WJ MF = Woodcock Johnson Math Fluency; WJ QC = Woodcock Johnson Quantitative Concepts; WRMT-R/NU = Woodcock Reading Mastery Tests -Revised/Normative Update.

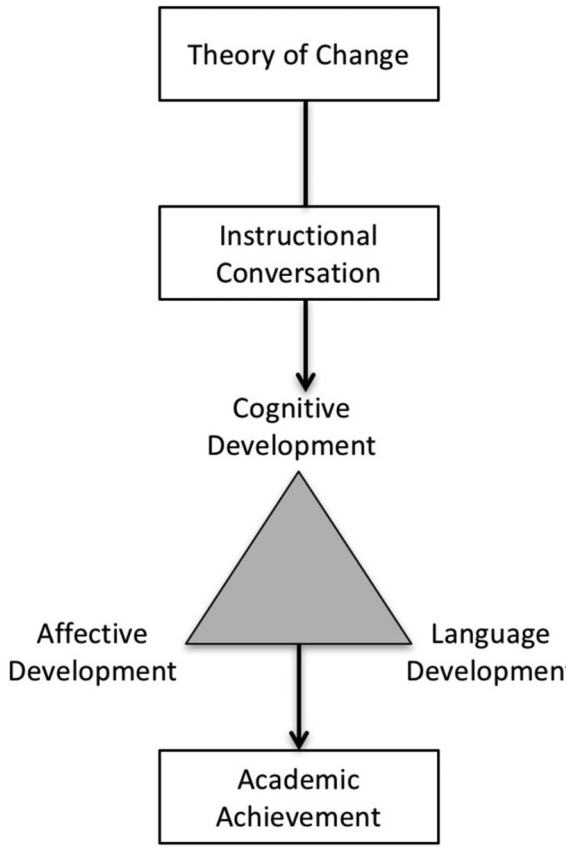

Figure 1. **Theory of change for the Instructional Conversation (IC) intervention. This figure illustrates how the intervention impact on academic achievement (reading) is mediated by cognitive, affective, and language development.**

affective, and second language development that is produced in regular conversational experiences framed around specific academic goals and contextualized instruction (see Figure 1). The definition of IC, as noted, is consistent with both cognitive sciences (see Bransford et al., 2005; Duschl & Hamilton, 2011; Tokuhama-Espinosa, 2014) and cultural historical theory (Cole, 1996; Vygotsky, 1978, 1986).

The five CREDE standards serve to integrate academic and language goals through the organization of student-centered learning activities in diverse classrooms. Research on CREDE practices has examined teachers' use of these standards in case studies of multiple classrooms, short-term

randomized designs, quasi-experimentation in single classrooms, and longitudinal studies of entire schools. Most studies have shown positive relationships between teachers' use of the five standards and student achievement and other favorable outcomes. The standards have also been associated with increased second language learning and positive affective development for emergent bilinguals (Estrada, 2004, 2005; García, 2009).

ELL students require ample opportunity for output to acquire correct forms and fluency (Swain, l985). As learners dialogue with their teachers and peers about meaningful content, they are provided time to write and research new words and skills to support spoken English. Through frequent small group discussions, new language skills, and metacognitive growth, students improve comprehension and focus competence in ways that translate to test situations. In other words, the transition from external to self-regulation in English is modeled and practiced regularly. In the IC model, the teacher is afforded sufficient time to model reasoning in standard English and offer strategic feedback that is necessary for a learner to master correct uses of language (DeKeyser, 1998; Krashen, 1985).

Finally, the IC model addresses important mediators that contribute to ELL engagement and development. Seixas and Peck (2004), among others, have noted that a truly student-centered model involves teaching the ability to see and understand the world from a perspective outside our own. For Latino ELLs in particular, Boutakidis, Rodríguez, Miller, and Barnett (2014) noted a significant interaction between academic engagement and grade point average not found for sampled non-Latino students. Latino-serving non-Hispanic teachers tend to be more effective when they have an interest in and/or knowledge of their students' cultures (Moll & Arnot-Hopffer, 2005). The IC model facilitates the latter through contextualization in learning activities throughout the school year. Teachers often prefer using pedagogies that are meaningful to all learners over those that are reported as useful primarily for Latino and/or ELL populations (Adesope et al., 2010; Rader-Brown & Howley, 2014).

In sum, this body of literature supports the IC pedagogy as compared to "teaching as usual" practices, particularly for culturally diverse students. Studies by the Saunders and Goldenberg group (Saunders, 1999; Saunders & Goldenberg, 1999; Saunders, Goldenberg, & Gallimore, 2009) found that ICs positively related to students' acquisition of thematic understanding of literature while the development of literal comprehension was not affected. Use of the five standards has also been linked to factors critical to school performance such as motivation, self-perceptions, attitudes, and inclusion (Estrada, 2004, 2005; Padron & Waxman, 1999).

Focus of Study: The Instructional Conversation Intervention

This study examines new evidence regarding how the IC pedagogy may impact diverse students in upper elementary grades. Specifically, it evaluates

the first part of a larger randomized efficacy trial of the IC model funded by the Institute of Educational Sciences (IES). Academic outcomes are examined using a common standardized test employed by the state where this research took place. Table 2 shows how our study differs from the original studies by Saunders (1999) and Saunders and Goldenberg (1999).

This design controls for gender and grade level to address some of the challenging issues found in replication studies that follow efficacy trials (Duncan, Engel, Claessens, & Dowsett, 2013; Stanovich, 2010). A main research question for this pilot study is:

> *Research Question:* Do students, and ELLs in particular, taught by teachers who implement the IC pedagogical model after a year of professional development and practice perform above those taught by control teachers?

This research question is operationalized as two null hypotheses:

> *Hypothesis 1:* Overall, experimental group students do not perform significantly different from controls on a standardized test measure of ELA.
> *Hypothesis 2:* Experimental ELL students taught by IC teachers do not perform significantly different from ELL control students on standardized test measures of ELA and other content areas.

In addition to these two hypotheses, we also explore (post hoc) the extent to which the intervention's impact on academic outcomes may vary by English language proficiency level.

Method

The students in this study (N = 1,521) were recruited from two cohorts of third and fifth graders (see Table 3). Teacher and district data suggested that the vast majority of ELLs spoke Spanish at home and were Hispanic (95%). Rates of participation in school free lunch programs for non-ELLs were 8% less than for ELLs participating in this study, indicating that the ELLs belonged to a lower socioeconomic status group than the non-ELLs. In our sample, both ELLs and non-ELLs participated in free lunch programs more than district averages (12% and 27% higher, respectively). Table 3 shows descriptive statistics for the combined two cohorts of students by experimental condition. Both treatment and control groups were comprised of similar proportions of female and male students, and both were roughly evenly distributed across grades. In both cohorts, slightly more than half of the students were ELLs who were monitored, already exited, or still served. Their English language proficiency (as measured by Assessing Comprehension and Communication in English State to State [ACCESS]test scores) ranged from low (N = 182) and intermediate (N = 178) to advanced (N = 179) levels. Student prior-year and current-year English language proficiency was measured using ACCESS[2] testing (WIDA

Table 2

Comparison of Current and Prior Instructional Conversation (IC) Intervention Studies

Characteristic		Saunders (1999)	Saunders and Goldberg (1999)	Current Study (2017)
Type		Quasi-experimental design	Randomized control trial	Randomized control trial
Design			Student-level random assignment within classroom	Student-level random assignment within classroom
Sample size	Districts	1 district	1 district	18 districts
	Schools	10 public elementary schools	1 school	21 schools
	Students	42 experimental	16 experimental	834 experimental
		42 control	16 control	693 control
Screening		ELL students all part of bilingual program	Both ELL and non-ELL students	None
		ELL students matched on individual and school characteristics	ELL students matched on language fluency measures. Non-ELL students participated without matching mechanism,	
Instructors		18 teachers—no specified control group other than population of nonexperimental teachers	5 experimental	40 experimental, 34 control
	Training/coaching	Intervention duration 5 years	Experimental group teachers part of research team	100 hours of professional development
		No specifics on training		Yearly refresh, follow-up coaching
	Intervention duration	5 years. Student-level treatment duration is unclear. Students may have received treatment throughout elementary school.	4 days	2 years per teacher cohort, 1 year per student. Measurements taken in efficacy year.

Note. ELL = English language learner.

Table 3
Descriptive Statistics for Students and Teachers

Characteristic		Treatment % (*N*)	Control % (*N*)
Students			
Gender	Female	51.26 (428)	51.46 (353)
	Male	48.74 (407)	48.54 (333)
Grade	Third	48.5 (405)	51.33 (359)
	Fifth	51.5 (430)	47.67 (327)
English language learner (ELL) status	Non-ELL	49.94 (411)	48.83 (335)
	ELL	50.06 (412)	51.17 (351)
Cohort***	1	52.69 (440)	53.21 (365)
	2	47.31 (395)	46.79 (321)
Teachers			
Gender	Female	88.24 (30)	95.00 (19)
	Male	11.76 (4)	5.00 (1)
Grade	3rd	55.88 (19)	60.00 (12)
	5th	44.12 (15)	40.00 (8)
Years teaching ELLs	0–2	8.33 (2)	14.29 (2)
	3–5*	41.66 (10)	7.14 (1)
	6–8	20.83 (5)	28.58 (4)
	9–11	16.66 (4)	7.14 (1)
	12+*	12.50 (3)	42.86 (6)
Total ELLs in classroom during efficacy year	0–5	16.67 (4)	0
	6–10	12.50 (3)	33.33 (6)
	11–15	20.83 (5)	38.89 (7)
	16–20	37.50 (9)	16.67 (3)
	21–25	4.17 (1)	5.56 (1)
	26+	8.33 (2)	5.56 (1)
Highest degree completed	BA	28.57 (4)	0
	MA	50.00 (7)	40.00 (2)
	EdS*	14.29 (2)	60.00 (3)
	PhD	7.14 (1)	0
Cohort	1	44.12 (15)	25.00 (5)
	2	55.88 (19)	75.00 (15)

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

Consortium, 2011). Scores were collected and analyzed to establish baseline equivalency for this study.[3]

The teachers participating in this study represented 22 schools and 14 school districts in a Southeastern state with Latino/ELL density ranging from 10% to 48% in elementary school during the 2013–2014 academic year. In total, 74 third- and fifth-grade teachers (39 and 35, respectively) took part in the study over a two-year period. We recruited teachers in these two grades by design because the academic outcomes of interest were

derived from the same standardized tests given by the state at the end of each respective year. State-level Criterion-Referenced Competency Tests (CRCT)[4] were administered to both third- and fifth-grade students for five content areas (ELA, reading, math, science, and social studies) as our outcome variables. Pre-intervention test scores for these same content areas were available for fifth graders only, who took these tests in the fourth grade.

Once the teachers were randomly assigned into treatment and control groups, treatment teachers (N = 40) completed a professional development practice year before implementing the IC model in the efficacy trial. While the first or "practice" year allowed teachers to develop expertise in the IC, our assessments of implementation fidelity and student test scores were collected in the second or "efficacy" year. Control teachers (N = 34) were also provided incentives to conduct "business-as-usual" teaching as we gathered fidelity of implementation videos and log teacher data (see Portes & González Canché, 2016). Information about teacher preparation and experience with ELLs, as well as other variables, were collected from teacher logs (Table 3).[5] Since the completion of many items in these logs was optional for teachers, some data about teachers' ESOL or bilingual education certifications can only be estimated. Generally speaking, treatment and control teachers were similarly distributed across grades and genders and had comparable prior experience working with ELLs.

The professional development provided to experimental teachers consisted of an intensive 100-hour program of assisted teaching experiences in the implementation of the IC. Trained coaches provided additional support and feedback and collected monthly evaluations of teachers' implementation of the pedagogy during a year of practice before the efficacy trial year. Control teachers also began to complete monthly activity logs during the efficacy year. Figure 2 depicts the research design after teachers were randomly assigned to experimental condition and completed the study.

Over the course of the study, experimental group teachers implemented small group IC instruction one to two times weekly, with each lesson averaging 30 minutes. Teachers were encouraged to enact the IC in all content areas while integrating vocabulary, phonemic awareness, and reading comprehension with science, math, and social studies. Counterfactual evidence was collected and monitored to document differences in small group instruction through logs and video-taped observations. Our approach was unique in that video-recorded lessons for both treatment and control group teachers were assessed by researchers trained in scoring a rubric-based method[6] for determining adherence and fidelity to the IC core elements and the other four CREDE standards.

Attrition

Based on the WWC (2014) standards handbook version 3.0, attrition is defined as study withdrawal due to participants' treatment or control statuses

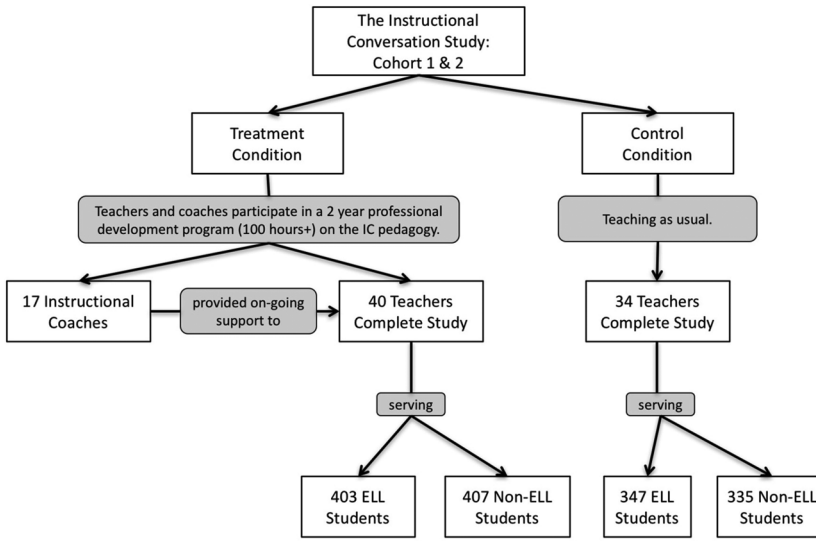

Figure 2. **Logic model. This figure depicts the randomized design after accounting for attrition.**

under circumstances that may be related to the outcome of interest. Participation that ceases for reasons that are not related to treatment status are not considered attrition, and the review team is responsible for deciding whether to apply the liberal or conservative attrition standard. In this study, we applied the conservative attrition standard to conduct our attrition analysis. Overall, we found that only 8 randomized teachers stopped enrollment, 4 treated and 4 controls. More specifically, Table 4 shows that of the 82 teachers who were initially randomized in these two cohorts, 74 remained until the end of the study. Following the WWC manual, we calculated overall and differential attrition figures (Table 4) and concluded that the overall attrition of 10.81%, rounded to 11%, would require a differential attrition of 6.2% and 10.9% under the conservative and liberal cutoffs provided by WWC, respectively. As Table 4 indicates, our differential attrition is 1.77%, which is below the conservative cutoff point.

The prevalent reason that teachers gave for withdrawing from the study was the time-consuming nature of activity logs. These logs were designed to provide evidence of fidelity of implementation and differences across treated and control teachers. To examine attrition bias at the student level, we also conducted within-cluster attrition analyses. In this case, all consented students remained with the treatment and control teachers who completed the study.

Table 4
Teacher Attrition Summary

	Cohort 1		Cohort 2		Total		
Status	Number	Attrited	Number	Attrited	Total	Attrited	Attrition
Treated	22	1	22	3	40	4	10.00
Control	20	2	18	2	34	4	11.77
Differential							1.77
Total	42	3	40	5	74	8	10.81

Defining ELL Status

The topic of ELL classification is important and has generated significant research (Abedi, 2008). Abedi (2008) both addressed the state of the literature and provided a quantitative analysis of how differing methods of classification related to different outcomes in achievement testing and completion. That author's final recommendation is a multitiered assessment that includes a home language survey, language proficiency test, standardized achievement test, and ultimately, teacher (or coordinator) recommendations. Kim (2011) analyzed how timing of classification affects achievement test and completion outcomes in an unnamed state system. The findings not only include the changes in classification procedures over a brief amount of time three years) but also illuminate the connection between early classification and outcomes. Umansky and Reardon (2014) analyzed the effects of different instruction programs on reclassification in a study of one large school district with multiple approaches to ELL progression. States, and therefore districts and schools, have a vested, varying interest in progression to English proficiency and can target shorter time to reclassification aggressively.

The definition of English language learners in research is of particular importance to this study. To determine eligibility for federally funded ESOL services, Georgia schools typically use a standard home language survey and the ACCESS placement exam to identify those students whose native home language is other than English (though overall those criteria vary from district to district and state to state) (Abedi, Hofstetter, & Lord, 2004). However, many students who are still learning English are not included in this narrow definition, such as those who are monitored after being exited from ESOL services and those whose parents refuse ESOL services but speak a language other than English at home. Assessing English learners' school performance and response to treatment requires attention to English language learning at different levels of development that are difficult to unpack from a single test score obtained midyear from districts. In this study, from school data provided for consented students by districts, we assessed ELL

students' English language proficiency to examine variations before and during the study.

Students not being served through ESOL services but qualified as ELLs as identified through the school systems' use of a home language survey and the WIDA-ACCESS Placement Test (W-APT) were included. The W-APT acts as a secondary "screener" that identifies students requiring further language assistance to provide them with the appropriate levels and amounts of language support services. This assessment is given whenever a language other than English is indicated as the home language on the state-mandated home language survey. The W-APT is also used as a guide for tier placement on the ACCESS for ELLs' annual assessment (Georgia Department of Education, 2015). We used both assessments to allow for a more accurate identification of ELL students.[7]

Using the W-APT adjusted sample produced a near-even split in ELL status (746 non-ELL, 764 ELL) among students in this study. Our ELLs represented different stages of English language proficiency. The sample of third- and fifth-grade ELL students was analyzed by experimental condition, grade, and gender using the aforementioned multitiered definition of ELL status. Most ELL students were receiving ESOL services or had recently exited into regular classrooms. Research questions regarding our intervention's goal of improving student academic achievement are examined mainly through ELA, reading, and other CRCT content area tests. We focused evaluation using the state mandated criterion test (CRCT) outcomes in this study before a new standardized test battery was adopted the following year.[8]

We begin our analyses for all students combined, then focus specifically on ELLs while considering grade level and gender. We then use follow-up analyses to explore if any potential intervention effect varies by ELLs' level of English language proficiency.

Baseline Equivalence

Baseline equivalency data were available only for fifth graders given state standardized testing policy for pre-intervention academic area test scores as shown in Table 5. The last column of this table includes Cohen's *d* effect sizes of the means, standard differences, and samples sizes observed. For the whole sample of consented fifth-grade students, the baselines for treatment and control groups were found to be equivalent in all academic achievement areas except social studies ($p < .001$). In this subject area, treatment students scored significantly higher than control students, displaying an average advantage of 8.33 points with a corresponding effect size of 0.28. Baselines were then examined separately for the ELL student group. Significant differences between treatment and control groups were found for all academic achievement areas, showing that fifth-grade ELLs in the treatment group started the efficacy trial with significant advantage

Table 5

Baseline Equivalencies for Student Achievement (CRCT) and English Language Proficiency (ACCESS)

	Treatment					Control					
Variable	*N*	*M*	*SD*	Minimum	Maximum	*N*	*M*	*SD*	Minimum	Maximum	Effect Size
All students											
ELA	384	831.22	46.78	311	930	308	830.43	28.96	745	930	0.02
Reading	384	836.58	41.79	317	920	309	834.38	30.45	772	920	0.06
Science	388	836.40	41.88	740	956	311	831.69	40.51	749	990	0.11
Social studies**	386	826.38	30.87	762	950	305	818.05	28.27	760	926	0.28
Math	384	835.18	43.12	735	990	310	828.53	46.07	339	940	0.15
ELLs											
ELA*	177	831.22	29.72	758	930	155	823.46	28.52	767	930	0.27
Reading*	177	836.58	33.96	765	920	155	827.99	30.24	775	920	0.27
Science*	179	834.43	44.37	748	956	157	823.20	39.64	749	956	0.27
Social studies**	177	824.89	31.36	763	926	154	812.36	28.79	760	926	0.42
Math*	177	837.69	49.90	735	990	156	826.28	38.27	739	940	0.26
ACCESS	285	325.48	26.57	233	395	254	328.99	25.57	236	393	−0.13
Non-ELLs											
ELA	207	831.21	57.56	311	930	153	837.49	27.75	745	930	−0.13
Reading	207	836.57	47.56	317	920	154	840.81	29.38	772	920	−0.10
Science	209	838.09	39.66	740	954	154	840.35	39.66	749	990	−0.06
Social studies	209	827.64	30.46	762	950	151	823.86	26.58	761	882	0.13
Math	207	833.04	36.32	743	940	154	830.82	52.83	339	940	0.05

Note. ACCESS means were calculated including both third and fifth grades. ACCESS = Assessing Comprehension and Communication in English State to State; CRCT = Criterion-Referenced Competency Tests; ELLs = English language learners; ELA = English language arts.
*$p < .05$. **$p < .01$.

over their control group counterparts. In ELA, the treatment group had a 7.76 point advantage ($p < .05$, effect size of 0.27); in reading, this advantage was 8.59 points ($p < .05$, effect size of 0.27); in science, the advantage was 11.23 points ($p < .05$, effect size of 0.27); for social studies, the advantage was 12.53 points ($p < .01$, effect size of 0.42); and in math, the advantage was 11.41 points ($p < .05$, effect size of 0.26). We also analyzed baseline equivalence for English language proficiency using ACCESS scores that were available for ELLs in both third and fifth grades (this test was administered to students halfway through the school year in February before and during the intervention). No significant difference was found for ELLs on this measure; however, the effect size yielded a disadvantage with a magnitude of 0.13 for ELLs in the treatment condition. In other words, while treatment and control groups were not significantly different concerning English proficiency level before the intervention, the effect size suggests an advantage for the control group. In sum, randomization at the teacher level did not produce baseline equivalency when data were disaggregated by ELL status. We further examined the problem of baseline equivalency by analyzing pre-ACCESS differences by grade. We noted the third-grade ELL sample size was considerably larger than the fifth-grade sample and found less English=proficient third-grade students were assigned to the experimental condition compared to those in the control group.

Data Analyses

Test score data were gathered from participating school districts for all consented experimental students. Table 6 shows CRCT posttest scores for the whole sample and by grade. Table 7 provides a more detailed set of descriptive statistics after disaggregation of post-intervention data by experimental and ELL status.

Given that CRCT test scores are not scaled across grade levels, these test data were standardized into *z*-scores stratified by grade level and content area. Our analytic strategy for this initial evaluation study focused first on an ordinary least squares (OLS) analysis of standardized scores that included experimental condition, gender, grade, and the interactions of both gender and grade with experimental condition as independent predictors of ELA test scores for all students (ELLs and non-ELLs). Our statistical model for this first analysis is as follows:

$$Y_i = \alpha + \beta_0 TREAT + \beta_1 FEMALE + \beta_2 THIRD + \varepsilon_i, \quad (1)$$

where Y_i represents the *z*-transformed ELA test score for student$_i$, β_0 is the coefficient of interest regarding impact of the IC treatment on student test scores, β_1 is student gender, β_2 is the effect of whether the student was in third or fifth grade, and ε_i is the error term.[9] Variable interactions with treatment status were also of interest in this exploratory study, leading to an

Table 6
Descriptive Statistics for Post-Intervention Content Area Standardized Tests (All Students and by Grade)

Variable	*N*	*M*	*SD*	Minimum	Maximum
All students					
English language arts (ELA)	1,496	826.49	55.09	273	972
Reading	1,491	829.59	54.58	286	932
Science	1,520	824.18	47.94	450	970
Social studies	1,512	824.18	41.97	464	950
Math	1,499	829.10	68.00	276	990
Third grade					
ELA	752	821.77	55.99	273	930
Reading	753	828.89	69.79	286	932
Science	764	821.64	35.28	731	926
Social studies	761	826.38	30.73	742	933
Math	755	825.95	68.49	276	990
Fifth grade					
ELA	744	831.26	53.78	283	972
Reading	738	830.30	32.41	301	920
Science	756	826.75	57.90	450	970
Social studies	751	821.96	50.81	464	950
Math	744	832.29	67.38	298	959

expanded model as depicted in Equation 2. Three-way interactions of treatment, grade, and gender were not significant and as such are not included in the models presented herein.

$$Y_i = \alpha + \beta_0 TREAT + \beta_2 FEMALE + \beta_3 THIRD + \beta_4 TREAT * FEMALE + \beta_5 TREAT * THIRD + \varepsilon_i. \quad (2)$$

Variables were entered into analyses in a stepwise fashion for the initial analyses. The results presented in Tables 8 through 13 include four models each: a model with treatment as the only predictor variable, a model that adds gender as a predictor, a model that adds grade as a predictor, and a fourth model that contains all predictor variables in the right side of Equation 2, including interactions.

As noted in Tables 8 and 9, we removed gender and the interaction between gender and treatment from analysis in the fifth OLS model to better understand the negative coefficient associated with the interaction between treatment status and grade (Treatment × Third Grade). The negative coefficient for the interaction between treatment and grade remained, which indicates that treated participants in the fifth grade consistently outperformed

Table 7

Descriptive Statistics for Post-Intervention Content Area Standardized Tests by Condition, English Language Learner (ELL) Status, and Grade

	Treatment Students					Control Students				
Variable	*N*	*M*	*SD*	Minimum	Maximum	*N*	*M*	*SD*	Minimum	Maximum
English language arts (ELA)	818	826.21	60.33	280	972	678	826.83	48.06	273	930
Reading	810	830.76	53.60	321	932	681	828.19	55.73	286	920
Science	834	822.93	55.01	450	970	686	825.71	37.56	727	945
Social studies	829	822.74	49.40	464	950	683	825.94	30.56	742	933
Math	820	828.28	71.28	276	990	679	830.08	63.84	284	990
Third Grade								Third Grade		
ELA	397	821.07	57.37	280	911	355	822.56	54.48	273	930
Reading	398	828.89	71.63	321	932	355	828.88	67.78	286	920
Science	405	821.22	35.93	740	926	359	822.11	34.58	731	914
Social studies	403	825.46	30.31	763	933	358	827.41	31.21	742	933
Math	399	826.98	72.33	276	990	356	824.81	64.00	284	990
Fifth Grade								Fifth Grade		
ELA	421	831.06	62.67	473	972	323	831.53	39.38	283	905
Reading	412	832.57	26.30	775	920	326	827.44	38.63	301	920
Science	429	824.54	68.31	450	970	327	829.66	40.27	727	945
Social studies	426	820.16	62.21	464	950	325	824.31	29.80	759	922
Math	421	829.51	70.34	298	959	323	835.90	63.25	300	959

(continued)

Table 7 **(continued)**

	Treatment Students					Control Students				
Variable	*N*	*M*	*SD*	Minimum	Maximum	*N*	*M*	*SD*	Minimum	Maximum
Treatment ELLs								Control ELLs		
ELA	403	831.59	27.66	761	930	347	822.71	38.15	283	930
Reading	403	828.60	25.38	765	920	348	821.51	53.59	301	920
Science	411	824.63	40.22	722	970	351	818.69	36.42	727	923
Social studies	407	824.03	31.94	753	950	349	821.85	31.52	742	922
Math	406	835.65	43.40	742	959	348	827.58	49.40	304	959
			Third Grade					Third Grade		
ELA	216	823.89	23.27	761	882	190	820.86	24.15	771	930
Reading	215	828.66	23.85	765	896	190	823.34	57.26	321	920
Science	220	818.11	33.64	740	914	192	815.20	30.69	731	914
Social studies	219	821.52	28.34	763	931	191	822.61	30.42	742	911
Math	217	830.65	45.80	742	955	191	822.49	38.66	725	958
Fifth Grade								Fifth Grade		
ELA	187	840.48	29.66	779	930	157	824.95	50.12	283	905
Reading	188	828.54	27.09	775	920	158	819.32	48.91	301	920
Science	191	832.14	45.61	722	970	159	822.91	42.04	727	923
Social studies	188	826.95	35.54	753	950	158	820.92	32.89	762	922
Math	189	841.39	39.82	758	959	157	833.78	59.47	304	959

(continued)

Table 7 **(continued)**

	Treatment Students					Control Students				
Variable	*N*	*M*	*SD*	Minimum	Maximum	*N*	*M*	*SD*	Minimum	Maximum
Treatment Non-ELLs						Control Non-ELLs				
ELA	403	830.35	59.37	280	972	331	831.15	56.35	273	908
Reading	407	832.90	71.25	321	932	333	835.17	57.12	286	915
Science	411	830.34	39.53	740	970	335	833.06	37.38	740	945
Social studies	410	830.62	31.12	765	933	334	830.22	28.96	759	933
Math	402	830.33	73.30	276	990	331	832.71	76.12	284	990
Third Grade						Third Grade				
ELA	181	817.70	81.08	280	911	165	824.52	75.68	273	908
Reading	183	829.16	102.58	321	932	165	835.27	77.85	286	915
Science	185	824.91	38.23	749	926	167	830.05	37.10	744	914
Social studies	184	830.14	31.96	765	933	167	832.91	31.29	763	933
Math	182	822.60	94.68	276	990	165	827.48	84.40	284	990
Fifth Grade						Fifth Grade				
ELA	222	840.66	28.63	779	972	166	837.75	23.90	774	905
Reading	224	835.95	25.19	785	920	168	835.07	23.11	758	885
Science	226	834.78	40.10	740	970	168	836.05	37.54	740	945
Social studies	226	831.01	30.49	768	922	167	827.52	26.24	759	922
Math	220	836.73	48.37	298	919	166	837.91	66.74	300	934

Table 8
Ordinary Least Squares Models for English Language Arts (ELA) for the Combined Sample

	Outcome: ELA				
	(1)	(2)	(3)	(4)	(5)
(Intercept)	−0.06	−0.21***	−0.21***	−0.30***	−0.15**
	(−0.04)	(−0.05)	(−0.05)	(−0.07)	(−0.06)
Treatment	0.12*	0.12*	0.12*	0.28**	0.27***
	(−0.05)	(−0.05)	(−0.05)	(−0.09)	(−0.07)
Female		0.27***	0.27***	0.28***	
		(−0.05)	(−0.05)	(−0.08)	
Third grade			0.01	0.18*	0.17*
			(−0.05)	(−0.08)	(−0.08)
Treatment × Female				−0.02	
				(−0.1)	
Treatment × Third Grade				−0.30**	−0.29**
				(−0.1)	(−0.1)
R^2	0	0.02	0.02	0.03	0.01
Adjusted R^2	0	0.02	0.02	0.02	0.01
N observations	1,476	1,476	1,476	1,476	1,476
Root mean square error of approximation	1	0.99	0.99	0.99	1

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

their treated counterparts in the third grade.[10] It should be noted that for the entire sample (Table 8, Model 5), control participants in the fifth grade performed 0.15 standard deviations worse than anyone else in the analytic sample (this is indicated by the model's intercept). Treated participants in the fifth grade performed 0.12 standard deviations better than their fifth-grade control participant counterparts. This estimate can be obtained by adding the intercept (−0.15) plus the treatment coefficient (0.27). Finally, treated participants in the fifth grade outperformed their treated counterparts in the third grade, also with a magnitude of 0.12 standard deviations. This estimate was calculated by adding all four coefficients (−0.15 + 0.27 + 0.17 + −0.29 = 0), from which the coefficients associated with treated status and fifth grade (0.27 + −0.15 = 0.12) were subtracted, yielding a value of −0.12 standard deviations for treated third graders with respect to treated fifth graders. For the ELL sample (Table 9, Model 5), the difference between treated fifth graders and treated third graders increased to 0.22 standard deviations. This result was obtained by applying a similar analytic procedure (−0.33 + 0.44 + 0.10 + −0.32 = −0.11 in the case of treated third graders and 0.44 + −0.33 = 0.11 in the case of treated fifth-grade students, with

Table 9
Ordinary Least Squares Models for English Language Arts (ELA) for English Language Learners

	Outcome: ELA				
	(1)	(2)	(3)	(4)	(5)
(Intercept)	−0.27***	−0.44***	−0.41***	−0.52***	−0.33***
	(−0.05)	(−0.06)	(−0.07)	(−0.09)	(−0.08)
Treatment	0.27***	0.25***	0.25***	0.45***	0.44***
	(−0.07)	(−0.07)	(−0.07)	(−0.13)	(−0.1)
Female		0.35***	0.34***	0.36***	
		(−0.07)	(−0.07)	(−0.1)	
Third grade			−0.06	0.12	0.1
			(−0.07)	(−0.1)	(−0.1)
Treatment × Female				−0.03	
				(−0.14)	
Treatment × Third Grade				−0.33*	−0.32*
				(−0.14)	(−0.14)
R^2	0.02	0.05	0.05	0.06	0.03
Adjusted R^2	0.02	0.05	0.05	0.05	0.02
N observations	749	749	749	749	749
Root mean square error of approximation	0.96	0.95	0.95	0.95	0.96

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

a difference of −0.11 – −0.11 = −0.22). In sum, these interaction coefficients do not indicate that the treatment had a negative effect on third graders relative to treated fifth graders but that treated fifth graders outperformed their treated third-grade counterparts.

To explore how the IC intervention may have impacted the academic achievement of ELL students in particular, our second step was to conduct similar analyses by ELL status. Although focusing specifically on ELA test scores as our outcome variable of interest for ELLs, we also conducted OLS analyses as described in Equations 1 and 2 for all five content areas measured by the CRCT: ELA, reading, math, science, and social studies. Finally, we conducted a series of exploratory analyses, described further in the following, that examine the relationship between pre-intervention English language proficiency and academic outcomes for ELLs. Due to the post hoc nature of these analyses, they are not powered to the same extent as our first two sets of analyses and should only be interpreted as exploratory in light of the lack of baseline equivalency among ELL students. This lack of baseline equivalency is additionally addressed by analyses summarized in Table 14

Table 10
Ordinary Least Squares Models for Reading for English Language Learners

	Outcome: Reading			
	(1)	(2)	(3)	(4)
(Intercept)	−0.24***	−0.34***	−0.37***	−0.44***
	(0.04)	(0.05)	(0.06)	(0.08)
Treatment	0.10	0.10	0.10	0.22*
	(0.06)	(0.06)	(0.06)	(0.11)
Female		0.22***	0.22***	0.22*
		(0.06)	(0.06)	(0.09)
Third grade			0.05	0.17
			(0.06)	(0.09)
Treatment × Female				0.01
				(0.12)
Treatment × Third Grade				−0.23
				(0.12)
R^2	0.00	0.02	0.02	0.03
Adjusted R^2	0.00	0.02	0.02	0.02
N observations	748	748	748	748
Root mean square error of approximation	0.82	0.81	0.81	0.81

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

that employ CRCT pretest scores as covariates, which were available for consented fifth-grade students.

Results

Results for our regression models are displayed in Tables 8 through 13. The first null hypothesis was rejected as the experimental students (ELLs and non-ELLs combined) performed significantly above controls in the ELA content area exam (see Table 8) for the sample of CRCT assessed cohorts (β = .28, $p < .01$). Since teacher-level randomization did not yield baseline equivalencies at the student level, the second hypothesis (see Tables 9–13) results are quasi-experimental and of heuristic value. They serve for comparison with those in Table 14, which include fifth-graders' pre-scores as covariates. ELL students taught by IC teachers appear to be significantly above ELL control students in ELA, reading, science, and social studies after controlling for grade, gender, and two-way interactions but not for pretests in these content areas. For example, ELA showed the largest intervention impact (β = .45) and reading the least (β = .22). All coefficients were significant at the $p <$ .05 level with ELA significant at the $p < .001$ level. Overall effect sizes

Table 11
Ordinary Least Squares Models for Math for English Language Learners

	Outcome: Math			
	(1)	(2)	(3)	(4)
(Intercept)	−0.15**	−0.25***	−0.21**	−0.23*
	(0.06)	(0.07)	(0.08)	(0.10)
Treatment	0.15*	0.15	0.15	0.20
	(0.08)	(0.08)	(0.08)	(0.14)
Female		0.19*	0.19*	0.27*
		(0.08)	(0.08)	(0.11)
Third grade			−0.07	−0.10
			(0.08)	(0.11)
Treatment × Female				−0.16
				(0.15)
Treatment × Third Grade				0.06
				(0.15)
R^2	0.01	0.01	0.01	0.02
Adjusted R^2	0.00	0.01	0.01	0.01
N observations	753	753	753	753
Root mean square error of approximation	1.04	1.03	1.03	1.04

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

(Cohen's *d*) were 0.28 for ELA (all students), 0.45 for ELA (ELLs only), 0.22 for reading (ELLs only), 0.20 for math (ELLs only), 0.20 for science (ELLs only), and 0.31 for social studies (ELLs only). Except for social studies, female students tended to score higher on standardized tests regardless of treatment status, and third graders (combined sample) tended to receive higher ELA test scores.

Robustness Check

The lack of pretreatment CRCT baseline equivalence found for older ELLs compromises the validity of the aforementioned findings regarding emergent bilinguals. Since these baseline equivalences corresponded to fifth-grade participants only, this second set of analyses relies on data for those ELL students only. These OLS analyses were conducted using the pretreatment indicator of each posttreatment academic outcome as a covariate in the models to control for significant differences in treatment and control students' pretest scores. This last step helps account for an observable source of bias in quasi-experimental fashion.

The results revealed that the treatment effect remained significant only for ELA with a magnitude of 0.17 ($p < .001$) favoring students in the

Table 12
Ordinary Least Squares Models for Science for English Language Learners

	Outcome: Science			
	(1)	(2)	(3)	(4)
(Intercept)	−0.15**	−0.25***	−0.21**	−0.23*
	(0.06)	(0.07)	(0.08)	(0.10)
Treatment	0.15*	0.15	0.15	0.20
	(0.08)	(0.08)	(0.08)	(0.14)
Female		0.19*	0.19*	0.27*
		(0.08)	(0.08)	(0.11)
Third grade			−0.07	−0.10
			(0.08)	(0.11)
Treatment × Female				−0.16
				(0.15)
Treatment × Third Grade				0.06
				(0.15)
R^2	0.01	0.01	0.01	0.02
Adjusted R^2	0.00	0.01	0.01	0.01
N observations	753	753	753	753
Root mean square error of approximation	1.04	1.03	1.03	1.04

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

experimental group (see Table 14). The treatment coefficients for models predicting math, science, and social studies test scores were not significant. Because these results apply only to a group of students for whom CRCT baseline data were available and equivalence was not established for ELL fifth graders, it may be that third graders also differed in baselines regarding relevant observed and unobserved characteristics. The models based on observed CRCT covariates (pretest scores) for fifth grade are shown in Table 14. These models indicate a limited benefit from the intervention overall (except for ELA).

ELL Analyses by English Language Proficiency

We conducted follow-up sensitivity analyses for each academic outcome area by disaggregating ELLs into English language proficiency tertiles to further explore how intervention may have varied in this regard. The tertiles were based on small samples of ELLs using pre-intervention ACCESS scores, which were available for 363 third-grade ELLs and 176 fifth-grade ELLs (N = 539). A total of 182 ELLs were in the low English language proficiency group, 178 were in the intermediate proficiency group, and 179 were in the high proficiency group. Given that this division of ELLs into tertiles resulted in three rather small groups of students, the analyses presented in this section

Table 13
Ordinary Least Squares Models for Social Studies for English Language Learners

	Outcome: Social Studies			
	(1)	(2)	(3)	(4)
(Intercept)	−0.16**	−0.17**	−0.15	−0.28**
	(0.05)	(0.07)	(0.08)	(0.10)
Treatment	0.07	0.07	0.07	0.31*
	(0.07)	(0.07)	(0.07)	(0.13)
Female		0.04	0.04	0.16
		(0.07)	(0.07)	(0.11)
Third grade			−0.05	0.08
			(0.07)	(0.11)
Treatment × Female				−0.23
				(0.15)
Treatment × Third Grade				−0.23
				(0.15)
R^2	0.00	0.00	0.00	0.01
Adjusted R^2	0.00	0.00	0.00	0.00
N observations	756	756	756	756
Root mean square error of approximation	1.02	1.02	1.02	1.02

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

are exploratory but useful. ACCESS scores for the low proficiency group ranged from 233 to 316, from 317 to 342 for the intermediate group, and from 343 to 395 for the high proficiency group. OLS regression analyses were then conducted to explore the effect of IC treatment on content area tests scores (ELA, reading, science, social studies, and math) for each language proficiency level rather than the ELL group as a whole.

The results summarized here are tentative due to uneven baselines and small sample sizes but help illustrate the extent to which intervention outcomes can be moderated by students' English language proficiency. They suggest the relation between IC intervention and ELL standardized academic achievement tests may vary by English language proficiency. For example, treatment status predicted higher ELA scores among low-proficiency ELLs ($p < .01$), and the intervention was also significant in math for the high-proficiency ELLs ($p < .05$).

Discussion

This study provides preliminary empirical support for the CREDE-based pedagogy and the study's socio-cultural theory of change with respect to ELA as a measure of academic language learning. Teachers who

Table 14
Ordinary Least Squares Robustness Check Including Pretreatment Indicators

	Outcomes			
	Post-ELA	Post-Math	Post-Science	Post–Social Studies
(Intercept)	−0.07	0.05	0.02	0.12
	(0.06)	(0.06)	(0.06)	(0.06)
Treatment	0.24**	−0.05	0.07	−0.11
	(0.08)	(0.09)	(0.08)	(0.09)
Pre-ELA	0.71***			
	(0.04)			
Pre-math		0.72***		
		(0.04)		
Pre-science			0.71***	
			(0.04)	
Pre–social studies				0.77***
				(0.04)
R^2	0.55	0.52	0.51	0.52
Adjusted R^2	0.55	0.52	0.51	0.52
N observations	325	325	331	327
Root mean square error of approximation	0.68	0.77	0.72	0.76

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

implemented this responsive pedagogical model facilitated upper elementary students' ELA outcomes generally. However, randomization of teachers in this study did not achieve baseline equivalency for ELL students in particular. As a result, our inferences are limited in a quasi-experimental design. It is important to note that the absence of baseline equivalence for ELL student measures generally indicates randomization design issues. These appear associated with lack of student-level randomization of ELLs across treated and control teachers, an issue we address in the following and in Figure 3.

The main finding, that the intervention positively influenced ELA test scores, holds after pretest scores are included in models for emergent bilingual students. The latter is important because the IC intervention's effect on ELA was significant for all students as stated in the first hypothesis where only one baseline (social studies) was not equivalent. These two findings are valuable in general for teacher preparation and education research. While the baselines for fifth-grade emergent bilinguals were not equivalent for reading, math, and science by condition, the intervention seems to be promising in these areas. In spite of its limitations, several aspects of this study are important especially because ELL students were mainstreamed

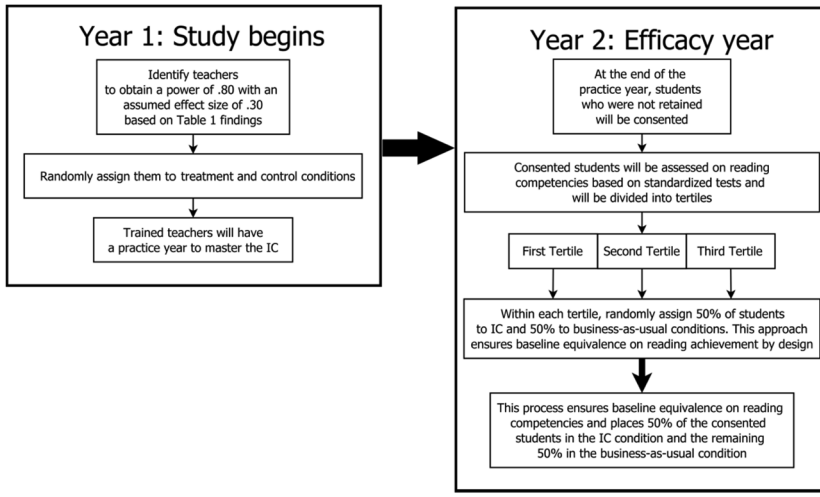

Figure 3. **Methodological framework to ensure baseline equivalence by design when conducting clustered randomized controlled trials.**

with other students in the classrooms participating in this study and both groups of students learned together through instructional conversations. The model appears to work for all students in an important content area where achievement gaps may be reduced.

It is evident that the randomization strategy in this study, while aimed to meet WWC standards, requires scrutiny. Given this study's focus on ELLs, the lack of baseline equivalence could have been avoided by randomizing students based on their pretreatment test performance. Future research may benefit from conducting randomization at the student level. This process is depicted in Figure 3, a model that includes randomization of both teachers and students for clustered RCTs. Randomizing within pretreatment academic performance can avoid limitations associated with baseline equivalence at the student level.

This matter is also related to the identification of students often classified as "joiners," (students who join a class after the beginning of the academic year) or "leavers" (students who were in the analytic sample before the intervention but not after). While we followed the WWC standards handbook (2014) for clustered RCTs, researchers generally should base their main analyses on "stayers," that is, students who remain in the treated or control conditions throughout the entirety of the study. According to WWC, studies relying only on stayers may meet WWC standards without reservations if attrition levels are within the acceptable limits. This point is

particularly relevant when considering parental consent to collect student level data. That is, skewed findings may result if not all students in the experimental condition are assessed (i.e., consent is not gained from all students) when learning outcomes are linked with the intervention.

Other Recommendations

This line of research requires larger samples in other grades along with both English and first language proficiency measures for examining response to intervention, particularly in relation to second language development. For example, advanced second language learners may respond to CREDE-like interventions in math or science while beginners are more likely to benefit in reading or ELA. While our preliminary analyses are encouraging, challenges remain including:

- exploring the extent to which this approach can be further developed to assist learners with limited academic language development at other proficiency levels or, more specifically, to assist all students at different levels of English language proficiency
- replicating and following up research outcomes with both teachers and students in other grades using rigorous designs that achieve baseline equivalence across relevant student characteristics and populations.

In terms of how this research informs the state of the art with regard to integrating a growing body of research on pedagogical strategies that work best for ELLs (Baker et al., 2014), several considerations may be in order. The CREDE strategies form part of a knowledge base for teacher education that requires comparative research with related models (Fuchs et al., 2005, 2013; Goldman, Snow, & Vaughn, 2016; May, 2015; Schwartz, 2005; Slavin & Madden, 2013). Beyond small group work (one of the four principles in the WWC Practice Guide; see Baker et al., 2014), advancing the CREDE pedagogy requires the integration of all five recommendations in achieving a comprehensive pedagogical approach for other content-specific teaching methods.

Limitations

Several drawbacks of the study are of concern. We investigated student assignment during the trial, and two main issues appear important to consider further. First, some sorting of students may have occurred with ELLs in some contexts. That is, students may have been purposely assigned to treatment teachers. For example, we found higher English language proficiency scores on the ACCESS test among third-grade control condition students yet lower CRCT pretest scores among the fifth-grade control group of ELLs. We asked teachers, principals, and coaches directly how students were being assigned to classrooms and if any irregular patterns such as

placing struggling students and students with low English language proficiency levels with treatment teachers occurred. The answers were negative in general, but some staff suspected that in some settings, struggling emergent bilinguals may have been assigned to treatment classrooms. This nonrandom assignment may have occurred after the IC practice year concluded and students were then being assigned to teachers for the study's efficacy trial second year. Nevertheless, such nonrandom assignment would not explain the advantage of treatment fifth-grade ELL students' test score baselines.

Second, an important source of bias concerns an active parental consent requirement that may be related to differences in student response rates. The latter can affect representativity as variations in consent form return rates may affect observed and unobserved student characteristics. As noted by Esbensen, Melde, Taylor, and Peterson (2008); Hussemann, Mortimer, and Zhang (2016); and others, active parental consent policies can lead to low participation rates and selection bias (i.e., loss of some students' representation) in field studies. While our active consent procedures produced an overall active consent rate of 80%, producing a sample of 1,476 consented students attending 74 schools in 16 districts, consent rates varied to some extent within and between schools.

Active student consent may be an important factor in designing studies such as this one, particularly among children of immigrants. For example, safety concerns and literacy may be in play for some Spanish-speaking parents. Even when consent forms are sent home in both languages, consent forms written in Spanish might be less likely to be signed if parents feel unsafe.

Broader Implications

Helping teachers identify and maximize the cultural resources that students bring to the classroom leads to creating a more competent and fair society. If current and preservice educators can better prepare academic language learners, lasting benefits from such investment might take into account each teacher's impact on new cohorts being multiplied over many cohorts. For a fast growing sector of the U.S. student population struggling with academic learning, teaching through conversation seems worthwhile. Thus, from a cultural-historical standpoint, implementing integrated pedagogical strategies provides incentives for restructuring teaching and learning in our schools. In spite of the intensive professional development required to enact the IC model and its related costs, the model seems promising for all students and for professional development of educators in multicultural settings.

Teaching emergent bilingual learners (or ELLs) with a culturally responsive dialogic model appears to have traction in general for all students in a gateway content area. The IC pedagogy was most strongly associated with ELLs and helped the least English proficient students most. This propensity appears reasonable as small group interaction allows multiple

opportunities for students to practice language with peers in joint productive activities that occur more frequently and by design in IC classrooms. CREDE standards need to be in effect before and after small group conversations during the school year to support teaching and learning in general. Together, they seem to help students enjoy various modeling opportunities in mastering academic language. These pedagogical changes may work to help learners in articulating their thinking through enriched discussions involving text and reading comprehension. Also, teachers use both language and academic lesson goals in teaching different content areas while explicitly showing correct uses of language. For example:

- In experimental classrooms, as teachers spoke less and listened to students more, they sharpened their awareness of students' language development and level of skills, inferencing, vocabulary, or other competencies and were thus able to better serve these students.
- Students in treatment classrooms increasingly benefitted from their interactions with other students through ICs as they voiced their reasoning and inferences of concepts that might have been advantageous when responding to standardized test questions assessed at the end of the school year.

Among ELLs, English language competency may develop through guided conversations as teachers challenge students' thinking. Teachers observed and became familiar with students' zones of proximal development and could assist them in transferring new skills across content areas (Lawrence et al., 2015; Palincsar & Brown, 1984). Theoretically, the IC discussions may help students better internalize teacher and peer learning processes dialogically and thus achieve greater self-regulation (Paris & Paris, 2001; Wertsch, 1979). Such progressions in reasoning may contribute to grasp other content areas where student-centered conversations can encourage learners. Teachers who employed the IC pedagogy were often observed facilitating cross-content learning in contextualized lessons.[11] These early findings appear encouraging but require further replication. The additional sensitivity analyses suggest the IC model may be particularly effective for dual language learners in ways that vary by English language proficiency level.

ELL students' learning potential may be advanced considerably once teachers' expertise is developed through an intensive learning-by-doing assisted experience that can (a) facilitate their students' discussions with peers (Howe et al., 2007) and (b) sustain teaching practices sensitive to students' language, culture, and literacy development (Portes & Smagorinsky, 2010). Teaching through conversation transforms the organization of classroom practices but is a practice not generally found in most regular elementary school classrooms (Lawrence et al., 2015; Tharp et al., 2000). Our findings underline the importance of activating a transformative teacher professional development process that is grounded in learning theory.

The interplay between improvement in academic language learning and increased higher order thinking can help account for benefits from cumulative instructional conversation experience. The role of expert-guided discussions may lead to better comprehension required of students in analyzing test questions keyed to learning objectives in content area outcomes. Instructional conversations, as defined in the study, encourage students to jointly share and evaluate ideas openly and critically (Howe, McWilliam, & Cross, 2005) and allow them to help each other using their native language. They represent a mediated learning approach (Bransford, Brown, & Cocking, 2000; Karpov & Bransford, 1995) that fosters engagement, motivation, and safety as students advance metacognitively (Corno & Anderman, 2016; King & Rosenshine, 1993; Pressley & Harris, 2008).

Summary and Conclusions

This study addressed important challenges and questions present in contemporary education research in evaluating the implementation of an established culturally oriented pedagogy. Some important questions included the extent to which students' ELA and other areas of academic performance can be improved through a student-centered cultural approach to teaching and learning. In spite of limitations, this intervention appears to benefit ELLs in an important foundational area without adversely affecting English monolingual students.

A two-step explanation is offered in accounting for these results. The first concerns the organization and procedures in the practice year that appear necessary and sufficient for enacting the IC teaching and learning model. That is, teaching expertise first has to be developed and sustained with a sufficient level of implementation fidelity to produce qualitative transformations in the classroom. Second, it seems reasonable that an intensive year of teacher professional development may be required for the IC model to take root and become associated with gains in students' learning (ELA) outcomes. Additional evidence of longer term or stronger impacts might also be explored in the future as teachers gain experience and their expertise grows in teaching emergent bilingual and monolingual students together. The aforementioned are critical aspects of the intervention model. They involve stimulating and mediating academic language as a mediational tool in ways that promote and challenge students' higher-order thinking and one that contextualizes learning through conversation. For the emergent bilingual populations in our schools today, the conditions brought about by this intervention represent a viable means for creating *the least restrictive* teaching and learning environments. In future research, the study of long-term effects for students taught by implementing contextualized and challenging conversations may be possible as schools gain capacity and focus on language development.

Notes

This study was funded by the Institute of Educational Sciences Grant No. R305 100670. The Institute of Educational Sciences is not responsible for ideas or information contained in this report. We would like to thank Paula J. Mellom, Jodi Weber, Rebecca Gokee, and Rolf Straubhaar for their contributions to this project.

[1]The terms *emergent bilingual* (EB) students and *English language learners* (ELLs) will be used interchangeably throughout this paper.

[2]Assessing Comprehension and Communication in English State to State (ACCESS) for ELLs is administered annually in February to all English learners in Georgia. ACCESS for ELLs is a standards-based, criterion-referenced English language proficiency test designed to measure English learners' social and academic proficiency in English. It assesses social and instructional English as well as the language associated with language arts, mathematics, science, and social studies within the school context across the four language domains. ACCESS for ELLs meets the federal requirements that mandates states to evaluate ELL students in grades K through 12 on their progress in learning to speak English (Georgia Department of Education, 2015).

[3]Student ACCESS test scores were not significantly different between the treated and control students prior to the intervention efficacy year. See Table 5.

[4]The Criterion-Referenced Competency Tests (CRCT) test reliability is reported to range from .86 to .91. This test was found valid with respect to its alignment with the curriculum, construction of items by content experts, reviews by educators, and correlations with the Iowa Test of Basic Skills (ITBS; Georgia Department of Education, 2008).

[5]In an attempt to protect teacher anonymity, teachers in Cohort 1 were not identified individually by name, only by grade and condition. As a result, some of their characteristics are missing, and the number of teachers representing each variable in Table 3 is quite small.

[6]This methodology is described in detail in a separate document examining implementation fidelity that is under review and available on request.

[7]The literature on ELL classification supports inclusion of multiple means of identification despite lack of consistency in school or district implementation. As this study's design was not intended to capture the subtleties of classification within our member districts and schools, we are limited to available data to address this issue. While there is not a known multitiered approach that precedes identification in Georgia schools, as suggested by Abedi (2008), the inclusion of the WIDA-ACCESS Placement Test (W-APT) for classification accuracy is reported.

[8]For this study, we reasoned that it was essential to investigate the intervention's early impact using a common outcome test before a new test (the Milestones) replaced the CRCT tests in the last year of our study.

[9]Multilevel models that account for the nested nature of the data are currently being fitted and will incorporate teacher-level indicators in the next analyses of all three cohorts, including outcomes from a new standardized test. Preliminary results taken from these models (null models) indicate that ELL students have an interclass correlation coefficient (ICC) of .30, whereas their non-ELL counterparts have an ICC of practically zero. This justifies the use of multilevel analyses for the ELL sample.

[10]Following the fundamentals of conditional expected values, the actual differences for treated and control participants across these interactions were decomposed.

[11]We base this assertion on video and log studies comparing small group interactions by condition as part of our final Institute of Educational Sciences evaluation report.

References

Abedi, J. (2008). Measuring students' level of English proficiency: Educational significance and assessment requirements. *Educational Assessment*, *13*, 2–3.

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

Abedi, J., & Linquanti, R. (2012). *Issues and opportunities in improving the quality of large scale assessment systems for English language learners*. Presented at the Understanding Language Conference, Palo Alto, CA. Retrieved from http://ell.stanford.edu/publication/issues-and-opportunities-improving-quality-large-scale-assessment-systems-ells

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*(2), 207–245.

Alfassi, M. (2004). Reading to learn: Effects of combined strategy instruction on high school students. *The Journal of Educational Research*, *97*(4), 171–185.

Apthorp, H., Randel, B., Cherasaro, T., Clark, T., McKeown, M., & Beck, I. (2012). Effects of a supplemental vocabulary program on word knowledge and passage comprehension. *Journal of Research on Educational Effectiveness*, *5*(2), 160–188.

Artiles, A. J., Rueda, R., Salazar, J., & Higareda, I. (2002). English-language learner representation in special education in California urban school districts. In D. J. Losen & G. Orfield (Eds.), *Racial inequality in special education* (pp. 265–284). Boston, MA: Harvard Education Press.

Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children*, *71*(3), 283–300.

August, D., Branum-Martin, L., Cárdenas-Hagan, E., Francis, D. J., Powell, J., Moore, S., & Haynes, E. F. (2014). Helping ELLs meet the Common Core State Standards for literacy in science: The impact of an instructional intervention focused on academic language. *Journal of Research on Educational Effectiveness*, *7*(1), 54–82.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C., Morris, J., & Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE 2014-4012). Washington, DC: National Center for Education Evaluation and Regional Assistance.

Boutakidis, I. P., Rodríguez, J. L., Miller, K. K., & Barnett, M. (2014). Academic engagement and achievement among Latina/o and non-Latina/o adolescents. *Journal of Latinos and Education*, *13*(1), 4–13.

Bransford, J., Barron, B., Pea, R. D., Meltzoff, A., Kuhl, P., Bell, P., & Reeves, B. (2005). Foundations and opportunities for an interdisciplinary science of learning. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 39–77). Cambridge, UK: Cambridge University Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn*. Washington, DC: National Academy Press.

Bruner, J. S. (1966). *Toward a theory of instruction* (Vol. 59). Cambridge, MA: Harvard University Press.

Bruner, J. S. (2009). *The process of education*. Cambridge, MA: Harvard University Press.

Bunch, G. C. (2013). Pedagogical language knowledge preparing mainstream teachers for English learners in the new standards era. *Review of Research in Education*, *37*(1), 298–341.

Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning* (2nd ed.). Portsmouth, NH: Heinemann.

Cheung, A. C., & Slavin, R. E. (2012). Effective reading programs for Spanish-dominant English language learners (ELLs) in the elementary grades a synthesis of research. *Review of Educational Research*, *82*(4), 351–395.

Clarke, S. N., Resnick, L. B., & Rosé, C. P. (2016). Dialogic instruction: A new frontier. In L. Corno & E.M . Anderman (Eds.), *Handbook of educational psychology* (pp. 378–389). New York, NY: Routledge.

Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Harvard University Press.

Corno, L., & Anderman, E. M. (2016). *Handbook of educational psychology* (3rd ed.) New York, NY: Routledge.

Coyne, M. D., Little, M., Rawlinson, D. A., Simmons, D., Kwok, O.-m., Kim, M., … Civetelli, C. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness*, *6*(1), 1–23.

Dalton, S. S. (2008). *Five standards for effective teaching: How to succeed with all learners, grades K–8*. New York, NY: John Wiley & Sons.

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42–63). New York, NY: Cambridge University Press.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, *24*(2), 81–112.

Dewey, J. (2007). *Experience and education*. New York, NY: Simon and Schuster.

Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2013). *The value of replication for developmental science* (Working paper). Retrieved from http://sites.uci.edu/gduncan/files/2013/06/Replication-paper-single-spaced.pdf

Duschl, R., & Hamilton, R. (2011). Learning science. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 78–107). New York, NY: Routledge.

Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities*, *46*(2), 166–181.

Ellis, R. (2008). *Principles of instructed second language acquisition*. Retrieved from http://www.cal.org/resource-center/briefs-digests/digests/principles-of-instructed-second-language-acquisition

Esbensen, F. A., Melde, C., Taylor, T. J., & Peterson, D. (2008). Active parental consent in school-based research: How much is enough and how do we get it? *Evaluation Review*, *32*(4), 335–362.

Estrada, P. (2004). Patterns of language arts instructional activity and excellence in first and fourth grade culturally and linguistically diverse classrooms. In H. Waxman, R. G. Tharp, & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 122–143). Cambridge, MA: Cambridge University Press.

Estrada, P. (2005) The courage to grow: A researcher and teacher linking professional development with small-group reading instruction and student achievement. *Research in the Teaching of English*, *39*(4), 320–364.

Filippini, A. L., Gerber, M. M., & Leafstedt, J. M. (2012). A vocabulary-added reading intervention for English learners at-risk of reading difficulties. *International Journal of Special Education*, *27*(3), 14–26.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology*, *90*(1), 37–55.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*, *97*(3), 493–513.

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., … Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology*, *105*(3), 683–700.

Gándara, P., & Orfield, G. (2010). *A return to the "Mexican room": The segregation of Arizona's English learners*. Retrieved from https://www.civilrightsproject.ucla.edu/research/k-12-education/language-minority-students/a-return-to-the-mexican-room-the-segregation-of-arizonas-english-learners-1

García, O. (2009). Emergent bilinguals and TESOL: What's in a name? *TESOL Quarterly*, *43*(2), 322–326.

García, O., & Kleifgen, J. A. (2010). *Educating emergent bilinguals: Policies, programs, and practices for English language learners*. New York, NY: Teachers College Press.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915–945.

Gay, G. (2000). *Culturally responsive teaching: Theory, research and practice*. New York, NY: Teachers College Press.

Genesee, F. (2006). *Educating English language learners: A synthesis of research evidence*. Cambridge, UK: Cambridge University Press.

Georgia Department of Education. (2008). *Validity and reliability for the CRCT 2008 CRCT*. Atlanta, GA: Assessment, Research, and Development Division.

Georgia Department of Education. (2015). *Access for ELLs*. Retrieved from https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Pages/ACCESS-for-ELLs.aspx

Gersten, R., Rolfhus, E., Clarke, B., Decker, L. E., Wilkins, C., & Dimino, J. (2015). Intervention for first graders with limited number knowledge large-scale replication of a randomized controlled trial. *American Educational Research Journal*, *52*(3), 516–546.

Gokee, R. K. (2017). *Improving reading achievement of ELLs one conversation at a time: Implementation of the IC model in upper elementary school classrooms—Voices from the field* (Unpublished doctoral dissertation). University of Georgia, Athens, GA.

Goldman, S. R., Snow, C., & Vaughn, S. (2016). Common themes in teaching reading for understanding: Lessons from three projects. *Journal of Adolescent & Adult Literacy*, *60*(3), 255–264.

Harrison, P. L., & Thomas, A. E. (Eds.). (2014). *Best practices in school psychology*. Bethesda, MD: National Association of School Psychologists.

Howe, C., McWilliam, D., & Cross, G. (2005). Chance favours only the prepared mind: Incubation and the delayed effects of peer collaboration. *British Journal of Psychology*, *96*(1), 67–93.

Howe, C., Tolmie, A., Thurston, A., Topping, K., Christie, D., Livingston, K., … Donaldson, C. (2007). Group work in elementary science: Towards organisational principles for supporting pupil learning. *Learning and Instruction*, *17*(5), 549–563.

Hussemann, J. M., Mortimer, J. T., & Zhang, L. (2016). Exploring the correlates of parental consent for children's participation in surveys: An intergenerational longitudinal study. *Public Opinion Quarterly*, *80*(3), 642–665.

Karpov, Y. V., & Bransford, J. D. (1995). LS Vygotsky and the doctrine of empirical and theoretical learning. *Educational Psychologist*, *30*(2), 61–66.

Kena, G., Aud, S., Johnson, F., Wang, X., Zhang, J., Rathbun, A., & Kristapovich, P. (2014). *The condition of education 2014* (NCES 2014-083). Washington, DC: National Center for Education Statistics.

Kim, J. (2011). *Relationships among and between ELL status, demographic characteristics, enrollment history, and school persistence* (CRESST Report 810). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

King, A., & Rosenshine, B. (1993). Effects of guided cooperative questioning on children's knowledge construction. *The Journal of Experimental Education*, *61*(2), 127–148.

Krashen, S. D. (1985). *Inquiries & insights: Second language teaching: Immersion & bilingual education, literacy*. Hayward, CA: Alemany Press.

Ladson-Billings, G. (2009). *The dreamkeepers: Successful teachers of African American children*. Hoboken, NJ: John Wiley & Sons.

Lawrence, J. F., Crosson, A. C., Paré-Blagoev, E. J., & Snow, C. E. (2015). Word generation randomized trial: Discussion mediates the impact of program treatment on academic word learning. *American Educational Research Journal*, *52*(4), 750–786.

Lawrence, J. F., Francis, D., Paré-Blagoev, J., & Snow, C. E. (2016). The poor get richer: Heterogeneity in the efficacy of a school-level intervention for academic language. *Journal of Research on Educational Effectiveness*. Advance online publication. doi:10.1080/19345747.2016.1237596

Lesaux, N. K., Kieffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of academic vocabulary instruction for linguistically diverse adolescents: Evidence from a randomized field trial. *American Educational Research Journal*, *51*(6), 1159–1194.

Lucas, T., Villegas, A. M., & Freedson-Gonzalez, M. (2008). Linguistically responsive teacher education preparing classroom teachers to teach English language learners. *Journal of Teacher Education*, *59*(4), 361–373.

May, H. (2015). *Evaluation of the i3 scale-up of reading recovery*. Paper presented at the 2015 Fall Conference: The Golden Age of Evidence-Based Policy.

McNamara, P., Burns, M. S., Griffin, P., Snow, C., & Strickland, D. (2002). *Preparing our teachers: Opportunities for better reading instruction*. Malden, MA: Joseph Henry Press.

Moll, L. C. (1992). *Vygotsky and education: Instructional implications and applications of sociohistorical psychology*. Cambridge, UK: Cambridge University Press.

Moll, L. C., & Arnot-Hopffer, E. (2005). Sociocultural competence in teacher education. *Journal of Teacher Education*, *56*(3), 242–248.

National Center for Education Statistics. (2015). *The condition of education*. Retrieved from http://nces.ed.gov/programs/coe/

Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, *14*, 1–175.

Padron, Y. N., & Waxman, H. C. (1999). Classroom observations of the five standards of effective teaching in urban classrooms with English language learners. *Teaching and Change*, *7*(1), 79–100.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117–175.

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist*, *36*(2), 89–101.

Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied*, *76*(27), 1–23.

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, *44*(4), 921–958.

Pinnell, G. S. (1989). Reading Recovery: Helping at-risk children learn to read. *Elementary School Journal*, *90*(2), 161–183.

Portes, P. R. (1996). Ethnicity and culture in educational psychology. In D. Berliner & R. Calfee (Eds.), *The handbook of educational psychology* (pp. 331–357). New York, NY: Macmillian.

Portes, P. R., & González Canché, M. (2016). *An evaluation of the instructional conversation model for English language learners and other students in elementary grades three and five: Executive summary*. Retrieved from https://coe.uga.edu/assets/downloads/clase/clase-grant-executive-summary.pdf

Portes, P. R., & Salas, S. (2015). Nativity shifts, broken dreams, and the new Latino south's post-first generation. *Peabody Journal of Education*, *90*(3), 426–436.

Portes, P. R., Salas, S., Baquedano-López, P., & Mellom, P. J. (Eds.). (2014). *U.S. Latinos and education policy: Research-based directions for change*. New York, NY: Routledge/Taylor & Francis.

Portes, P. R., & Smagorinsky, P. (2010). Static structures, changing demographics: Educating teachers for shifting populations in stable schools. *English Education*, *42*(3), 236–247.

Powers, J. M. (2014). From segregation to school finance: The legal context for language rights in the United States. *Review of Research in Education*, *38*(1), 81–105.

Pressley, M., & Harris, K. R. (2008). Cognitive strategies instruction: From basic research to classroom instruction. *The Journal of Education*, *189*(1/2), 77–94.

Rader-Brown, L., & Howley, A. (2014). Predictors of the instructional strategies that elementary school teachers use with English language learners. *Teachers College Record*, *116*(5), 1–34.

Rios-Aguilar, C., & Gándara, P. (2012). *Horne v. Flores* and the future of language policy. *Teachers College Record*, *114*(9), 1–13.

Rios-Aguilar, C., González Canché, M., & Sabetghadam, S. (2012). Evaluating the impact of restrictive language policies: The Arizona 4-hour English language development block. *Language Policy*, *11*(1), 47–80.

Rolstad, K., Mahoney, K. S., & Glass, G. V. (2005). Weighing the evidence: A meta-analysis of bilingual education in Arizona. *Bilingual Research Journal*, *29*(1), 43–67.

Rueda, R., & Windmueller, M. P. (2006). English language learners, LD, and overrepresentation a multiple-level analysis. *Journal of Learning Disabilities*, *39*(2), 99–107.

Salas, S., & Portes, P. R. (Eds.). (2017). *U.S. Latinization: Education and the new Latino south*. Albany, NY: SUNY Press.

Saunders, W. M. (1999). Improving literacy achievement for English learners in transitional bilingual programs. *Educational Research and Evaluation*, *5*(4), 345–381.

Saunders, W. M., & Goldenberg, C. N. (1999). Effects of instructional conversations and literature logs on limited-and fluent-english-proficient students' story comprehension and thematic understanding. *The Elementary School Journal*, *99*(4), 277–301.

Saunders, W. M., Goldenberg, C. N., & Gallimore, R. (2009). Increasing achievement by focusing grade-level teams on improving classroom learning: A prospective, quasi experimental study of Title I schools. *American Educational Research Journal*, *46*(4), 1006–1033.

Schwartz, R. M. (2005). Literacy learning of at-risk first-grade students in the reading recovery early intervention. *Journal of Educational Psychology*, *97*(2), 257–267.

Seixas, P., & Peck, C. (2004). Teaching historical thinking. In A. Sears & I. Wright (Eds.), *Challenges and prospects for canadian social studies* (pp. 109–117) Vancouver: Pacific Educational Press.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, *78*(1), 40–59.

Short, D. J., & Fitzsimmons, S. (2007). *Double the work: Challenges and solutions to acquiring language and academic literacy for adolescent English language learners: A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.

Slavin, R. E., & Madden, N. A. (2013). Success for all at 27: New developments in whole-school reform. *Journal of Education for Students Placed at Risk*, *18*(3–4), 169–176.

Smith, T. M., Cobb, P., Farran, D. C., Cordray, D. S., & Munter, C. (2013). Evaluating math recovery: Assessing the causal impact of a diagnostic tutoring program on student achievement. *American Educational Research Journal*, *50*(2), 397–428.

Stanovich, K. E. (2010). *How to think straight about psychology*. Boston, MA: Allyn & Bacon.

Swain, M. (1985). Large-scale communicative language testing: A case study. In Y. P. Lee, A. C. Y. Y. Fok, R. Lord, & G. Low (Eds.), *New directions in language testing* (pp. 35–46). Oxford, UK: Pergamon.

Tharp, R. G., Estrada, P., Dalton, S., & Yamauchi, L. (2000). *Teaching transformed: Achieving excellence, fairness, inclusion, and harmony*. Boulder, CO: Westview Press.

Tharp, R. G., & Gallimore, R. (1989). Rousing schools to life. *American Educator: The Professional Journal of the American Federation of Teachers*, *13*(2), 20–*25*, 46–52.

Thomas, W., & Collier, V. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement* (Report to the Office of Educational Research and Improvement). Washington, DC: CREDE.

Tokuhama-Espinosa, T. (2014). *Making classrooms better: 50 practical applications of mind, brain, and education science*. New York, NY: W.W. Norton & Company.

Umansky, I. M., & Reardon, S. F. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal*, *51*(5), 879–912.

Unrau, N. J., & Alvermann, D. E. (2013). Literacies and their investigation through theories and models. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 47–90). Newark, DE: International Reading.

Vaughn, S., Wanzek, J., Murray, C. S., & Roberts, G. (2012). *Intensive interventions for students struggling in reading and mathematics. A practice guide*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Vygotsky, L. (1978). *Mind in society: The development of higher mental process*. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1986). *Thought and language* (Newly revised and edited by Alex Kozulin). Cambridge, MA: The Massachusetts Institute of Technology.

Wertsch, J. V. (1979). From social interaction to higher psychological processes. A clarification and application of Vygotsky's theory. *Human Development*, *22*(1), 1–22.

What Works Clearing House. (2007). *WWC intervention report: Reading Recovery*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

What Works Clearing House. (2014). *Procedures and standards handbook version 3.0*. Retrieved from http://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

WIDA Consortium. (2011). *Annual technical report for ACCESS for ELLS English language proficiency test, series 201, 2009–2010 administration*. Retrieved from: https://www.wida.us/get.aspx?id=391

Wlodkowski, R. J., & Ginsberg, M. B. (1995). A framework for culturally responsive teaching. *Educational Leadership*, *53*(1), 17–21.

Manuscript received November 21, 2015
Final revision received October 3, 2017
Accepted October 9, 2017