

English and Spanish Acquisition by Hispanic Second Graders in Developmental Bilingual Programs

A 3-Year Longitudinal Randomized Study

Fuhui Tong

Texas A&M University

Beverly J. Irby

Sam Houston State University

Rafael Lara-Alecio

Texas A&M University

Patricia G. Mathes

Southern Methodist University

The authors studied a longitudinal English intervention in a 70/30 (Spanish/English) developmental bilingual education (DBE) program ($n = 302$) and compared this model with a typical-practice, 80/20, late-exit transitional bilingual education program ($n = 187$) in terms of students' language and literacy acquisition from kindergarten to second grade in an urban school district in southeastern Texas. The findings revealed significant differences (p values $< .05$) in favor of DBE students on (a) English measures of oral language, preliteracy skills, and reading fluency and comprehension (effect sizes of .12 to .71) and (b) Spanish measures of letter name and sound, preliteracy skills, and reading comprehension (effect sizes of .19 to .38). It was concluded that quality English instruction that incorporates direct and focused instruction, context-embedded vocabulary learning, and ongoing professional training in DBE programs is much needed to promote bilingualism and biliteracy.

Keywords: *developmental bilingual program; ESL intervention; bilingualism; biliteracy; quality of instruction*

In general, bilingual education entails various forms of models that are distinctive from one another, such as early-exit transitional bilingual education (TBE), two-way immersion, and developmental bilingual education

(DBE). Early-exit TBE is the most common first language (L1) instructional model in the United States, with the goal being a strong command of English and quick mainstreaming into English-only classrooms in a certain period of time (Genesee, 1999; Ramírez, Yuen, & Ramey, 1991). Unlike early-exit TBE, two-way immersion aims for bilingualism (i.e., to communicate in both languages) and biliteracy (i.e., to read and write in both languages). It involves both native English speakers and native speakers of another language in the same classroom, with the goals of (a) full academic language proficiency in L1 and the second language (L2) and (b) high levels of academic performance and cross-cultural understanding (Christian, 1994).

In contrast to two-way immersion, DBE, also used interchangeably with one-way dual language, one-way developmental, maintenance bilingual, and late-exit TBE (Genesee, 1999; Lindholm-Leary, 2000; Ramírez et al., 1991; Thomas & Collier, 2002), includes classrooms of only or primarily language-minority students (Genesee, 1999; Gomez, Freeman & Freeman, 2005; Mora, Wink, & Wink, 2001). This type of late-exit TBE provides strong grade-level schooling throughout the elementary school years and in most cases gradually increases the amount of instruction in English with each year until 50% of the content instruction is in English by fourth grade. The allocation of instructional time in L1 and L2 may vary across program types, and Collier and Thomas (2004) identified 90/10 and 50/50 as two major patterns of one-way dual-language or DBE programs. In a 90/10 model, L1 is used from kindergarten to second grade, for about 90% of the school day, and English as the medium of instruction is gradually increased until it reaches roughly 50% in the upper grades in elementary school (fifth or sixth grade). In a 50/50 model, instruction in L2 and L1 is balanced from the beginning throughout the elementary years (Christian, 1994; Lindholm-Leary, 2000).

Supportive theoretical frameworks for DBE programs have suggested that students reach high levels of academic achievement when the following components are developed: (a) Subject matter is taught in L1, and the two languages of instruction are separated; (b) literacy is developed in L1 and L2; (c) comprehensible input is provided in L1 and L2; (d) teachers are well trained and have high expectations of their students; and (e) instructional

Authors' Note: The research described in this article is supported by Grant R305P030032 from the Institute of Education Sciences (IES). The contents, findings, and opinions expressed in this article are those of the authors and do not necessarily represent the positions or policies of IES. We thank our IES project coordinator, our grant project coordinators, and the graduate assistants, teachers, paraprofessionals, parents, and district officials who made this research possible.

delivery is appropriate and well paced, focusing on developing students' higher order thinking skills (Genesee, 1999; Texas Education Agency, 2005). As a result, DBE is an additive model in that it focuses on continued instruction in a student's native language while he or she learns English, and this type of late-exit TBE was the basis for its use in our study.

Effectiveness of Developmental Bilingual Program

Several researchers have determined the effectiveness of DBE programs that promote the language, literacy, and academic achievement of English language learners (ELLs). Medina and Escamilla (1992a) evaluated the long-term impact of a 3-year 80/20 Spanish/English maintenance bilingual education (MBE) program on the development of English proficiency among 291 native Spanish-speaking students who were all identified as having limited English proficiency in kindergarten. Language use in this MBE program followed the following format: 80% Spanish and 20% English in kindergarten, 75% Spanish and 25% English in first grade, and 60% Spanish and 40% English in second grade. Data were collected from 24 elementary schools in a single school district in southern Arizona during kindergarten and second grade and were then divided into two linguistically homogeneous and independent samples according to Spanish oral language proficiency levels. The results indicated that all MBE participants had acquired significant levels of English, and students with lower Spanish oral proficiency demonstrated the greatest gains in their oral English development compared with their fluent counterparts. In a related study to investigate students' L1 and L2 oral language development in this 80/20 MBE model compared with an early-exit TBE model, Medina and Escamilla (1992b) reported that although students placed in both MBE and transitional bilingual models made statistically significant improvement in oral English proficiency from kindergarten to second grade, additive effects of MBE emerged in that students' oral proficiency in their native language was maintained. The authors concluded that MBE models might be the best alternative to serve ELLs, and yet they were the least implemented.

A government-funded evaluation of English immersion, early-exit TBE, and late-exit bilingual models conducted by Ramírez et al. (1991) revealed no difference among the three program types serving Spanish ELLs in their English reading and math by the end of first grade. However, afterward (i.e., from third grade to sixth grade), students in late-exit TBE demonstrated a significantly faster growth rate than students in the other two models in

the areas of English language and math and reached comparable levels with their English monolingual peers. The National Research Council reviewed Ramírez et al.'s findings and argued that conclusions beyond first grade were tenuous because Ramírez et al. did not (a) differentiate differences among students before they entered any program from differences produced by the instructional program, (b) control for family background (i.e., parental involvement, home literacy, social economic status), or (c) distinguish among program types (Meyer & Fienberg, 1992).

Thomas and Collier (2002) presented the results of a 5-year ex post facto longitudinal study to compare eight major program types serving language-minority students in five large districts nationwide. They stated that only students in two-way bilingual and 50/50 DBE models reached the 50th percentile or higher levels in both L1 and L2 in all subjects after 4 to 7 years of participation in the program. In their later longitudinal gap-closure research over 18 years, Collier and Thomas (2004) found that students attending either DBE (one-way dual-language) or two-way dual-language models achieved at grade level or higher on standardized English and Spanish reading tests, with an annual effect size of .14 or higher. They further asserted that students with former schooling in dual-language programs outperformed those with former schooling in English immersion. Although Thomas and Collier's study had a well-constructed longitudinal design with a rigorous methodology, it lacked random assignment and statistical pretesting; consequently, their study was not included in the most recent meta-analyses of bilingual education programs (e.g., August & Shanahan, 2006; Rolstad, Mahoney, & Glass, 2005; Slavin & Cheung, 2005).

More recently, de Jong (2004) identified the absence in literature regarding English proficiency development for ELLs placed in various bilingual program types and studied three cohorts in a total of 51 Spanish-speaking ELLs who entered DBE in kindergarten and stayed throughout third, fourth, or fifth grade. The DBE students made significant progress in English oracy, reading, and writing at each grade level and were just below fluency by the end of second grade in reading. By fourth grade, the DBE students had reached an average of nativelike academic English proficiency in comprehension, production, and writing.

Concerns Related to DBE Studies

Notwithstanding the positive results that emerged from the aforementioned empirical studies addressing programmatic differences and the effectiveness of educational settings for ELLs, questions still arise. First, few of

the studies randomized participants or followed the same groups of participants across time, with some of them using cross-sectional (e.g., de Long, 2004) or ex post facto (e.g., Thomas & Collier, 2002) data. Second, specific attention should be drawn to various DBE models, because although 90/10 and 50/50 models are more widely implemented and hold promise (Thomas & Collier, 2002), other models also have been identified with regard to language distribution during instructional activities, for instance, an 80/20 Spanish/English DBE program (e.g., Medina & Escamilla, 1992a, 1992b). In cases in which a particular program type is not offered in schools because of reasons such as limited resources, a lack of appropriate certified personnel, or district policy, there should be a focus on the quality of instruction within the classroom (August & Hakuta, 1997; Cheung & Slavin, 2005; Goldenberg, 1996). Unfortunately, existing literature on evaluative or comparative studies related to L2 learning has been conducted with a lack of indication of quality instruction.

Moreover, a chronological gap in research has been identified on the effectiveness of DBE, because most of the aforementioned studies were conducted in the early 1990s (e.g., Medina & Escamilla, 1992a, 1992b), with only two studies conducted after the new millennium (de Jong, 2004; Thomas & Collier, 2002). The paucity of such literature corroborates Medina and Escamilla's (1992a) claim that DBE is the least studied type of bilingual education program. Finally, among those studies reviewed, not a single one documented language and literacy development in both L1 and L2 simultaneously.

Therefore, the purpose of our study was to evaluate the effectiveness of student outcomes in language and literacy acquisition in both Spanish and English longitudinally from kindergarten through second grade, under controlled conditions, of an enhanced 70/30 Spanish/English late-exit TBE model (specifically, a DBE instructional model, which will be referred to as such hereafter) implementing an intervention of systematic English curriculum compared with an 80/20 Spanish/English late-exit TBE model implementing typical practice in the same district. Specifically, we sought to determine if treatment students (DBE intervention) would outperform control students (TBE typical practice) in both Spanish (their native language) and English (their L2) language development; second, we intended to answer if treatment students would outperform control students in Spanish and English literacy development. We were able to compare the two types of programs because they were both TBE program types, with one being enhanced as the experimental program, and because of the intervention's being transformed to a DBE program under the umbrella of TBE.

Method

Sampling and Research Design

The present study was derived from an ongoing 4-year longitudinal randomized trial targeting Hispanic ELLs' English language and literacy acquisition in a large urban school district in southeastern Texas. The district provides services to over 45% of students whose L1 is Spanish. The majority of students in this school district have low socioeconomic status, and those students qualify for free or reduced-price lunch. All student participants were identified as ELLs and had home language surveys indicating that Spanish was the primary language spoken at home.

Schools were selected on the basis of the availability of TBE programs in the school district and were then randomly assigned to either treatment (implementing an enhanced transitional bilingual program model of DBE) or the control group (remaining in the late-exit TBE model). In accordance with the state law (Texas Education Code, 1995) that prohibits the random assignment of individual students to bilingual programs or to English immersion programs, and to minimize contamination of the intervention, initial randomization was performed at the school level among 28 elementary schools that offered TBE programs in the district. Because we encountered a low return rate of parental consent forms in some schools, the final sample resulted in 19 schools and 32 classrooms, with 10 schools randomly assigned to receive enhanced treatment and 9 schools randomly assigned to receive typical practice in TBE. Teachers at each grade level were randomly assigned to treatment or nontreatment groups within the experimental or control campuses. Therefore, this project was a quasi-experimental design at the student level and an experimental design at the school level.

Student and Teacher Participants

We had a total of 502 initial participating students; by the end of kindergarten, there were 489 students (experimental $n = 302$, control $n = 187$), and by the end of second grade, there were 262 students (experimental $n = 141$, control $n = 121$). These 262 students' scores were matched to their respective scores at the initiation of the project. Table 1 depicts the specific tests administered with the pretest and posttest time frames for students from kindergarten to second grade. There was an attrition rate of approximately 48% over the 3 years, which is an expected result of most studies implemented in urban areas. The mean age of the final sample ($n = 262$) at the

Table 1
Pretest and Posttest Measures and Time of Test Administrations

Measure	Pretest	Posttest
Letter name/letter sound; Rapid Object Naming, Rapid Letter Naming (CTOPP/TOPP-S)	Beginning (first month) of kindergarten	End (last month) of kindergarten
Blending Phonemes Into Words, Segmenting Words (CTOPP/TOPP-S); Picture Vocabulary, Listening Comprehension (WLPB-R)	Beginning (first month) of kindergarten	End (last month) of second grade
Memory for Sentences, Oral Vocabulary, Letter Word Identification, Passage Comprehension, Word Attack (WLPB-R)	Beginning (first week) of first grade	End (last month) of second grade
Oral Reading Fluency (DIBELS)	Middle of first grade	End (last month) of second grade

Note: All measures were available in English and Spanish. CTOPP = Comprehensive Test of Phonological Processing; TOPP-S = Test of Phonological Processing–Spanish; WLPB-R = Woodcock Language Proficiency Battery–Revised; DIBELS = Dynamic Indicators of Basic Early Literacy Skills.

pretest (1st week) in second grade was 90.11 months ($SD = 4.35$ months). The experimental sample did not differ from control sample in class size, with an average of 10 across the 3-year span. There were 77 teachers involved, with an average of 14 teachers per grade level in the experimental group with an average number of total years of teaching of 9, and 11 teachers per grade level in the control group with an average number of total years of teaching of 7. Just as schools were randomly assigned to the experimental or control group, teachers were randomly assigned at the grade level to participate in the project.

Intervention: 70/30 DBE Model

Treatment students in the DBE model received a structured English intervention (75 minutes in kindergarten and 90 minutes in first and second grade daily), with the objective of developing their English language and literacy skills. The language distribution was 70/30 (Spanish/English) in kindergarten, 60/40 in first grade, and 50/50 in second grade. In kindergarten, Spanish instruction was in all content areas (language arts, math, science, and social studies). In first grade, Spanish instruction continued in all content areas, and formal English instruction was initiated by the spring semester of first grade.

In second grade, Spanish instruction continued in all content areas, with English language arts added, while English math was introduced during the second semester. The instruction followed a developmental approach in language development: first with oral and listening skills in Spanish to English, then moving gradually to reading and writing skills from Spanish to English. The curriculum is aligned with the Texas state standards. At each grade level, students transition from Spanish to English through oral language development that teachers provide and through music, art, and physical education beginning in kindergarten; however, the model maintains their native Spanish language. For each grade level, a detailed scope and sequence was provided to guide teachers in the implementation of the specific instructional model interventions. In addition, structured lesson plans for the intervention components were provided to all experimental teachers. These lesson plans reflected the curricular alignment between the national, state, district, and instructional program academic standards and objectives.

The intervention contained two levels. Level I was professional development. Teachers and paraprofessionals were provided with biweekly professional development workshops by research coordinators for 3 hours per session to (a) review and practice upcoming lessons, (b) reflect on and discuss student learning, (c) assess pedagogical progress as a teacher in the intervention, and (d) be instructed on the following strategies in English as a second language (ESL) that were incorporated into the researcher-developed lessons for kindergarten through second grade: language scaffolding, bridging, communication games, realia, flexible grouping, shared reading, leveled questions, manipulatives, preview and review, partner work and tutoring, vocabulary word dramatization, word walls, story reenactment, language experience approach, total physical response, and free voluntary reading. Teachers in the treatment group followed a script with minimal language-of-instruction clarifications, specifically L2 (English) clarified by L1 (Spanish) from kindergarten to second grade; otherwise, the language of instruction for the intervention components was L2. Teachers were trained not to translate or code-switch between the two languages; the languages remained separate, which is a premise of DBE or one-way dual-language programs.

Level II was student instruction, which was composed of three tiers. The Tier I intervention was the regular language arts, math, science, and social studies instruction discussed above, whether the language of instruction was Spanish or English.

Tier II was the English intervention, including three strands: (a) a research-based curriculum in teaching Spanish speakers content areas implemented in kindergarten and first grade using Santillana Intensive

English for 40 minutes (Ventriglia & González, 2000) and replaced by Early Intervention in Reading (EIR) Level II (Mathes, Torgesen, Menchetti, Wahl, & Grek, 2004) for 45 minutes in second grade; (b) Story Retelling and Higher Order Thinking for English Language and Literacy Acquisition (STELLA; Irby, Lara-Alecio, Quiros, Mathes, & Rodriguez, 2004), designed to be delivered in a 25-minute lesson during an ESL block; and (c) academic oral language (AOL), teacher-conducted daily oral language to develop students' oral language, for 10 minutes in kindergarten (Lakeshore Learning Materials, 1997), modified to AOL in science (AOLS) in first grade and academic oral and written language in science (AOWLS) in second grade to elicit students' writing.

Tier III of the instructional intervention was provided for the very lowest performing students identified by teachers via students' classroom functionality. Highly trained paraprofessionals delivered communication games (Quiros, Irby, Lara-Alecio, & Mathes, 2003) for an additional 20 minutes. During the second semester of first grade, a more aggressive reading intervention, EIR Level I, replaced communication games and was continued in second grade. Following language developmental patterns, the first 2 years (kindergarten and first grade) of the intervention focused on oral English language acquisition. In second grade, the intervention focused on the direct teaching of English reading fluency and comprehension. An in-depth description of Tiers II and III follows.

Santillana Intensive English (kindergarten and first grade) and EIR Level II (second grade). Santillana Intensive English is a research-based program that encourages L2 learners to be actively involved in their own learning. The curriculum is based on academic content such as math, science, and social studies. In addition, one major objective is to increase the students' English vocabulary banks. Each topic covered a 4-day period. Students were introduced to vocabulary words from lesson cards representing one teaching unit in the curriculum. During the structured lessons that the research team developed from the lesson cards, the students listened to stories read by the teacher, answered appropriately leveled questions for comprehension, and practiced new words with the teacher. The students also were grouped in pairs or small groups to role-play conversation and to work on their accompanying Santillana workbooks, individually and then collaboratively with partners. Day 5 was a make-up or extension day, when the teacher selected extended activities developed from the lesson cards or reviewed components from the lesson cards that the students had difficulty understanding.

After laying the foundation for oral language development in kindergarten and first grade, the research team decided to incorporate EIR Level II for 45 minutes to ensure that students became competent and fluent readers with strong comprehension skills during second grade. Although EIR Level II was developed to be used with small groups, one of its developers and one of the researchers in this project assisted in adjusting EIR Level II to be taught in a whole-class situation. Building on the principles of direct instruction (Becker, 1971), the program had 120 carefully integrated lessons that build student mastery skills through activities along five central content strands: phonemic awareness, letter-sound correspondence, word recognition and spelling, fluency, and comprehension.

The phonemic awareness strand included two types of activities: phoneme discrimination and phoneme segmentation and blending. In early activities, students were required to isolate initial sounds in words or to tell if words started with particular sounds. Later, these activities moved to isolating final and medial sounds. The letter-sound correspondence strand was introduced from the first day of instruction and continued through all lessons, with a new letter-sound correspondence being introduced every 2 to 3 days. Prior to presenting the symbol representing a particular phoneme, that phoneme was manipulated auditorily during segmenting and blending activities in preceding lessons. The primary objective was to ensure that children made automatic associations between letter-symbols and the most common sounds they represented so that this information could be used to decode unknown words. The word recognition strand included both phonetically regular and irregular or high-frequency words. The actual teaching or word recognition strategies were accomplished using lists of words that were either presented by the teacher or located in the student activity book. For the fluency strand, the word recognition strategy was practiced through the reading of decodable connected text.

Beyond decoding and fluency, a major objective of EIR Level II was to ensure that children were making meaning as they read. Thus, for the comprehension strand, prior to reading a story each day, teachers engaged in browsing the story, during which they asked children to look at the pictures in the story and to predict what they thought the story would be about. Teachers then set a purpose for reading, which usually was stated as finding out if these predictions were true or not. When the story was expository, teachers activated prior knowledge by asking students to tell what they already knew about the topic. After reading the story, students then engaged in a number of activities depending on the student competence and text structure. Following these prescriptive lesson plans, teachers delivered

explicit instruction designed to assist students in the integrated and fluent use of alphabetic knowledge and comprehension strategies.

STELLA. In kindergarten, *STELLA* was designed to be delivered in a 20-minute lesson during the 75-minute ESL block. In first grade, *STELLA* remained 20 minutes of the 90-minute ESL block and was then increased to 35 minutes during second grade. It was completely scripted instruction. Teachers received scripts prior to the week of story introduction, allowing them time to practice. Teachers introduced one book a week, accompanied by a script that included vocabulary words (3 in kindergarten, 6 to 8 in first grade, and 10 to 12 in second grade) per book, preselected ESL strategies (one to two in kindergarten and four to seven in first and second grades) aligned to the story, and a set of different leveled questions identified as easy, moderate, and difficult according to Bloom's taxonomy. Most of the activities included a mixture of dramatization and music to allow students to use their motor skills. During the instruction, teachers introduced the vocabulary, provided students opportunities to participate in discussion, and asked leveled questions; students were encouraged to work in pairs and/or small groups.

AOL (kindergarten), AOLS (first grade), and AOWLS (second grade). *AOL*, used in kindergarten, was selected for the daily oral language component. The chart has preprinted questions that help spark student discussion on a variety of topics. The teacher placed the preselected question of the day in a pocket chart along with three or four answer choices. The students addressed the question by answering in complete sentences. The students then placed cards with their names under the selected column that matched their answer choices. This created an instant graph that the teacher used to make comparisons and generalizations and to ask the students further questions. The preselected questions and answer choices were chosen on the basis of the Santillana themes. Therefore, the questions and answer choices were aligned to Santillana.

In first grade, *AOLS* was created on the basis of the same concept as Lakeshore Learning Materials' (1997) Question of the Day. However, because of the need for students to pass the science Texas Assessment of Knowledge and Skills by the end of fifth grade, the district requested that embedding the content area of science into the literacy component would be important. This allowed for academic language to be developed in the area of science.

In second grade, we continued integrating academic language through *AOWLS*. This science-based component was created using science-related

visuals to scaffold concept development, encourage oral language, and stimulate written languages.

Communication games and EIR Level I. Struggling students were pulled out from the Santillana instruction for the purpose of receiving paraprofessional-delivered small-group oracy instruction. Such small-group support is undergirded by evidence-based research in preventing reading failure among at-risk students (Foorman & Torgesen, 2001). Communication games, developed by the research team, expanded students' vocabulary and practice on phonemic awareness, increasing listening and speaking skills in the target language, English. The games included vocabulary introduced in Santillana, practice in segmenting and blending words, and the identification of the initial and ending sounds of a word. Starting from February of first grade, EIR Level I was implemented with the struggling students. It is a published curriculum providing intensive small-group instruction for struggling readers to develop phonemic awareness, reading fluency, and comprehension. This curriculum continued to be used with struggling students in second grade.

Control: 80/20 Transitional Bilingual Model

Control students received their regular TBE program with ESL instruction approximately 45 minutes daily. On the basis of observations by the research team, there was great variation across classrooms taught by bilingual teachers who were certified or permitted from kindergarten to second grade. Language distribution was 80/20 (Spanish/English) in kindergarten, 70/30 in first grade, and 60/40 in second grade. The Spanish instruction was identical to that of treatment, with differences being (a) more time allocated to Spanish language of instruction, (b) lack of monitored separation of the languages during instructional time, and (c) no specified clarifications of L2 clarified by L1 during ESL time. For the ESL component, the difference was that the treatment group received the minimum state- and district-required amount of English instruction, and teachers had no specific structured lessons. However, all curricula were aligned with Texas state standards and used Houghton-Mifflin (Spanish) series in language arts and Harcourt (Spanish) series in reading and content areas of science and social studies.

In kindergarten and first grade, all content areas (including language arts, math, science, and social studies) were taught in Spanish. In the second semester of first grade, students who passed Spanish language arts began formal English reading instruction. In second grade, Spanish instruction

continued in all content areas, with English language arts added, while English math was introduced during the second semester.

All ESL curricula were aligned with state objectives and district ESL benchmarks. A basic lesson included a focus, a lesson (what was being taught), guided practice, and independent practice, with the same materials as were used in regular classrooms. In a typical ESL lesson, the teacher first showed students the table of contents of a story, discussed the title and author, read and predicted the story, and conducted choral and partner reading. The teacher then checked students' understanding by asking comprehension questions. The instruction of vocabulary occurred in a variety of ways within the TBE typical-practice classrooms. For example, some of the instruction involved incidental words, flash cards, automaticity, and an activity page. In some classrooms, direct instruction, which was scripted, was used, whereas in other classrooms, the district-adopted ESL or English language arts curriculum was implemented. Other activities included oral reading of sentences, decoding words, counting syllables, and reading syllables and words. In general, the 45-minute ESL instruction varied across classrooms, with teachers in this TBE typical-practice model observed by the research team using code switching to clarify and to explain English concepts. No support was provided by the research team to the treatment teachers. The typical practice model is a late-exit TBE because students are not exited at least by the end of third grade.

Measures

The same comprehensive battery of measures in language and literacy was administered in both English and Spanish to all participants by well-trained professionals. Data analyzed in our current study were collected at the beginning and end of kindergarten (2004-2005), the beginning of first grade (2005-2006), and the end of second grade (2006-2007). First-time administration at the beginning of kindergarten took place during the 1st month of school, and the total length of time involved in the testing was between 2 and 3 weeks for both samples. Teams of well-trained testers were sent to schools, and individual students were pulled for no more than 30 minutes of testing at a time. Students were pulled on several occasions until all of the assessments were given in both English and Spanish. All tests were administered outside the classroom to minimize distraction. Table 1 presents the time of administration of each of the subtests in our study. The constructs tested included oral language proficiency, phonological processing, rapid naming, decoding skills, basic reading, reading comprehension,

and reading fluency. Pretest performance was used as the covariate for posttest performance on the same measure. To simplify the presentation of the outcomes, we describe all the measures with these constructs embedded within each test.

Letter naming and letter sound. Students were required to identify and provide at least one name or one sound for each of the 26 English alphabetic letters and each of the 30 Spanish alphabetic letters. These measures were untimed, and dependent variables were the total raw scores for each measure.

Comprehensive Test of Phonological Processing (CTOPP). CTOPP (Wagner, Torgesen, & Rashotte, 1999) assesses phonological awareness, phonological memory, and rapid naming, which are indicators of reading skills. It aids in the identification among children through college students with deficits in phonological processing to provide appropriate instructional activities to improve such abilities. Norming was based on a sample of 1,636 representatives of the U.S. population as a whole with regard to gender, race, ethnicity, residence, family income, educational attainment of parents, and geographic regions, ranging in age from 5 through 24 years. Elementary school representatives accounted for over half of the norming sample because CTOPP is expected to have its widest use through fifth grade. The average internal consistency and alternate-forms reliability coefficients are reported to be over .80. The test-retest reliability coefficients range from .70 to .92. Detailed information on content validity and concurrent and predictive validity are also provided in the CTOPP technical manual.

Four subtests were used in our current study: (a) Blending Phonemes Into Words (phonological awareness construct), a 20-item subtest that involves listening to sounds produced on an audiocassette recording and then combining the phonemic sounds of strings into words; (b) Rapid Object Naming (72 items, rapid naming construct), which presents a series of six objects, randomly arranged in a 4×9 table, to be named as quickly as possible (Forms A and B); (c) Rapid Letter Naming (72 items, rapid naming construct), which presents strings of six letters to be reported, randomly arranged in a 4×9 table (Forms A and B); and (d) Segmenting Words (20 items, phonological awareness construct), a supplementary subtest that requires an examinee to identify the separate phonemes that make up a target word. Although standard scores and percentile ranks are available in CTOPP, raw scores were used for the purpose of comparing with the Spanish version of CTOPP introduced below.

Test of Phonological Processing–Spanish (TOPP-S). The TOPP-S (Branum-Martin et al., 2006) was developed as the Spanish version of CTOPP (with phonemes and syllables suited for Spanish), parallel in linguistic complexity. The same subtests were administered, and each subtest had the same number of test items. Rasch analysis yielded a reliability coefficient of .83 for the entire test. We used raw scores comparable with those calculated for CTOPP in data analyses.

Woodcock Language Proficiency Battery–Revised (WLPB-R). The WLPB-R (Woodcock, 1991; Woodcock & Munoz-Sandoval, 1995) is a standardized instrument assessing broad language proficiency in oral, language, reading, and written language in English and Spanish. The English form was normed on a sample of 6,359 native English-speaking participants from age 2 to 99 (3,245 in K-12). Construct, content, and concurrent validity and related information can be found in the test manual (Woodcock, 1991). The Spanish form was normed on a sample of 3,911 native Spanish-speaking individuals from 22 countries, including 1,325 from the United States and 1,512 from Mexico, who were close to monolingual Spanish speakers; median coefficient alpha values ranged from .84 to .92 across all age ranges and from .68 to .95 at ages 6 and 9 (Woodcock & Munoz-Sandoval, 1995). The two forms are parallel. *W* scores (a special transformation of the Rasch ability scale) are used to determine a student's age- and language-specific standard score for that measure (Woodcock, 1991; Woodcock & Munoz-Sandoval, 1995). The Spanish tests have been calibrated on the English norm so that scores are comparable across languages.

The Picture Vocabulary, Listening Comprehension, Memory for Sentences, and Oral Vocabulary subtests were used to assess oral language. The Letter Word Identification, Passage Comprehension, and Word Attack subtests were used to assess decoding and comprehension. All subtests were administered individually by trained professionals outside the classroom to minimize distraction.

The Picture Vocabulary subtest requires test takers to name familiar and unfamiliar pictured objects. It is an expressive semantic task on a single word level that assesses test takers' familiarity with vocabulary. In the Listening Comprehension subtest, test takers are presented with a passage auditorily and are required to supply the single word missing at the end of the passage. The test focuses on a number of semantic operations. It begins with simple verbal analogies and associations and progresses to a higher level of comprehension involving the ability to discern implications. The Memory for Sentences subtest requires a student to repeat phrases or sentences provided

auditorily. The text increases in length. In the Letter Word Identification subtest, test takers are asked to match a rebus to a picture of an object (beginning items), then to read aloud individual letters, and then to read aloud words that are less frequently written. Word Attack measures test takers' ability to read aloud nonsense or unfamiliar words that are linguistically logical. Finally, the subtest of passage comprehension consists of multiple-choice questions that require test takers to point to the picture represented by a phrase. The remaining tasks measure test takers' skill in reading a short passage and identifying a missing key word. *W* scores, derived from item response theory with equal intervals, are well suited to assessing change in language development and were used for statistical analyses. However, to better inform students' performance for comparison across studies, as well as the relative standing in the norm, descriptive statistics presented were age-based standard scores, although analyses of *W* scores yielded very similar results.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS)/Indicadores Dinámicos del Exito en la Lectura. DIBELS (Good, Bank, & Watson, 2003; Good & Kaminski, 2002) assessments were designed as benchmarks administered to the whole class three times per year to determine progress in "critical skills that underlie early reading success" from kindergarten to third grade. The assessments are available in both English and Spanish. They can be used for evaluative purpose on individual students or student groups. For the Oral Reading Fluency subtest, examinees are asked to read grade-level fictional passages aloud, and the score is the number of words correctly read in 1 minute. The score for the Oral Retelling Fluency measure is based on the number of words used by the student that indicate understanding of the passage. In this study, Oral Reading Fluency was administered to students the first time in the middle of first grade and periodically through end of second grade. For each time of administration, three stories were tested, and the middle score was used for analysis.

Intervention Fidelity

To ensure the quality of implementation in both models, intervention validity checks were completed three times a year (beginning, middle, and end) by two observers trained by the research team of this project. The quantitative check was based on a 4-point, Likert-type rating scale on the following areas: (a) knowledge of the content and script, (b) materials usage and student involvement, (c) teacher talk versus student talk, (d) leveled questions, and (e) classroom management. The total possible score was 96.

The mean scores were 81.3 ($SD = 14.2$) in kindergarten, 80.7 ($SD = 14.1$) in first grade, and 81.2 ($SD = 13.9$) in second grade. The average observation time for each teacher was 77 minutes in kindergarten, 61 minutes in first grade, and 79 minutes in second grade.

Results

Pretest and posttest performance results for English and Spanish measures are presented in Tables 2 and 3, respectively. Because the intervention was in English, we did not anticipate significant findings in Spanish; however, considering instructional time difference in Spanish between two models, we decided to administer Spanish measures to determine the potential cross-linguistic transfer between Spanish and English.

English Letter Naming and Letter Sounds

Intervention and control students did not differ in their ability on the results of the naming of English letters ($p > .05$) after controlling for pretest levels, and the effect size was small (.23).

CTOPP. In the CTOPP measure Rapid Object Naming, treatment students scored higher than control students, with a marginally significant difference and small effect size, $F(1, 488) = 3.89, p = .049$, effect size = .12, whereas no difference was found for the Rapid Letter Naming subtest ($F < 1, p > .05$).

Spanish Letter Naming and Letter Sounds

Adjusting for pretest performance, treatment students outscored control students in naming Spanish letters, $F(1, 488) = 10.75, p < .001$, effect size = .25. In addition to the main effect, there was also an interaction effect between treatment and pretest performance, $F(1, 483) = 12.421, p = .001, \eta^2 = .02$. Plots of the interaction indicated that the increase in performance in two groups decreased as students' pretest performance increased, and students with lower pretest performance tended to develop faster; this phenomenon was more apparent within treatment group.

TOPP-S. The performance levels of the two groups on the TOPP-S Rapid Object Naming and Letter Naming subtests were not significantly different (F values $< 2, p$ values $> .05$).

Table 2
Pretest and Posttest Performance: English Measures

Measure	Model	n	Pretest		Posttest		d	F	p	Gain	η^2
			M	SD	M	SD					
Letter name ^a	T	302	5.71	6.77	13.04	7.71	0.23	1.62	<i>ns</i>	7.33	
	C	186	6.93	7.32	14.76	7.44				7.83	
Letter sound ^a	T	302	7	7.24	15.65	6.77	0.01	0.47	<i>ns</i>	8.65	
	C	186	7.3	7.06	15.55	7.06				8.25	
Rapid Object Naming ^{a,b}	T	302	208.15	54.72	146.55	64.87	0.12	3.89	<.05	-61.6	
	C	187	199.51	59.63	154.78	68.52				-44.73	
Rapid Letter Naming ^{a,b}	T	302	226.81	41.71	190.84	73.58	0.12	0.8	<i>ns</i>	-35.97	
	C	187	221.81	49.52	181.64	78.36				-40.17	
Blending Phonemes Into Words ^a	T	138	11.96	3.57	15.39	3.01	0.44	7.5	<.01	3.43	
	C	123	10.97	3.39	13.89	3.74				2.92	
Segmenting Words ^a	T	138	7.75	3.76	11.32	4.19	0.71	38.38	.00	3.57	
	C	123	7.69	4.63	8.5	3.76				0.81	
Picture Vocabulary ^c	T	138	60.25	16.43	76.8	16.83	0.56	9.05	<.00	16.55	
	C	122	62.61	30.94	66.59	19.58				3.98	.006
Listening Comprehension ^{c,d}	T	138	54.6	19.52	74.41	12.23	0.48	15.35	<.00	19.81	
	C	121	47.27	22	67.88	15.07				20.61	
Memory for Sentences ^c	T	133	75.08	12.74	87.36	14.35	0.27	4.75	<.03	12.28	
	C	119	74.2	13.78	83.21	12.53				9.01	

(continued)

Table 2 (continued)

Measure	Model	n	Pretest		Posttest		d	F	p	Gain	η^2
			M	SD	M	SD					
Oral Vocabulary ^c	T	133	81.33	14.12	87.55	13.04	0.3	5.01	<.03	6.22	
	C	120	74.11	17.52	83.42	12.64				9.31	
Letter Word Identification ^c	T	131	97.22	23.35	121.11	19.81	0.52	16.84	<.00	23.89	
	C	119	99.91	22.13	111.14	17.67				11.23	
Passage Comprehension ^c	T	133	98.04	14.8	102.26	9.36	0.46	13.72	<.00	4.22	
	C	120	97.63	15.01	97.07	11.1				-0.56	
Word Attack ^c	T	133	100.96	25.96	116.05	16.48	0.4	10.33	<.00	15.09	
	C	120	99.63	28.14	109.12	16.37				9.49	
Oral Reading Fluency ^a	T	118	21.13	22.63	91.51	29.25	0.16	5.16	<.02	70.38	
	C	116	31.11	23.61	86.72	34.28					

Note: n = number of students with scores available at posttest; d = Cohen's d (adjusted for sample overestimation bias) of the between-groups difference at posttest on the basis of unadjusted means; F = statistical test of the difference between adjusted group means; gain = average change in performance from pretest to posttest; T = treatment developmental bilingual education (70% Spanish/30% English) model; C = control transitional bilingual education (80% Spanish/20% English) model.

a. Raw score: Oral Reading Fluency on Dynamic Indicators of Basic Early Literacy Skills is the number of words read correctly within the specified time limit. Rapid Object Naming and Rapid Letter Naming on the Comprehensive Test of Phonological Processing are the total time spent in reading words and letters. All other raw scores are total correct.

b. A higher score corresponds to lower performance.

c. Standard scores: All are standard scores derived from the Woodcock Language Proficiency Battery-Revised.

d. These measure demonstrated a statistically significant interaction between pretest performance and treatment group, as indicated by partial η^2 .

Table 3
Pretest and Posttest Performance: Spanish Measures

Measure	Model	n	Pretest		Posttest		d	F	p	Gain	η^2
			M	SD	M	SD					
Letter name ^{a,d}	T	302	17.91	8.8	26.57	4.33	0.25	10.75	<.00	8.66	.02
	C	187	18.33	9.37	25.17	6.55				6.84	
Letter sound ^{a,d}	T	302	17.13	9.73	27.12	4.02	0.26	8.5	<.00	9.99	.017
	C	187	18.58	9.52	25.8	6.08				7.22	
Rapid Object Naming ^{ab}	T	299	141.03	71.28	111.23	42.39	0	0.89	ns	-29.8	
	C	179	159.63	82.09	111.05	44.5				-48.58	
Rapid Letter Naming ^{a,b}	T	264	206.16	107.64	106.49	68.2	0.08	1.57	ns	-99.67	
	C	153	224.27	102.1	100.91	67.77				-123.36	
Blending Phonemes Into Words ^a	T	138	12.87	3.8	16.16	2.85	0.38	9.14	<.00	3.29	
	C	123	12.52	4.14	15.01	3.18				2.49	
Segmenting Words ^a	T	138	10.34	6.65	16.03	3.86	0.21	4.14	<.04	5.69	
	C	123	10.88	6.23	15.21	3.91				4.33	
Picture Vocabulary ^c	T	140	94.97	27.25	100.06	22.15	0.11	0.12	ns	5.09	
	C	120	90.93	26.04	97.13	22.13				6.2	
Listening Comprehension ^c	T	140	88.29	15.92	94.85	10.07	0.19	4.36	<.04	6.56	.015
	C	120	88.6	13.88	92.4	12.17				3.8	
Memory of Sentences ^c	T	140	81.97	12.16	84.76	9.36	-0.24	3.51	<.06	2.79	
	C	120	82.5	13.23	86.49	9.62				3.99	
Oral Vocabulary ^c	T	140	86.16	18.78	97.58	18.51	-0.03	1.96	ns	11.42	
	C	120	80.67	26.32	97.59	19.09				16.92	

(continued)

Table 3 (continued)

Measure	Model	<i>n</i>	Pretest		Posttest		<i>d</i>	<i>F</i>	<i>p</i>	Gain	η^2
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
Letter Word Identification ^c	T	140	139.83	25.7	151.21	20.9	-0.6	21	<.00	11.38	11.38
	C	120	143.38	24.94	161.4	21.22				18.02	18.02
Passage Comprehension ^c	T	140	109.22	16.33	107.76	8.67	0.32	12.69	<.00	-1.46	-1.46
	C	120	111.89	15.61	104.72	9.51				-7.17	-7.17
Word Attack ^c	T	140	123.34	21.85	133.3	19.7	-0.51	11.51	<.00	9.96	9.96
	C	120	129.11	20.56	141.97	20.3				12.86	12.86
Oral Reading Fluency ^a	T	118	42.78	19.33	84.21	26.92	-0.26	0.56	<i>ns</i>	41.43	41.43
	C	116	48.38	21.84	92.08	31.86				43.7	43.7

Note: *n* = number of students with scores available at posttest; *d* = Cohen's *d* (adjusted for sample overestimation bias) of the between-groups difference at posttest on the basis of unadjusted means; *F* = statistical test of the difference between adjusted group means; gain = average change in performance from pretest to posttest; T = treatment developmental bilingual education (70% Spanish/30% English) model; C = control transitional bilingual education (80% Spanish/20% English) model.

a. Raw score: Oral Reading Fluency on Dynamic Indicators of Basic Early Literacy Skills is the number of words read correctly within the specified time limit. Rapid Object Naming and Rapid Letter Naming on the Comprehensive Test of Phonological Processing are the total time spent in reading words and letters. All other raw scores are total correct.

a. Raw scores.

b. A higher score corresponds to lower performance.

c. Standard scores.

d. These measure demonstrated a statistically significant interaction between pretest performance and treatment group, as indicated by partial η^2 .

English Phonological Processing

Adjusting for pretest performance, the results indicated that treatment students outperformed control students on the English CTOPP Blending Phonemes Into Words subtest, $F(1, 257) = 7.50, p < .01$, effect size = .44, and Segmenting Words subtest, $F(1, 235) = 38.38, p < .001$, effect size = .71.

Spanish Phonological Processing

After the adjustment of pretest levels, a statistically significant difference was identified between the two groups on Spanish TOPP-S Blending Phonemes Into Words subtest, $F(1, 259) = 9.14, p < .001$, effect size = .38, and Segmenting Words subtest, $F(1, 259) = 4.14, p = .04$, effect size = .21.

English Oral Language

The performance levels of the two groups on the WLPB-R Listening Comprehension subtest were significantly different (after adjustment for pretest performance), $F(1, 257) = 15.35, p < .001$, effect size = .48. In addition to the main effect between the two groups, there was also an interaction effect detected between group and pretest performance, $F(1, 257) = 7.73, p = .006, \eta^2 = .03$. Plots of interaction demonstrated that students with lower pretest performance tended to develop faster, and this was more apparent within the treatment group. Significant difference were also found on other three WLPB-R oral language subtests (i.e., Picture Vocabulary, Memory for Sentences, and Oral Vocabulary; F values > 4.75 , p values $< .03$, effect sizes $> .27$; see Table 2).

Spanish Oral Language

Treatment students did not differ from their control peers on most of the WLPB-R oral language subtests (p values $> .05$), controlling for pretest performance, with small effect sizes (.03 to .24) (see Table 2). The only difference was found for the Listening Comprehension subtest, on which treatment students outperformed control students, $F(1, 257) = 4.36, p = .038$, with a small effect size of .19. In addition to the main effect between the two groups, there was also a marginally significant interaction effect between treatment and pretest performance, $F(1, 257) = 3.89, p = .049, \eta^2 = .015$. Plots of interaction demonstrated that students with lower pretest performance tended to develop faster, and this was more apparent within the treatment group.

English Reading

Students were administered three English text-connected word-reading fluency stories from DIBELS, and the middle score was recorded for analysis. After adjusting performance levels using a DIBELS pretest at the middle of first grade, a significant difference was found in favor of the treatment group on the basis of posttest DIBELS in second grade, $F(1, 234) = 5.16$, $p = .02$, with a small effect size of .16. Performance on the English WLPB-R measures of decoding skills and reading comprehension indicated that treatment students outperformed their peers on the posttest at end of second grade, after adjustment for pretest performance levels at the beginning of first grade (F values > 10.33 , p values $< .001$), with moderate effect sizes of over .40.

Spanish Reading

Treatment and control students did not differ in word-reading fluency after controlling for pretest performance in the middle of first grade. The effect size for difference was small ($-.26$). Performance on the Spanish WLPB-R measures of reading comprehension indicated that by the end of second grade, treatment students outperformed control students after adjustment for pretest performance levels at the beginning of first grade, $F(1, 251) = 12.69$, $p < .001$, with a moderate effect size of .38. However, control students scored significantly higher than treatment students on WLPB-R word-level posttest measures (Letter Word Identification and Word Attack), with moderate effect sizes (.60 and .51, respectively).

Discussion

We were interested in the effectiveness of an English intervention implemented in an enhanced 70/30 Spanish/English TBE program, which was defined more specifically as a DBE program among Spanish-speaking ELLs from kindergarten to second grade. Statistically significant differences were apparent between treatment and control groups on both English and Spanish measures, with the magnitude being relatively smaller on Spanish measures.

English Oral Proficiency and Reading

For English measures, all significant findings were in favor of the treatment group, DBE, reflecting a range of oral and reading proficiency in

English, including phonological processing, letter-sound identification, vocabulary knowledge, and comprehension, with effect sizes ranging between .12 and .71, suggesting the effectiveness of this English intervention for students participating in a DBE program model. Such an intervention reflected the quality of English instruction, which is highly demanded, particularly among ELLs regardless of language of instruction (August & Hakuta, 1997; Cheung & Slavin, 2005; Goldenberg, 1996). It is recognized that phonological awareness is a critical determinant of reading, while the ultimate goal of reading includes high reading fluency and comprehension (Vaughn et al., 2006). In these two critical areas, treatment students demonstrated statistically higher performance than control students on contextualized word reading fluency. Although control students had scored an average of words read correctly higher than that of treatment students when the Oral Reading Fluency measure was first administered during the middle of first grade, at posttest in second grade, the treatment group on average correctly read 92 words per minute, higher than the control group (86 words per minute). According to screening level designation on DIBELS, performance of 90 words per minute or higher is considered reading with low risk; therefore, the control group would have been classified on average as at a higher reading risk than the treatment group. It should also be noted that students' levels of decoding and comprehension skills were higher compared with their oral English proficiency in both the treatment and control groups. Although it was not until the end of second grade that performance levels on measures of oral language improved modestly over the course of 2 years in both the treatment (gaining on average 14 standard score points) and control groups (gaining on average 11 standard score points), on average, students in both groups remained more than 1 standard deviation below the test norms of 100 (1.23 standard deviations for treatment and 1.65 standard deviations for control group), indicating that the parity of these students at the expected levels of English among monolingual English speakers with regard to English oral proficiency was not yet evident. Such findings confirm the literature regarding the number of years required before ELLs can fully master academic language in English (August & Hakuta, 1997; Collier, 1987; Cummins, 1984; Genessee & Riches, 2006).

Spanish Oral Proficiency and Reading

For Spanish measures, on average, treatment students outperformed control peers in letter naming, phonological processing, listening comprehension and reading comprehension, with effect sizes ranging between

.19 and .38. The performance levels of both groups were higher than for English measures at school entry, which can be attributed to their home language background. On measures of oral language, no difference was found between the two groups except for the Listening Comprehension subtest. Posttest performance in second grade for both groups approximated the average norm value by less than 0.4 standard deviations. Control students demonstrated higher levels in Spanish decoding skills than treatment students. Such results might be misleading, because both groups had already demonstrated outstanding decoding skills, with nearly 2 standard deviations above the expected value at the beginning of first grade and almost 3 standard deviations above the norm value at the end of second grade, indicating that on average, students had fully developed the foundational skills necessary for successful subsequent reading in their native language. No difference was found between two groups of students in reading fluency in Spanish.

The findings of this study are consistent with the preliminary results of 2 years of implementation (i.e., kindergarten and first grade) of the same project, in that students receiving treatment in TBE demonstrated a faster rate in oral English acquisition than their peers in typical practice TBE (Tong, Lara-Alecio, Irby, Mathes, & Kwok, in press). The findings are also congruent with studies that support DBE as an additive model maintaining L1 and promoting L2 (de Jong, 2004; Medina & Escamilla, 1992a). Little is known about the effect of literacy instruction in one language on the literacy development in the other language for ELLs (Vaughn et al., 2006). In our study, gains in English literacy acquisition appeared to influence gains in Spanish literacy, particularly for students in treatment DBE models, compared with students in control DBE models. Significant gains were observed in Spanish rapid naming, phonological processing, and reading comprehension in DBE students. One might argue for the decreased Spanish instructional time in the adapted 70/30 DBE model that reduced students' exposure to their native language; however, well-established bilingual theories and researchers have demonstrated the reciprocal language learning process, that the knowledge of L1 assists L2 acquisition, and this L2 acquisition process facilitates subsequent L1 development (August, 2003; Cummins, 1989, McLaughlin, 1985; Snow, 1992; Vaughn et al., 2006). This cross-linguistic transfer explains greater skills in Spanish on tests of early literacy among treatment students in our study. With the enhanced curriculum and teacher and paraprofessional training, students in the 70/30 DBE model made substantial progress over their 80/20 TBE typical practice counterparts on both Spanish and English oracy and comprehension skills, indicating that they benefited from this longitudinal English intervention.

Limitations and Conclusions

There are two limitations of this study. First, we were not able to study the least stable part of the student population because of a high mobility rate (nearly 35% every year). This is an inherent limitation embedded in most longitudinal quasi-experimental and experimental research of this kind. Second, we were unable to control for Spanish instruction in both groups, and it is unclear if students' English language development was a result of Spanish instruction, given the extensive literature on cross-linguistic transfer between Spanish and English. Even so, it is evident that the decrease of Spanish instruction and increase in English instruction did not prevent students' development in Spanish; instead, quality L2 instruction positively affected L1 outcomes. On the other hand, students in both treatment and control classrooms receiving Spanish instruction achieved meaningful gains in both English and Spanish, with their performance approximating or above the average, which corroborates with the converging evidence that L1 instruction is positively associated with L2 reading outcome (Francis, Lesaux, & August, 2006; Rolstad et al., 2005; Slavin & Cheung, 2005).

Even though existing literature has supported the positive learning attainment when students were placed in 90/10 and 50/50 models, this 70/30 DBE model has also proved to be effective in promoting Spanish-speaking ELLs' bilingualism and biliteracy. As Collier and Thomas (2004) proposed, "a next major step for researchers to take is to produce the next generation of bilingual education researchers who will conduct program evaluation research, to refine what particular forms of dual language programs are most effective" (p. 18). Future research investigating the best time allocation in language of instruction is recommended.

Our results lead us to conclude that the ELL students who were provided a structured transitional bilingual program, specifically DBE, with monitored, quality instruction and an aligned English-as-L2 curriculum including targeted L2-to-L1 clarifications, along with typical practice Spanish instruction and aligned curricula, responded favorably for English oral and beginning reading skills, including comprehension. It is important to note that these students were not proficient language speakers in English (pretest scores averaged more than 2 standard deviations below the mean) and were instructed most of the time in their native language. We found that students achieved substantial gains in English when the following strands were integrated into instruction: phonemic awareness, letter-sound correspondence, word recognition and spelling, fluency, and comprehension. Such findings are supported by published research (e.g., Ehri et al., 2001). In addition, the

intervention facilitated students' learning by incorporating structured and direct instruction, ESL strategies, and context-embedded vocabulary learning. It is also suggested that students achieved greater gains when students were engaged in story-retelling activities that deliberately aimed to develop higher order thinking skills. Strategies reflected in our intervention are supported by research that has determined that structured and phonetic reading programs emphasizing language development are most successful for beginning readers (Cheung & Slavin, 2005) and that direct vocabulary instruction is effective for Hispanic ELLs (Carlo et al., 2004). The ongoing professional development was also critical to ensure consistency between objectives and implementation of this intervention. One issue that remains unanswered is the value of each intervention component, because each was delivered as an integral package. On one end, we claim that it is the combined components of the intervention that have promoted students' English and Spanish language and reading skills in treatment DBE model at early grade levels. On the other end, the disentanglement of these components is possible, and specifically, we, along with Quiros (2008), identified the positive effect of the story retelling activity (STELLA) on oral and reading comprehension in English.

We have finalized the intervention at the third grade level, which emphasized content reading in science. This component was based on the foundation of the previous three years' implementation. The entire longitudinal study with the integrated components over the 4 years should provide us with a better understanding of the appropriate instruction in language and literacy development that facilitates English content area learning for Hispanic ELLs. Future studies on the effectiveness of DBE serving ELLs for long-term academic, cognitive, and social attainment beyond elementary grades are much needed.

References

- August, D. (2003). *Supporting the development of English literacy in English language learners: Key issues and promising practices* (Report No. 61). Baltimore: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.
- August, D., & Hakuta, K. (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Research Council.
- August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum.
- Becker, W. C. (1971). *Teaching concepts and operations; or how to make kids smart*. In W. C. Becker, S. Engelmann, & D. Thomas (Eds.), *Teaching: A course in applied psychology* (pp. 401-423). Chicago: SRA Associates.

- Branum-Martin, L., Metha, P., Fletcher, J. M., Carlson, C. D., Ortiz, A., Mario, C., et al. (2006). Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking kindergartners in transitional bilingual education classrooms. *Journal of Educational Psychology, 98*(1), 170-181.
- Carlo, M., August, D., McLaughlin, B., Snow, C., Dressler, C., Lippman, D., et al. (2004). Closing the gap: Addressing the vocabulary needs of English language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39*(2), 188-215.
- Cheung, A., & Slavin, R. E. (2005). Effective reading programs for English language learners and other language-minority students. *Bilingual Research Journal, 29*(2), 241-267.
- Christian, D. (1994). *Two-way bilingual education: Students learning through two languages* (Educational Practice Report No. 12). Washington, DC: National Center for Cultural Diversity and Second Language Learning.
- Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly, 21*(4), 617-640.
- Collier, V. P., & Thomas, W. P. (2004). The astounding effectiveness of dual language education for all. *NABE Journal of Research and Practices, 2*(1), 1-20.
- Cummins, J. (1984). Wanted: A theoretical framework of relating language proficiency to academic achievement among bilingual students. In C. Rivera (Ed.), *Language proficiency and academic achievement* (pp. 2-19). Clevedon, UK: Multilingual Matters.
- Cummins, J. (1989). *Empowering minority students*. Sacramento: California Association for Bilingual Education.
- de Jong, E. (2004). Second language (L2) proficiency development in a two-way and developmental bilingual program. *NABE Journal of Research and Practice, 2*(1), 57-77.
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*(3), 250-287.
- Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities: Research & Practice, 16*(4), 203-212.
- Francis, D., Lesaux, N., & August, D. (2006). Language of instruction. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth* (pp. 365-413). Mahwah, NJ: Lawrence Erlbaum.
- Genesee, F. (Ed.). (1999). *Program alternatives for linguistically diverse students* (Educational Practice Report No. 1). Santa Cruz, CA: Center for Research on Education, Diversity, & Excellence.
- Genesee, F., & Riches, C. (2006). Literacy: Instructional issues. In F. Genesee, K. Lindholm-Leary, W. M. Saunders, & D. Christian (Eds.), *Educating English language learners: A synthesis of research evidence* (pp. 109-175). New York: Cambridge University Press.
- Goldenberg, C. (1996). Commentary: The education of language-minority students: Where are we, and where do we need to go? *Elementary School Journal, 96*, 353-361.
- Good, R. H., Bank, N., & Watson, J. M. (Eds.). (2003). *Indicadores Dinámicos del Exito en la Lectura*. Eugene, OR: Institute for the Development of Educational Achievement.
- Good, R. H., & Kaminiski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.
- Gomez, L., Freeman, D., & Freeman, Y. (2005). Dual language education: A promising 50-50 model. *Bilingual Research Journal, 29*(1), 145-164.

- Irby, B. J., Lara-Alecio, R., Quiros, A. M., Mathes, P. G., & Rodriguez, L. (2004). *English Language and Literacy Acquisition evaluation research program (Project ELLA): Second annual evaluation report*. Washington, DC: Institute for Educational Sciences, U.S. Department of Education.
- Lakeshore Learning Materials. (1997). *Question of the Day*. Carson, CA: Author.
- Lindholm-Leary, K. (2000). *Biliteracy for a global society: An idea book on dual language education*. Retrieved November 22, 2004, from <http://www.ncela.gwu.edu/pubs/ideabook/dual/biliteracy.pdf>
- Mathes, P. G., Torgesen, J. K., Menchetti, J. C., Wahl, M., & Grek, M. K. (2004). *Proactive early intervention in reading*. Columbus, OH: SRA/McGraw-Hill.
- McLaughlin, B. (1985). *Second language acquisition in childhood, Vol. 2: School-age children*. Hillsdale, NJ: Lawrence Erlbaum.
- Medina, M., & Escamilla, K. (1992a). English acquisition by fluent- and limited-Spanish-proficient Mexican-Americans in a 3-year maintenance bilingual program. *Hispanic Journal of Behavioral Sciences, 14*(2), 252-267.
- Medina, M., & Escamilla, K. (1992b). Evaluation of transitional and maintenance bilingual programs. *Urban Education, 27*(3), 263-290.
- Meyer, M., & Fienberg, S. (Eds.). (1992). *Assessing evaluation studies: The case of bilingual education strategies*. Washington, DC: National Academy Press.
- Mora, J. K., Wink, J., & Wink, D. (2001). Dueling models of dual language instruction: A critical review of the literature and program implementation guide. *Bilingual Research Journal, 25*(4) 435-460.
- Quiros, A. (2008). *Structure story reading and retelling related to listening comprehension and vocabulary acquisition among English language learners*. Unpublished doctoral dissertation, Texas A&M University.
- Quiros, A., Irby, B. J., Lara-Alecio, R., & Mathes, P. G. (2003). Communication games. Unpublished intervention for Project ELLA.
- Ramírez, J. D., Yuen, S. D., & Ramey, D. R. (1991). *Final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children, Vol. 1* (Report No. 300-87-0156). San Mateo, CA: Aguirre International.
- Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy, 19*(4), 572-594.
- Slavin, R., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research, 75*(2), 247-284.
- Snow, C. E. (1992). Perspectives on second-language development: Implications for bilingual education. *Educational Researcher, 21*(2), 16-19.
- Texas Education Agency. (2005). *Dual language immersion programs: Two-way/one-way developmental bilingual programs for Texas*. Austin: Author.
- Texas Education Code, 74th Leg., Ch. 260, § 1 (1995, May 30).
- Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Retrieved October 20, 2004, from http://www.crede.ucsc.edu/research/llaa/1.1_final.html
- Tong, F., Lara-Alecio, R., Irby, B. J., Mathes, P., & Kwok, O. (in press). Accelerating early academic oral English development in transitional bilingual and structured English immersion programs. *American Educational Research Journal*.
- Vaughn, S., Mathes, P. G., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., et al. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *Elementary School Journal, 107*(2), 153-180.

- Ventriglia, L., & González, L. (2000). *Santillana Intensive English*. Miami, FL: Santillana USA.
- Wagner, R. K., Torgesen, J., & Rashotte, C. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: PRO-ED.
- Woodcock, R. W. (1991). *Woodcock Language Proficiency Battery-Revised, English and Spanish Forms: Examiner's manual*. Itasca, IL: Riverside.
- Woodcock, R. W., & Muñoz-Sandoval, A. F. (1995). *Woodcock Language Proficiency Battery-Revised, Spanish Form: Supplemental manual*. Itasca, IL: Riverside.

Fuhui Tong is an assistant professor in the Department of Educational Psychology at Texas A&M University. She received her PhD in educational psychology from Texas A&M University and a master's degree in applied linguistics from Shanghai Jiao Tong University in China. Her research interests include second language and literacy development and assessment for students in bilingual and ESL settings. Her current research focuses on program and teacher factors on language minority students' academic success and quantitative methodology in bilingual and ESL education, and her teaching emphasizes advocacy for diversity and student-centered learning. She has also been involved in several federal and state projects that seek to reduce educational inequality for language minority students. She enjoys music and fashion.

Beverly J. Irby is a professor and chair of the Department of Educational Leadership and Counseling at Sam Houston State University. She received her MEd and EdD from the University of Mississippi. Her current research involves issues regarding social justice activism, such as curricular development and the administration of bilingual and ESL instruction, early childhood interventions, gifted education, and women's leadership issues. She is the codeveloper of the Synergistic Leadership Theory, the first purposefully gender inclusive leadership theory. Additionally, she is a cofounder and coeditor of *Advancing Women in Leadership Journal*, the first international online, refereed journal advocating women's leadership issues in all disciplines. Her recreational activities are writing and watching documentaries and meaningful movies.

Rafael Lara-Alecio is a professor and director of the bilingual programs in the Department of Educational Psychology at Texas A&M University. He received his PhD in educational psychology from the University of Utah and his master's degree in measurement, evaluation and research in education from University del Valle in Guatemala. His current research involves language development and evaluation issues surrounding bilingual students. His recreational activities involve spending time on his ranch in Texas and raising Red Brangus cattle.

Patricia G. Mathes is the Texas Instruments Chair of Reading, a professor of literacy and language learning, and director of the Institute for Reading Research at Southern Methodist University. She received her PhD in education and human development from the Peabody College for Teachers of Vanderbilt University. Her research focuses on intervening with struggling readers, including students who are English-language learners, are being raised in poverty, or have disabilities. She is the author of numerous articles, chapters, books, and curricular materials related to learning and reading disabilities and that address issues of academic diversity and best practices for struggling readers. She likes to spend time with her children.