Early education of dual language learners: An efficacy study of the Nuestros Niños School Readiness professional development program

Cristina Gillanders, School of Education and Human Development at the University of Colorado Denver

Available at: https://works.bepress.com/cristina-gillanders/16/
Early education of dual language learners: An efficacy study of the Nuestros Niños School Readiness professional development program

Dina C. Castro a,†, Cristina Gillanders a,†, Ximena Franco a,†, Donna M. Bryant a,†, Marlene Zepeda b,†, Michael T. Willoughby a,†, Lucía I. Méndez a,†

a University of North Carolina at Chapel Hill, United States
b California State University, Los Angeles, United States

A R T I C L E   I N F O
Article history:
Received 30 May 2015
Received in revised form 27 January 2017
Accepted 9 March 2017

Keywords:
Dual language learners
Bilingual children
Professional development
Preschool

A B S T R A C T
The purpose of this experimental study was to assess the efficacy of the Nuestros Niños School Readiness (NNSR) Professional Development Program, a 2-year program that includes an integrative approach to teacher professional development (PD) and a research-based, systematic intervention component aimed to promote language, literacy, and social-emotional development, and mathematics learning in pre-kindergarten Spanish-English dual language learners (DLLs). Across cohorts and experimental conditions, 56 preschool teachers and 340 Spanish-English DLLs from early childhood programs in California, Florida, and North Carolina participated in the study. Results indicate that the NNSR program had positive effects on the overall quality of early childhood classroom practices and on practices specifically focused on DLLs. Positive results were also found for children’s outcomes. DLLs in treatment classrooms showed greater gains in expressive vocabulary in English than DLLs in control classrooms, and, when assessed in Spanish, gains were higher in receptive vocabulary, alphabet knowledge, writing and early mathematics. Issues of implementation fidelity and implications for using both languages of DLL children in instruction and assessment are discussed.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

The number of families in the U.S. who speak a language other than English has been steadily increasing during the past two decades. Nationally, 26% of people ages 5 and older spoke a language other than English in 2011, 62% of them spoke Spanish (U.S. Census Bureau, 2013). The proportion of young children aged birth to five years who are considered dual language learners (DLLs) is even higher. We define DLLs as children under five who are learning English as their second language (L2) while continuing to learn their first language (L1) (Office of Head Start, 2008). Over a third of the children enrolled in Head Start programs nationwide are DLLs, and 84% of these are from Spanish-speaking homes. Among children enrolled in Early Head Start, 26% are DLLs and 91% of these speak Spanish as their first or second language (Office of Head Start, 2008).

These demographic trends are important for several reasons. The proportion of the population entering early education programs speaking a language other than English is increasing, and is projected to grow at even greater rates. Second, the educational success of these children is critical to the overall progress of our educational system. Third, increased risk factors associated with poorer school performance have been identified within the DLL population, particularly for DLLs from low socioeconomic status homes (Hernández, Denton, & Macartney, 2008). Finally, targeted, effective and well implemented educational approaches during the early years of schooling have the potential to improve the academic achievement of a large and growing group of diverse learners, including DLLs. This paper reports the results of one such study.

† Corresponding author. Current address: University of North Texas, 1155 Union Circle #310740, Denton, TX 76203-5017, United States.
E-mail addresses:
dina.castro@unt.edu (D.C. Castro), cristina.gillanders@ucdenver.edu (C. Gillanders), ximena.franco@unc.edu (X. Franco), donna.bryant@unc.edu (D.M. Bryant), mzp@calstatela.edu (M. Zepeda), mw@rti.org (M.T. Willoughby), mendezl@ecu.edu (L.I. Méndez).

1 The author is now affiliated with the University of Colorado, Denver, 1380 Lawrence Street, Denver, CO 80217, United States.
2 The author is now affiliated with the University of North Carolina at Chapel Hill, CB#8180, Chapel Hill, NC 27599-8180, United States.
3 The author is now affiliated with the California State University, Los Angeles, 5151 State University Drive, Los Angeles, CA 90033, United States.
4 The author is now affiliated with the RTI International, 3040 East Cornwallis Rd., Research Triangle Park, NC 27709-2194, United States.
5 The author is now affiliated with the East Carolina University, 3310 Health Sciences Building, Mail Stop 668, 1001 E 5th St, Greenville, NC 27858, United States.

http://dx.doi.org/10.1016/j.ecresq.2017.03.002
0885-2006/© 2017 Elsevier Inc. All rights reserved.
Although a range of instructional strategies are currently used to support young DLLs, there is limited research that systematically examines how to provide effective professional development (PD) programs to prepare teachers of DLLs in the implementation of these strategies (Buyse, Peisner-Feinberg, Pérez, Hammer, & Knowles, 2014). One such study focused on improving teaching practices to support Spanish-English DLLs’ language and literacy development is the Nuestros Niños program (Buyse, Castro, & Peisner-Feinberg, 2010; Castro, Peisner-Feinberg, Buyse, & Gillanders, 2010). This was a one-year PD intervention program evaluated in a randomized controlled trial, whose findings revealed significant gains in the quality of language and literacy classroom practices at the end of the intervention. Benefits from the intervention were also observed in children’s phonological awareness skills in their primary language, Spanish, but not in a number of other child outcomes domains. These limited findings suggest that the Nuestros Niños PD intervention was more effective in improving teaching practices than in enhancing children’s outcomes across learning domains. This may be due to the need for additional time before changes in teaching practices affected child learning outcomes. Based on these findings, we hypothesized that extending the length of the PD intervention from one to two years, and expanding the content of the PD intervention to include other school readiness domains such as mathematics learning and social-emotional development, in addition to language and literacy, may result in both improved classroom outcomes and greater gains on various child development outcomes.

The study reported in this paper extends the previous work described above by presenting findings from an experimental study evaluating the efficacy of the new Nuestros Niños School Readiness PD Program (NNSR), a research-based, systematic intervention aimed to promote language, literacy, mathematics learning and social-emotional development in Spanish-English DLLs. This program includes extensive content and an integrative approach to teacher PD encompassing institutes, consultation and professional learning communities delivered over two school years.

We first review the research evidence and theory that informed the design of the NNSR professional development program and practices to promote development and learning among young DLLs. We then describe the NNSR program, followed by results and a discussion of the implications of the research findings for the quality of practices and policy decisions regarding the education and care of young DLLs.

1.1. Developmental characteristics of young dual language learners

To be effective in promoting learning, instructional practices should be tailored to address children’s developmental characteristics and educational needs (Copple & Bredekamp, 2009). Thus, when designing classroom instruction for DLLs, an important consideration should be the diversity of experiences and unique characteristics of these children’s development, which differ in several ways from those of their monolingual peers. Recent neuroimaging research has shown that bilingualism, particularly in simultaneous bilinguals (children who are learning two languages before the age of three), changes the organization and functioning of the brain (Conboy, 2013; Conboy & Kuht, 2011). This helps explain differences found between bilinguals and monolinguals across several developmental domains. The most current research on the cognitive, language, literacy and social-emotional development of DLLs provides relevant information on how bilingualism can influence children’s development, in both simultaneous and successive/sequential bilinguals (children who begin to acquire L2 approximately after the age of three) (Barac, Bialystok, Castro, & Sanchez, 2014; Halle et al., 2014; Hammer et al., 2014). For instance, variation in the rate and patterns of language development have been found between monolinguals and successive/sequential bilinguals. This variation is often associated with the age of onset of L2 acquisition, and the amount and quality of exposure to each language (Hammer et al., 2014; Paradis, Genesee, & Crago, 2011). Children growing up in bilingual environments develop vocabularies for each of their two languages that can overlap or be unique to a particular context or experience. For example, sequential bilinguals in the U.S. tend to learn words associated with the academic experience in their L2 in the early education setting, and words associated with everyday life in their first language at home. Because bilinguals’ opportunities to learn words are distributed between their two languages, it is not unusual for them to have smaller vocabularies in each of their languages as compared to their monolingual peers (Oller, Pearson, & Cobo-Lewis, 2007). Yet their combined conceptual vocabulary size (the number of concepts they have words/labels for) across languages may be similar to that of monolingual children (Bialystok, Luk, Peets, & Yang, 2009; Hoff et al., 2012). Children’s language and cognitive development are closely related. Bilingualism has been found to be associated with a number of cognitive advantages, in particular in the areas of executive function, and the advantages are larger in children with advanced skills in their two languages (Barac et al., 2014).

Furthermore, having access to linguistic skills in two different languages may positively impact the overall development of bilingual children (Cummins, 2005; MacSwan & Rostad, 2005). Research evidence suggests that shared linguistic knowledge from the first and second language is available to bilingual children to support certain aspects of language and cognitive/academic skills across languages (Cummins, 2005; Dunn Davison, Hammer, & Lawrence, 2011; Hammer, Lawrence, & Miccio, 2007; Quiroga, Lemos-Britton, Mostafapour, Abbott, & Berninger, 2002). Based on the accessibility of shared linguistic information, sequential bilinguals in the early stages of L2 vocabulary acquisition may use conceptual knowledge available to them in L1 to facilitate the learning of new L2 words (Goodrich, Lonigan, & Farver, 2013; Lugo-Neris, Wood Jackson, & Goldstein, 2010; Méndez, Crais, Castro, & Kainz, 2015). Similarly, positive cross-language correlations have been reported using measures of syntactic complexity and mean length of utterance across languages. Several studies also report a stronger, solid cross-language relation between phonological awareness skills in DLLs (Hammer et al., 2014).

Further evidence supports the notion that shared language skills could also potentially support literacy development. Several studies in this area suggest that language skills in L1 support literacy in L2 (August & Shanahan, 2006; Cummins, 2005; Dunn Davison et al., 2011; Quiroga et al., 2002). Initial levels of emergent writing also seem to develop in parallel in both languages (Gillanders, Franco, Seidel, Castro, & Méndez, 2016).

In the area of social-emotional development, studies have shown that DLLs tend to obtain higher scores on measures of self-control and interpersonal skills, and fewer internalizing and externalizing behaviors than their monolingual English-speaking peers (Halle et al., 2014; Han, 2010; Luchtel, Hughes, Luze, Bruna, & Peterson, 2010). However, when compared to their own peer language group, lower rates of English acquisition may lead to lower levels of social-emotional competence in early elementary school. In a reanalysis of the Early Childhood Longitudinal Study data, it was found that Asian and Latino kindergarten children who remained categorized as limited English proficient into first grade were rated by teachers as lower on approaches to learning, self-control and interpersonal skills when compared to peers who had attained English proficiency (Kang, Haddad, Chen, & Greenberger, 2014). Conversely, young Spanish-speaking 4-year-olds who demonstrate greater initiative, self-control, secure attachment and fewer behavior problems show higher English language attainment at the end of
kindergarten (Winsler, Kim, & Richards, 2014). An important moderating effect in the social-emotional development of young DLLs is the use of the child’s primary language in the classroom. For example, one study found that the use of Spanish by teaching staff was positively associated with teacher reported ratings of frustration tolerance, assertiveness, task orientation, and peer social skills among Spanish–English DLLs (Chang et al., 2007).

1.2. Instructional strategies to promote DLLs’ school readiness

Considering the impact and benefits of bilingualism on DLLs’ development described above, programs to improve DLLs’ school readiness outcomes should focus on the quality of their educational experiences. Although some research suggests that high-quality early education (as designed for monolingual children) promotes DLLs’ development and school readiness (Buyse et al., 2014), other research suggests that those practices may not be sufficient to support the same level of performance in DLLs as in their monolingual peers (August & Shanahan, 2006).

High-quality early education for these children involves attending to their particular developmental characteristics and using bilingualism as a resource for learning (Castro, Espinosa, & Páez, 2011). Research evidence indicates that instruction in children’s first language leads to higher social, cognitive, and academic achievement levels for bilingual learners (Barnett, Yarosz, Thomas, Jung, & Blanco, 2007; Burchinal, Field, López, Howes, & Pianta, 2012; García, Kleifgen, & Falchi, 2008; Genesee & Lindholm-Leary, 2012). The manner in which the first language is used in instruction, however, will vary depending on children’s proficiencies in their two languages, the language of instruction approaches (i.e., English-only, bilingual transitional, two-way instruction), and the availability of bilingual staff. For monolingual English-speaking teachers accustomed to providing instruction only in English, the prospect of supporting DLLs in their L1 may initially seem daunting. However, a number of instructional strategies to support DLLs have been identified that can be used by both monolingual English and bilingual speakers.

Drawing from research, the NNSR Program utilized five instructional strategies that formed the foundation of its approach. Each is summarized below.

1. Ongoing and frequent assessments. Close monitoring of DLLs’ language development helps inform instructional planning so that practices are tailored to children’s levels of proficiency and target specific areas in which they may need additional support (Lesaux & Siegel, 2003). Monitoring progress in other developmental and academic domains is also important. A limited number of valid and reliable assessments normed on DLLs exist. Thus, the careful selection of assessment tools and their application is important to avoid misinterpretations, especially when standardized instruments are used. Recommendations to assess DLLs include using a combination of standardized measures and systematic observational methods and portfolios of children’s work to obtain the most accurate information about DLLs’ academic performance and development. Also, gathering information across settings and types of interactions (i.e., teacher–child, peers) will be informative (Espinosa & Gutierrez-Clellen, 2013; National Association for the Education of Young Children (NAEYC), 2009).

2. Focused small-group activities. DLLs need opportunities for additional exposure to and use of new concepts and words in their second language. Randomized controlled trials (RCTs) of reading interventions for struggling DLLs in grades K–5 have indicated that small-group and peer-assisted interventions allow children multiple opportunities to respond to questions, practice reading skills, and receive explicit instruction on vocabulary and phonological awareness (Gunn, Smolkowski, Biglan, & Black, 2002; McMaster, Kung, Han, & Cao, 2008; Vaughn et al., 2006).

Based on this research evidence, small-group activities with no more than four or five children and planned in conjunction with classroom-wide activities have been proposed to be included in interventions with preschool-aged DLL children (Castro, Gillanders, Franco, & Zepeda, 2010).

3. Explicit vocabulary instruction. For monolingual children, vocabulary learning is largely incidental, accomplished through conversations and by listening to words in their everyday routines (Childers & Tomasello, 2002). Children who are learning a second language, however, may not benefit equally from this strategy because their vocabulary learning is distributed across two linguistic contexts (Hoff et al., 2012; Oller et al., 2007). Consequently, differences in the quality and quantity of word exposure available in each of the two languages may significantly reduce opportunities for word learning (Bohman, Bedore, Peña, Mendez-Peres, & Gillam, 2010). Given these different conditions for vocabulary learning, teachers need to create effective and efficient word-learning situations for DLL children. This includes purposeful planning of opportunities for repeated word exposure and use in a variety of settings (Gillanders, Castro, & Franco, 2014). Explicit instruction can also be used to accelerate vocabulary learning for DLLs and can be delivered through read-alouds. Direct teaching of core vocabulary including a combination of tier one words (frequently used common words that may be unfamiliar to DLLs, such as bags, trousers, fences) and tier two words (general academic and multiple meaning words that are important and often used across content areas of the curriculum such as compare, measure, etc.) is also important to support language and literacy development (Beck, McKeown, & Kucan, 2002). Additionally, the strategic use of the primary language to scaffold the learning of new words across the curriculum can also contribute to DLLs’ overall school readiness (Carlo et al., 2004; Gillanders & Castro, 2011; Méndez et al., 2015).

4. Development of academic English language. To be successful in school, DLLs need to develop the academic discourse that can effectively support their learning of content knowledge (August, Carlo, Dressler, & Snow, 2005). Lack of proficiency in English academic discourse – which differs from conversational English language skills used in social interaction – can interfere with the learning of academic content (Cummins, 2008). The learning of mathematics provides an example of a content area that employs academic language as it uses specific words in a particular context. Although children might learn mathematical concepts and skills using manipulatives, they also need to learn academic vocabulary related to mathematics in order to understand math-related concepts, follow mathematics instruction and experience overall success in school (Sarama, Lange, Clements, & Wolfe, 2012). Examples of mathematics academic language include the names and parts of shapes, words to describe patterns, and terms to describe quantities (e.g., more and less). Therefore, curricula should incorporate opportunities that provide explicit instruction and support the development of academic language to facilitate the learning of early mathematics concepts and other content areas (Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006).

5. Strong focus on social-emotional development. Positive teacher–child relationships and facilitation of DLLs’ participation with their classroom peers can increase DLLs’ adaptation and learning (Castro, Gillanders, et al., 2010, Castro, Peisner-Feinberg, et al., 2010; Gillanders, 2007; Howes & Ritchie, 2002). The preschool environment may be the first time that a young DLL experiences a different cultural environment. For children receiving instruction only in English, there will be the additional challenges of having difficulties communicating, following
directions, expressing ideas and feelings, and responding to questions. DLLs may feel withdrawn, insecure, and under stress (Dawson & Williams, 2008; Santos & Ostrosky, 2002). Limited attention has been given to this aspect of DLLs’ development (Halle et al., 2014), but research with monolingual children indicates that children who feel rejected by their peers in their early years face higher risk of academic and delinquency problems in adolescence (Jimerson, Egeland, Strofe, & Carlson, 2000).

In addition to adequate policies and resources, the implementation of these instructional practices requires the provision of ongoing professional development for teaching staff (Castro et al., 2011). Teachers – bilingual and non-bilingual – need to be knowledgeable about how development unfolds in DLL children and about effective instructional practices to support development and learning among DLLs (Zepeda, Castro, & Cronin, 2011; Zepeda, 2014). Finally, the curriculum needs to support both languages of DLLs, with opportunities to incorporate instructional enhancements targeting DLL children in an intentional and systematic manner.

1.3. Professional development in early childhood

As noted above, progress has been made in identifying classroom strategies that can promote DLLs’ development and learning. However, these practices can only be effective if teachers are willing and able to implement them in the classroom. Research on professional development in early childhood classrooms has pinpointed several features that seem to be effective for promoting changes in teachers’ knowledge, skills and dispositions, and consequently in children’s outcomes (Desimone, 2009; Fukuik & Lont, 2007; Sheridan, Edwards, Marvin, & Knoche, 2009; Zaslouf, Tout, Halle, Whittaker, & Lavelle, 2010). Among these are: (a) specialized content with an emphasis on summaries of existing research on children’s development in the specific area and instructional strategies; (b) links between theory and practice with opportunities for practice in the job setting, and immediate feedback after the rehearsal of skills; (c) collective participation and structures for staff collaboration; (d) multidimensional methods of delivery; (e) opportunities for learning to conduct child assessments for monitoring in the specialized content area, and (f) intensity and duration that are appropriate to the complexity of the targeted skills.

A range of approaches has been used to deliver these strategies to classroom teachers, including specialized trainings, coaching/consultation, and communities of practice (also called professional learning communities) (Sheridan et al., 2009), but no single strategy has been determined most effective. Specialized trainings often take the form of short-term workshops that provide participants with general information about particular topics. In general, these trainings are less effective if they are not followed up by opportunities for feedback and practice (Schachter, 2015; Sheridan et al., 2009). Consultation and coaching have been used for developing knowledge, changing dispositions and improving practice (Downer, Locasale-Crouch, Hamre, & Pianta, 2009; Gupta & Daniels, 2012; Isbell et al., 2013). Although it is not yet clear what critical features can make consultation or coaching more effective (Schachter, 2015), several authors agree that it should be a collaborative problem-solving process in which the consultant/coach has frequent contact with the teacher to share information, encourage, model, assess, provide feedback on the implementation of classroom practices and promote reflection (Buyss & Wesley, 2005).

Finally, communities of practice, or professional learning communities, have been defined as groups of teachers who collaboratively examine issues of practice with the goal of improving children’s learning (Vescio, Ross, & Adams, 2008). Key characteristics are shared values and norms, consistent focus of children’s learning and reflective dialog and collaboration (Vescio et al., 2008). A particular form of professional learning community is the lesson study approach (Lewis, 2002), which is modeled after the Japanese form of professional development. A distinct feature of the lesson study is “observation of live classroom lessons by a group of teachers who collect data on teaching and learning and collaboratively analyze it” (Lewis, Perry, & Murata, 2006, p. 3).

An aspect common both to consultation and communities of practice is that teachers become active participants in the problem-solving process of implementing strategies. Rather than following a prescriptive curriculum with little understanding of the pedagogical principles underlying the practices, through a collaborative process, teachers become both co-constructors and implementers of the curriculum. Indirect evidence of the value of this approach comes from Justice, Mashburn, Hamre, and Pianta (2008) who found that although teachers could implement a prescribed language and literacy curriculum with high levels of procedural fidelity, only a few teachers provided quality instruction. It is impossible to cover all the possibilities that might arise in a classroom in a prescribed curriculum. Therefore, only if teachers can truly understand the underlying principles and purposes of a set of given strategies, can they adjust them to the diversity of contexts and children they will encounter while teaching.

The NNSR PD Program reported here used all three professional development approaches to help teachers with specific content and practices that should facilitate the development of DLLs. In the context of previous early childhood professional development research, three aspects of the study are worth noting. Because a previous study of the Nuestros Niños Program (Buyss et al., 2010), delivered through one year of teacher PD, found changes in teacher practices but few changes in children’s outcomes, the present study extended the PD for two years. Also, considering that the majority of professional development interventions have focused on only language and literacy (Schachter, 2015), the current intervention aims to address the areas of language, literacy, mathematics, and social-emotional development. In a review of 73 PD programs in early childhood education targeting monolingual populations, 54% of studies addressed language and literacy, 27% social-emotional development, 5% studied mathematics instruction, and 10% general instruction (Schachter, 2015). Given that early childhood teachers are expected to promote all areas of development, we are interested in determining if it is possible for a PD program to facilitate changes in all these areas.

Finally, a limited number of published studies have examined the impact of PD that addresses the improvement of teachers’ skills in working with children from linguistically and culturally diverse backgrounds (e.g. Pollard-Durodola et al., 2012; Silverman, 2007). The present study aims to address this gap in research. Silverman (2007) developed a reading aloud curriculum for promoting vocabulary development in both monolingual English-speaking kindergarteners and Spanish-English bilinguals. Teachers received a 5-week training in the curriculum that was to be implemented during a 14-week period. The curriculum was implemented only in English, so it is unknown if the intervention could have had an effect in the DLLs’ Spanish vocabulary development. At the end of the intervention, Spanish-English bilinguals learned the target words at a faster rate than the English monolinguals. Pollard-Durodola et al. (2012) conducted a qualitative study examining the Spanish shared book reading practices of seven preschool teachers of DLLs before and after receiving professional development supports in the form of instructional prompts encouraging teachers to provide an ideal vocabulary lesson and the implementation of a prescribed Spanish shared book reading intervention. Results indicated that after a 15-week curriculum implementation and professional development of one half-day and two 1-h sessions, five of the seven teachers
who initially showed low levels of teaching quality improved their ratings in general teaching, vocabulary of instruction, and interactive teaching practices. Interestingly, teachers with initial lower levels of quality reported that the provided lesson materials did not increase their effectiveness as teachers while those with initial higher levels of quality indicated that the materials improved their teaching. As discussed, although these studies examined the impact of PD on teacher practices working with DLLs, they focused only on language and literacy.

1.4. The present study

Based on research summarized above, more clarity is needed in how interventions targeting DLLs can generate effective school readiness outcomes (Barnett et al., 2007; Goldenberg, Nemeth, Hicks, Zepeda, & Cardona, 2013). Using an experimental design, this study examines the efficacy of the NNSR PD Program, designed specifically to address the needs of DLL children, on teachers’ classroom practices and on language, literacy, mathematics and social-emotional outcomes for Spanish-English DLL preschoolers. We hypothesize that:

General classroom quality and the quality of practices specifically targeting DLLs will be higher in classrooms of teachers participating in the NNSR program (treatment group) (Buysse et al., 2010; Buysse et al., 2014).

DLLs’ language, literacy, mathematics and social-emotional development outcomes will be higher among DLLs in classrooms of teachers participating in the NNSR program (treatment group) (Carlo et al., 2004; Dawson & Williams, 2008; Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006; Gillanders & Castro, 2011; Méndez et al., 2015).

In addition to the main effect of the NNSR program treatment on classroom and child level outcomes, the study also investigates whether classroom and child-level outcomes differ depending on whether a teacher was in her first or second year of participation in the NNSR (Desimone, Porter, Garet, Suk Yoon, & Birman, 2002), and also whether child outcomes differ depending on whether both teacher and assistant teacher were Spanish-English bilinguals or only one of them was bilingual (Barnett et al., 2007; Hammer et al., 2014).

2. Method

2.1. Participants

This study involved 340 Spanish-English DLL preschoolers who were nested in 56 classrooms from 33 preschool centers (mean = 1.8, mode = 1.0 classrooms/center). Participants were from early childhood programs in California (CA), Florida (FL), and North Carolina (NC). In NC and FL, programs had to have demonstrated relatively high quality by having earned at least four stars on the state’s five-star quality rating system and, in CA, by participating in the Los Angeles Universal Preschool five-star quality assessment and improvement program. Classrooms had to enroll at least 25%, but not more than 75%, 4-year-old, Spanish-English DLLs. Lead teachers either needed to possess a bachelor’s degree or an Associate of Arts (2-year) degree while working toward a bachelor’s degree. Although not a requirement of the study, all classrooms happened to use either the Creative Curriculum or High Scope Curriculum as their core curricula. Random assignment to treatment (teachers received the NNSR PD intervention in addition to their curriculum) or control (only their curriculum) conditions was conducted at the program/site level to maximize independence and ensure adequate statistical power. One cohort of programs/teachers in CA and NC began participation in 2010 and a second cohort recruited in NC and FL began participation in 2011.

After teachers agreed to participate, Spanish-English DLL children and families were recruited via personal contacts made at the program. The children were of Mexican (47%), and Central American (53%) descent, and had at least one foreign-born parent. Given the enrollment requirements of the NC More at Four and the FL and CA Head Start programs, participant children were from low-income families. In order to ensure that the effects of the intervention condition would not be confounded by extreme differences in individual development, children with identified disabilities and an Individualized Education Program (IEP) were excluded from participating in this study. Teacher participation in the study was expected to last two years, whereas children were expected to participate for only one year.

Classroom lead teachers assigned to the intervention condition received intervention materials with detailed classroom activities and home-school activities for parents, and educational materials (e.g., children’s books, posters) and $200 as a token of appreciation. Their assistant teachers received $100. In addition, lead teachers and assistant teachers in the intervention condition who completed the professional development requirements were given the opportunity to receive continuing education credits. Control group lead teachers received $100 and the opportunity to participate in a training institute at the end of the study. Parents who returned a signed consent form received a $30 gift certificate after completing a demographic form, and participating in a structured interview in the fall of the year in which their child participated in the study.

2.2. Data collection procedures

Children were assessed near the beginning and end of their pre-kindergarten year to examine changes in their development and learning across the language, literacy, mathematics and social-emotional domains. Data collectors were bilingual in English and Spanish, knowledgeable about early childhood education and assessment practices and trained in standardized data collection measures and procedures by a study investigator. Children were assessed in both languages, with the administration in English and Spanish counterbalanced to control for order effects. Classrooms were observed for two days in the fall and two days in the spring for the two years of the intervention to gather data about classroom practices. Data collectors were blind to group assignments (i.e., control and/or intervention groups). Teacher questionnaires and parent interviews were conducted in the fall of each year to obtain both demographic and attitudinal measures.

2.2.1. Classroom practice

Three observational measures of classroom quality that have been associated with children’s outcomes were used in the fall and spring of each school year to gather data about classroom practices. To ensure high-quality, reliable data, the data collectors were trained as stipulated by the measure developers. After group training that consisted of reviewing and discussing the measures, field training included two practice observation visits and one reliability visit. To minimize drift during the data collection phases, inter-rater reliability visits were conducted by two observers every 10th observation.

2.2.1.1. Quality of teacher–child interactions. The Classroom Assessment Scoring System (CLASS; Pianta, LaParo, & Hamre, 2006) is a system for observing and assessing the qualities of interactions between teachers and students in classrooms. The CLASS measures instructional and social-emotional interactions proven to contribute to students’ academic achievement and social competencies during early childhood years and elementary grades. The CLASS describes three broad areas of classroom quality and specific dimensions of classroom quality within each of these broad areas:
Emotional Support (positive climate, negative climate, teacher sensitivity, and regard for student perspective), Classroom Organization (behavior management, productivity, and instructional learning formats), and Instructional Support (concept development, quality of feedback, language modeling, and literacy focus). The authors report that after receiving training, observers can achieve an inter-rater reliability of 87%. The dimensions of Emotional Support, Classroom Organization, and Instructional Support show alphas of .81 to .89 (LaParo et al., 2009). Criterion validity with the Early Childhood Environment Rating Scale (ECERS) and Snapshot showed moderate to high correlations (.3 to .63). In terms of predictive validity, in an 11-state pre-k study, the CLASS was associated with children's performance at the end of the preschool year (Early et al., 2013). In this study, CLASS observers were certified as reliable when they demonstrated a proficiency of 85% or higher (within one point) inter-rater reliability; during data collection, inter-rater reliability ranged from .77 to .98.

2.2.1.2. Quality of language and literacy practices (specific to DLLs). The Early Language and Literacy Classroom Observation Dual Language Learners (ELLCO-DLL; Castro, 2005) was initially developed as a supplement to the ELLCO (Smith & Dickinson, 2002). It was developed with permission from the authors and validated in a previous Nuestros Niños study (Buyse et al., 2010) that assessed specific language and literacy classroom practices for children who are DLLs. The ELLCO-DLL is used as a stand-alone instrument and includes three subscales: Literacy Environment (10 items with a total possible score of 16) that rates the amount and quality of literacy materials (e.g., books, word cards, puzzles) available in Spanish and English in the classroom; Classroom Observation (8 items, total possible score of 40); and Literacy Activities (8 items, total possible score of 14) that rate the appropriateness of the classroom environment and curriculum for DLLs. The internal reliability of the ELLCO-DLL derived from the initial study data was quite high, ranging from 91.1% (Classroom Observation Scale) to 100% (Literacy Activities Rating scale). Internal consistency was .78 for the Classroom Observation Scale, .57 for the Literacy Environment Checklist, and .30 for the Literacy Activities Rating Scale. For the present study, data collectors were trained to a reliability standard of 90% (within one point), and during data collection, their inter-rater reliability ranged from .78 to 1.0.

2.2.1.3. Quality of language environment (specific to DLLs). The Language Interaction Snapshot (LiS; Sprachman, Caspe, & Atkins-Burnett, 2009) was designed to examine how the language environment differs for different children, particularly in classrooms that include DLLs. With the LiS, an observer focuses on an individual child and the language provided to that child by the lead teacher and other adults in the classroom as well as the language used by the child. The authors report that after receiving training, observers can achieve an inter-rater reliability of 88% to 90%. Initial evidence of validity was obtained by examining relationships with the observations of the same classroom using the CLASS (Pianta et al., 2006). The strongest relationships were found between the LiS variables about requests for language, giving both contextualized and decontextualized information, repeating or confirming child language, and the total talk with the CLASS Quality of Feedback ($r = .64$ to $r = .72$). In an effort to reduce the number of tests and to create variables that exhibit better variation, the codes were collapsed into two quality talk categories, basic and complex. The basic talk variable consists of: (1) Repeats of or confirms, (2) Gives directions, (3) Provides information, labels, names, and (4) Sings. The complex talk variable consists of: (1) Request language contextualized, (2) Decontextualized Talk, and (3) Read. For this study, data collectors were trained to a reliability standard of 90% (within one point), and during data collection their reliability ranged from .72 to .99.

2.2.2. Child outcome measures

Six measures were used to assess children's language development, literacy and mathematics learning, and social-emotional development.

2.2.2.1. Receptive vocabulary. The Receptive One-Word Picture Vocabulary Tests (ROWPVT; Martin & Brownell, 2010) assess receptive vocabulary by asking an individual to match an object, action, or concept with its name. In this study we used the English and the Spanish Bilingual versions. The Spanish Bilingual version is used to assess individuals who are bilingual in Spanish and English. The measure assesses total acquired vocabulary by permitting responses in both languages. It measures an individual's ability to match an object, action, or concept with its name. The bilingual forms are administered in the child's dominant language, though responses can be accepted in either English or Spanish. Norms are based on a sample of 2,327 Spanish-bilingual students between 2–0 and 18–11 years of age, residing in the U.S. Internal consistency of test items showed coefficient alphas ranging from .95 to .98 by age group for all individuals participating in the standardization study. In terms of temporal stability, the corrected test-retest correlation range from .78 to .93.

2.2.2.2. Expressive and receptive vocabulary. Capturing small changes in oral language after short periods of instruction using norm-referenced tools can be challenging since standardized tests are designed to resist short-term changes (Silverman, 2007). To address this concern, we developed two picture vocabulary probes to measure changes in expressive and receptive target vocabulary. The words used were selected from the corpora of words in the instructional storybooks used in the intervention. Only unfamiliar words identified at a rate not greater than chance (25%) by DLLs during a pilot study were included in the probes. To reduce the likelihood of familiarity with the probe pictures, they were not used in instruction. Children's responses were scored as correct or incorrect using a 1 or a 0. We developed a standard training for data collectors and certification procedures were conducted for each probe. Below we describe each probe.

The NNSR Receptive Word Knowledge Probe (RWKP) is a 28-item English receptive picture vocabulary task designed to measure changes in receptive vocabulary knowledge. Each picture test-item is presented in an array of four pictures per test plate. One of the four pictures in each test plate is the target or test-item (e.g., a dog), and the other three pictures are distractors. The placement of the target pictures in the test plates was randomized. For each picture plate, the children are asked to look at the four pictures and point to the one corresponding to the target word spoken, following the general format of the PPVT-4 (e.g., “Show me_____”)

The NNSR Expressive Word Knowledge Probe (EWKP) is a 21-item expressive vocabulary task designed to measure change in expressive word knowledge of words selected from the corpora of the instructional storybooks. Each test-item is presented in a picture in a test plate for a total of 21 test plates. For each test item the child is shown a picture and is asked a question. To elicits labels for nouns the child is asked, “What is this?”. For pictures denoting action verbs the child is asked, “What is she doing?” or “What is he doing?”

2.2.2.3. Alphabet knowledge and expressive vocabulary. Drawing from measures employed in other national studies, we used the Letter-Word Identification (LWI) and the picture vocabulary (PV) subtest of the Woodcock Language Proficiency Battery-Revised (WLPB-R) – English. In addition, the Spanish version of the same
2.2.2.6. Correlations between various clusters (PAT; Tardibuono, 2004; Tardibuono & Hammer, 2004) were administered. The Spanish version consists of one more subscale (i.e., First Syllable Matching).

2.2.2.5. Writing. The Name Writing Task was used to measure children’s emergent name writing in Spanish and English. Following the task designed by Yaden and Tardibuono (2004), children were asked to write their name and to read what they had written. Subsequently, the data collector covered the first part of what the child had written, whether or not it was their name, and asked them to read the uncovered part. Next, the data collector covered the last part and asked again to read the uncovered part. Data collectors wrote children’s responses and scored them according to the Yaden and Tardibuono (2004) scoring scheme, with potential scores ranging from 0 to 6. Data collectors scored the writing tasks and these scores were compared with those of the master coder. Correlations calculated at assessment time point ranged from .68 at time 2 to .74 at time 4.

2.2.2.6. Early mathematics. The Test of Early Mathematics Ability-Third Edition (TEMA-3; Ginsburg & Baroody, 2003, English and Spanish versions) measures the mathematics performance of children between the ages of 3 and 9. It measures both informal and formal concepts and skills in several domains: numbering skills, number-comparison facility, numeral literacy, mastery of number facts, calculation skills, and understanding of concepts. It has two parallel forms each containing 72 items. The standardization sample was composed of 1,219 children. Test results are reported as standard scores, percentile ranks, and age and grade equivalents. Internal consistency reliabilities were all above .92; immediate and delayed alternative form reliabilities were in the .80s and .90s.

2.2.2.7. Social competence and behavior. The Social Competence and Behavior Evaluation Preschool Edition (SCBE; LaFreniere & Dumas, 1996) is a standardized teacher or parent rating instrument designed to assess patterns of social competence, affective expression, and adjustment difficulties. It is intended for children aged 30–78 months and was standardized on a sample of 1,263 children. Each of the 80 items is rated on a 6-point scale (1 = almost never occurs, 2–3 = sometimes occurs, 4–5 = often occurs, 6 = always occurs). Eight scales can be created and in the standardization were shown to have a high degree of internal consistency, ranging from .80 to .89 with reliability estimates ranging from .72 to .89. In this study, summary scales of social competence, internalizing problems, externalizing problems, and general adaptation are examined. Only the internalizing and externalizing scales are included in both the teacher and parent ratings; Cronbach’s alphas for these scales at all four time points ranged from .79 to .83 for the internalizing problems subscale and .87 to .92 for the externalizing subscale.

2.2.2. Demographic and descriptive measures

Other measures were used to gather demographic and descriptive data about programs, classrooms, teachers, parents and children. Program and classroom information forms were used at the beginning of the school year to document characteristics of each program including type of license, license level (on star rating, an index of global program quality), national accreditation status, classroom staff-child ratio, group size, percentage of Spanish-English DLL children enrolled in the center (children from households where Spanish was spoken), and the presence of Spanish-English bilingual staff (lead teachers, assistant teachers, specialists, administrative staff). Teacher questionnaires completed at the beginning of the first year of participation gathered information regarding their level of education, major, specialized training in working with culturally and linguistically diverse students (particularly Latino children), professional experience, race/ethnicity, bilingual status, age, and gender. Family data were gathered through a structured parent interview conducted in the fall of the year in person or via phone. The purpose was to gather demographic information (e.g., child gender, child birth date, parental income and education, and household composition), as well as the children’s exposure to languages at home and home literacy activities.

2.2.4. Fidelity of implementation measure

Teachers’ implementation fidelity was evaluated through an Implementation Fidelity Observation Guide developed by the research team and used by data collectors. Data collectors were trained to use this observation guide during two visits in the classroom. The guide includes eleven core instructional strategies that distinguish the NNSR Program from a basic classroom curriculum. These instructional strategies were in the areas of oral language development (i.e., reading and vocabulary development practices) and literacy development (i.e., phonological awareness and early writing practices). Examples include “teacher reads a book in Spanish before reading it in English;” “teacher reads a book using props, illustrations, gestures and facial expressions when reading in English to DLLS;” and teacher uses Spanish to introduce a mathematical activity during small groups.” Data collectors observed the treatment group classrooms during fall and spring of each intervention year on two separate days for 3 h per session, checking Yes or No when these strategies were observed. Results indicated that at the end of the two years of participation, 12 out of the 22 intervention teachers who completed the full professional development program implemented 55% or more of the 11 core strategies items included in the fidelity of implementation measure; 10 teachers implemented 45% or fewer of the items. Six teachers did not complete two years of intervention for various reasons, including resignation from their position, reassignment to non-participant classrooms, or center closure due to a change in administration.

2.3. The Nuestros Niños program intervention

The NNSR program consists of a PD approach and content designed to promote school readiness in DLLs. The PD was delivered by project staff over two years via institutes, consultation, and professional learning communities. The institutes were conducted at the beginning of the school year, three days in the first year and two days in the second year. Following the institutes, teachers received support over an 8-month period from bilingual consultants twice a month, and professional learning community meetings also twice a
month (in alternate weeks) to assist in the implementation of new and enhanced classroom practices. During the second year, teachers were expected to receive consultation meetings twice a month and participate once a month in a professional learning community meeting. The purpose of the professional development intervention was to teach and support teachers in the use of ongoing and frequent assessments to monitor children’s English acquisition (Gillanders & Castro, 2015); to conduct reading aloud sessions and follow-up small group activities for explicit vocabulary instruction (Gillanders & Castro, 2011; Gillanders et al., 2014); to develop children’s phonological skills through small group activities (Adams, Foorman, Lundberg, & Beeler, 1998) and emergent writing (Vernon & Ferriero, 1999); to increase children’s mathematical knowledge through an emphasis on academic vocabulary, introducing concepts in number and informal operations, geometry, patterns and measurement (Sorkin, Pappas, Ginsburg, & Gillanders, 2010); and to promote effective teacher–child and peer relationships that create positive social-emotional environments (Gillanders, 2007).

The content of the professional development institutes is documented in a handbook given to all teachers in the treatment group (Castro et al., 2011). The handbook contains six modules that include core concepts, teaching strategies, and classroom resources for promoting language, literacy and social-emotional development, and mathematics learning among Latino DLLs: (1) foundations for teaching young dual language learners, (2) social-emotional development, (3) bilingual oral language development, (4) early reading, (5) early writing, and (6) early mathematics. Each module describes the research evidence for DLLs in that content area and effective strategies for its development. The institutes were designed and led by the investigators.

The consultation process used was adapted from the 8-stage consultation model for early childhood programs developed by Buysse and Wesley (2005). During consultation, the consultant engages the teacher in a collaborative problem-solving process aimed at changing attitudes and dispositions, developing knowledge and improving practice. Consultants were part of the project staff, had a BA in ECE or related field, were bilingual, and had extensive experience in early education as teachers or directors. At the beginning of the project, they received three days of intensive training from one of the investigators on the consultation model, and also on the NNSR Program’s theoretical background and instructional strategies. Each consultant carried a caseload of four to five teachers. The consultants received ongoing support through biweekly reflective supervision meetings with a project investigator. To increase fidelity of implementation of the consultation model, consultants documented their work daily using a consultation contact summary form and consultation action plan which were used during supervision.

The professional learning communities attended by the teachers used the principles of situated learning and reflective practice of the Japanese Lesson Study approach (Lewis, 2002) which offers teachers a structured process for developing and refining instructional strategies through shared inquiry and learning. Through regular professional learning community meetings led by the bilingual consultants, teachers received guidance, working together to create a lesson around a read aloud session, activities of phonological awareness and early mathematical concepts. Teachers engaged in ongoing inquiry in each of these areas of practice introduced in the institutes and reinforced through consultation. Once the lesson plans were created, a volunteer teacher implemented the group’s lesson plan and videotaped different aspects of the lesson. During the second year of the professional development, intervention teachers videotaped the implementation of strategies and received feedback during the consultation and professional learning community sessions. Consultants documented their work in a summary of each professional learning community session which was then discussed in supervision with a project investigator.

2.4. Analytic strategy

The general form of the model that was used to test the effects of NNSR intervention on child outcomes was:

$$\text{post}_{ij} = b_0 + b_1(\text{childpre} - \text{centerpre}) + b_2\text{centerpre} + b_3\text{group} + b_4\text{bilingual} + b_5\text{Tx} + b_6\text{Tx} \times \text{group} + b_7\text{Tx} \times \text{bilingual} + b_8\text{Tx} \times \text{group} \times \text{bilingual} + r_j + e_{ij}$$

The pretest scores (i.e., center mean and child scores deviated from center means) served to increase statistical power (Bloom, 2006; Bloom, Bos, & Lee, 1999). The interaction terms involving treatment condition, treatment year (i.e., whether children had teachers in her 1st or 2nd year of intervention), and teachers’ bilingual status (i.e., whether both lead and assistant teachers were able to speak in Spanish—as reported in the teacher questionnaire) were included to test questions about whether the NNSR Program was differentially efficacious based on those classroom characteristics.

We present results from models that tested the main effect of treatment (i.e., that did not include interaction terms) and from models that did include significant interactions.

The general form of the model that was used to test the effects of the NNSR intervention on classroom outcomes was:

$$\text{post}_{ij} = b_0 + b_1\text{classpre} + b_2\text{time} + b_3\text{bilingual} + b_4\text{Tx} + b_5\text{Tx} \times \text{time} + b_6\text{Tx} \times \text{bilingual} + b_7\text{Tx} \times \text{time} \times \text{bilingual} + r_j + e_{ij}$$

Whereas models for child outcomes involved up to two assessments (pre, post) per child and a designation of whether children were being taught by a teacher in her first vs. second year in the NNSR Program (i.e., treatment year variable), models for classroom outcomes involved four assessments (pre and post in years 1 and 2) for the same classrooms/teachers. The initial pretest score in year 1 was used as a covariate to improve statistical power. Time represented the subsequent three assessments during which treatment was active (post-test year 1, pre and post-test year 2), and bilingual teachers’ status was defined as above (i.e., whether both lead and assistant teachers were able to speak in Spanish). We present results for the main effect of treatment (models without interaction terms) and any significant interactions.

The terms $r_j$ and $e_{ij}$ represent variance components that accounted for the non-independence of observations and residual variation, respectively. In models not presented here, we determined the covariance structure that was optimal for each outcome, including whether each outcome was best represented as a 2-level (child in classroom or child in center) or 3-level (child in classroom, classroom in center) model. For child academic outcomes, this was typically a 2-level random effect (intercept) model in which children were nested within center. For child behavioral outcomes, this was typically a 2-level random effect (intercept) model in which classrooms were nested in centers. For some outcomes, a covariance pattern model parameterization (Hedeker & Gibbons, 2006), which does not involve separate random intercept and residual variances but rather imposes a structure
on a combined covariance matrix, was determined to provide the best fit to the observed data.

All models were estimated using PROC MIXED in SAS version 9.3. Full information maximum likelihood estimation, which assumes that data are missing at random (MAR) and which is considered a statistical best practice (Enders, 2010; Schafer & Graham, 2002), was used to accommodate missing outcome data. In models not presented here, we determined that children with complete versus incomplete data (typically 10–15% of cases primarily with missing data at post-test) were equally distributed across treatment conditions and that there was no consistent pattern of differences between these groups on pretest scores. These results were consistent with the MAR assumption. Random effect and covariance pattern parameterizations were accomplished through the RANDOM and REPEATED statements, respectively, in PROC MIXED. Kenward-Roger (KR) degrees of freedom were used throughout because they accommodate unbalanced and small N (i.e., centers, at level 2) data (Kenward & Roger, 1997). The use of different parameterizations for the covariance structure of the data, along with the use of KR degrees of freedom, helps to explain variation in the reported degrees of freedom (including some that were not integers). We present Cohen’s d effect sizes to characterize the magnitude of treatment effects using a common metric. Effect sizes were computed by taking the difference between the least squares (i.e., adjusted for covariates, including pretest scores) means for treatment and comparison conditions and dividing this difference by the observed standard deviation of the raw data.

3. Results

3.1. Sample characteristics

Table 1 provides a summary of the main characteristics of the NNSR participants overall and by treatment condition. No significant differences were found between treatment and control groups on child, teacher and classroom characteristics. Across both cohorts and treatment conditions, the majority of the programs (44%) were from the Miami, Florida area, followed by North Carolina (37%), and California (15%). The auspice of most of these programs was Head Start (77%). The curriculum used in most classrooms was High Scope (67%); the rest used Creative Curriculum. Seventy-five percent of the lead teachers and 40% of the assistant teachers of the total sample participated in both years of the study. In 39% of the classrooms, both the teacher and assistant teacher spoke Spanish as well as English. Participating teachers had an average of 13.5 years of experience in early education and had held their current position for 5.5 years on average. Sixty-two percent of the participating teachers were Latinas and about 77% held a BA/MA degree and 21% an Associate degree. Some lead teachers and assistant teachers were replaced during the course of two years resulting in a classroom attrition rate of 19% for both control and intervention classrooms. At the child level, about half the sample were girls and ethnicity was slightly more Mexican (55%) than Central American (45%), although the treatment and control groups did not differ on these characteristics.

3.2. Effects on classroom outcomes

One of the aims of this study was to evaluate the efficacy of the NNSR Program on classroom practices overall and those to specifically support DLLs’ development and learning. Table 2 displays the test statistics and effect size estimates for classroom outcome analyses. Sample sizes for the classroom outcome analyses were 30 at pre-test, 25 at post-test and 22 at follow-up for the treatment group, and 26 at pre-test, 22 at post-test and 20 at follow-up for the control group. Descriptive statistics of classroom outcome variables for treatment and control groups can be found in eTable 1. Treatment classrooms had greater gains than control classrooms on the quality of instructional support (d = .34) as an indicator of general classroom quality, and on the quality of specific practices to support DLLs’ language and literacy development (d = .32 and d = .29). Regarding the quality of language interactions in the classroom, there was suggestive evidence indicating that treatment classrooms showed greater gains than control classrooms on both basic (d = .30) and complex (d = .39) language directed to children in Spanish. Although language interactions in English also increased, those changes were not significant.

Additional analyses were conducted to examine whether there were differences in the effect of the NNSR Program when teachers were in the first or second year of their participation, and/or when classrooms had two or only one bilingual teacher. No effects were found on these analyses for outcomes related to general classroom quality (measured by the CLASS) or the quality of specific practices to support DLLs (measured by the ELLCO-DLL). However, there was evidence showing gains in the quality of language interactions in Spanish (measured by the LIsn) in classrooms with two bilingual teachers in the second year of their participation in the NNSR Program. These effects were found for both basic and complex language interactions. In order to better characterize these effects, we tested the treatment effect separately at each time point in teachers’ participation in the professional development program (i.e., post year 1, pre- and post-test year 2) considering the two levels of teachers’ bilingual status. Treatment classrooms showed greater gains than control classrooms on the use of both basic and complex language in Spanish at the pre- and post-test assessments in classrooms with two bilingual teachers, during their 2nd year in the NNSR Program (see Table 2 for test statistics). Fig. 1 displays the effect size comparisons of treatment and control classrooms for basic and complex language in Spanish as a function of classroom teachers’ bilingual status and treatment year. These results clarify the marginally significant main effects of treatment that were reported above; the apparently small overall treatment effect appears to have primarily resulted from a large treatment effect in classrooms with two bilingual teachers during the 2nd year of treatment.

3.3. Effects on child outcomes

A second aim of the study was to evaluate the efficacy of the NNSR Program on promoting DLLs’ language, literacy, mathematics learning and social-emotional development. Table 3 displays the test statistics and effect size estimates for the child outcomes analyses. Sample sizes for child outcome analyses were 178 at pre-test and 159 at post-test for the treatment group, and 161 at pre-test and 144 at post-test for the control group. Descriptive statistics of child outcome variables for treatment and control groups can be found in Tables 2 and 3 of the online supplemental materials. Results below are for the analyses of child academic outcomes in Spanish and English, and child behavioral outcomes rated by teachers and parents.

3.3.1. Academic outcomes assessed in Spanish

As shown in the bottom panel of Table 3, there were greater gains for children in treatment than control classrooms on mathematics ability, writing and on receptive vocabulary (when assessed bilingually-Spanish and English answers were accepted). In addition, there was evidence that children in classrooms with two bilingual teachers and whose teachers were in the second year of treatment, outperformed children in the control classrooms on alphabet knowledge (d = .59). Finally, children in the treatment classrooms showed greater gains in expressive vocabulary when
Table 1
Sample description.

<table>
<thead>
<tr>
<th></th>
<th>Treatment M (SD)</th>
<th>Control M (SD)</th>
<th>Total M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children/Classroom</td>
<td>19.3 (1.1)</td>
<td>19.3 (1.1)</td>
<td>19.3 (1.1)</td>
</tr>
<tr>
<td>Latino children/Classroom</td>
<td>10.1 (4.1)</td>
<td>10.4 (3.9)</td>
<td>10.3 (3.9)</td>
</tr>
<tr>
<td>Lead teacher</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in early childhood</td>
<td>11.0 (8.1)</td>
<td>16.1 (28.7)</td>
<td>13.5 (20.7)</td>
</tr>
<tr>
<td>Years in position</td>
<td>4.5 (4.8)</td>
<td>6.6 (6.5)</td>
<td>5.5 (3.7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Site</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>California</td>
<td>16</td>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>North Carolina</td>
<td>39</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td>Florida (Miami)</td>
<td>44</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>Program type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head Start</td>
<td>77</td>
<td>76</td>
<td>77</td>
</tr>
<tr>
<td>NAEYC accredited</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>Classroom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curricular (Creative)</td>
<td>32</td>
<td>34</td>
<td>33</td>
</tr>
<tr>
<td>Curriculum (High scope)</td>
<td>68</td>
<td>66</td>
<td>67</td>
</tr>
<tr>
<td>Bilingual (lead &amp; assistant)</td>
<td>33</td>
<td>46</td>
<td>39</td>
</tr>
<tr>
<td>Lead teacher</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (Female)</td>
<td>97</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>Ethnicity (Latino)</td>
<td>61</td>
<td>63</td>
<td>62</td>
</tr>
<tr>
<td>Education (CDA)</td>
<td>6</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Education (Associate)</td>
<td>29</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>Education (BA/MA)</td>
<td>71</td>
<td>83</td>
<td>77</td>
</tr>
<tr>
<td>Participated 2 years v. 1</td>
<td>77</td>
<td>72</td>
<td>75</td>
</tr>
<tr>
<td>Child</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (female)</td>
<td>50</td>
<td>45</td>
<td>48</td>
</tr>
<tr>
<td>Central American</td>
<td>41</td>
<td>50</td>
<td>45</td>
</tr>
<tr>
<td>Mexican</td>
<td>59</td>
<td>50</td>
<td>55</td>
</tr>
</tbody>
</table>

Table 2
Synopsis of treatment effects for classroom outcomes.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Tests</th>
<th>Treatment effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tx = Time x Bilingual F (ndf, ddf)</td>
<td>Tx LSM</td>
</tr>
<tr>
<td>Quality of teacher–child interactions (CLASS)</td>
<td>Emotional support 0.2 (2, 92.3) 1.7 (1, 115)</td>
<td>5.7</td>
</tr>
<tr>
<td>Instructional support 0.3 (2, 107) 3.7 (1, 133)*</td>
<td>2.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Classroom organization 0.4 (2, 89.4) 2.0 (1, 122)</td>
<td>5.2</td>
<td>5.0</td>
</tr>
<tr>
<td>Quality of language and literacy practices for DLLs (ELLCO-DLL)</td>
<td>Classroom environment 1.0 (2, 96.6) 3.8 (1, 119)*</td>
<td>23.8</td>
</tr>
<tr>
<td>Language and literacy curriculum 1.1 (2, 96.6) 4.2 (1, 120)*</td>
<td>17.8</td>
<td>16.7</td>
</tr>
<tr>
<td>Literacy activities 0.2 (2, 109) 0.3 (1, 132)</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>Quality of language environment for DLLs (LISN)</td>
<td>Basic quality English 1.4 (2, 103) 0.2 (1, 124)</td>
<td>17.4</td>
</tr>
<tr>
<td>Complex quality English 1.7 (2, 100) 0.7 (1, 126)</td>
<td>7.8</td>
<td>8.6</td>
</tr>
<tr>
<td>Complex quality Spanish 3.5 (2, 103)* 3.0 (1, 126)*</td>
<td>4.1</td>
<td>2.8</td>
</tr>
<tr>
<td>Complex quality Spanish 7.2 (2, 96.1)** 3.1 (1, 128)*</td>
<td>2.8</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Note: * p < .10, * * p < .05, *** p < .001; Tx = Treatment; Ct = Control; ndf = numerator degrees freedom; ddf = denominator degrees freedom; LSM = least squares mean; CLASS = Classroom Assessment Scoring System; ELLCO-DLL = Early Language and Literacy Classroom Observation for DLLs; LISN = Language Interaction Snapshot.

their teachers were in the second year of participation in the NNSR program.

3.3.2. Academic outcomes assessed in English
As summarized in the top panel of Table 3, one child outcome showed evidence of a treatment effect. Children in treatment classrooms outperformed children in control classrooms on expressive vocabulary (d = .39). Other effect sizes were generally small magnitude (ds = -.15 to .19; see top panel of Table 3).

3.3.3. Behavioral outcomes rated by teachers and parents
The final set of analyses considered treatment effects on teacher and parent ratings of various domains of children’s social behavior. Sample sizes for teacher ratings of behavioral outcomes analyses were 178 at pre-test, and 152 at post-test for the treatment group, and 162 at pre-test, and 140 at post-test for the control group. For parent ratings, sample sizes were 156 at pre-test, and 147 at post-test for the treatment group, and 141 at pre-test, and 122 at post-test for the control group. As summarized in Table 4, there was no evidence for main effects of treatment on teacher or parent ratings of social behavior. However, for teacher-rated child outcomes, the treatment was found to have a positive effect depending on teachers’ year of participation in the program and whether there were two bilingual teachers in the classroom. Specifically, effects were statistically significant for two of the teacher-reported outcomes – internalizing and adaptive behaviors. For internalizing behaviors, children in classrooms with teachers in their first year of participation and with two bilingual teachers were rated as better functioning (fewer internalizing behaviors) than children in control classrooms (d = .53). On the other hand, children in classrooms
with two bilingual teachers who were in their second year of participation in the treatment, were rated as showing worse functioning (more internalizing behaviors) \(d = -.47\). An identical pattern of results was evident for adaptive behaviors. Children in classrooms with teachers in their first year of participation and that had two bilingual teachers showed better functioning (more adaptive behaviors) than children in control classrooms \(d = .30\). However, children in classrooms with two bilingual teachers who were in their second year of participation in the treatment, were rated as showing worse functioning (less adaptive behaviors) \(d = -.45\). Differences in the direction of treatment effects for children whose teachers were in their first versus second year of participation in the treatment (and exclusively in classrooms with two bilingual teachers) were most likely responsible for the statistical significance of this interaction effect.

Among parent-rated child behavioral outcomes, treatment effects were found for anger-aggression based on teachers’ year of participation in the treatment. Similar to the teacher-rated outcomes, children whose teachers were in their first year of participation in the NNSR program were rated by their parents as showing less anger-aggression than children in the control group \(d = .29\). However, children whose teachers were in their second year of participation showed worse outcomes (more anger-aggression) than children in the control group \(d = -.23\). Again, the differences in the direction of effects across condition, not the simple main effects of treatment, were primarily responsible for the statistical significance of this interaction term.

**4. Discussion**

In this efficacy study of the NNSR Program, positive effects of the intervention were found on selected classroom practices and children’ school readiness outcomes. We present an interpretation of those positive findings and possible explanations for the non-significant results. Also, study limitations and future directions for this research program are discussed.

**4.1. Classroom practice outcomes**

As hypothesized, the classrooms of teachers participating in the NNSR program were observed to use significantly higher quality instructional practices with all children, and, in addition, there was an improvement in the use of specific instructional strategies to support DLLs, in particular, those related to promoting DLLs’ language and literacy abilities. Although the treatment effects were small, they were of approximately equal magnitude to those of other systematic interventions using teacher professional development for promoting oral language in young children (e.g., Piasta et al., 2012; Wasik & Hindman, 2011), indicating that the NNSR can have a positive impact on the overall quality of early childhood classroom practices, and on those specifically focusing on DLLs. These findings are consistent with those from a study conducted by Buysse et al. (2010) in which teachers participating in a PD program focused on instructional strategies to promote DLLs’ language and literacy development showed significant improve-
Table 3
Synopsis of treatment effects for child academic outcomes.

<table>
<thead>
<tr>
<th>Tests</th>
<th>Treatment effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tx x Grp x Bilingual</td>
</tr>
<tr>
<td></td>
<td>F(ndf, df)</td>
</tr>
<tr>
<td>English</td>
<td>Receptive vocabulary</td>
</tr>
<tr>
<td></td>
<td>BWKP</td>
</tr>
<tr>
<td>Expressive vocabulary</td>
<td>PV</td>
</tr>
<tr>
<td></td>
<td>EWKP</td>
</tr>
<tr>
<td>Phonological awareness</td>
<td>PAT</td>
</tr>
<tr>
<td>Alphabet knowledge</td>
<td>LWI</td>
</tr>
<tr>
<td>Writing</td>
<td>Write name</td>
</tr>
<tr>
<td>Early mathematics</td>
<td>TEMA</td>
</tr>
<tr>
<td>Spanish</td>
<td>Receptive vocabulary</td>
</tr>
<tr>
<td></td>
<td>Expressive vocabulary</td>
</tr>
<tr>
<td>Phonological awareness</td>
<td>PAT</td>
</tr>
<tr>
<td>Alphabet knowledge</td>
<td>LWI</td>
</tr>
<tr>
<td>Writing</td>
<td>Write name</td>
</tr>
<tr>
<td>Early mathematics</td>
<td>TEMA</td>
</tr>
</tbody>
</table>

Note: * p < 0.10, "p < 0.05, "p < 0.01; Tx = Treatment; Ct = control; Grp = Treatment year; Bilingual = # bilingual teachers in classroom; ndf = numerator degrees freedom; df = denominator degrees freedom; LSM = least squares mean; LWI = Letter Word Identification; PV = Picture Vocabulary; ROWPVT = Repetitive One Word Picture Vocabulary Test; TEMA = Test of Early Mathematics; PAT = Phonological Awareness Test; RWKP = Repetitive Word Knowledge Probe; EWKP = Expressive Word Knowledge Probe.

Table 4
Synopsis of treatment effects for child behavioral outcomes.

<table>
<thead>
<tr>
<th>Tests</th>
<th>Treatment effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tx x Grp x Bilingual</td>
</tr>
<tr>
<td></td>
<td>F(ndf, df)</td>
</tr>
<tr>
<td>SCBE Teacher</td>
<td>Social competence behaviors</td>
</tr>
<tr>
<td></td>
<td>Internalizing behaviors</td>
</tr>
<tr>
<td></td>
<td>Externalizing behaviors</td>
</tr>
<tr>
<td></td>
<td>Social adaptive behaviors</td>
</tr>
<tr>
<td>SCBE Parent</td>
<td>Social competence behaviors</td>
</tr>
<tr>
<td></td>
<td>Internalizing behaviors</td>
</tr>
<tr>
<td></td>
<td>Externalizing behaviors (anger-aggression)</td>
</tr>
</tbody>
</table>

Note: * p < 0.05, "p < 0.001; Tx = Treatment; Ct = control; Grp = Treatment year; Bilingual = # bilingual teachers in classroom; ndf = numerator degrees freedom; df = denominator degrees freedom; LSM = least squares mean; SCBE = Social Competence and Behavior Evaluation.

ments in the quality of instructional practices overall and on those specifically targeting DLLs. The present study contributes evidence to the emerging body of knowledge indicating that improving the quality of practices for DLLs can result in an improvement of overall classroom quality which will be beneficial for all children in the classroom.

Because research shows that exposure to instruction in Spanish is beneficial for DLLs’ cognitive and academic skills (Barnett et al., 2007; Burchinal et al., 2012; García, Kleifgen, & Falchi, 2008), increasing classroom language interactions in Spanish is one of the goals of the NNPSR program. It is noteworthy that more Spanish was spoken by teachers in the treatment group in their interactions with DLL children, using both basic- and complex-quality language and this effect increased from year 1 to year 2 of their participation in the program. This effect seems to have resulted mainly from a large treatment effect among classrooms in which the lead and assistant teachers were bilingual and had participated in the intervention for two years. This may indicate that, even among teachers who are somewhat primed to speak with a child in their first language, the professional development program increased their willingness and/or ability to intentionally do it more often. This is important, especially, considering that the intervention was implemented in classrooms that were not using a bilingual instructional approach (e.g., dual language immersion, bilingual transitional).
4.2. Child school readiness outcomes

It is perhaps optimistic to expect that significant child outcomes might result from an intervention with their teachers in the same year in which the professional development is being conducted, however, this study did find positive child outcomes in the treatment classrooms. Children’s expressive vocabulary in English was higher among children in the treatment group. Similarly, when assessed in Spanish, children in the treatment group scored higher in receptive vocabulary, early mathematics, writing and alphabet knowledge. In particular, treatment group children showed a higher performance on receptive conceptual vocabulary when assessed bilingually. The fact that findings were more noticeable in Spanish than in English supports the need to assess DLLs in both of their languages to capture children’s full developmental growth. Moreover, if we had not assessed them in Spanish we would have not identified that they were making progress on mathematical concepts, vocabulary, and writing their names. In particular, considering that for mathematics and name writing the only difference in the assessment was the language of administration, these findings suggest that giving DLLs the opportunity to respond in Spanish allowed them to demonstrate their competence and also underscores the fact that bilingual children learn using both of their languages. Also, we interpret the fact that positive child outcomes were found mostly on children’ Spanish abilities as part of a developmental progression in which the benefits of improved classroom practices will first be observed in children’ strongest language, and will be transferred into children’s L2 as their proficiency increases. This interpretation is supported by studies on the transferability of certain language and literacy abilities between L1 and L2 in bilinguals (Hammer et al., 2014), and has relevance to the importance of using a child’s first or home language in instruction.

Differences in children’s phonological awareness abilities were not found. That could be explained by the limited implementation of phonological awareness activities by teachers in the treatment group, as documented by the implementation fidelity data. Also, teacher consultants reported that phonological awareness activities seemed to be somewhat difficult to implement for teachers in the treatment group. Results related to DLLs’ social-emotional development are also challenging to explain. They seem to indicate that bilingual teachers (working with a bilingual assistant) had a more positive perception of improvements in DLLs’ internalizing and adaptive behaviors only in the first year of treatment but not in the second year. A possible explanation is that in the first year of implementation teachers received information and discussed strategies for creating a positive social-emotional climate in the classroom that supports DLLs. This aspect was not reinforced or promoted in the second year, at least not as much as strategies supporting other areas of development. Other explanations include the possible limitations of the measure in detecting changes in child behavior when used with very young children and the challenges inherent to using teacher-reported measures. Parent perceptions of their children’s behaviors differed from those of their children’s teachers. Parents did not identify any changes in their children’s internalizing and adaptive behaviors. However, parents had a more positive perception of their children’s improvements in levels of anger and aggression when they were in classrooms of teachers who were in the first year of treatment. It is not unusual to find disagreements between teacher and parent reports of children’s abilities and behaviors (Rescorla et al., 2014) because their ratings are based on children’s experiences in different contexts.

4.3. Professional development findings

Several conclusions can be drawn from this study in relation to approaches to teachers’ PD. Positive findings for teachers in the second year of their participation in the NNSR intervention program support our hypothesis of the need for at least two years of intervention to attain changes in teacher practices. It should also be considered that the intervention was intensive and consisted of ongoing individual and peer support, facilitated by a trained bilingual professional, knowledgeable about early childhood education in general, and bilingual development and education of bilingual children, in particular. All consultants had experience as early childhood teachers or directors. In spite of the support available to teachers, implementation fidelity data revealed that only 55% of teachers had implemented more than 50% of the core strategies that comprise the NNSR intervention. One of the challenges most often reported by teachers was the amount of paperwork they had to complete to be in compliance and to fulfill the accountability requirements of their preschool program. Also, teacher turnover and administrative changes, for example, preschool center closings, teachers’ contracts ending and re-hiring processes and teachers moving to a different classroom in the middle of the school year, were some of the challenges faced during the project period. It is remarkable that despite the challenges, positive significant changes were found at the classroom and child school readiness outcome levels (with effect sizes of up to 0.67). An assumption could be made that in conditions more conducive to teachers’ higher levels of implementation fidelity, the impact of the NNSR intervention could be larger.

It is also possible that addressing several areas of children’s school readiness in one intervention study made it difficult to demonstrate teachers’ acquisition of the knowledge, skills and dispositions for facilitating all of those areas. As teachers gain competence in their abilities, it might be easier to observe the consolidation of capabilities and how instructional strategies develop over time. In fact, most teacher PD research studies have examined interventions that focus on one developmental or learning domain or one instructional strategy (e.g., small groups, storybook reading). In creating a multi-dimensional, comprehensive intervention, we added complexity and potential challenges to the teachers and to the study.

Other conclusions relate to the assessment of quality of practices in classrooms enrolling DLLs. Findings indicate that although a measure of general classroom quality (i.e., CLASS) provided useful information, it was not sufficient in identifying high-quality instructional practices specifically targeting the needs of DLLs. Thus, both a general classroom quality measure and a specific measure of classroom practices focusing on DLLs were needed. This was relevant given the purpose of the study and the nature of the intervention. As a result, we are able to report on the effect of the NNSR program to improve practices for DLLs. Our study findings support the conclusion of a recent review of the literature on the use of program quality measures in early childhood classrooms enrolling DLLs (Peisner-Feinberg et al., 2014) that the combination of general quality measures and measures focusing more specifically on aspects of the environment related to supports for DLLs would better capture and represent the experiences of DLL children in ECE settings.

4.4. Limitations and future directions

Attrition was a limitation of the study – a 19% classroom attrition rate over the two years of intervention. Although, these teachers were replaced and we provided PD for them to catch up, these staff changes may have affected the study results. The attrition rate may explain the low level of implementation fidelity of about half of the teachers, as noted above. On the other hand, our fidelity results may have been affected by the way we collected the data. More precise forms of collecting fidelity of implementation which have been currently proposed take into account different aspects of fidelity such
as adherence, dosage, quality of delivery, and participant responsiveness (Guo et al., 2016). Our measure focused on adherence to the NNSR strategies but did not provide a measure of dosage, quality of delivery or participant responsiveness, information that could have been useful for further interpreting the results. On the other hand, the fact that we found a number of positive changes at both classroom and child outcome levels provides evidence of the potential of the NNSR intervention to generate even larger improvements under better conditions. Nevertheless, given the overall lower than desired fidelity of implementation, meaning that results do not represent a test of the complete intervention, findings from this study should be interpreted with caution.

The challenges we confronted in conducting this study confirm that conducting professional development and intervention research is one of the most difficult types of research in the early education field. It requires an understanding of the conditions under which early care and education occurs, including policies, resources, curricula, and staffing. For example, in the United States, the increased in accountability requirements in early education have meant that teachers’ availability to participate in PD or PD research has been reduced. We tried to address this challenge by accompanying teachers and accommodating PD activities to their schedules and individual needs.

Another limitation of the study relates to the challenges to find measures that are sensitive to assess changes in DLLs’ development. It continues to be difficult to find assessment instruments normed on a DLL sample that reflect the experiences of these children across settings and are equivalent in both languages. Specifically, we had difficulties in identifying a measure of social-emotional development that could capture various aspects of social-emotional development of DLLs. The study findings in this domain, with teacher ratings of children’s social-emotional development showing negative outcomes in the second year of the intervention are difficult to explain. In addition to limitations of the instrument itself, it may also be related to the limitations of self-reports. For instance, it is possible that teachers’ perceptions of children’s behavior could have changed after their participation in the intervention as they became better observers of children’s behaviors and might have paid more attention to aspects they were not observing before.

A longer-term follow-up of the quality of classroom practices and its association with children’ developmental and learning outcomes in future school years would be instructive. Longitudinal studies on early childhood interventions are few and they have not included a focus on children who are growing up while exposed to two languages. Furthermore, studies on the benefits of targeted interventions with this population are scarce, inconclusive and have quite a few methodological challenges (Buyssse et al., 2014). Next steps in the analysis of data from this study include analyzing consultants’ records to further identify factors that facilitated or hindered the implementation of the NNSR strategies. As noted earlier, it is important to determine the processes involved in the different forms of delivery of PD (Schachter, 2015; Sheridan et al., 2009). Analysis of these data might shed light on the critical features that facilitate change in teachers’ knowledge, skills and dispositions. Also, additional analyses will describe and examine the role of home language and literacy experiences in moderating the effect of the intervention on children’s outcomes.

Acknowledgments

Funding for this study was provided by the National Institute of Child Health and Development (NICHD), U.S. Department of Health and Human Services (Grant Number 5U01HD060299-05). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of NICHD. The authors would like to thank all members of the project team, teacher consultants, data collectors, administrative assistant, and graduate students, for their dedication and hard work. Special thanks to the teachers, program directors, parents and their children who participated in the study, for allowing us to learn through their experiences.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecresq.2017.03.002.

References

Castro, D. C., Peisner-Feinberg, E., Buyssse, V., & Gillanders, C. (2010). Language and literacy development of Latino dual language learners: Promising instructional practices. In O. Nasaro, & B. Spodek (Eds.), Contemporary perspectives on...
language and cultural diversity in early childhood education [Contemporary Perspectives in Early Childhood Education Series] (pp. 65–93), Charlotte, NC: Information Age Publishing.


