MIND, BRAIN, AND EDUCATION

Early Second Language Learning through SparkLing™: Scaling up a Language Intervention in Infant Education Centers

Naja Ferjan Ramírez[1] and Patricia K. Kuhl[2,3]

ABSTRACT— Infancy (0–3 years) represents a unique time for language learning. Previous research shows that infants' second language (L2) learning advances rapidly in early education centers, through a research-based method and curriculum delivered by native-speaking language tutors. Here we test the potential of this method for broad application, through an interactive online tool, SparkLing™, which trains and certifies language tutors to implement the program in infant education centers. Intervention infants ($n = 168$) from families of predominantly low socioeconomic status (SES) backgrounds in Madrid, Spain, experienced 18 or 36 weeks of L2 (English) exposure, through daily, 45-min, group sessions led by SparkLing™-trained tutors and were compared to a matched Current Practice Comparison (CPC) group of peers ($n = 72$). Intervention children outperformed the CPC group and showed rapid gains in English comprehension and production. SES was *not* a significant factor in learning. Infants' L2 skills advance quickly using this research-to-practice, scalable program for play-based infant L2 instruction.

The ability to speak two languages provides benefits to individuals across the life span (Bialystok, Craik, & Luk, 2012). In addition to the well-understood communicative (Fan, Liberman, Keysar, & Kinzler, 2015), social (Genesee, 2009; Lieberman, Woodward, Keysar, & Kinzler, 2017), and economic benefits (Callahan & Gándara, 2014), early exposure to more than one language is associated with enhanced metalinguistic awareness, and the ability to learn another language (Bartolotti & Marian, 2017; Cenoz, 2003; Petitto et al., 2001). There is also evidence that active use of two languages provides protection against cognitive decline with aging (Gold, Johnson, & Powell, 2013) and has been associated with delaying the onset of Alzheimer's disease (Schweizer, Ware, Fischer, Craik, & Bialystok, 2012). While a number of behavioral studies with bilingual infants, children, and adults have demonstrated advantages in executive functioning (Bialystok, Craik, & Luk, 2008; Carlson & Meltzoff, 2008; Martin-Rhee & Bialystok, 2008), these findings have been challenged in a recent meta-analysis (Paap, Johnson, & Sawi, 2015), and more research is needed. However, a recent study using magnetoencephalography (MEG) shows that, compared to monolingual peers, 11-month-old infants raised in bilingual households show enhanced brain activity in areas related to executive functioning (Ferjan Ramírez, Ramírez, Clarke, Taulu, & Kuhl, 2016).

Considering this body of work, it is not surprising that communities around the world aspire to create school-based programs for early second language (L2) instruction. The vast majority of such programs focus on L2 learning in preschool- or elementary school-aged children who already have a well-established first language (L1; see Muñoz, 2006). With the exception of limited private programs, attempts to introduce L2 learning in school settings in infancy, when L1 skills are less well developed, have been

[1]Department of Linguistics, University of Washington
[2]Institute for Learning & Brain Sciences, University of Washington
[3]Department of Speech and Hearing Sciences, University of Washington

Address correspondence to Naja Ferjan Ramírez, University of Washington, Department of Linguistics Guggenheim Hall 415B, Box 352425 3940 Benton Ln NE Seattle, WA 98195-2425; e-mail: naja@uw.edu

limited. This is surprising, given the well-established finding that infants have the ability to rapidly and naturally acquire one or multiple languages through social interactions (for review, see Kuhl, 2007; Werker & Byers-Heinlein, 2008). Research with infants from bilingual households indicates that the human brain is adept at acquiring two languages simultaneously (Ferjan Ramírez et al., 2016; Garcia-Sierra, Ramírez-Esparza, & Kuhl, 2016), and that infants who grow up in bilingual households efficiently monitor and control their two languages, as evidenced by pupil size measurements during real-time language processing (Byers-Heinlein, Morin-Lessard, & Lew-Williams, 2017). Studies also show that acquiring two languages in infancy, compared to acquiring an L2 in later childhood or adolescence, results in more efficient neural language processing (Berken, Gracco, & Klein, 2017; Klein, Mok, Chen, & Watkins, 2014) and higher brain-tissue density in regions supporting language, memory, and attention (Mechelli et al., 2004). This research is consistent with seminal behavioral studies, which demonstrate clear and strong benefits of early L2 exposure which are most prominent for the acquisition of phonology (Flege, MacKay, & Meador, 1999; Mackay & Flege, 2004), but have also been demonstrated in morphology (Verissimo, Heyer, Gunnar, & Clahsen, 2018), and syntax (Clahsen & Felser, 2006; Johnson & Newport, 1989).

Foreign Language Learning in Infancy

In light of these findings, an interesting question to consider is *how much* and *what kind* of L2 exposure is necessary for infant L2 learning. Laboratory studies show that nine-month-old infants exposed to a foreign language through playful interactions with a live tutor learn to discriminate foreign language sounds at levels equivalent to infants exposed to that language from birth with a total of 6 h of foreign language experience (Kuhl, Tsao, & Liu, 2003). Eighteen-month-old infants detect foreign language words through brief exposure sessions when transitional probabilities and stress-related cues are available (Hay, Pelucchi, Graf Estes, & Saffran, 2011), and short-term foreign language exposure through play sessions in infancy modulates neural responses to foreign language sounds (Conboy & Kuhl, 2011).

While these results are promising, a key question to consider is how they translate to real-life environments, such as infant education centers. Can infant foreign language learning be ignited in a group setting, in the context of an infant school? We recently tested this question through a foreign language intervention in four public infant education centers in Madrid, Spain, and demonstrated that Spanish infants' L2 (English) skills advance rapidly through a play-based, highly social and interactive method and curriculum based on a theoretical model of infant language development (Ferjan Ramírez & Kuhl, 2017). Intervention children ($N = 126$, ages 7–33.5 months) experienced 18 weeks (72 instruction days) of daily, hour-long, group English sessions with native-speaking tutors, trained to execute the research-based method and curriculum. Intervention infants were compared to a matched group of peers in the same schools who received Madrid's standard bilingual Spanish-English program (Current Practice Comparison, CPC). Intervention children showed rapid gains on measures of English comprehension and production, significantly outperforming the CPC group, while Spanish (L1) continued to grow equally in both groups. These results demonstrated that infants are capable of making rapid gains in L2 learning in school settings, when the method of language exposure incorporates key evidence-based features that have been demonstrated to be effective in laboratory research. These findings led to the present study, the goal of which is to test the potential of this program to be applied broadly.

The Present Study

The present study, like the original study (Ferjan Ramírez & Kuhl, 2017, henceforth referred to as the "2017 Study"), was conducted in the Community of Madrid's public infant education centers (schools) serving children 0–3 years of age. The goal was to use the same research-based method and curriculum as the 2017 Study (see Methods), but to conduct the Intervention tutors' training, which was administered in person for the 2017 Study, using an interactive online training and certification tool, SparkLing™. In the present study, all language tutors were trained and certified with SparkLing™ prior to delivering the Intervention. Intervention children received daily group English sessions with native-speaking SparkLing™-trained tutors and were compared to a matched group of peers attending the same schools who received Madrid's standard bilingual Spanish-English program (CPC).

The program was implemented in 13 schools, and language data was collected in 7 schools varying in size and number of classrooms receiving the Intervention. Four schools were selected for the initial analyses reported here: two received the Intervention over the full school year (36 weeks; 140 instruction days) and two over half of the school year (18 weeks; 71 instruction days). The four schools were selected to obtain an age match with the participants in the 2017 Study, an age match between the Intervention and CPC children within the present study, and an age match between the 18- and 36-week Intervention (Figure 1). Some differences between the 2017 Study and the present study implementations are summarized in Table 1. Of note, SES in the present study was lower than that of the 2017 Study. Furthermore, several steps were taken to ensure that the research procedures imposed as little interference as possible with infants' day-to-day needs, activities, and

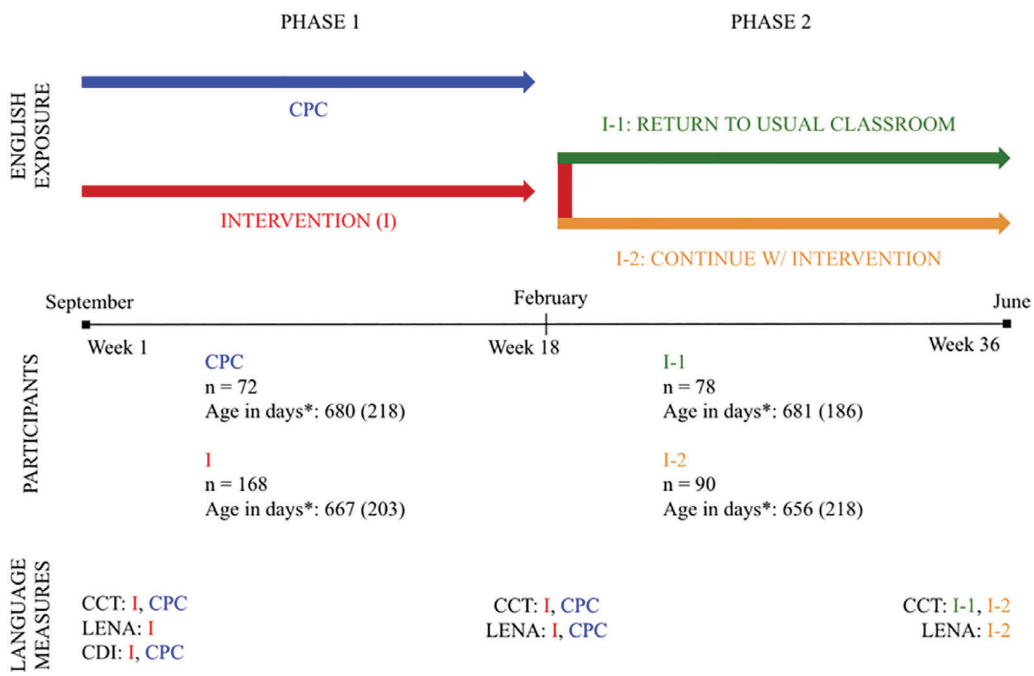

Fig. 1. Outline of study design. I, intervention; CPC, current practice comparison; I-1, Intervention1; I-2, Intervention2. CTC, computerized comprehension task; CDI, MacArthur-Bates Communicative Developmental Inventory. *Age on the first day of Intervention sessions—mean (SD).

Table 1
Implementation differences between present study and 2017 Study

2017 Study	*Present Study*	*Reason for change*
Tutor training in person	Tutor training online with SparkLing™	Scalability, accessibility, transferability
Daily English sessions 60 min long	Daily English sessions 45 min long	Reduces interference with schools' usual curriculum; Allows tutor teams to execute three daily sessions at each school each morning
4 tutors with up to 12 children	3 tutors with up to 14 children (in classrooms for ages 9–21 months); 3 tutors with up to 20 children (in classrooms for ages 21–33 months)	Reduces cost; more children can receive program
Children pulled out of usual classroom for Intervention sessions, which take place in a separate room	Tutor teams enter children's usual classroom	No additional space needed; transition easier for children and school staff
Half-year (18 week) program with CPC group	Half-year (18 week) program with CPC group and full-year (36 week) program	Assess the effect of program length
Two schools of mid SES, two of low SES	One school of mid SES, two schools of low SES, one school of very low SES	Ability to generalize findings across a broader range of SES backgrounds, with a focus on low SES

Abbreviations: CPC, current practice comparison; SES, socioeconomic status.

the schools' curricula. For example, rather than dividing individual children into the I and the CPC group by random assignment, as would be required by a classical Randomized Control Trial design, all participating children in a given classroom were assigned to the same group, either Intervention or CPC, which did not differ on any demographic or language characteristics at the start of the study (see Results, Figure 1, Table 2, and Supplemental Information for details).

The present study was structured in two phases (Figure 1): Phase 1 started with data collection in September, ended with data collection in February, and included 18 weeks of English exposure, like the 2017 Study. Intervention children in all four schools participated and were compared against CPC children, who attended two of the four participating schools (School1 and School2). The purpose of Phase 1 was to compare the Intervention and CPC children within the present study, and draw comparisons with the 2017 Study. We predicted that the general pattern of results from the 2017 Study would be replicated with this new set of SparkLing™-trained tutors, schools, and children. Specifically, we hypothesized that Intervention children would exhibit rapid growth in English comprehension and production, outperforming the CPC group. As in the 2017 Study, Spanish comprehension was hypothesized to grow at the same rate in both groups. Based on the well-documented relationship between language input and child language outcomes, we hypothesized that we would see attenuation in L2 growth compared to the 2017 Study, as a result of shortening the sessions and reducing the tutor/child ratio, and overall lower Spanish levels as a result of lower SES in the present study.

Phase 2 began in February. The Intervention group was split into Intervention1 (I-1, children in School1 and School2) and Intervention2 (I-2, children in School3 and School4). I-1 children returned to their usual classroom activities and did not receive the Intervention between February and June. I-2 children continued to receive the Intervention for another 18 weeks, until June. I-1 and I-2 children participated in the February and June language testing. Phase 2 allowed us to assess how Spanish and English learning continues after having received 18 weeks of Intervention, upon experiencing their usual classroom activities (I-1), or upon receiving an additional 18 weeks of Intervention (I-2). We hypothesized that participation in the full-year Intervention would result in enhanced English outcomes compared to participation in the half-year Intervention.

METHODS

Participants and Schools

Parents of 252 children between 9 and 33 months of age at the start of the study signed informed consent. Enrollment criteria included no major birth or postnatal complications, and no significant exposure to English from live human speakers outside of school. Data from 12 children were excluded because of parental report of suspected developmental delay ($n = 2$) or English exposure through live speakers outside of the school ($n = 10$). A total of 240 children (114 girls) had usable data for analyses. The ethics committee at the University of Washington approved the study.

The four participating schools served families who lived in the neighborhoods, which ranged in SES from mid to very low. In each school, the percentage of children receiving free meals was determined and used as an SES-proxy. These percentages were: School1: 11% (mid SES), School2: 48% (low SES), School3: 35% (low SES), and School4: 76% (very low SES). While the relationship between SES and language development is well documented, there is currently no consensus on the most effective measure of SES, and measures of parental education, parental occupation, and income-based measures are supported with evidence that validates their application as SES proxies (for review, see Pace, Luo, Hirsh-Pasek, & Golinkoff, 2017). The SES-proxy used here was income based, and was used as a covariate in statistical analyses (details in Supplemental Information).

English Exposure

Intervention children experienced daily, 45-min-long English play sessions led by teams of three tutors, who interacted with a group of 14 children (in classrooms for ages 9–21 months) or 20 children (in classrooms for ages 21–33 months). Tutors were undergraduate students or recent graduates, and had successfully completed the SparkLing™ program. They were recruited within the University of Washington, by announcements on bulletin boards, sharing through social media, departmental email lists, and verbal announcements in University of Washington classes, which ensured that the applicants came from varying backgrounds. The only two eligibility requirements were being a current or past undergraduate student at the University of Washington, and being a native speaker of English (self-report, confirmed with an in-person interview).

SparkLing™ is an interactive online program, consisting of a training, a certification, a set of lesson plans, and a collection of classroom materials. The training consists of five units describing the six-point method (Ferjan Ramírez & Kuhl, 2017; outlined below) through interactive activities and video examples of tutor behaviors and children's responses (Figure 2). Quiz questions appear throughout the training to provide knowledge checks. Each unit ends with a certification test that must be passed to progress to the next unit. In the present cohort of tutors, 77% of unit certifications

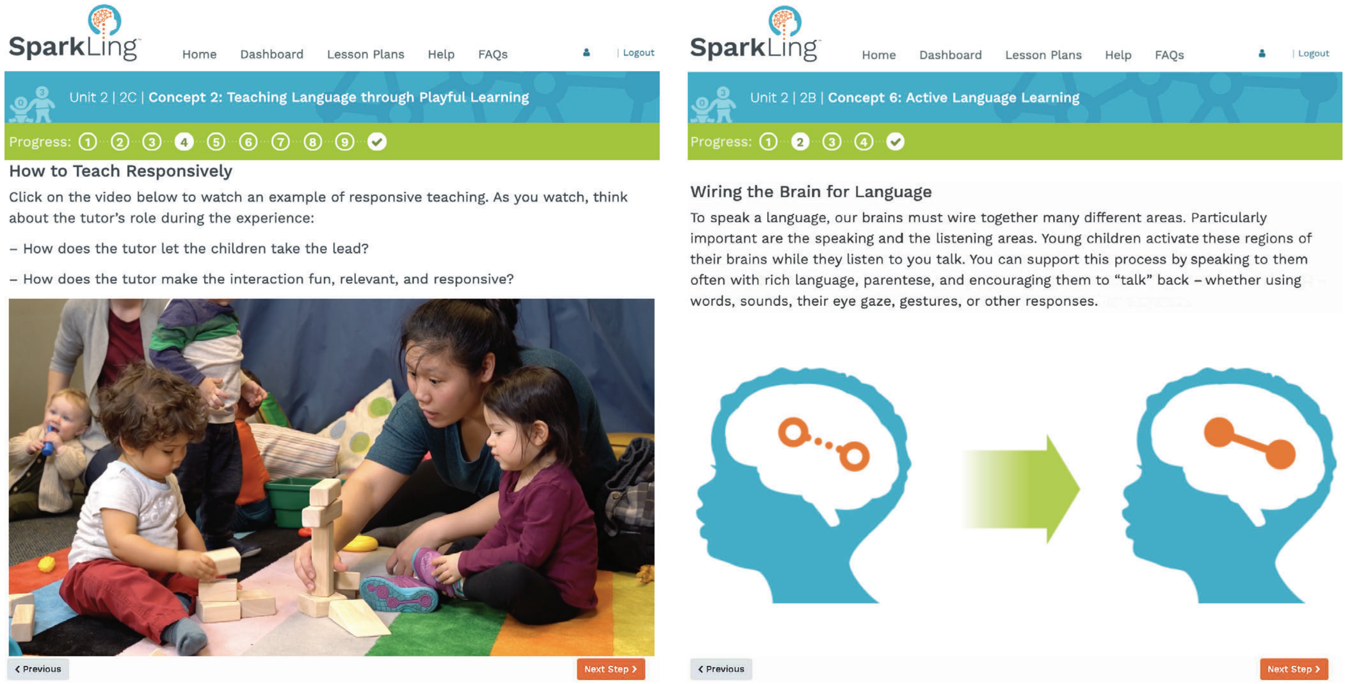

Fig. 2. Two example pages from the SparkLing™ training. The training describes the six-point method through interactive activities, video examples, and frequent knowledge checks.

were passed on the first attempt, and 23% were passed on the second or third attempt. Upon passing all five certification tests, tutors received access to the 36-week curriculum (lesson plans), describing how to implement the research principles through 45-min daily sessions, frequently referencing the research principles. The SparkLing™ materials are a collection of children's books and classroom manipulatives to support the classroom activities.

A detailed description of the six key principles of the Intervention method and the research behind each is given in Ferjan Ramírez & Kuhl, 2017. Briefly, (1) Tutors addressed children often to achieve a high quantity of English input. (2) Tutors addressed children in "parentese," which is characterized as having higher pitch, slower tempo, and exaggerated intonation contours. (3) The learning context was highly social with weekly themes, games, and activities that prompt face-to-face interaction. (4) Children were encouraged to "talk" and interact. Tutors were trained to provide prompt, contingent responses, and engage the children in frequent back-and-forth exchanges. (5) Children heard English from multiple native speakers. (6) The daily sessions were presented through adult-scaffolded play and language content was delivered in a way that ensured distributed exposure. A description of the CPC program can be found in the 2017 Study and in Supplemental Information. Detailed information about enrolment, participation, and attrition for each measure can be found in Supplementary Tables S1, S2, and S3.

Language Measures.

Prior to Week1 (September) children's Spanish level was assessed with the European Spanish MacArthur-Bates Communicative Developmental Inventory (CDI; López Ornat et al., 2005). Prior to Week1 (September), after Week18 (February), and after Week36 (June) children who were at least 16 months of age participated in a Spanish and English vocabulary comprehension assessment with the Computerized Comprehension Task (CCT; Friend & Keplinger, 2003), an assessment of word comprehension administered on a touch screen. Children respond to paired pictures on the screen via a sentence prompt (e.g., Where is the *dog*? Which one is *big*?). Intervention children completed this assessment in September, February, and June. CPC children completed it in September and February.

Children's language production was assessed with the LENA technology (LENA Pro Version 3.4.0, 2015). Each child wore a LENA vest, with a digital language processor (DLP): a small, lightweight recorder placed into the vest's pocket, designed to record the child's voice and the language they hear. Children's language production contained English and Spanish, in addition to other sounds, such as babble, raspberries, and imitations of animal sounds. To assess children's English production, manual coding was performed, and all statistical analyses were conducted on manually coded data. The fully analyzed dataset reported here includes the LENA recordings from Week1 (September) and Week18 (February) for Intervention children, Week18 (February) recordings for CPC children, and Week36 (June) recordings for I-2 children. Details about LENA data collection, coding, and analysis procedures are described in Ferjan Ramírez and Kuhl (2017) and in Supplemental Information. In brief, coders identified all English vocalizations in a 25-min segment of each child's recorded file. As in the 2017 Study, statistical analyses were performed on the total number of English vocalizations per child per hour, obtained by multiplying the number of English vocalizations in the 25-min segment by 2.4. All LENA data was logarithmically transformed to meet the assumptions for parametric testing, determined by the skewness and kurtosis values, both of which were between −2 and + 2, which is considered acceptable in order to prove normal univariate distribution (George & Mallery, 2010; Gravetter & Wallnau, 2014; Trochim & Donnelly, 2006).

RESULTS

Preliminary Analyses and Approach

Preliminary analyses confirmed that there were no differences between boys and girls on any variables of interest (all *p*s > .1). Data from boys and girls was therefore collapsed. For measures obtained at multiple time points, only children with complete data at the considered time points were included in analyses of any given measure.

September Measures of Spanish (CDI)

September Spanish productive vocabularies ranged between 0 and 546 words, with the mean productive vocabulary of 131.7 ($SD = 154.6$) words. These scores placed children between 1st and 95th percentile for their age, with the average being the 32nd percentile. The mean CDI productive vocabularies and percentile scores did not differ between the Intervention and the CPC group, $t(190) = -0.045$, $p = .96$, 95% CI [−48.7, 46.5] and $t(190) = -0.31$, $p = .76$, 95% CI [−9.5, 6.9] and these results did not change when controlling for SES: mean productive vocabularies, $F(1, 189) = 0.12$, $p = .93$, percentile scores, $F(1, 189) = 0.007$, $p = .73$. As expected based on the lower SES scores, when compared to the 2017 Study, the mean CDI productive vocabularies were lower in the current cohort by an average of 51 words or 10 percentile points.

Phase 1: September–February

Spanish and English Comprehension (CCT)

Table 2 shows mean English and Spanish CCT scores and standard deviations for Intervention and CPC children in September and February. For English, a repeated measures ANOVA with Time (September/February) and Group (I/CPC) as independent variables showed a significant main effect of Time, $F(1, 93) = 68.83$, $p < .0001$ (Greenhouse–Geisser), partial $\eta^2 = .43$ and an interaction of Time and Group, $F(1, 93) = 8.34$, $p = .005$ (Greenhouse–Geisser), partial $\eta^2 = .08$, indicating the impact of the Intervention on the pattern of change over time. The interaction remained significant after controlling for SES (covariance analysis): $F(1, 92) = 6.38$, $p = .013$, partial $\eta^2 = .07$. Follow-up *t*-tests revealed that the Intervention and the CPC groups' English scores were not significantly different in September, $t(93) = -0.68$, $p = .50$, 95% CI [−1.1, 1.7]. In February, the Intervention group scores were significantly higher than the CPC group scores, $t(93) = 3.39$, $p = .001$, $d = 0.74$, 95% CI [1.5, 5.8].

For Spanish, there was also a main effect of Time, $F(1, 105) = 86.9$, $p < .0001$ (Greenhouse–Geisser), partial $\eta^2 = .45$, but the interaction of Time and Group was not significant, $F(1, 105) = 0.46$, $p = .50$ (Greenhouse–Geisser), partial $\eta^2 = .00$. The Spanish CCT scores in the two groups, as predicted, increased in both groups equally over the course of the 18-week Intervention period after starting at similar levels.

Raw scores from the 2017 Study (Ferjan Ramírez & Kuhl, 2017) are presented in Supplemental Information (Table S4). As expected based on SES comparisons in the two studies, the initial (preintervention) scores for Spanish and English vocabulary were lower in the present cohort compared to the 2017 Study. However, the English point increase in the Intervention group was comparable:

Table 2
Mean English and Spanish Computerized Comprehension Task (CCT) Scores and Standard Deviations in September and February for Intervention and Current Practice Comparison (CPC) children with valid scores at both time points

	Spanish CCT Score M (SD)			*English CCT Score* M (SD)		
	N; age in days M (SD)	*September*	*February*	*N; age in days* M (SD)	*September*	*February*
Intervention	72; 771 (133)	21.2 (10.6)	28.1 (8.2)	62; 794 (123)	11.8 (7.6)	20.9 (3.9)
CPC	35; 778 (156)	20.6 (10.8)	28.6 (8.7)	33; 791 (151)	12.5 (7.8)	17.0 (6.5)

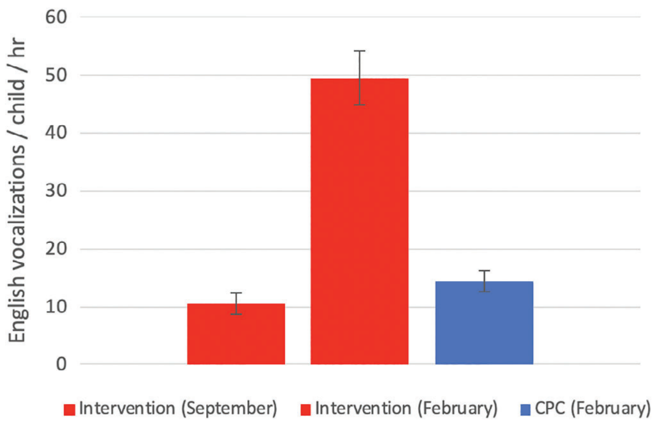

Fig. 3. Mean number of English vocalizations per child per hour for Intervention group (red; $n = 146$) in September and in February, and for the CPC group (blue; $n = 69$) in February. Error bars represent standard error.

9.1 points in the present study and 8.7 points in the 2017 Study.

English Production (LENA)

As shown in Figure 3, the rate of English vocalizations per child per hour changed significantly for the Intervention group from Week1 ($M = 10.6$, $SD = 22.3$) to Week18 ($M = 49.4$, $SD = 56.2$), $t(145) = 12.3$, $p < .001$, $d = 1.19$, 95% CI [0.62, 0.86]. In Week18 the I group also produced more vocalizations per hour than the CPC group ($M = 14.4$, $SD = 16.1$), $t(213) = 5.01$, $p < .0001$, $d = 0.75$, 95% CI [0.29, 0.66]. Both effects remained significant after controlling for SES; Intervention group Week1 to Week18, $F(1, 144) = 41.7$, $p < .0001$, partial $\eta^2 = .22$; Week1 Intervention to Week18 CPC comparison, $F(1, 212) = 14.4$, $p < .0001$, partial $\eta^2 = .06$.

Compared to the 2017 Study, the increase in English production was lower in the present cohort (38.8 vocalizations, or 4.7-fold increase in the present study, vs. 61.7 vocalizations, or 6-fold increase in the 2017 Study). In part, this may be as a result of a lower tutor/child ratio, shorter Intervention sessions (45 vs. 60 min/day), or overall lower initial Spanish or English levels. However, and as confirmed statistically by a large effect size (Cohen's $d = 1.19$), the increase in English production in the present study was substantial.

Phase 2: February–June

Spanish and English Comprehension (CCT)

In February, the Intervention group was split into I-1 and I-2 group. I-1 children returned to experiencing their usual classroom activities, while I-2 children continued receiving the Intervention for another 18 weeks, until June. Figure 4 shows the mean English and Spanish CCT scores and standard errors for I-1 and I-2 children in February and June.

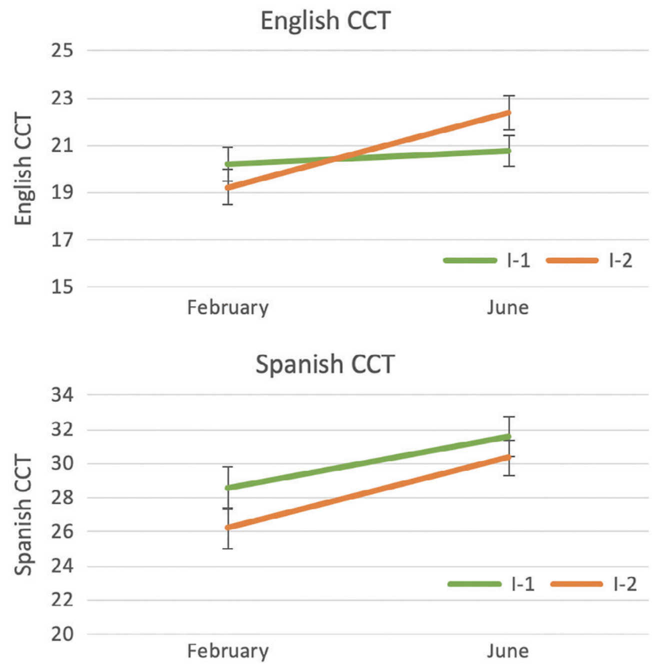

Fig. 4. Mean English and Spanish Computerized Comprehension Task (CCT) Scores in February and June, for I-1 (English $n = 45$; Spanish $n = 49$) and I-2 (English $n = 46$; Spanish $n = 50$) children who completed the assessments at both time points. Error bars represent standard error.

For English, the main effect of Time was significant, F (1, 89) = 12.09, $p = .001$ (Greenhouse–Geisser), partial $\eta^2 = .12$. More interestingly, the interaction of Time and Group was also significant, F (1, 89) = 6.07, $p = .016$ (Greenhouse–Geisser), partial $\eta^2 = .06$, indicating the impact of the Intervention on the pattern of change in English comprehension over time. This interaction remained significant after controlling for SES (covariance analysis): $F(1, 89) = 6.72$, $p = .019$, partial $\eta^2 = .06$. Two post hoc paired sampled t-tests were conducted to assess the changes in English CCT scores between February and June in each group. In the I-1 group, the English CCT scores did not change significantly, $t(44) = .93$, $p = 0.36$, 95% CI [−0.62, 1.69], indicating that this group of children did not advance their English comprehension between February and June, but retained what they had learned upon returning to their usual classroom activities. In I-2, however, the change between February and June was significant, $t(45) = 3.56$, $p = .001$, $d = 0.60$, 95% CI [1.36, 4.90], showing that English comprehension continued to improve.

For Spanish, a repeated measures ANOVA with Time (February/June) and Group (I-1/I-2) as independent variables showed that the main effect of Time was significant, F (1, 97) = 47.63, $p < .0001$ (Greenhouse–Geisser), partial $\eta^2 = .33$, but the interaction of Time and Group was not significant, F (1, 97) = 1.143, $p = .288$ (Greenhouse–Geisser),

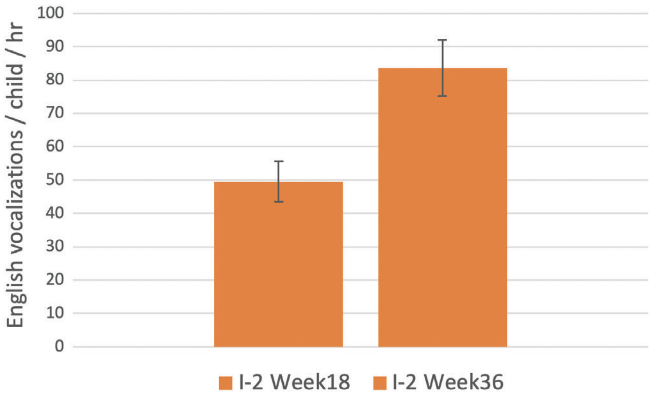

Fig. 5. Mean number of English vocalizations per child per hour for the Intervention2 (I-2) children who contributed LENA recordings in February and June ($n = 85$). Error bars represent standard error.

partial $\eta^2 = .01$, indicating that the Spanish CCT scores increased equally in both groups.

English Production (LENA)

LENA recordings were obtained in Week36 (June) for I-2 children. In Week18 (February), this group of children produced an average of 49.5 ($SD = 56.4$) English vocalizations per child per hour. By Week36, the mean number of English vocalizations per child per hour increased to 83.6 ($SD = 77.9$; Figure 5). A paired sampled t-test revealed that this was a significant increase, $t(84) = 4.372$, $p < .0001$, $d = 0.47$, 95% CI [0.19, 0.49]. This effect remained significant after controlling for SES, $F(1, 83) = 12.57$, $p = .001$, partial $\eta^2 = 0.13$.

DISCUSSION

The present Intervention was designed as a scale-up of a previous language intervention in Madrid, Spain (Ferjan Ramírez & Kuhl, 2017), with the goal of replicating the pattern of results using an online training, SparkLing™, a new set of language tutors, and a new set of children from public infant schools serving predominantly families of low SES backgrounds. Given that in the current study, when compared to the 2017 Study, the language sessions were shortened, the tutor/child ratio was reduced to lower the costs and the impact on the schools' usual curriculum, and that the participating children were of predominantly low SES backgrounds, the current study also served as a test of whether the research-based program was sufficiently robust to reproduce the original results under less ideal, but more realistic, educational settings.

The general pattern of results replicates the findings of the 2017 Study: Phase 1 results show that Intervention children exhibited rapid growth in English comprehension and production, outperforming the CPC group. As in the 2017 Study, Spanish comprehension grew equally in both groups, indicating that it was not affected by the Intervention. Phase 2 results replicate the 2017 Study findings with regard to retention: I-1 children, who participated in the Intervention over the first 18 weeks (September–February) and experienced their usual classroom activities over the second 18 weeks (February–June), retained what they had learned between September and February at least until June. Extending the findings of the 2017 Study, Phase 2 also shows that I-2 children, who received the Intervention over 36 weeks, continued to advance their English comprehension and production, demonstrating that the program continues to enhance L2 skills if administered over an entire school year.

Compared to the 2017 Study, and as hypothesized, the English comprehension and production scores at the end of the 18-week period (in February) were lower in the present sample. This may be as a result of a lower tutor/child ratio, shorter Intervention sessions, overall lower initial Spanish or English levels, or a combination of these factors. Considering these important implementation differences, the patterns of English growth in the two studies were remarkably similar, and demonstrate the potential of the current program to be scaled up and applied more broadly, in other countries or communities, including the USA. The online training and certification tool is designed to train the language tutors on the science behind the six principles. The daily curriculum, available upon completion of the certification, provides a set of interactive activities for each day of instruction, with frequent reference to the research principles.

Because the intervention was based on six principles, we cannot yet specify the minimal conditions that are necessary for producing the observed effects. We also cannot predict how the results may be affected by modifying any one of the six parameters (for example, by replacing the native-speaking tutors with nonnative English speakers). Follow-up studies focusing on a single variable at a time will be needed to test such questions. Future studies will also further explore children's learning across other domains of linguistic structure.

Communities around the world aspire to create effective school-based programs for L2 instruction. In most cases, L2 instruction is introduced in preschools or elementary schools, therefore not harnessing children's full potential for native-like language acquisition through natural social interaction. The results of the present study underscore the importance of bridging the gap between laboratory and applied research, and demonstrate that infants, across SES backgrounds, can begin acquiring an L2 rapidly in a social environment that engages them through high-quality and -quantity language input.

Acknowledgments—The authors thank Julia Mizrahi, Bo Woo, Anna Kunz, Leanne Kehoe, and Taylor Carrasco-Hermerding of the University of Washington's Institute for Learning & Brain Sciences (UW I-LABS) for valuable assistance throughout data collection. This research was supported by the Madrid Regional Ministry of Education, Youth, and Sports, the UW I-LABS Ready Mind Project with additional support from the Bezos Family Foundation, and the UW CoMotion Innovation Gap Fund.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1: Supporting Information

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Bartolotti, J., & Marian, V. (2017). Bilinguals' existing languages benefit vocabulary learning in a third language. *Language Learning*, *67*, 110–140.

Berken, J. A., Gracco, V. L., & Klein, D. (2017). Early bilingualism, language attainment, and brain development. *Neuropsychologia*, *98*, 220–227.

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*, 240–250.

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 859–873.

Byers-Heinlein, K., Morin-Lessard, E., & Lew-Williams, C. (2017). Bilingual infants control their languages as they listen. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 9032–9037.

Callahan, R., & Gándara, P. (2014) *The bilingual advantage: Language, literacy, and the labor market*. Bristol, England: Multilingual Matters.

Carlson, S., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*, 282–298.

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism*, *7*, 71–87.

Conboy, B. T., & Kuhl, P. K. (2011). Impact of second-language experience in infancy: Brain measures of first- and second-language speech perception. *Developmental Science*, *14*, 242–248.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*, 564–570.

Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological Science*, *26*, 1090–1097.

Ferjan Ramírez, N., Ramírez, R. R., Clarke, M., Taulu, S., & Kuhl, P. K. (2016). Speech discrimination in 11-month-old bilingual and monolingual infants: A magnetoencephalography study. *Developmental Science*, *20*. https://doi.org/10.1111/desc.12427

Ferjan Ramírez, N., & Kuhl, P. K. (2017). Bilingual baby: Foreign language intervention in Madrid's infant education Centers. *Mind, Brain, and Education*, *11*, 133–143.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*, 2973–2987.

Friend, M., & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behavior Research Methods Instruments, & Computers*, *35*, 302–309.

Garcia-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2016). Relationships between quantity of language input and brain responses in bilingual and monolingual infants. *International Journal of Psychophysiology*, *110*, 1–17.

Genesee, F. (2009). Early childhood bilingualism: Perils and possibilities. *Journal of Applied Research on Learning*, *2*, 1–21.

George, D., & Mallery, M. (2010) *SPSS for Windows Step by Step: A Simple Guide and Reference, 17.0 update*. (10a ed.). Boston: Pearson.

Gold, B. T., Johnson, N. F., & Powell, D. K. (2013). Lifelong bilingualism contributes to cognitive reserve against white matter integrity declines in aging. *Neuropsychologia*, *51*, 2841–2846.

Gravetter, F., & Wallnau, L. (2014) *Essentials of statistics for the behavioral sciences*. (8th ed.). Belmont, CA: Wadsworth.

Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, *63*, 93–106.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.

Klein, D., Mok, K., Chen, J.-K., & Watkins, K. (2014). Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain and Language*, *131*, 20–24.

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Developmental Science*, *10*, 110–120.

Kuhl, P. K., Tsao, F. -M., & Liu, H. -M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 9096–9101.

LENA Pro (Version 3.4.0) [Apparatus and software]. (2015). Boulder, CO: LENA.

Lieberman, Z., Woodward, A. L., Keysar, B., & Kinzler, K. D. (2017). Exposure to multiple languages enhances communication skills in infancy. *Developmental Science*, *20*, e12420. https://doi.org/10.1111/desc.12420

López Ornat, S., Gallego, S., Gallo, P., Karousou, A., Mariscal, S., & Martínez, M. (2005) *MacArthur: Inventario de desarrollo comunicativo. Manual y Cuadernillos*. Madrid: TEA Ediciones.

Mackay, I. R. A., & Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied PsychoLinguistics*, *25*, 373–396.

Martin-Rhee, M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*, 81–93.

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Neurolinguistics: Structural plasticity in the bilingual brain. *Nature*, *431*, 757.

Muñoz, C. (Ed.) (2006) *Age and the Rate of Foreign Language Learning*. Clevedon: Multilingual Matters.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278.

Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics*, *3*, 285–308.

Petitto, L. A., Katerelos, M., Levy, B., Gauna, K., Tetreault, K., & Ferraro, V. (2001). Bilingual signed and spoken language acquisition from birth: Implications for the mechanisms underlying early bilingual language acquisition. *Journal of Child Language*, *28*, 453–496.

Schweizer, T. A., Ware, J., Fischer, C. E., Craik, F. I., & Bialystok, E. (2012). Bilingualism as a contributor to cognitive reserve: Evidence from brain atrophy in Alzheimer's disease. *Cortex*, *48*, 991–996.

Trochim, W. M., & Donnelly, J. P. (2006) *The research methods Knowledge Base*. (3rd ed.). Cincinnati, OH: Atomic Dog.

Verissimo, J., Heyer, V., Gunnar, J., & Clahsen, H. (2018). Selective effects of age of acquisition on morphological priming: Evidence for a sensitive period. *Language Acquisition*, *25*, 315–326.

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism: First steps in perception and comprehension. *Trends in Cognitive Science*, *12*, 144–151.