Title

Patterns of Development in Spanish-English Conceptually Scored Vocabulary Among Elementary-Age Dual Language Learners.

Journal

Journal of Speech, Language, and Hearing Research, 63(9)

Authors

Mancilla-Martinez, Jeannette
Hwang, Jin
Oh, Min
et al.

Publication Date

2020-09-15

DOI

10.1044/2020_JSLHR-20-00056

Peer reviewed

JSLHR

Research Article

Patterns of Development in Spanish–English Conceptually Scored Vocabulary Among Elementary-Age Dual Language Learners

Jeannette Mancilla-Martinez,[a] Jin Kyoung Hwang,[b] Min Hyun Oh,[a] and Elena Lauren Pokowitz[a]

Purpose: Elementary-age dual language learners (DLLs) from Spanish-speaking homes in the United States are often characterized as at risk for low vocabulary in both Spanish and English. This longitudinal study examined conceptually scored receptive and expressive vocabulary development among DLLs from Spanish-speaking, low-income homes and investigated patterns of language responses on the conceptually scored measures.
Method: DLLs in kindergarten and second grade (N = 118) attending school in the Southeast region of the United States were followed for three consecutive academic years and assessed on measures of receptive and expressive conceptually scored vocabulary.
Results: Individual growth modeling, using raw scores, revealed positive vocabulary growth over time, with *above average* performance relative to national norms. However, initial conceptually scored receptive vocabulary scores were higher than conceptually scored expressive vocabulary scores. Furthermore, DLLs' conceptually scored receptive—but not expressive—vocabulary rate of growth was more rapid compared to national norms. Finally, DLLs' patterns of language responses revealed a shift toward more English over time, but Spanish continued to be used through fourth grade.
Conclusion: Results contradict deficit-driven views about DLLs' vocabulary skills and underscore the utility of conceptually scored vocabulary measures to assess vocabulary development among elementary-age DLLs to more comprehensively account for the linguistic assets they bring to learning.

By definition, dual language learners (DLLs)—hereon referring to children from households in which a language other than English is used—receive input in more than one language, and their language environments are thus fundamentally different from those of monolingual children (i.e., monolinguals hear and use only one language across contexts while DLLs hear and use more than one language, to varying extents, across contexts). As such, DLLs' vocabulary knowledge is distributed across their languages (Bedore et al., 2005; Mancilla-Martinez et al., 2011; Pearson et al., 1993). Indeed, researchers have long cautioned against the conceptualization of bilinguals as two monolinguals in one person (e.g., Grosjean, 1989). Yet, in the United States, DLLs continue to be assessed with monolingual measures that are designed for and normed on monolingual children. Given typical assessment practices, it is not surprising that many studies report that DLLs from Spanish-speaking homes—the largest and fastest growing segment of the school-age population (McFarland et al., 2019)—often demonstrate low Spanish and low English vocabulary knowledge (e.g., Hoff, 2018; Mancilla-Martinez & Lesaux, 2011). Of concern, theoretical and empirical studies suggest important links between vocabulary knowledge and key educational outcomes, such as reading comprehension (e.g., Gough & Tunmer, 1986; RAND Reading Study Group, 2002; Scarborough, 2001). Perhaps more importantly, among children who struggle with reading comprehension, DLLs struggle with language comprehension—most typically proxied via vocabulary measures—significantly more than non-DLLs

[a]Peabody College of Education and Human Development, Vanderbilt University, Nashville, TN
[b]School of Education, University of California, Irvine

Correspondence to Jeannette Mancilla-Martinez: jeannette.mancilla-martinez@vanderbilt.edu

Editor-in-Chief: Stephen M. Camarata
Editor: Sudha Arunachalam

Received February 10, 2020
Revision received March 27, 2020
Accepted June 8, 2020
https://doi.org/10.1044/2020_JSLHR-20-00056

Disclosure: The authors have declared that no competing interests existed at the time of publication.

(Spencer & Wagner, 2017). While work to date raises concerns about DLLs' vocabulary, most studies have been cross-sectional. Thus, gaining a better understanding of DLLs' longitudinal vocabulary *development* is not only a timely issue but also a matter of educational equity.

Given that understanding DLLs' vocabulary development during the formal school years is imperative to providing appropriate instructional support for reading comprehension, an open empirical question of basic—yet foundational—importance centers on conceptually scored vocabulary development among DLLs from Spanish-speaking homes across the elementary school years. Conceptually scored vocabulary approaches aim to proxy DLLs' overall understanding in an integrated way by tapping Spanish and English vocabulary simultaneously rather than utilizing a monolingual Spanish vocabulary measure and then a separate monolingual English vocabulary measure. This longitudinal study, spanning kindergarten through fourth grade, was designed to contribute to our understanding of conceptually scored receptive (i.e., comprehension) and expressive (i.e., production) vocabulary development among elementary-age DLLs from Spanish-speaking homes in a new destination state in which instruction is provided only in English.

Though DLLs are by no means a new population of learners in U.S. schools, in recent years, new destination states (e.g., Georgia, South Carolina, and Tennessee) have experienced unprecedented DLL population growth (Gándara & Mordechay, 2017; McFarland et al., 2019). Following national trends, the majority of DLL populations in new destination states come from Spanish-speaking, low-income homes and receive English-only instruction in schools (Gándara & Mordechay, 2017; Migration Policy Institute, 2015). Amidst the rapidly changing school-age demographics, educators in new destination states who have not historically served DLLs are comparatively less prepared to meet and support DLLs' language and literacy development and academic success.

DLLs are a heterogeneous group of students, even though a distinguishing feature is that they hear (and often use) more than one language. Indeed, DLLs vary in many aspects, including proficiency in their first language (e.g., Spanish), proficiency in their second language (i.e., English in this case), and years of instruction they received in their first and second languages (e.g., August & Shanahan, 2006). While some DLLs enter kindergarten with full proficiency in English, others enter with limited but developing English proficiency needed to succeed in U.S. schools. Understanding vocabulary developmental trajectories based on measures designed to simultaneously proxy Spanish and English vocabulary can offer valuable insight for researchers and educators seeking assessments for evaluating language skills of DLLs who vary in Spanish and English proficiency. This insight can, in turn, potentially serve as a catalyst for efforts aimed at increasing the sheer feasibility of administering conceptually scored vocabulary assessments (e.g., addressing the shortage of bilingual teachers nationwide). Additionally, patterns of language use are dynamic. Some children evidence stable language use patterns between kindergarten and eighth grade, while others shift toward more use of English or their home language (Mancilla-Martinez & Kieffer, 2010). Thus, an analysis of DLLs' relative use of Spanish compared to English on the conceptually scored vocabulary measures across the elementary grade years was conducted in the current study. By examining the patterns of DLLs' language responses, we aimed to document potential shifts in language use across this critical developmental period in which DLLs' language environments become increasingly English based. More uniquely, this analysis would also shed light on the extent to which Spanish continues to be used, speaking to the utility (or lack thereof) of conceptually scored measures.

Vocabulary Development Using Single-Language Measures

Extant longitudinal studies have typically found that elementary-age DLLs from low-income, Spanish-speaking homes tend to evidence Spanish and English vocabulary levels that raise concerns (Mancilla-Martinez & Lesaux, 2011, 2017). For example, in a sample of DLLs from Spanish-speaking homes instructed in English and followed from preschool to fifth grade, Mancilla-Martinez and Lesaux (2011) found positive but decelerating Spanish and English vocabulary growth. Specifically, DLLs' Spanish vocabulary was not only initially lower than their English vocabulary, but their vocabulary growth began to decelerate sooner in Spanish (8 years of age) compared to English (10 years of age). By fifth grade, DLLs evidenced considerably lower Spanish than English vocabulary levels, and DLLs' Spanish and English vocabulary levels remained well below the national average. More recently, Mancilla-Martinez and Lesaux (2017) found a similar pattern in a 6-year longitudinal study that investigated growth rates of Spanish-speaking DLLs' word reading and vocabulary skills and their predictive roles on English reading comprehension achievement. Using single-language vocabulary measures in Spanish and English, the study found a developmental shift in the contribution of word- and language-based skills, wherein the English language–based skills (e.g., vocabulary knowledge) emerged as a robust contributor to reading comprehension in eighth grade, but not fifth grade. However, word- and language-based skills in Spanish were not robust predictors of reading comprehension at either fifth grade or eighth grade. Although these studies accounted for DLLs' achievement in both languages, the measures used were designed for and normed on Spanish and English monolinguals, not Spanish–English DLLs. Thus, DLLs have typically been treated as two monolinguals in one person, despite the fact that their vocabulary knowledge is distributed across their languages.

More research is needed to comprehensively understand the relationships between DLLs' language development in both Spanish and English and their academic trajectories. Empirical evidence underscores the robust contribution of vocabulary knowledge to later reading comprehension outcomes for both monolingual learners and DLLs (e.g., Catts et al., 2005; Ouellette & Beers, 2010; Paris & Hamilton, 2009), necessitating a better understanding of vocabulary development patterns among DLLs to help inform instruction and equitable assessment practices for this growing group of learners. However, a major limitation of work to date continues to be the scarcity of valid measures designed for and normed on DLLs in the United States (Espinosa, 2010; Peña & Halle, 2011). This poses a challenge in not only gaining an accurate understanding into this population's vocabulary achievement and development but also in distinguishing a language difference from a language disorder (Ortiz & Artiles, 2010; Paradis et al., 2011).

Vocabulary Development Using Conceptually Scored Vocabulary Measures

Vocabulary assessment of DLLs—who have more than one linguistic resource—is challenging and complex, but crucial. A long line of research has cautioned against expecting DLLs to show the same level of vocabulary knowledge in each of their languages as that of monolinguals in those languages (e.g., Bedore et al., 2005; Grosjean, 1989). Indeed, DLLs' vocabulary knowledge tends to be distributed between the two languages, given variations in exposure, use, and proficiency (Oller & Pearson, 2002; Oller et al., 2007). Nonetheless, measures designed for monolingual speakers of either Spanish or English continue to be used with Spanish-speaking DLLs. This raises concerns, as there is general research consensus that DLLs in the United States with limited vocabulary knowledge are likely to experience long-term difficulties with English reading comprehension (e.g., Kieffer, 2008; Mancilla-Martinez & Lesaux, 2017; Spencer & Wagner, 2017). Given the distributed nature of DLLs' vocabulary knowledge, the home language must be considered, in addition to English, when assessing their vocabulary.

A growing body of research reveals that DLLs' conceptually scored vocabulary generally offers a more comprehensive account of their language skills (Anaya et al., 2018; Hwang et al., 2019; Mancilla-Martinez et al., 2018; Peña & Halle, 2011). Conceptually scored vocabulary approaches attend to DLLs' proficiency in an integrated way, such that the focus is on students' knowledge of known concepts, not on language-specific labels. In other words, rather than utilizing a monolingual Spanish vocabulary measure and then a separate monolingual English vocabulary measure, conceptual scoring taps Spanish and English vocabulary simultaneously. In this way, DLLs can respond in either language, such that they receive credit for known concepts regardless of whether the response is provided in Spanish or English. Conceptual scoring itself is not new, with researchers such as Pearson and colleagues (e.g., Pearson et al., 1993, 1995) being among the first to propose adapting monolingual measures designed for infants and toddlers to derive a conceptual vocabulary score. Conceptual scoring aims to proxy DLLs' overall understanding, which is a foundational skill in comprehension. Vocabulary knowledge is not simply about learning individual word labels and associated meanings. Instead, vocabulary knowledge reflects learners' *conceptual* knowledge that is built around the labels and meanings of words (Stahl & Bravo, 2010). In addition to addressing the distributed nature of DLLs' vocabulary knowledge, conceptually scored vocabulary approaches are grounded in scientific understandings of bilingual language acquisition.

The revised hierarchical model (Kroll & Stewart, 1994) suggests that a conceptual system is shared between two languages and is likely represented by separate labels, in which proficiency in one language supports proficiency in another as influenced by an individual's experiences with each language. Thus, conceptually scored vocabulary approaches represent a promising alternative compared to the stubborn reliance on single-language assessments to understand DLLs' vocabulary knowledge. Indeed, recent evidence suggests that conceptual development might be similar between monolingual and bilingual children, wherein encounters with *concepts,* rather than *labels,* contribute to early vocabulary development (Jardak & Byers-Heinlein, 2019). This is aligned with empirical evidence suggesting that bilingual and monolingual children possess comparable levels of conceptual vocabulary (e.g., Bedore et al., 2005; Mancilla-Martinez et al., 2018; Pearson et al., 1993, 1995; Thordardottir et al., 2006), contradicting a misled, deficit orientation of DLLs' vocabulary achievement and learning capability. However, this work has been conducted with toddlers and preschoolers, leaving open questions about the utility of conceptually scored approaches for the large and growing population of elementary-age DLLs. In other words, extant research has not examined conceptually scored vocabulary among school-age children, even though this developmental stage represents a critical time during which high-stakes decisions are made (e.g., special education placement).

Furthermore, research with DLLs from Spanish-speaking homes generally finds stronger receptive—compared to expressive—vocabulary skills when each language domain is considered separately (e.g., Gibson et al., 2012; Gross et al., 2014; Oller et al., 2007; Windsor & Kohnert, 2004), in line with findings from English monolinguals (Benedict, 1979). However, findings are somewhat mixed when considering conceptually scored receptive compared to expressive vocabulary. Mancilla-Martinez et al. (2018) found that, among children from Spanish-speaking homes attending English-only preschool, conceptually scored vocabulary knowledge in both the receptive and expressive domains was on par with that of their monolingual peers. In contrast, in a sample of primary grade children receiving a mix of formal instruction in English and Spanish, Gross et al. (2014) reported that conceptually scored vocabulary only removed the significant difference between monolinguals and DLLs in receptive, but not expressive, vocabulary. Gross et al. (2014) hypothesize that this might be attributed to the receptive–expressive gap commonly identified in the literature, domestically and internationally, in which DLLs perform better receptively than expressively (e.g., Gibson et al., 2012; Keller et al., 2015; Oller et al., 2007; Sheng et al., 2011). More research comparing receptive and expressive conceptually scored vocabulary is warranted. There is practical utility in engaging in these investigations as testing time is constrained in schools. Thus, it is important to examine whether DLLs tend to need more support with the receptive or expressive domain as this has direct implications for the selection of assessments. Additionally, conceptually scored vocabulary offers a unique opportunity to examine potential shifts in DLLs' use of Spanish and English across the elementary-grade years, which can directly help inform the extent to which use of measures that allow DLLs to respond in either language is warranted (i.e., do DLLs continue to use Spanish over time?). That is, if DLLs do not use Spanish at all over time, the case for conceptually scored measures is markedly reduced. In contrast, learning that DLLs indeed use Spanish would suggest that use of conceptually scored measures during the school-age years are needed to more comprehensively understand their vocabulary development.

Language Use Shifts

Research shows general shifts toward English preference at school entry among DLLs (Lutz, 2008; Tse, 2001). Furthermore, increased English proficiency in later childhood and adolescence can lead to attrition of the native language (Anderson, 2012; Tran, 2010), as opportunities to hear and use one's home language become limited for DLLs (Hammer et al., 2014). The possibility of language loss is heightened for children for whom the use of English at school and lack of instructional support in the home language lead to a more drastic language shift at the time of school entry (Castilla-Earls et al., 2019; Ijalba, 2015), a common occurrence in the U.S. context. Associating the first language with specific environments, such as the home or smaller communities, can also hinder concept development in the child's native language (Anderson, 2012). Likewise, with increased exposure to formal schooling in English, students from Spanish-speaking homes with more experience in English tend to show declines in Spanish expressive and receptive vocabulary, as they become more exposed to hearing and speaking English at school (Anderson, 2012; Castilla-Earls et al., 2019) and show overall improvements in English proficiency (Castilla-Earls et al., 2019; Guo & Schneider, 2016). Some work finds that, by second grade, DLLs have generally transitioned to English as their primary and preferred language (Castilla-Earls et al., 2019). At the same time, we also know that patterns of language use are dynamic, such that a majority of children evidence stable language use patterns between kindergarten and eighth grade, while others shift toward more use of English or of their home language (Mancilla-Martinez & Kieffer, 2010). It seems obvious that continued development and proficiency in both languages is not impossible, but the current sociolinguistic environment in the United States appears to pressure individuals to focus solely on the acquisition of English to optimize academic and occupational outcomes (Anderson 2012; Castilla-Earls et al., 2019).

In summary, most studies on conceptual vocabulary scoring to date have been cross-sectional and have focused on the early childhood level (Anaya et al., 2018; Core et al., 2013; Gross et al., 2014; Hoff et al. 2012; Mancilla-Martinez et al., 2018; Peña & Halle, 2011). Very little is known about DLLs' vocabulary *development* patterns using conceptual scoring at the elementary school level, when high-stakes educational decisions—including special education recommendations—are typically made. We know that DLLs from Spanish-speaking homes tend to utilize both Spanish and English to varying extents and conceptually scored vocabulary measures leverage this and allow DLLs to respond in either language (Hwang et al., 2019; Mancilla-Martinez et al., 2019; Oller & Pearson, 2002; Oller et al., 2007; Peña et al., 2002). By utilizing measures that are grounded in scientific understandings of bilingual, rather than monolingual, language acquisition to document patterns of vocabulary development among elementary-age DLLs from Spanish-speaking homes in the United States, this study offers unique insight into this population's vocabulary knowledge. Additionally, by documenting receptive and expressive vocabulary, this study offers both instructional and assessment implications. Finally, this study affords an opportunity to examine potential shifts in language use among DLLs across the elementary grades, allowing for a more nuanced understanding of whether measures that employ bilingual scoring, and namely that allow students to respond in Spanish if they prefer to, have utility during this critical developmental period. This study is guided by the following two research questions:

1. What are Spanish–English DLLs' patterns of conceptually scored receptive and expressive vocabulary growth across the elementary academic years and how do students' patterns of growth compare to national norms?
2. Over the course of the elementary grades, to what extent do Spanish–English DLLs provide correct responses in Spanish compared to English on conceptually scored receptive and expressive vocabulary? Specifically, is there more Spanish use over time, more English use over time, or relatively stable use of both languages over time?

Based on the extant research reviewed focused on younger populations, we hypothesize that elementary-age Spanish–English DLLs' conceptually scored vocabulary will evidence patterns of growth that are on par with national norms, which would contrast typical findings when monolingual measures are used. Despite the limited research base comparing receptive and expressive conceptually scored vocabulary, we further hypothesize that receptive vocabulary levels will likely exceed expressive vocabulary levels. This hypothesis is largely drawn from previous work that emphasizes the need to afford DLLs greater opportunities to use language (e.g., Mancilla-Martinez et al., 2019). Finally, given DLLs' English-only instructional context, we hypothesize that more English will be used across the elementary grade years (e.g., Castilla-Earls et al., 2019). At the same time, we expect DLLs to continue to use Spanish to some extent, which would lend support for the utility of conceptually scored measures during this developmental stage.

Method

Participants

The sample, taken from a larger study, consists of kindergarten and second-grade students (N = 118, 55% female) recruited from three elementary schools in a large urban school district in the Southeastern region of the United States in the beginning of the 2016–2017 academic year (Year 1) and followed through the end of the 2018–2019 academic year (Year 3). In the larger study, all students in kindergarten, second grade, and fourth grade at the three participating schools received recruitment letters in August 2016. Of the 266 students who returned letters, 190 students (59 kindergartners, 59 second graders, 72 fourth graders) indicated that they spoke Spanish at home and were eligible to participate in the study.

For this longitudinal study, we followed students in kindergarten and second grade for three consecutive years (2016–2019). Fourth-grade students from Year 1 were not included in the current study because they graduated from their elementary schools and went to middle schools (i.e., fifth grade) in Year 2 (i.e., 2017–2018 academic year). In the spring of Year 1, five students withdrew from the study. In Year 2, six students moved schools, leaving 107 possible participants. Seventy-three students' parents provided continuing consent for their children to participate in this study, with 38 in first grade (i.e., kindergarten cohort) and 35 in third grade (i.e., second-grade cohort). In the fall of Year 3 (i.e., 2018–2019 academic year), 52 students' parents provided consent for their children to continue participating in the study, with 27 in second grade (i.e., kindergarten cohort) and 25 in fourth grade (i.e., second-grade cohort). Early in the spring of Year 3, we distributed a last round of parent consent forms to re-invite participants. We were able to recruit 60 students in total, with 34 in second grade and 26 in fourth grade for the fall of Year 3.

The number of students we were able to assess at each wave of data collection are reported in Table 1. Attrition varied by year, at 38% from Year 1 to Year 2 and 29% from Year 2 to Year 3. However, there was an increase in sample size (by 15%) from fall of Year 3 to spring of Year 3. We conducted a series of t tests to examine whether there were vocabulary differences in test scores for students who did and did not participate in more than one study year. There was no statistically significant difference in fall of Year 1 conceptually scored vocabulary scores for those who only participated in Year 1 and those who continued to participate in Year 1 and Year 2: receptive vocabulary, $t(116) = -0.16$, $p = .87$, and expressive vocabulary, $t(116) = 1.04$, $p = .30$. In addition, there was no statistically significant difference in fall of Year 2 conceptually scored vocabulary scores for those who participated in Year 2 and those who participated in both Year 2 and Year 3: receptive vocabulary, $t(63) = -0.09$, $p = .93$, and expressive vocabulary, $t(63) = -0.44$, $p = .66$. Finally, 70% percent of our participants (n = 83) were formally classified as limited English proficient (LEP) by their school district based on WIDA Consortium's Assessing Comprehension and Communication in English State-to-State for English Language Learners placements tests (http://www.wida.us/assessment/ACCESS20.aspx) in fall of Year 1.

Eighty-nine percent of our participants' parents (n = 105) provided demographic information via a parent questionnaire. According to the parent reports, 86% of students were born in the United States, and the rest were born outside the United States: Mexico (3%), Cuba (2%), El Salvador (2%), and other countries in Latin America (7%). In contrast, only 5% of their parents were born in the United States, with 54% from Mexico, 13% from El Salvador, 9% from Guatemala, 6% from Cuba, and the rest (13%) from other countries in Latin America. Through this parent questionnaire, parents also reported on eight questions about patterns of language use in the home among their children and other family members. Parents rated each question on a 5-point scale: 1 = *only Spanish*, 2 = *mostly Spanish*, 3 = *English and Spanish equally*, 4 = *mostly English*, and 5 = *only English*. On average, parents reported that the language their children heard in the home (i.e., language exposure) was Spanish-dominant (M = 2.32, SD = 0.83). The average rating for the language children used in the home with other family members was slightly higher (M = 2.64, SD = 1.02), indicating that more English was used, though children's language use in the home still remained Spanish-dominant. Concerning parental education, 69% of parents reported having less than a high school education, 19% reported completing high school or high school equivalency certificate, and 12% reported having more than a high school education. Finally, of the parents who reported their family income level (n = 63, 60%), the families had an income-to-need ratio at the poverty level (0.95), on average. The official U.S. definition of poverty is based on a comparison of a household's income to an income threshold level that varies by family size and composition (see https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html). We computed the income-to-need ratio by dividing total family income by the poverty threshold for the appropriate family size.

Measures

Students' conceptually scored vocabulary knowledge was assessed one-on-one by trained Spanish–English bilingual research assistants 5 times during the 3-year period: fall and spring of Year 1, fall of Year 2, and fall and

Table 1. Sample means and standard deviations by grade-level cohort and limited English proficient (LEP) status.

			Conceptually scored receptive vocabulary					Conceptually scored expressive vocabulary				
				Raw scores		Standard scores			Raw scores		Standard scores	
Cohort		Year	*n*	*M*	*SD*	*M*	*SD*	*n*	*M*	*SD*	*M*	*SD*
Kindergarten cohort	Total	Fall Year 1	59	67.37	13.17	108.46	10.08	59	51.64	10.79	106.08	10.56
		Spring Year 1	55	82.80	16.73	116.31	13.57	55	58.93	10.03	109.25	10.04
		Fall Year 2	30	82.07	15.31	111.40	13.27	30	61.83	11.43	107.53	12.09
		Fall Year 3	26	97.85	17.71	115.69	12.99	27	72.63	10.53	109.59	10.51
		Spring Year 3	33	107.58	17.23	120.18	11.99	33	80.82	10.95	114.85	10.97
	Non-LEP students	Fall Year 1	19	67.79	10.32	108.58	7.97	19	53.63	10.31	108.26	10.45
		Spring Year 1	17	83.41	14.04	116.47	11.52	17	61.00	9.24	111.59	8.28
		Fall Year 2	6	89.33	12.23	116.67	9.89	6	71.00	15.79	116.33	15.56
		Fall Year 3	8	99.75	8.99	116.50	8.11	8	72.63	11.76	107.63	9.24
		Spring Year 3	11	115.55	19.95	125.55	12.16	11	82.00	11.20	114.00	10.18
	LEP students	Fall Year 1	40	67.18	14.45	108.40	11.04	40	50.70	11.02	105.05	10.58
		Spring Year 1	38	82.53	17.97	116.24	14.53	38	58.00	10.35	108.21	10.66
		Fall Year 2	24	80.25	15.67	110.08	13.84	24	59.54	9.12	105.33	10.31
		Fall Year 3	18	97.00	20.62	115.33	14.85	19	72.63	10.32	110.42	11.13
		Spring Year 3	22	103.59	14.60	117.50	11.22	22	80.23	11.04	115.27	11.55
Second-grade cohort	Total	Fall Year 1	59	91.03	17.65	111.88	13.61	59	69.86	11.97	108.24	12.25
		Spring Year 1	57	101.23	15.93	115.25	12.12	57	75.75	13.09	110.70	13.37
		Fall Year 2	35	100.60	16.37	111.74	13.40	35	78.17	12.97	110.20	13.54
		Fall Year 3	25	112.52	19.03	113.08	15.63	25	83.48	14.88	106.00	15.12
		Spring Year 3	26	120.96	20.31	115.88	18.24	26	90.27	15.32	107.85	16.40
	Non-LEP students	Fall Year 1	16	92.63	17.07	111.25	13.55	16	71.81	15.50	108.69	14.72
		Spring Year 1	16	102.31	13.24	114.88	10.25	16	81.69	13.13	116.06	12.89
		Fall Year 2	7	103.43	12.92	112.14	10.95	7	85.71	13.17	116.29	14.73
		Fall Year 3	5	120.20	20.47	116.20	17.02	5	80.80	17.34	101.60	15.66
		Spring Year 3	5	125.00	20.19	120.80	19.10	5	93.60	11.57	110.60	12.62
	LEP students	Fall Year 1	43	90.44	18.02	112.12	13.79	43	69.14	10.49	108.07	11.40
		Spring Year 1	41	100.80	17.00	115.39	12.90	41	73.44	12.48	108.61	13.12
		Fall Year 2	28	99.89	17.26	111.64	14.12	28	76.29	12.45	108.68	13.06
		Fall Year 3	20	110.60	18.70	112.30	15.64	20	84.15	14.63	107.10	15.19
		Spring Year 3	21	120.00	20.71	114.71	18.31	21	89.48	16.22	107.19	17.38

Note. Two students in fall Year 1 and four students in spring Year 3 had standard scores of > 145. They were arbitrarily given a score of 146 for this descriptives table. LEP students' status was determined in the beginning of the study.

spring of Year 3. We could not assess students in spring of Year 2 due to conflicts with the district-wide standardized assessment schedule.

Spanish–English Conceptually Scored Receptive and Expressive Vocabulary

The Receptive One-Word Picture Vocabulary Test–4: Spanish-Bilingual Edition (ROWPVT-4: SBE; Martin, 2013b) was used to assess children's Spanish–English conceptually scored receptive vocabulary knowledge, while the Expressive One-Word Picture Vocabulary Test–4: Spanish-Bilingual Edition (EOWPVT-4: SBE; Martin, 2013a) was used to assess children's Spanish–English conceptually scored expressive vocabulary knowledge. There are an equal number of items on both measures (180 in each). For both assessments, the standardized administration procedure was strictly followed. Children were presented with the first item of the age-recommended starting point, per the established protocol from the publishers, in Spanish or English. Children's language dominance was determined based on parent or teacher report. Incorrect response in one language (e.g., Spanish) prompted the examiner to ask the child again in the other language (e.g., English). Thus, potential switches in prompting occurred as soon as the child made an error. The publisher reports the median internal consistency reliability coefficients as .95, and raw scores were used in the analyses to make direct comparisons feasible.

The ROWPVT-4: SBE (Martin, 2013a) allowed for the assessment of DLLs' receptive knowledge in either language. The items are ordered by increasing difficulty, beginning with the easiest concepts and ending with less frequently encountered concepts. Each item displays four pictures. The child is asked which of the four pictures is the target word and prompted to point to the correct picture. The task is discontinued when the child makes four errors within six consecutive responses. Similarly, the EOWPVT-4: SBE (Martin, 2013b) allowed for the assessment of DLLs' expressive vocabulary knowledge in either language. Children were presented with a target picture and prompted in Spanish or English to name the item. The items are ordered by increasing difficulty, beginning with the easiest concepts and ending with less frequently encountered concepts.

Each item displays a picture. The child is asked, "What is this?" or "¿Qué es esto?" depending on language dominance. The task is discontinued when the child fails six consecutive items.

Both ROWPVT-4: SBE and EOWPVT-4: SBE include items (180 items per test) that have been arranged in a developmental sequence normative to the Spanish–English bilingual sample. Of note, the test items are not designed to be the same between the receptive expressive versions, and they consist of primarily nouns, with some gerunds, verbs, and modifiers. Additionally, the ROWPVT-4: SBE and EOWPVT-4: SBE are standardized assessments with standardized protocols (as described above), which allows us to generate standard scores to understand DLLs' performance relative to the Spanish–English bilingual population in the United States. The publisher reports that the two assessments have been normed on more than 1,200 Hispanic individuals, ranging from 2 to 93 years of age and closely approximating the demographics (e.g., region, gender, parent education, and residence) of the U.S. Hispanic population. Lastly, previous studies have utilized different methods to generate conceptual scores, most often adapted from monolingual measures (e.g., Core et al., 2013; Mancilla-Martinez & Vagh, 2013), which offer no normative information specifically for elementary-age Spanish–English DLLs in the United States. As such, utilizing standardized assessments—designed for and normed on Spanish–English bilinguals in the United States—can offer valuable insight into the language skills of this growing population of learners. We underscore that the findings from this longitudinal study cannot be gleaned by simply referencing the published norms that rely on a cross-sectional sample.

Time

For the multilevel analysis of change analyses, we coded each wave of data collection in months (i.e., fall Year 1 = 0 month, spring Year 1 = 7 months, fall Year 2 = 12 months, fall Year 3 = 24 months, and spring Year 3 = 31 months). Time is a Level 1 variable in our analyses that indicates the time since the start of the study (i.e., fall Year 1).

English Proficiency Status

To examine whether DLLs' proficiency in English influenced their conceptually scored vocabulary development, we created a student-level (Level 2) dummy variable for LEP students for our analyses. We treated language status as time-invariant, reflecting students' language status at study entry (i.e., fall Year 1) because we could not track language status information of students who did not participate in the study in Years 2 and 3.

Grade-Level Cohort

To control for different grade levels in our analyses, student-level (Level 2) dummy variables were created for each grade-level cohort (i.e., kindergarten cohort, second-grade cohort). The reference group in our analyses were the kindergarten cohort students.

Analytic Approach

To answer our research questions, we conducted multilevel models for change (Singer & Willett, 2003) and descriptive analyses. To address our first research question, we ran statistical models separately for conceptually scored receptive and expressive vocabulary. These models allowed us to use all waves of data from each student to create a model of conceptually scored vocabulary growth over three academic years. The data were prepared in a person-period data set, such that each student had up to five rows of data. All data analyses were conducted using STATA 14 software with the XTMIXED command. The hypothesized multilevel model for change for our first research question was the following:

Level 1:

$$\text{Conceptually } \widehat{\text{scored}} \text{ vocabulary} = \pi_{0i} + \pi_{1i}\text{Time}_{ij} + \varepsilon_{ij} \quad (1)$$

Level 2:

$$\pi_{0i} = \gamma_{00} + \gamma_{01}\text{Second Grade}_i + \zeta_{0i} \quad (2)$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11}\text{Second Grade}_i \quad (3)$$

where $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, and $\zeta_{0i} \sim N(0, \sigma_\varsigma^2)$.

The coefficient γ_{00} represents the average score for students in the kindergarten cohort at the first measurement point (fall Year 1); γ_{10} represents the average initial slope for students in the kindergarten cohort. The random effect ε_{ij} is a Level 1 residual for student i at time j and is assumed to be drawn from a normal distribution with a mean of zero and variance s_ε^2. Random effect z_{0i} refers to Level 2 residual for the intercept and is hypothesized to be drawn from a normal distribution with a mean of zero, unknown variance s_ς^2. The parameter γ_{01} refers to the baseline difference in conceptually scored vocabulary scores between the kindergarten cohort and the second-grade cohort of students. The parameter γ_{11} represents the difference in rate of growth between the kindergarten cohort and the second-grade cohort of students. Students' conceptually scored receptive and expressive vocabulary scores from five waves were used as dependent variables to answer our first research question.

Based on the multilevel modeling results, we plotted prototypical growth trajectories for students in both grade-level cohorts and compared them to the nationally normed median scores reported in the technical manual of the vocabulary measures (Martin 2013a, 2013b). To quantify and interpret the differences in students' predicted scores and national norms, we calculated effect sizes for all time points. We obtained the effect sizes by dividing the difference between predicted conceptually vocabulary outcomes and nationally normed median scores by the standard deviation of our sample at each wave. This way, we were able to examine how many standard deviation units the predicted means of our sample were apart from the national norms.

To address our second research question, we examined DLLs' correct responses in Spanish and in English on the conceptually scored receptive and expressive vocabulary measures, attending to potential changes in the language of correct responses over time.

Results

Preliminary Descriptive Analyses

Table 1 displays means and standard deviations of conceptually scored receptive and expressive vocabulary scores by students' grade-level cohort. The average standard scores reveal that DLLs performed above the mean of 100, which is in the average to above average range and thus on par with national norms. As shown in Table 1, DLLs evidenced different patterns of conceptually scored receptive and expressive vocabulary, with generally higher raw and standard receptive scores.

Figure 1 illustrates the observed receptive–expressive gap across grades. Furthermore, DLLs in the second-grade cohort had higher scores on both conceptually scored vocabulary measures compared to the kindergarten cohort. In comparing DLLs' vocabulary scores in the fall and spring of Year 1, as well as in the fall and spring of Year 3, DLLs evidenced improvement in their conceptually scored vocabulary knowledge during the academic year. Finally, DLLs' scores in the spring of Year 1 and fall of Year 2 suggest they experienced vocabulary summer setback/plateau, particularly in conceptually scored receptive vocabulary. On conceptually scored expressive vocabulary, DLLs' scores did not appear to regress, though growth during the summer was slower compared to growth during the academic year.

Research Question 1: Conceptually Scored Receptive and Expressive Vocabulary Growth

Tables 2 and 3 show results from a series of multilevel models for change predicting conceptually scored receptive and expressive vocabulary, respectively. Model E in Tables 2 and 3 accounts for linear time, grade-level cohorts, the linear time and second-grade interaction term, and LEP status. As shown, on average, the rate of growth in students' conceptually scored receptive and expressive vocabulary is positive. We tested for a quadratic term of time to examine whether the rate of growth accelerates or decelerates over time, but it was not statistically significant and inclusion did not improve the model fit (model results not shown). Furthermore, the second-grade cohort students had higher baseline scores than the kindergarten cohort students, on average. Model E in Tables 2 and 3 also indicates that the second-grade cohort students' rate of growth in conceptually scored vocabulary was slower compared to that of the kindergarten cohort students. Finally, the LEP status was also tested in our models. As shown in Table 2, Model E, for conceptually scored receptive vocabulary, the LEP status was not a statistically significant predictor, while the LEP status was negative and statistically significant in predicting students' conceptually scored expressive vocabulary (see Table 3, Model E; $B = -4.75$). This indicates that, on average, LEP students had lower conceptually scored expressive vocabulary baseline scores than non-LEP students at the beginning of the study. We tested for Time × LEP interaction term, but it was not statistically significant (model results not shown).

Figures 2 and 3 display prototypical graphs of conceptually scored receptive and expressive vocabulary growth,

Figure 1. Sample means in standard scores of conceptually scored receptive and expressive vocabulary. The blank bar graph refers to conceptually scored receptive vocabulary, and the dotted bar graph refers to conceptually scored expressive vocabulary. Red line indicates average standard score.

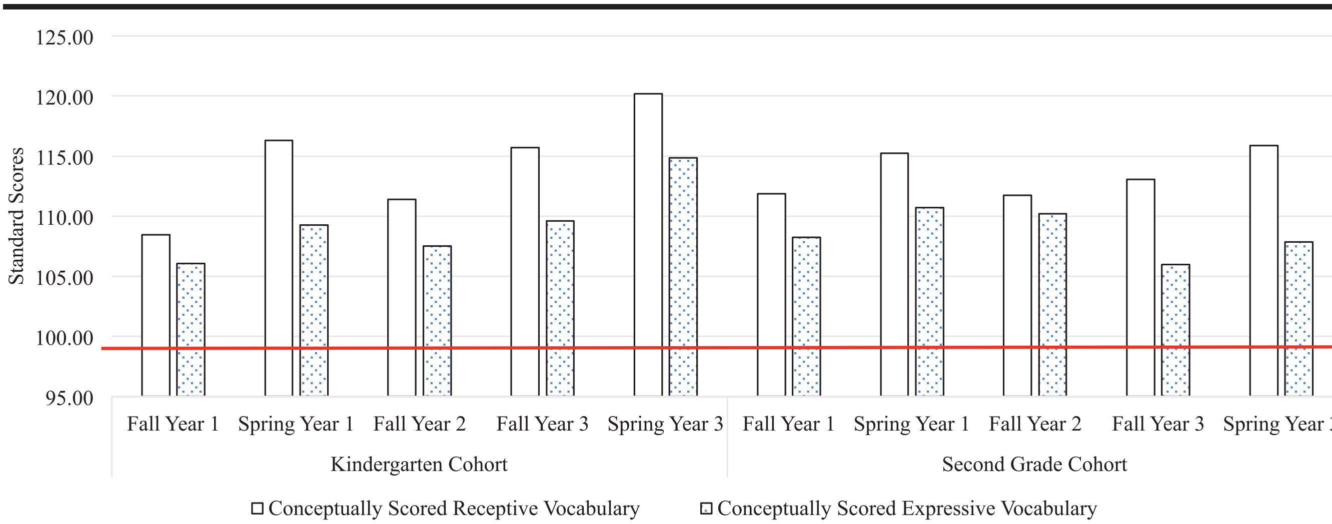


Table 2. Multilevel models for change predicting conceptually scored receptive vocabulary.

		Model A Unconditional means	Model B Unconditional growth	Model C Conditional growth with grade-level cohorts	Model D Conditional growth with interaction term	Model E Conditional growth with all predictors
Fixed effects	Intercept	90.99***	81.17***	71.07***	69.99***	71.41***
		(1.47)	(1.65)	(1.75)	(1.84)	(2.53)
	Time		1.06***	1.07***	1.19***	1.19***
			(0.07)	(0.07)	(0.09)	(0.09)
	Second grade			19.92***	22.18***	22.27***
				(2.32)	(2.59)	(2.60)
	Time × Second				−0.26*	−0.26*
					(0.13)	(0.13)
	LEP					−2.08
						(2.53)
Level 1 variance component	Residual	182.06***	176.79***	176.59***	176.23***	176.34***
		(17.83)	(16.87)	(16.82)	(16.77)	(16.78)
Level 2 variance component	Intercept	311.90***	211.59***	88.97***	89.42***	89.41***
		(76.06)	(43.37)	(27.96)	(27.79)	(27.92)
	Time	1.13	0.03	0.02	0.01	0.01
		(0.24)	(0.07)	(0.06)	(0.06)	(0.06)
Goodness of fit statistics	Deviance (−2LL)	3543.17	3429.80	3373.5	3369.69	3369.03
	N (students)	118	118	118	118	118
	N (observations)	405	405	405	405	405

Note. Raw scores were used for conceptually scored receptive vocabulary. Time × Second = interaction term between variables Time and Second Grade; LEP = limited English proficient.

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

Table 3. Multilevel models for change predicting conceptually scored expressive vocabulary.

		Model A Unconditional means	Model B Unconditional growth	Model C Conditional growth with grade-level cohorts	Model D Conditional growth with interaction term	Model E Conditional growth with all predictors
Fixed effects	Intercept	67.32***	61.32***	52.72***	52.13***	55.37***
		(1.20)	(1.29)	(1.35)	(1.38)	(1.91)
	Time		0.76***	0.77***	0.86***	0.86***
			(0.05)	(0.05)	(0.06)	(0.06)
	Second grade			17.08***	18.32***	18.54***
				(1.85)	(1.95)	(1.91)
	Time × Second				−0.19*	−0.19*
					(0.09)	(0.094)
	LEP					−4.75*
						(1.98)
Level 1 variance component	Residual	50.88***	51.54***	51.28***	51.49***	51.37***
		(5.18)	(5.01)	(4.99)	(5.00)	(5.00)
Level 2 variance component	Intercept	201.50***	163.68***	79.55***	79.92***	75.20***
		(38.49)	(25.78)	(14.93)	(14.94)	(14.32)
	Time	0.66*	0.07***	0.07***	0.07***	0.07***
		(0.12)	(0.03)	(0.03)	(0.03)	(0.03)
Goodness of fit statistics	Deviance (−2LL)	3184.20	3071.29	3008.73	3004.74	2999.09
	N (students)	118	118	118	118	118
	N (observations)	406	406	406	406	406

Note. Raw scores were used for conceptually scored expressive vocabulary. Time × Second = interaction term between variables Time and Second Grade; LEP = limited English proficient.

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

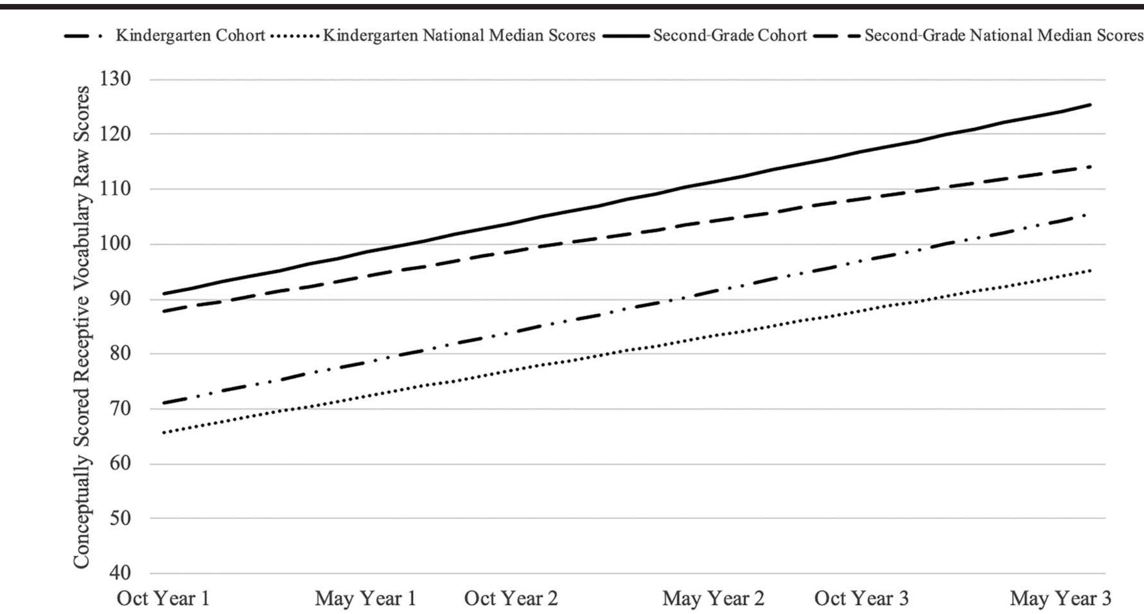
respectively, for the kindergarten cohort (dashed line with two dots) and second-grade cohort (solid line) in comparison to the national median scores (dotted line for kindergarten and long dashed line for second grade). The differences at each wave are expressed in effect sizes and are presented along the *x*-axis. We used results from Model C from Tables 2 and 3 to plot these graphs. We also plotted graphs with the significant Time × Second Grade interaction terms (i.e., results from Model D) and for conceptually scored expressive vocabulary with LEP status (i.e., results from Model E), but the trajectories were essentially indistinguishable. We thus display the trajectories without the interactions and without LEP status given our focus on DLLs' normative development on conceptually scored vocabulary.

There were two notable differences between the conceptually scored receptive and expressive vocabulary growth trajectories. First, the baseline scores (i.e., intercept) for conceptually scored receptive vocabulary were higher than those for the conceptually scored expressive vocabulary. Although there were an equal number of items in each measure, students tended to have higher scores on the receptive conceptually scored vocabulary measure compared to the expressive. The national median scores also reflected this trend. Second, for conceptually scored expressive vocabulary, the projected growth trajectories of our participants were on par with the national norm (kindergarten cohort average effect size = .10; second-grade cohort average effect size = .04). However, the growth trajectory for conceptually scored receptive vocabulary evidenced a different pattern. The initial performance for our sample was higher compared to the national norm (kindergarten cohort average effect size = .40; second-grade cohort average effect size = .18), and our sample also evidenced a more rapid rate of growth compared to the national norm. The vocabulary growth among our sample exceeds those of the national norms across all time points, and the gap between the two widened over time (kindergarten cohort average effect size = .46; second-grade cohort average effect size = .35).

Research Question 2: Student Language Use Patterns on Conceptually Scored Vocabulary

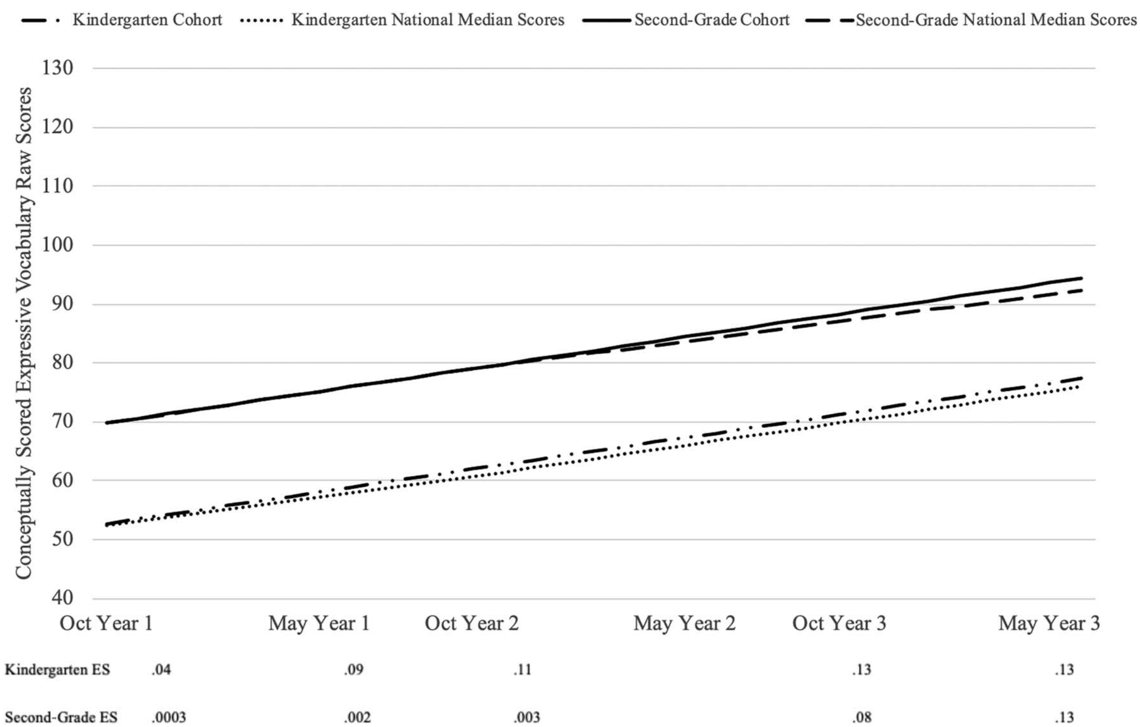
Table 4 displays the percentage of DLLs' correct responses in Spanish and English on the conceptually scored receptive and expressive vocabulary measures. When administering the vocabulary measures, the research assistants always started the first administered item using students' reported dominant language (Spanish or English). Of 118 students who contributed to at least one wave of data, 50 (42%) started the assessment in Spanish and 68 (58%) started the assessment in English, revealing a relatively even split in language dominance. As Table 4 shows, across all five data points for both conceptually scored receptive and expressive vocabulary, DLLs' accurate responses were predominantly in English, with one exception. Among the kindergarten cohort, accurate receptive vocabulary responses since study entry (fall of Year 1) were generally evenly

Figure 2. Prototypical graph of conceptually scored receptive vocabulary growth trajectories for kindergarten and second-grade cohort students in comparison to the national median scores. ES = effect size.



Kindergarten ES	.40	.37	.45	.51	.59
Second-Grade ES	.18	.27	.32	.45	.53

Figure 3. Prototypical graph of conceptually scored expressive vocabulary growth trajectories for kindergarten and second-grade cohort students in comparison to the national median scores. ES = effect size.



balanced in Spanish and English. Furthermore, at every wave of data collection, there was a notable shift toward fewer accurate responses in Spanish and toward more accurate responses in English (see changes in Table 4, indicated as Δ). The only exception to this shift was that the second-grade cohort was English-dominant since study entry.

To examine the robustness of our results, we conducted the same descriptive analyses constraining the sample to students who contributed at least three waves of data (n = 86). The percentages for fall and spring of Year 1 remain virtually unchanged, with a difference of less than 2% and with the overall pattern remaining the same. Finally, to understand the extent to which the option of responding in Spanish

Table 4. Percentage of students' responses in English in conceptually scored receptive and expressive vocabulary measures, with change (i.e., increase) in average percentages of students' responses in English from fall of Year 1 to spring of Year 3.

		Receptive vocabulary							Expressive vocabulary						
			Spanish			English				Spanish			English		
Cohort	Year	*n*	*M*	*SD*	Δ	*M*	*SD*	Δ	*n*	*M*	*SD*	Δ	*M*	*SD*	Δ
Kindergarten cohort	Fall Year 1	59	46.61	34.33	−36.17	53.39	34.33	36.17	59	37.05	37.91	−18.57	62.95	37.91	18.57
	Spring Year 1	55	37.65	31.38		62.35	31.38		55	24.55	35.18		75.55	35.25	
	Fall Year 2	30	34.87	29.57		64.99	29.61		30	25.57	29.79		72.47	31.15	
	Fall Year 3	26	40.99	30.39		59.01	30.39		27	13.88	16.16		86.28	15.60	
	Spring Year 3	33	10.44	6.81		89.56	6.81		33	18.48	16.09		81.52	16.09	
Second-grade cohort	Fall Year 1	59	30.71	25.79	−25.61	69.29	25.79	25.61	59	22.20	26.64	−5.66	77.80	26.64	5.66
	Spring Year 1	57	32.30	23.56		65.96	24.78		57	16.35	26.65		81.91	28.71	
	Fall Year 2	35	34.64	25.22		65.36	25.22		35	13.38	19.49		86.62	19.49	
	Fall Year 3	25	40.84	27.27		59.16	27.27		25	17.75	28.48		82.25	28.48	
	Spring Year 3	25	5.10	5.61		94.90	5.61		25	16.54	6.82		83.46	6.82	

Note. Students who contributed to at least one wave of data were included. Δ refers to change.

has utility for elementary-age DLLs, we further examined the extent to which DLLs responded only in English. Among the kindergarten cohort, on average, only one student responded 100% in English on the conceptually scored receptive measure across all waves. On the conceptually scored expressive vocabulary measure, on average, nine students did so. A very similar pattern emerged among the second-grade cohort. On average, only two students responded 100% in English on the conceptually scored receptive measure across all waves, and nine did so on the conceptually scored expressive vocabulary.

Discussion

Findings from this longitudinal study documenting the vocabulary achievement and development of elementary-age DLLs from Spanish-speaking homes offer an understanding of the linguistic resources DLLs bring to learning that has typically been overlooked. This study also helps shed light on patterns of Spanish and English use over the course of the elementary-grade years. Three key findings emerged that have timely implications for both theory and practice, particularly considering the growing number of DLLs from Spanish-speaking homes in new destination states. First, results revealed that DLLs not only achieved positive vocabulary growth but that their achievement levels were either on par with or surpassed national norms. Second, different patterns of development emerged for conceptually scored receptive and expressive vocabulary. Finally, an examination of language use patterns revealed increased use of English across the three study years but, more uniquely, also continued use of Spanish.

To our knowledge, this is the first study to longitudinally document elementary-age DLLs' vocabulary skills using measures specifically designed for and normed on DLLs in the United States with varying levels of proficiency in Spanish and English. Specifically, this study utilized conceptually scored measures in which DLLs could respond in either Spanish or English, given the distributed nature of vocabulary knowledge among bilinguals (Bedore et al., 2005; Oller & Pearson, 2002; Pearson et al., 1993). Our results suggest that the use of conceptually scored vocabulary assessments to measure DLLs' vocabulary knowledge affords DLLs the opportunity to showcase their linguistic knowledge more comprehensively than can be captured via reliance on monolingual measures. That is, DLLs were able to respond in Spanish or English rather than requiring them to provide responses in only one of their two languages (as is required when monolingual vocabulary measures are used). As our results illustrate, DLLs effectively utilized *both* Spanish and English. This, in turn, paints a starkly different—and positive—developmental understanding of their language skills that may help shift the prevalent, deficit-oriented view of DLLs' language achievement and growth. Rather than starting below and falling behind in their language development over time, conceptually scored measures help to better elucidate DLLs' linguistic knowledge. The conceptually scored measures we utilized are normed on Spanish–English bilinguals in the United States but rely on cross-sectional samples for deriving the norms. Our results illustrate the potential these standardized measures have for tracking DLLs' vocabulary development over time across the elementary-age years, an area that has been sorely understudied and that has limited our understanding of this growing population's linguistic abilities.

Our second finding offers additional insight, raising questions about the extent to which there is a receptive–expressive vocabulary gap, as some previous work suggests (e.g., Gross et al., 2014; Sheng et al., 2011), and, importantly, about the selection of vocabulary measures in school settings. Although there were the same number of items on the conceptually scored receptive and expressive vocabulary measures, students' raw scores (i.e., the number of correct items) on the receptive vocabulary measure were always higher than the scores on the expressive vocabulary measure (see Table 1). In line with previous studies in English-only instructional contexts that show DLLs' higher receptive, compare to expressive, vocabulary knowledge (e.g., Gibson et al., 2012; Kan & Kohnert, 2005), both the kindergarten and second-grade cohorts showed higher initial performance on conceptually scored receptive vocabulary when compared to conceptually scored expressive vocabulary. However, the growth trajectories of conceptually scored expressive vocabulary were still indistinguishable between our sample and the national norm. Notwithstanding, the DLLs in our study not only demonstrated higher initial performance on conceptually scored receptive vocabulary than the national norm but their rate of growth also exceeded that of the national norm. This means that, although DLLs' expressive vocabulary and achievement by no means raised concerns, their receptive vocabulary achievement and development outpaced national norms. In fact, the gap between DLLs' growth trajectories and national norms on receptive vocabulary continued to widen over time. Thus, DLLs' in this study evidenced comparatively higher achievement and development receptively than expressively.

The receptive–expressive vocabulary gap was also apparent when we compared the standard scores (see Figure 1), suggesting DLLs likely *understand* (i.e., receptive) more of their languages than they can *produce* (i.e., expressive). This finding is not unexpected and is in fact considered normative among young bilingual children (Keller et al., 2015). Among our elementary-age sample of DLLs, this also appears to be the case. Gibson et al. (2012) further suggest that the receptive–expressive vocabulary gap may form among Spanish–English DLLs either abruptly at entry into English-only instructional context or gradually over time as students are exposed increasingly to their second language (i.e., English). Of note and with implications for the selection of measures, more recent studies suggest that the expressive language domain may be especially important for DLLs (Mancilla-Martinez et al., 2019; Ribot et al., 2018). For example, Mancilla-Martinez and colleagues found that elementary-age DLLs' expressive, but not receptive, vocabulary skills predict their English reading comprehension outcomes. Though not the focus of this study, these findings suggest that increased opportunities for DLLs to use language may be especially valuable for DLLs' language and literacy achievement. At a practical level, rather than assessing both receptive and expressive vocabulary given testing time constraints in schools, attending to expressive vocabulary may be particularly useful.

A final key contribution of this study is that we were able to document patterns of Spanish and English use over the course of three academic years. The DLLs in our study responded in English more than 50% of the time, on average, since study entry (i.e., kindergarten and second grade). They also demonstrated a clear transition to English dominance, on average, on both conceptually scored receptive and expressive vocabulary. Specifically, over this time, there was an 18%–36% increase in English responses. This notable shift in patterns of English language use aligns with previous research, which has documented a shift to English dominance over time (Castilla-Earls et al., 2019; Kohnert, 2010; Tse, 2001). This trend may reflect current policies in place that advocate for English-only instruction, beginning at the preschool level (Gándara & Hopkins, 2010; Menken, 2013), and that limit formal instruction in the home language (Spanish, in this case), particularly in new destination states. At the same time—and we argue more importantly—on average, nearly 50% of students in the kindergarten cohort responded in Spanish on receptive vocabulary and almost 40% did so on expressive vocabulary. While a smaller percentage of responses were in Spanish among the second-grade cohort (30% receptive and 20% expressively), this was nonetheless a sizable percentage. Without the option of being able to showcase their vocabulary knowledge in either Spanish or English, we can expect that our results would have likely underestimated DLLs' vocabulary. In other words, these results point to the utility of conceptually scored measures that allow DLLs to draw in either Spanish or English, which sharply contrasts with use of single-language assessments.

In summary, our findings converge with extant empirical evidence that suggest comparable conceptual vocabulary knowledge between monolingual and bilingual children (e.g., Bedore et al., 2005; Core et al., 2013; Mancilla-Martinez et al., 2018). Most importantly, our study extends developmental research in this area by examining the utility of conceptual scoring among elementary-age students longitudinally, attending to receptive and expressive vocabulary and examining trends in language use. We also underscore that our study was conducted in a new destination state experiencing unprecedented growth of DLLs, where DLLs' language skills are more likely to be misunderstood due to English-only instruction and assessment.

We also caution that results from our study may not generalize to the broader population of DLLs in the United States whose specific demographic characteristics and developmental contexts can be expected to differ. The DLLs in our study came from low-income, Spanish-speaking homes, and the majority were of Mexican origin. A disproportionate number of children from Spanish-speaking homes in the United States live in poverty (Gennetian et al., 2019), and the DLLs in our study similarly lived in homes at the poverty level. It is unclear whether the present vocabulary trajectories, which reveal DLLs' age-appropriate receptive and expressive vocabulary trajectories, would look similar or different if greater numbers of DLLs from middle- and high-income homes were represented. Furthermore, immigrant group differences in children's verbal trajectories have been reported, such that children of Mexican origin did not evidence persistent verbal growth relative to their nonimmigrant peers (e.g., Leventhal et al., 2006), underscoring the need for additional research in this area that includes DLLs from more diverse countries of origin.

Furthermore, the standardized procedure for conceptual scoring does not prompt the child in the other language (e.g., Spanish or English) if they respond correctly in the starting language (e.g., English or Spanish). While 42% of the DLLs in our study started the assessment in Spanish, it may be the case that the pattern of English-dominant responses we found was influenced by the slightly greater percentage of DLLs who began the assessments in English. Though this pattern of language preference is becoming increasingly prevalent in the school-age DLL population in the United States (e.g., Lutz, 2008), research with DLLs with a greater mix of Spanish and English language use is warranted. We also relied on conceptually scored vocabulary as a measure of language ability, as this is the only set of standardized bilingual measures currently available for this age group. Studies that utilize other bilingual scoring approaches (e.g., total vocabulary) are a natural next step, though such standardized measures do not currently exist. Future research would also benefit from following DLLs beyond the elementary school years.

Finally, summer setback was not modeled in our analyses given that we were unable to collect data in spring of Year 2, leaving only one summer during which we could examine this phenomenon. Thus, summer setback or plateau for both conceptually scored receptive and expressive vocabulary was only suggested in our descriptive analyses (i.e., difference in scores between spring of Year 1 and fall of Year 2; see Table 1), and it appeared that the effect of summer was more pronounced for conceptually scored receptive vocabulary. Previous studies have shown that students, especially those from low-income homes such as the DLLs in our study, tend to experience loss in their reading skills during the summer months when school is not in session (e.g., Alexander et al., 2001). Research that specifically examines this question among DLLs is needed.

Notwithstanding the limitations and avenues for next steps noted, our results support the use of conceptually scored vocabulary measures in accounting for the linguistic knowledge of elementary-age DLLs in a way that aligns with scientific understandings of bilingual language acquisition and that does not privilege monolingualism as the norm. DLLs' growth trajectories and achievement levels show that

DLLs do in fact harness rich linguistic knowledge and resources, which contradicts deficit-driven views about DLLs' vocabulary skills and highlights the linguistic assets they bring to learning.

Acknowledgments

This research was supported by a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development awarded to Vanderbilt University (Grant P20 HD075443).

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis, 23*(2), 171–191. https://doi.org/10.3102/01623737023002171

Anaya, J. B., Peña, E. D., & Bedore, L. M. (2018). Conceptual scoring and classification accuracy of vocabulary testing in bilingual children. *Language, Speech, and Hearing Services in Schools, 49*(1), 85–97. https://doi.org/10.1044/2017_LSHSS-16-0081

Anderson, R. T. (2012). First language loss in Spanish-speaking children: Patterns of loss and implications for clinical practice. In B. A. Goldstein (Ed.), *Bilingual language development and disorders in Spanish–English speakers* (2nd ed., pp. 193–212). Brookes.

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second language learners: Report of the National Literacy Panel on language-minority children and youth*. Erlbaum.

Bedore, L. M., Peña, E. D., Garcia, M., & Cortez, C. (2005). Conceptual versus monolingual scoring. *Language, Speech, and Hearing Services in Schools, 36*(3), 188–200. https://doi.org/10.1044/0161-1461(2005/020)

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language, 6*(2), 183–200. https://doi.org/10.1017/S0305000900002245

Castilla-Earls, A., Francis, D., Iglesias, A., & Davidson, K. (2019). The impact of the Spanish-to-English proficiency shift on the grammaticality of English learners. *Journal of Speech, Language, and Hearing Research, 62*(6), 1739–1754. https://doi.org/10.1044/2018_JSLHR-L-18-0324

Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In H. W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 25–40). Erlbaum. https://doi.org/10.4324/9781410612052

Core, C., Hoff, E., Rumiche, R., & Señor, M. (2013). Total and conceptual vocabulary in Spanish–English bilinguals from 22 to 30 months: Implications for assessment. *Journal of Speech, Language, and Hearing Research, 56*(5), 1637–1649. https://doi.org/10.1044/1092-4388(2013/11-0044)

Espinosa, L. (2010). *Getting it right for young children from diverse backgrounds: Applying research to improve practice*. National Association for the Education of Young Children.

Gándara, P., & Hopkins, M. (2010). *English learners and restrictive language policies*. Columbia University, Teachers College.

Gándara, P., & Mordechay, K. (2017). Demographic change and the new (and not so new) challenges for Latino education. *The Educational Forum, 81*(2), 148–159. https://doi.org/10.1080/00131725.2017.1280755

Gennetian, L., Guzman, L., Ramos-Olazagasti, M. A., & Wildsmith, E. (2019). *An economic portrait of low-income Hispanic families: Key findings from the first five years of studies from the National Research Center on Hispanic Children & Families. Report 2019-03*. National Research Center on Hispanic & Families. https://www.hispanicresearchcenter.org/research-resources/an-economic-portrait-of-low-income-hispanic-families-key-findings-from-the-first-five-years-of-studies-from-the-national-research-center-on-hispanic-children-families

Gibson, T. A., Oller, D. K., Jarmulowicz, L., & Ethington, C. A. (2012). The receptive–expressive gap in the vocabulary of young second-language learners: Robustness and possible mechanisms. *Bilingualism: Language and Cognition, 15*(1), 102–116. https://doi.org/10.1017/S1366728910000490

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*(1), 3–15. https://doi.org/10.1016/0093-934X(89)90048-5

Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech-Language Pathology, 23*(4), 574–586. https://doi.org/10.1044/2014_AJSLP-13-0026

Guo, L.-Y., & Schneider, P. (2016). Differentiating school-aged children with and without language impairment using tense and grammaticality measures from a narrative task. *Journal of Speech, Language, and Hearing Research, 59*(2), 317–329. https://doi.org/10.1044/2015_JSLHR-L-15-0066

Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly, 29*(4), 715–733. https://doi.org/10.1016/j.ecresq.2014.05.008

Hoff, E. (2018). Bilingual development in children of immigrant families. *Child Development Perspectives, 12*(2), 80–86. https://doi.org/10.1111/cdep.12262

Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language, 39*(1), 1–27. https://doi.org/10.1017/S0305000910000759

Hwang, J. K., Mancilla-Martinez, J., McClain, J. B., Oh, M. H., & Flores, I. (2019). Spanish-speaking English learners' English language and literacy skills: The predictive role of conceptually scored vocabulary. *Applied Psycholinguistics, 41*(1), 1–24. https://doi.org/10.1017/S0142716419000365

Ijalba, E. (2015). Effectiveness of a parent-implemented language and literacy intervention in the home language. *Child Language Teaching and Therapy, 31*(2), 207–220. https://doi.org/10.1177/0265659014548519

Jardak, A., & Byers-Heinlein, K. (2019). Labels or concepts? The development of semantic networks in bilingual two-year-olds. *Child Development, 90*(2), e212–e229. https://doi.org/10.1111/cdev.13050

Kan, P. F., & Kohnert, K. (2005). Preschoolers learning Hmong and English: Lexical semantic skills in L1 and L2. *Journal of Speech Language and Hearing Research, 48*(2), 372–383. https://doi.org/10.1044/1092-4388(2005/026)

Keller, K., Troesch, L. M., & Grob, A. (2015). A large receptive–expressive gap in bilingual children. *Frontiers in Psychology, 6,* 1284. https://doi.org/10.3389/fpsyg.2015.01284

Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology, 100*(4), 851–868. https://doi.org/10.1037/0022-0663.100.4.851

Kohnert, K. J. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders, 43*(6), 456–473. https://doi.org/10.1016/j.jcomdis.2010.02.002

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Leventhal, T., Xue, Y., & Brooks-Gunn, J. (2006). Immigrant differences in school-age children's verbal trajectories: A look at four racial/ethnic groups. *Child Development, 77*(5), 1359–1374. https://doi.org/10.1111/j.1467-8624.2006.00940.x

Lutz, A. (2008). Negotiating home language: Spanish maintenance and loss in Latino families. *Latino(a) Research Review, 6*(3), 37–64.

Mancilla-Martinez, J., Greenfader, C. M., & Ochoa, W. (2018). Spanish-speaking preschoolers' conceptual vocabulary knowledge: Towards more comprehensive assessment. *NHSA Dialog, 21*(1), 22.

Mancilla-Martinez, J., Hwang, J. K., Oh, M. H., & McClain, J. B. (2019). Early elementary grade dual language learners from Spanish-speaking homes struggling with English reading comprehension: The dormant role of language skills. *Journal of Educational Psychology*. Advance online publication. https://doi.org/10.1037/edu0000402

Mancilla-Martinez, J., & Kieffer, M. J. (2010). Language minority learners' home language use is dynamic. *Educational Researcher, 39*(7), 545–546. https://doi.org/10.3102/0013189X10383168

Mancilla-Martinez, J., & Lesaux, N. K. (2011). The gap between Spanish speakers' word reading and word knowledge: A longitudinal study. *Child Development, 82*(5), 1544–1560. https://doi.org/10.1111/j.1467-8624.2011.01633.x

Mancilla-Martinez, J., & Lesaux, N. K. (2017). Early indicators of later English reading comprehension outcomes among children from Spanish-speaking homes. *Scientific Studies of Reading, 21*(5), 428–448. https://doi.org/10.1080/10888438.2017.1320402

Mancilla-Martinez, J., Pan, B. A., & Vagh, S. B. (2011). Assessing the productive vocabulary of Spanish–English bilingual toddlers from low-income families. *Applied Psycholinguistics, 32*(2), 333–357. https://doi.org/10.1017/S0142716410000433

Mancilla-Martinez, J., & Vagh, S. B. (2013). Growth in toddlers' Spanish, English, and conceptual vocabulary knowledge. *Early Childhood Research Quarterly, 28*(3), 555–567. https://doi.org/10.1016/j.ecresq.2013.03.004

Martin, N. (2013a). *Expressive One-Word Picture Vocabulary Test–4: Spanish-Bilingual Edition (EOWPVT-4: SBE)*. Academic Therapy Publications.

Martin, N. (2013b). *Receptive One-Word Picture Vocabulary Test–4: Spanish-Bilingual Edition (ROWPVT-4: SBE)*. Academic Therapy Publications.

McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., Diliberti, M., Forrest Cataldi, E., Bullock Mann, F., & Barmer, A. (2019). *The condition of education 2019 (NCES 2019-144)*. National Center for Education Statistics.

Menken, K. (2013). Restrictive language education policies and emergent bilingual youth: A perfect storm with imperfect outcomes. *Theory Into Practice, 52*(3), 160–168. https://doi.org/10.1080/00405841.2013.804307

Migration Policy Institute. (2015). *ELL information center fact sheet series: Top languages spoken by English language learners nationally and by state*. https://www.migrationpolicy.org/research/top-languages-spoken-english-language-learners-nationally-and-state

Oller, D. K., & Pearson, B. Z. (2002). Assessing the effects of bilingualism: A background. In D. K. Oller & R. E. Eilers (Eds.), *Language and literacy in bilingual children* (pp. 3–21). Multilingual Matters. https://doi.org/10.21832/9781853595721-002

Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics, 28*(2), 191–230. https://doi.org/10.1017/S0142716407070117

Ortiz, A., & Artiles, A. J. (2010). Meeting the needs of English language learners with disabilities: A linguistically and culturally responsive model. In G. Li & P. Edwards (Eds.), *Best practices in ELL instruction* (pp. 247–272). Guilford Press.

Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual word recognition complicate the story. *Reading and Writing: An Interdisciplinary Journal, 23,* 189–208. https://doi.org/10.1007/s11145-008-9159-1

Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning*. Brookes.

Paris, S. G., & Hamilton, E. E. (2009). The development of children's reading comprehension. In *Handbook of research on reading comprehension* (pp. 56–77). Routledge.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*(1), 93–120. https://doi.org/10.1111/j.1467-1770.1993.tb00174.x

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of Child Language, 22*(2), 345–368. https://doi.org/10.1017/S030500090000982X

Peña, E. D., Bedore, L. M., & Zlatic-Guinta, R. (2002). Category-generation performance of bilingual children: The influence of condition, category, and language. *Journal of Speech, Language and Hearing Research, 45*(5), 938–947. https://doi.org/10.1044/1092-4388(2002/076)

Peña, E. D., & Halle, T. G. (2011). Assessing preschool dual language learners: Traveling a multiforked road. *Child Development Perspectives, 5*(1), 28–32. https://doi.org/10.1111/j.1750-8606.2010.00143.x

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R & D program in reading comprehension*. RAND.

Ribot, K. M., Hoff, E., & Burridge, A. (2018). Language use contributes to expressive language growth: Evidence from bilingual children. *Child Development, 89*(3), 929–940. https://doi.org/10.1111/cdev.12770

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis) abilities: Evidence, theory, and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). Guilford Press.

Sheng, L., Lu, Y., & Kan, P. F. (2011). Lexical development in Mandarin–English bilingual children. *Bilingualism: Language and Cognition, 14*(4), 579–587. https://doi.org/10.1017/s1366728910000647

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Spencer, M., & Wagner, R. K. (2017). The comprehension problems for second-language learners with poor reading comprehension despite adequate decoding: A meta-analysis. *Journal of Research in Reading, 40*(2), 199–217. https://doi.org/10.1111/1467-9817.12080

Stahl, K. A., & Bravo, M. A. (2010). Contemporary classroom vocabulary assessment for content areas. *The Reading Teacher, 63*(7), 566–578. https://doi.org/10.1598/RT.63.7.4

Thordardottir, E. T., Rothenberg, A., Rivard, M. E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages. *Journal of Multilingual Communication Disorders, 4*(1), 1–21. https://doi.org/10.1080/14769670500215647

Tran, V. C. (2010). English gain vs. Spanish loss? Language assimilation among second-generation Latinos in young adulthood. *Social Forces, 89*(1), 257–284. https://doi.org/10.1353/sof.2010.0107

Tse, L. (2001). Resisting and reversing language shift: Heritage language resilience among U.S. native biliterates. *Harvard Educational Review, 71*(4), 676–709. https://doi.org/10.17763/haer.71.4.ku752mj536413336

Windsor, J., & Kohnert, K. (2004). The search for common ground: Part I. Lexical performance by linguistically diverse learners. *Journal of Speech, Language, and Hearing Research, 47*(4), 877–890. https://doi.org/10.1044/1092-4388(2004/065)