

School-Year Grammatical Development in Narrative Retells: Spanish-English Dual Language Learners

Communication Disorders Quarterly
2019, Vol. 40(4) 228–238
© Hammill Institute on Disabilities 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1525740118791138
cdq.sagepub.com



Mary Claire Wofford, MS, CCC-SLP¹  and Carla L. Wood, PhD¹ 

Abstract

Oral narrative retells are rich sources of information for language development. Investigators collected English-language oral narrative retells during the fall and spring from 65 Spanish-English-speaking dual language learners (DLL) in kindergarten and first grade. Investigators examined transcripts of oral narratives for (a) inclusion and accuracy of microstructural elements using the Narrative Assessment Protocol (NAP; Pence, Justice, & Gosse, 2007), (b) percentage of grammatical utterances, and (c) types of verb errors. Prepositional phrases, elaborated noun phrases, irregular past tense verbs, and copula verbs were the most prevalent grammatical forms. Omission errors were the most prevalent verb error type. DLLs' narrative retells revealed significantly increased number of total NAP codes and diversity of NAP codes. Grammaticality of utterances increased from approximately 77% to 87% from fall to spring. All verb errors types decreased over the academic year. Direct feature coding approaches are useful for tracking developmental progress in DLLs' retells.

Keywords

cultural/linguistic, diversity, as a second language (ESL)/bilingualism/dialects, English/languages, acquisition/development, language/linguistics, Speech-Language Pathologists (SLPs), elementary school, Age, language, Assessment, syntax

As population trends evidence a greater presence of dual language learners (DLLs) entering school systems nationally, there is an increasing need for accurate and culturally relevant assessment that reflects the variability in DLLs' language development. Often speech-language pathologists (SLPs) and other educators take primary responsibility for assessment of DLLs to identify needs for additional support by using a variety of assessment techniques. In particular, knowledge and use of language sampling procedures is considered best practice when assessing children from minority-language backgrounds for language services (American Speech-Language-Hearing Association [ASHA], 2004).

Language sampling has been applauded as a mode of assessment that is naturalistic, ecologically valid, and culturally relevant (Bedore, Peña, Gillam & Ho, 2010; Cleave, Girolametto, Chen, & Johnson, 2010; Gutierrez-Clellen, Restrepo, Bedore, Peña, & Anderson, 2000; J. F. Miller & Iglesias, 2010; L. Miller, Gillam, & Peña, 2001). In addition, children newly introduced to academic settings are not disadvantaged in the assessment process by lack of familiarity with typical test-taking procedures (Bedore et al., 2010). Results from language sampling may echo poor language performance determined by norm-referenced measures, or conversely language sampling measures may demonstrate that a child has other linguistic forms not evident in standardized performance. For this reason, inclusion of language

sampling measures can significantly enhance eligibility determination and ongoing assessment. Given the flexibility of language sample analysis across context, language, task, and derived measures, greater use and exploration of different dimensions of language sampling practices is promising for the profession.

Importantly, assessments that do not fully explore a DLL's language abilities are at risk of misrepresenting a child's language ability and may disproportionately disadvantage students from culturally and linguistically diverse (CLD) backgrounds (Kapantzoglou, Restrepo, & Thompson, 2012). Historically, evidence has reflected a bias toward "mainstream" students. Increasing need for research on children coming from varied socioeconomic and/or CLD backgrounds has prompted increased investigation of these populations' typical and atypical language performance. However, common assessment practices may not change quickly (Arias & Friberg, 2017; Caesar & Kohler, 2007). DLLs' two languages are continually developing on separate continua. As such,

¹Florida State University, Tallahassee, USA

Corresponding Author:

Mary Claire Wofford, School of Communication Science & Disorders, Florida State University, 201 W. Bloxham, Tallahassee, FL 32306-1200, USA.

Email: mcw15e@my.fsu.edu

wide variability is expected in young DLLs' language performance contingent upon their unique language backgrounds (Hoff et al., 2012; Jackson, Schatschneider, & Leacox, 2014). Differentiation of language disorder from difference in DLL children depends on accurate, thorough assessment practices, as well as greater understanding of typical language development on language sample measures (Bedore et al., 2010; Gutierrez-Clellen & Simon-Cerejido, 2009).

Trajectories for DLLs' language development can be monitored through language sampling tasks. Oral narrative retells are well-established as one procedural option within language sample analysis that can provide plentiful information about a child's language development (Heilmann, Nockerts, & Miller, 2010). The tendency in the extant literature on oral narratives is that measures are categorized as macrostructural or microstructural. Macrostructure refers to the inclusion of conceptual content of a retell, commonly referred to as story grammar elements, while microstructure refers to the internal organization and mechanics of how story elements are presented; both measures are expected to grow as a child matures (Brooks & Kempe, 2014; Paul & Norbury, 2012; Westby, 2005). Oral narrative retells have been shown to be sensitive indicators of macrostructural and microstructural growth in both monolinguals (Heilmann, Miller, Nockerts, & Dunaway, 2010; Hewitt, Hammer, Yont, & Tomblin, 2005) and DLLs (Bitetti & Hammer, 2016; Rezzonico et al., 2015; Uccelli & Paez, 2007) and can differentiate typical from atypical language development in DLLs (Squires et al., 2014).

Characterizing the trajectory of language growth can be complicated by the effect of exposure to two languages. Naturally, language development depends on the relative amounts of exposure a bilingual child has to his or her two languages, which may be dynamic (Hoff et al., 2012). Consistent with a constructivist background, DLLs make use of diverse structures as they grow over an academic year (Tomasello, 2003). By identifying patterns in their linguistic environment, they increasingly become aware of and subsequently use certain forms in their linguistic endeavors. Important particularly for the development of morphosyntax, or morphological and syntactic units that support grammar and sentence organization, structures expected to emerge should not differ from typical monolingual development, although DLLs may experience delays due to the additional time necessary to detect linguistic patterns and extract relevant information (Gathercole, 2007). Typical Spanish-English-speaking DLLs also may demonstrate differences in certain morphosyntactic elements (e.g., subject omission, overgeneralization of regular tense forms, prepositional phrase use for possessive markers) that result from borrowing of rules from the native Spanish language (Paradis, Genesee, & Crago, 2011; Paul & Norbury, 2012). For Spanish-English DLLs with language impairment, morphosyntactic knowledge in particular tends to be an area of

relative weakness (Bedore & Peña, 2008; Restrepo, 1998). To the untrained observer, native-language transfer patterns can be mistaken for atypical language development or conversely remain undetected. Of course, perceived errors can also be the result of patterns that reflect typical development for a child's age group. For this reason, language sample measures that provide a snapshot of morphosyntax can serve as indicators of typical growth, and an educator's understanding of the type of error produced can help to determine their source.

Many narrative-derived measures represent distinct aspects of microstructure. Of particular utility in oral narrative retell analysis are options for observing morphosyntax. Some commonly reported measures include mean length of utterance (MLU) and its variants (e.g., mean length of longest five utterances in sample, MLU-words, MLU-morphemes), subordination index (ratio of total clauses to total communication units), and number of different words (NDW), which is commonly used as a measure of lexical diversity. However, these individual measures of productivity, complexity, and/or specific features distinctly depict morphosyntactic ability. Each of these measures provides its own nuanced representation of structural organization. MLU, as an example, reflects the productivity of a child's retell, as more mature structures typically accompany longer utterances (Paul & Norbury, 2012). Contrastively, measures such as subordination index or other composite measures of complexity provide an estimate of how frequently complex structures are used throughout the retell. It should be noted that microstructural measures may tap elements of vocabulary knowledge (e.g., NDW, use of connectives, tier-two nouns/verbs). The role of vocabulary is not to be underestimated as its codevelopment with grammar and syntax is critical in DLLs' early language development (Paradis et al., 2011).

Often reported alongside measures of microstructure in narrative retell measures is percentage of grammatical utterances, which is a measure that can indicate the overall coherence of a retell, as well as the linguistic maturity of the child. Although grammaticality does not provide information about length of utterance or productivity, its inclusion in narrative-based analyses is valuable considering that typically developing DLLs tend to demonstrate decreased grammaticality compared to monolingual-English peers (Bedore et al., 2010). DLLs with typical development demonstrate capacity to develop complex narratives before they attain mastery of all structures, and as such fully intact grammaticality in early DLL retells is not expected (Bedore et al., 2010; Gutierrez-Clellen, 2002).

Grammatical analysis of a language sample can consist of both aggregate measures and individual grammatical form analysis of accuracy. Aggregate measures of morphosyntactic ability such as MLU, subordination index, percent grammaticality, and other variants are quickly derived, making

them convenient for clinical use (Heilmann, Nockerts, & Miller, 2010). However, fine-grained analyses of specific structures also provide meaningful information that can aid in development of short-term and long-term goals for intervention. Direct coding of specific morphosyntactic features provides an indication of the mastery of critical morphosyntactic forms (e.g., third-person plural *-s*, copula *be*, regular past tense *-ed*). Grammatical structures that are less perceptually salient tend to be vulnerable for later mastery or inconsistent use by DLLs (Jacobson & Walden, 2013).

In the existing literature base, researchers at times opt to derive a verb morphology composite by observing the percentage of correct use of specific grammatical features out of all obligatory contexts (Bedore & Leonard, 1998; Gutierrez-Clellen & Simon-Cerejido, 2009). Examples of structures that have been selected in composite scoring systems include regular past tense *-ed*, present third person singular *-s*, auxiliary verb *do*, auxiliary forms of *be*, and copula *be*. Other dedicated assessments, such as the *Narrative Assessment Protocol* (NAP; Pence, Justice, & Gosse, 2007), use a grammatical feature coding approach to provide information regarding syntactic, morphological, and lexical structures that constitute microstructural elements of a narrative. The presence of these elements in an oral retell context can reflect the microstructural quality of a narrative, as well as the child's mastery of certain structures indicated by the frequency with which elements occur in the retell. The NAP was designed to be scored in real time. It demonstrated reasonable specificity, validity, and reliability for use with monolingual English-speaking preschool-aged children (Justice, Bowles, Pence, & Gosse, 2010); a Spanish-language version (NAP-S) also demonstrated appropriate preliminary psychometric properties for use with young Spanish-speaking children living in the United States (Gorman, Bingham, Fiestas, & Terry, 2016). Microstructural analysis that relies on microstructural feature coding correlated to standardized measures of language for both English speakers and Spanish speakers (Gorman et al., 2016; Justice et al., 2010).

Close analysis of grammaticality can also yield information that informs intervention. Categorization of verb errors allows clinicians and researchers to gain an understanding of typical variations in dual language development, as certain errors are expected when learning two languages. In a review of research on typically developing Spanish-dominant speakers, error analysis revealed common patterns of transfer to include overgeneralization of regular verb forms, substitution of person, and infinitive use for inflected forms (see Jackson-Maldonado, 2004). Verb errors might be productive or unproductive in nature. Productive errors are those in which a speaker applies an incorrect inflection or rule to a verb, while unproductive errors are those in which no inflection is made in an obligatory context (Jacobson & Schwartz, 2005). Some types of verb errors are of more concern for

clinicians. In a study of 48 typically developing and language-impaired bilingual children who provided Spanish and English-language narratives, omission errors were found to be the most predictive of language impairment (Jacobson & Walden, 2013). However, omission errors are characteristic of typical development as well, although less prevalent and persistent. Typical verb errors must be well understood to predictably recognize the influence of native language transfer.

Research Aims

Greater understanding of narrative performance in typical English- and Spanish-language narratives is necessary for better assessing and intervening with this population. The current study focuses on the English-language development of young DLLs. The current study contributes a unique approach in observing the codevelopment of microstructural features and grammaticality on an English-language oral narrative retell task as DLLs transition through an academic year.

Research questions for this current study include the following:

Research Questions 1: What microstructural features from the NAP do DLLs demonstrate in their oral narrative retells at the beginning and end of a school year?

Research Questions 2: Do Spanish-English DLLs demonstrate patterns of microstructural growth on NAP features in their performance on an oral narrative retell over the course of an academic year? If so, does rate of change differ by grade?

Research Questions 3: What is the average accuracy of verb use by DLLs in the fall and spring of the school year? Specifically, what types of verb errors are demonstrated by DLLs in kindergarten and first grade during oral narrative retells?

Research Questions 4: Do Spanish-English DLLs demonstrate change in verb grammaticality in their performance on an oral narrative retell over the course of the school year? If so, what types of verb errors remain? Does change differ by grade?

As children progress through the academic year and during the oral narrative task itself, they are exposed to grammatical forms of varying complexity, which may or may not be reflected in production of their oral narrative retell. DLLs may attempt to mimic advanced syntactic structures heard during the narration, but it is predicted that they are more likely to rely on known structures to retell their story. It is hypothesized that over the course of the year, exposure to English-language input grows, and participants will demonstrate greater usage of microstructural elements. As a greater number and increasingly diverse set of grammatical forms are mastered, verb errors are expected to decrease.

Method

All children in the current study were participants in a larger study, funded by the Institute of Education Sciences, designed to develop a vocabulary intervention program to facilitate English vocabulary learning for DLLs (Wood et al., in press). The larger project did not involve direct instruction on narratives or explicitly teach verbs or syntax. As part of a larger randomized control trial intervention study, children listened to recorded ebook readings 3 days a week for approximately 20 weeks during the school year. Each weekly recorded reading provided elaborations of four target words. The project did not include oral narrative retell activities between test points. Investigators collected narrative retells at four participating schools. The narratives constituted one piece of a larger assessment battery. All study procedures were approved by the university's committee on research involving human subjects (HSC#: 2016.18265).

Participants

Students completing narrative retells in both the fall and spring were a subset ($n = 74$) within a larger multisite, school-based vocabulary intervention designed for Spanish-English DLLs. Exposure to both Spanish and English was an inclusion criterion, as reported by parents and teachers in interviews (see Appendix). Average performance on an English-language vocabulary test was described by a standard score of 86.23 ($SD = 10.65$; $n = 65$) using the Peabody Picture Vocabulary Test—Fourth (PPVT; Dunn & Dunn, 2007). Participants demonstrated Spanish-language vocabulary scores at an average standard score of 89.20 ($SD = 15.93$; $n = 41$) using the Test de Vocabulario en Imagenes Peabody (TVIP; Dunn, Lugo, Padilla, & Dunn, 1986). An exclusion criterion of Spanish-language performance within 2.5 standard deviations resulted in the elimination of 9 participants from the larger data set. Within this subsample, the average nonverbal intelligence measured by the Primary Test of Nonverbal Intelligence (PTONI) was 97.85 ($SD = 20.44$; Ehrler & McGhee, 2008). Average age of all participants was 7.35 years ($M = 88.25$ months, $SD = 8.12$ months). Participants were approximately evenly split between females ($n = 31$) and males ($n = 34$) and slightly more representative of first grade students ($n = 41$) than kindergarteners ($n = 24$). All parents who participated in a demographic interview reported that children qualified for free or reduced lunch ($n = 56$).

Procedures

Trained undergraduate and graduate research assistants administered the narrative retell task in English early in the fall semester and late in the spring semester for a total of two time points. Participants first viewed a 3-min animated

Table 1. Video Narrative Retell Story Characteristics.

Story title	Total T-units	MLU	NTW	NDW	Time
Billy Bear & The Balloon	30	9.3	279	130	2:35
You Can't Catch Me	29	10.14	294	151	2:15
Pingu Goes Fishing	32	11.41	365	148	2:19
That's My Hat	41	8.95	367	158	2:35

Notes. MLU = Mean Length of Utterance; NTW = Number of Total Words; NDW = Number of Different Words.

video with English-language narration on a laptop computer with headphones. They were asked to retell the story to a stuffed animal hidden from view during the viewing. The videos and administration protocol were identical to procedures described in other studies (Wood, Wofford, Gabas, & Petscher, 2018; Wood, Wofford, & Schatschneider, 2018). Researchers constructed videos to be balanced in time, difficulty, and number of microstructural elements, as well as story grammar elements (see Table 1). Video narrations of videos contained at least three instances of each *Narrative Assessment Protocol* (NAP; Pence et al., 2007) microstructural element. Participants viewed and retold the same story at both time points, except in the case of two participants that were administered different stories.

All stories contained at least three examples of each microstructure element scored on the NAP. Research assistants then asked participants to retell the story back to a stuffed animal that was hidden from view during viewing. Responses were audio recorded and transcribed by trained undergraduates in Systematic Analysis of Language Transcripts (SALT; J. F. Miller & Iglesias, 2010). Interrater agreement was calculated by dividing word level instances of agreement by total opportunities for agreement, yielding a percentage agreement at 83.31% on a randomly selected 20% of transcripts. The discrepancies were flagged by a research assistant and resolved by the primary researcher.

Transcripts underwent coding by trained research assistants for NAP structures and verb errors. Interrater reliability using a "within one" procedure (La Paro, Pianta, & Stuhlman, 2004) was utilized consistent with previous NAP studies (Gorman et al., 2016; Justice et al., 2010), meaning that reliability was obtained on any one element out of 18 when coders were within one point of each other. In relation to the maximum sum score acquired on the NAP, one point is equivalent to 1.85% of the total range of points (range 0–54). The current study also involved one procedural difference from NAP coding in that participants were given credit for more than three instances of an element if they were produced in the narrative retell. To be systematic in the assignment of credit when scoring narratives, the researchers decided to fully describe NAP elements included in the narrative by counting all instances of

an element, rather than a maximum of three instances per element. The scores were thus reflective of the proposed descriptive research questions. Reliability percentages are reported below in parentheses. Sentence types included sentences that were compound (100%), complex (100%), interrogative (100%), and/or negative (89.7%). Phrase level elements included elaborated noun phrases (94.4%), compound nouns (100%), and prepositions (94.8%). Modifier elements included advanced modifiers (100%) and adverbs (100%). Noun elements included pluralized nouns (100%), possessive forms (100%), and tier-two status nouns (100%). Verb structures observed included auxiliary + main verb (100%), copula *be* verbs (100%), irregular past tense (92.9%), regular past tense (89.3%), tier-two status (100%), and compound verbs (100%).

All utterances received a code for correct verb use or a verb error, including errors of substitution, omission, overgeneralization, or other, adapted from a coding scheme developed by Bedore and colleagues (2010). Substitution errors constituted verbs in which an incorrect inflected form was applied, while omission verb errors encompassed uninflected verbs, omitted morphemes, or omitted auxiliary verbs. Overgeneralization verbs were those in which participants applied a verb with an overgeneralized regular form instead of an irregular verb. Issues not easily identified by other categories received an *Other* verb error code. Coders demonstrated acceptable interrater reliability for correct (94.3%) and incorrect verbs (87.6%).

Analyses

Descriptive statistics were compiled to determine the inclusion of individual NAP features from fall to spring. Paired sample *t* tests were conducted on total number of NAP elements and the diversity of NAP elements. For a more statistically valid comparison, only these two elements were subjected to inferential statistical tests, as many children were expected to produce zero instances of certain microstructural forms. Use of aggregate performance across structures met assumptions for normality, while individual element distributions tended to be skewed. Data were divided by grade to observe descriptive differences in performance.

Results

The first research question sought to determine the frequency and diversity of NAP elements used in fall and spring narrative retells (see Table 2). Total NAP and Diversity of NAP element usage demonstrated an increasing trend (see Figures 1 and 2). Most frequently occurring elements on average included prepositional phrases, elaborated noun phrases, copula *be* verbs, irregular past tense verbs, and regular past tense verbs.

The second research question addressed change in NAP structures from fall to spring. The mean difference across all participants was calculated (see Table 2). Between fall ($M = 17.60$, $SD = 14.95$) and spring ($M = 21.14$, $SD = 14.05$) total NAP codes, a paired sample *t* test revealed a statistically significant increase ($t = 2.006$, $df = 64$, $p = .049$). For diversity of NAP codes, fall performance ($M = 6.20$, $SD = 3.28$) was significantly greater than spring performance ($M = 7.37$, $SD = 2.66$; $t = 3.266$, $df = 64$, $p = .002$). Positive increases were seen in 10 of 18 total NAP elements including: compound, complex, and negative sentences; all phrase type codes; advanced modifiers; possessive nouns; and irregular past and regular past tense verbs. The largest increases in grammatical forms comprised irregular past tense verbs ($MD = 1.68$) and elaborated noun phrases ($MD = .91$). Direction of change did differ by grade in the case of complex sentences, interrogative sentences, compound noun phrases, adverbs, plural nouns, tier-two nouns, auxiliary + main verbs, copula verbs, and compound verbs.

The third research question addressed the patterns of verb usage in fall and spring narrative retells (see Table 3). On average, participants' retells were accurate in greater than half of total verbs, and verb accuracy showed an increasing trend across the year by grade (see Figure 3). Omission-type errors were the most prevalent types of errors in both the fall and spring time points. Substitution, overgeneralization, and other errors accounted on average for less than 0.5 of total errors. All verb error types decreased over the year for overall sample and by grade.

The fourth research question addressed change in verb grammaticality over the academic year. Participants had a significantly higher rate of correct verb accuracy in the spring retell, increasing from 76.9% to 87.0% ($t = 3.991$, $df = 64$, $p = .000172$). In addition, negative mean differences indicated that all verb error types decreased throughout the academic year.

Discussion

The current study sought to describe the microstructural forms used by early elementary, Spanish-English DLL students in oral narrative retells and to observe growth in microstructural elements over the course of an academic year. Oral narrative retells were collected with the aim of describing verb grammaticality and verb errors. Grammaticality developed as evidenced by increased verb accuracy and decreased verb errors from the fall to the spring.

Key Findings

Microstructural elements were present in the oral narrative retells as identified by the NAP. On average, participants

Table 2. NAP Element Usage in Fall and Spring Narrative Retells.

NAP element	Fall				Spring				Mean difference	Direction of change	
	M	SD	Min	Max	M	SD	Min	Max		KG	G1
Sentence											
Compound	.26	.64	0	3	.58	.83	0	3	0.32	▲	▲
Complex	.49	1.21	0	7	.54	.94	0	4	0.05	▲	▼
Interrogative	.20	.64	0	4	.18	.46	0	2	-0.02	▲	▼
Negative	.88	1.08	0	4	1.14	1.48	0	10	0.26	▲	▲
Phrase											
Elaborated noun	1.69	2.05	0	10	2.60	2.73	0	15	0.91	▲	▲
Compound noun	.06	.24	0	1	.09	.28	0	1	0.03	▲	▼
Prepositional	2.80	2.74	0	13	3.29	2.89	0	11	0.49	▲	▲
Modifiers											
Adverb	.18	.49	0	2	.06	.30	0	2	-0.12	-	▼
Advanced modifier	.23	.63	0	4	.32	.69	0	3	0.09	▲	▲
Nouns											
Pluralized	.75	1.10	0	4	.62	1.09	0	5	-0.13	▲	▼
Possessive	.09	.34	0	2	.22	.63	0	4	0.13	▲	▲
Tier-two	.12	.48	0	3	.08	.27	0	1	-0.04	▲	▼
Verbs											
Auxiliary + Main	1.06	1.30	0	6	1.03	1.31	0	6	-0.03	▲	▼
Copula	2.09	2.23	0	12	1.82	1.80	0	9	-0.27	▲	▼
Irregular past	4.63	4.75	0	18	6.31	5.00	0	21	1.68	▲	▲
Regular past	1.55	2.35	0	11	2.06	1.94	0	9	0.51	▲	▲
Tier two	.49	1.12	0	6	.20	.54	0	3	-0.29	▼	▼
Compound	.02	.12	0	1	.00	.00	0	0	-0.02	▼	-
NAP total codes	17.60	14.95	0	63	21.14	14.05	1	56	3.54	▲	▲
Diversity NAP codes	6.20	3.28	0	13	7.37	2.66	1	13	1.17	▲	▲

Notes. *n* = 65. NAP = Narrative Assessment Protocol. ▲ = Positive change. ▼ = Negative change. - = No change.

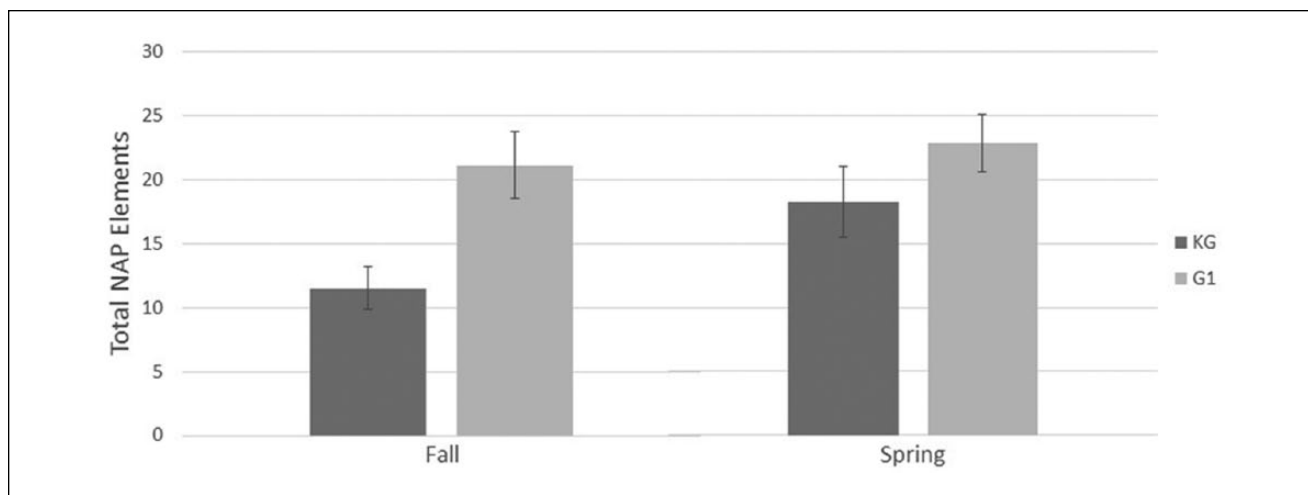


Figure 1. Total NAP element usage on average across academic year by grade (*n* = 65).

Note. Error bars represent standard error. NAP = Narrative Assessment Protocol.

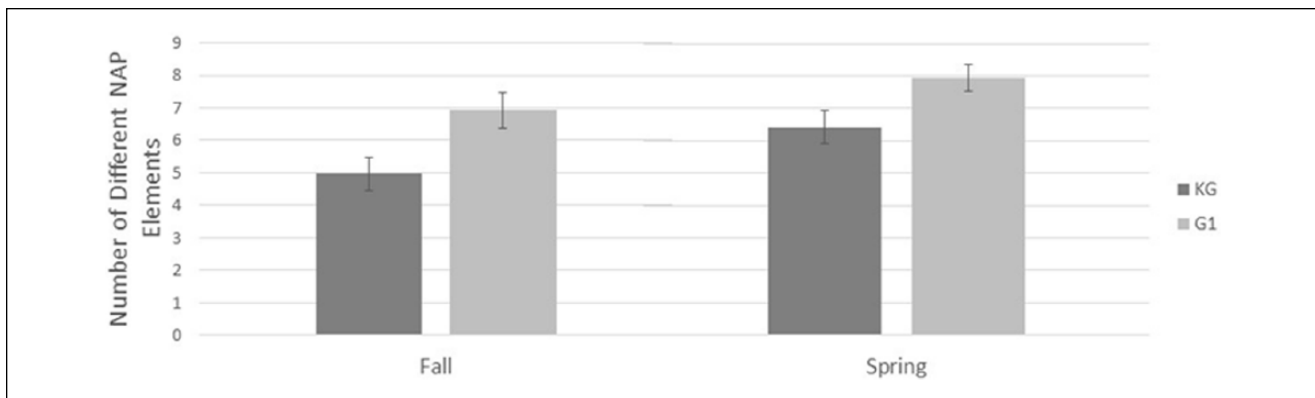


Figure 2. Average diversity of NAP element usage across academic year by grade. $n = 65$

Note. Error bars represent standard error. NAP = Narrative Assessment Protocol.

Table 3. Verb Usage in Fall and Spring Narrative Retells.

Verb usage	Fall				Spring				Mean diff.	Direction of change	
	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>		KG	G1
Correct verb use (%)	76.88	20.48	0	100	87.07	15.30	31	100	10.19	▲	▲
Raw correct	12.46	10.13	0	46	13.82	8.09	1	36	1.36	▲	▼
Omission	2.60	2.59	0	11	1.52	2.17	0	11	-1.08	▼	▼
Substitution	.25	.73	0	3	.14	.35	0	1	-0.11	▼	▼
Overgeneralization	.17	.42	0	2	.09	.34	0	2	-0.08	▼	▼
Other	.06	.24	0	1	.02	.13	0	1	-0.04	▼	▼

Note. $n = 65$. Negative mean difference values indicate decreasing verb error types from the fall to spring time point. ▲ = Positive change.

▼ = Negative change. - = No change.

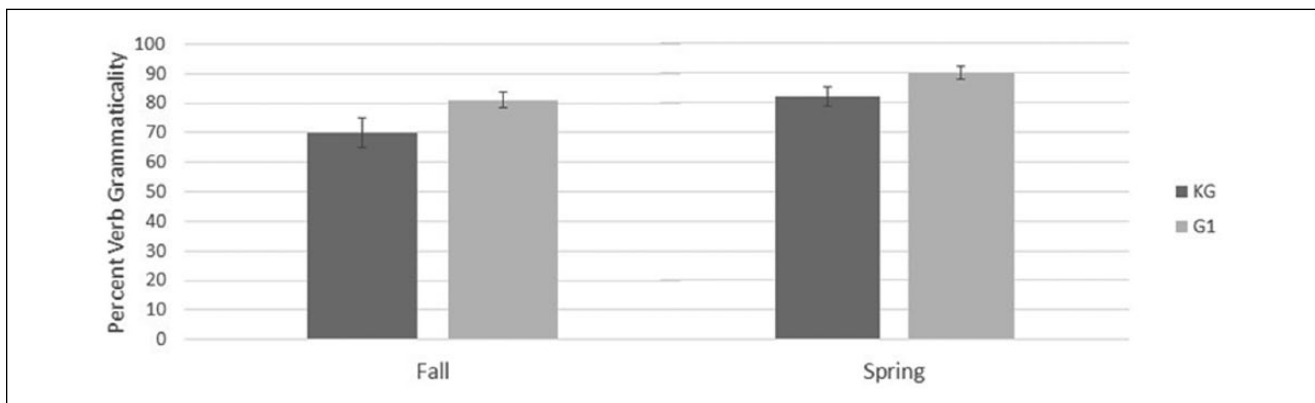


Figure 3. Grammatical accuracy rate on average across the academic year by grade.

Note. Bars represent a percentage of correct verbs out of total verbs used in narrative retell; error bars represent standard error.

produced 17.60 total NAP elements at the beginning of the year, across 6.20 different codes. At the end of the year, the average number of NAP elements significantly increased in both number and type, amounting to 21.14 total elements spanning 7.37 different elements. Totally, 10 out of 18 NAP elements increased, with larger mean differences evidenced

in elaborated noun phrases, prepositional phrases, and irregular past tense verbs. Total NAP elements included in the retell increased on average by 3.54 elements, and participants averaged at least one more type of NAP element by the end of the school year. All participants that used zero total NAP elements or zero types of NAP elements in the

fall retell ($n = 2$) increased their total NAP element use and diversity of NAP elements use in the spring retell.

Average grammatical accuracy of oral narrative retells in the fall was 76.88% across all participants and increased to 87.07% in the spring. In both time points, omission errors tended to be the most prevalent. Substitution, overgeneralization, and other verb errors constituted a small percentage of verb errors in both time points. Increasing verb accuracy and decreasing verb error types confirm that early elementary school is a time of changing linguistic proficiency for DLLs.

As expected with DLLs, variability in inclusion of NAP elements and overall verb accuracy was present in the retells. Results of more developmentally advanced individual NAP structures (e.g., advanced modifiers, adverbs, complex sentences, tier-two verbs, etc.) were skewed toward zero. Naturally, it is possible that DLLs who are in the initial stages of exposure to English might not choose to exhibit more developmentally advanced or cross-linguistically distinct structures in a spontaneous task such as an oral narrative retell. Furthermore, while variability in performance amounted to significant differences in aggregate measures, close to half (49.2%, $n = 32$) of the sample demonstrated fewer or the same total NAP elements. For diversity of NAP elements, fewer participants (40.0%, $n = 26$) demonstrated negative or no change from fall to spring. With regard to percent accuracy of verbs, approximately one-third of the sample ($n = 23$) was less accurate or equally accurate in their spring retell.

Consistency With Literature Base

Although few studies have utilized the NAP in analysis of DLLs' narrative retells, several points of comparison were made with the current study. In a sample of 262 preschool-aged monolingual children, Justice and colleagues (2010) found the average NAP score was approximately 14.06 ($SD = 8.19$) elements used in a fall retell of a wordless storybook. While the current study's target population was older and classified as DLLs, the marginally better performance in the fall retell ($M = 16.76$) indicates that for the current study's target population, performance averages were similar to NAP results of typically developing monolingual children. Terry, Mills, Bingham, Mansour, and Marencin (2013) collected oral narrative retells from 142 typically developing preschool-aged ($mean\ age = 4.34$ years, $SD = 5.5$ months) African American children and reported an average NAP score of 18.99 ($SD = 7.03$). The younger, nonmainstream dialect-speaking sample in Terry et al. (2013) performed slightly better than DLL participants in the current study. Furthermore, both Terry and colleagues and the current study found that the NAP was sensitive to the developmental change that occurred throughout a school year.

Relative to extant research on grammaticality, the current study confirmed that verb grammaticality is shown to increase in accuracy during this period of DLL development (Bedore et al., 2010; Gusewski & Rojas, 2017; Jacobson & Walden, 2013). The present study's findings were consistent with other studies showing that young Spanish-English DLLs did not tell completely grammatical stories at either time point (Bedore et al., 2010; Gutierrez-Clellen, 2002). When considering types of individual verb errors reported in the existing literature, there are some considerations that coincide with the current study's findings. In a slightly older sample ($mean\ age = 8$ years), typical DLLs committed 1.42 omission errors in a retell, a trajectory consistent with the present study if verb errors of omission continue to decrease (Jacobson & Walden, 2013). Another possible explanation is that the current study's definition for omission errors was slightly broader in its inclusion of bare stem verbs, which tend to be more prevalent in DLLs' narratives (Jacobson & Schwartz, 2005). Grammatical utterances in another study (Bedore et al., 2010) of kindergarten DLLs ($mean\ age = 5.7$ years, $SD = 4.39$ months) revealed lower accuracy on grammaticality in English narrative retells ($M = 46.09$, $SD = 25.83$) when compared to the current study. Notably Bedore's study did not focus on verb accuracy in isolation, but on other grammatical elements (e.g., article, possessive, or preposition omissions or substitutions; pronoun case substitution), possibly resulting in fewer fully grammatical utterances. English grammaticality was found to correlate significantly with performance on a standardized bilingual measure (i.e., Bilingual English Spanish Assessment, Peña, Gutierrez-Clellen, Iglesias, Goldstein, & Bedore, 2013) along with English MLU, English lexical diversity, and Spanish grammaticality. Findings from the literature base and the current study provide impetus to observe closely the dynamics between these variables in early elementary DLL development.

Grade Differences

Although the examination of grade differences was not a primary aim, descriptive analyses revealed differences in performance that might be explained by grade in future studies. Inconsistency in individual NAP elements may be explained by individual differences in child characteristics, as well as school, classroom, and teacher differences not measured in this study. However, growth for both grades was demonstrated in several elements, including compound sentences, negative sentences, elaborated noun phrases, prepositional phrases, advanced modifiers, possessive nouns, irregular past tense verbs, and regular past tense verbs.

Implications

The current study provides support for the use of language sampling with DLLs (Gutierrez-Clellen & Simon-Cerejido, 2009). Specific implications for practicing clinicians pertain to the utility of direct feature coding and grammaticality analysis in language sampling. Global measures of morphosyntactic competence (e.g., MLU, subordination index, TNW) provide quick measures of production derived from DLLs' retells. Direct coding of microstructure enhances evaluation reports as they provide rich linguistic information that can inform SLPs and other educators regarding specific microstructural elements that can be short- or long-term goals of intervention. These less perceptually salient features may otherwise go overlooked if there is limited opportunity for observation. Furthermore, direct feature coding of retells can be utilized in concert with standardized assessment to characterize sentence organization or demonstrate mastery of specific morphological elements in a highly authentic task. Synthesis of information from both sources can provide the clinician with a sense of the overall linguistic ability of the child (Ebert & Pham, 2017). Furthermore, SLPs and other educators working with DLLs should not anticipate absolute mastery of grammaticality during this period of development. It is possible that spontaneous use of individual structures may be unstable or interacting with other developmental variables, indicated by descriptive differences in performance by grade.

Morphosyntactic learning is of concern to SLPs because of its own relevance for written and oral expressive language, and also because of its connection to vocabulary learning. Morphosyntax development is finite, while vocabulary development is open, such that morphosyntactic forms once acquired support the further acquisition of vocabulary (Gathercole, 2007; Paradis et al., 2011). When morphosyntactic development is delayed, vocabulary knowledge likely will be affected also, which in turn has an impact on other academic competencies. Spanish-English DLLs' knowledge of morphological patterns ultimately have effects on reading comprehension (Kieffer & Lesaux, 2008). Vocabulary learning is well-established as being critical in literacy and general academic outcomes. The current study's findings apply to the notion that morphosyntactic learning occurring across the early elementary years supports understanding of vocabulary growth in later grades.

Limitations

Several limitations should be considered in interpreting the current findings. Only one narrative retell per time point was collected in this study. Greater stability of results could be gleaned from multiple samples in both languages. Furthermore, the participants in this study came from a lower socioeconomic background, indicated by 100% of

participants receiving free/reduced lunch. This homogeneity in the sample may prevent direct comparison to other groups, and consideration of the effect of low socioeconomic status cannot be overlooked. Procedural differences such as the use of video stimuli must be considered when comparing the present study's findings to the literature base; however, inclusion of diversity of NAP codes and alternative sources for narrative retell content also constitute innovative approaches that broaden the understanding of DLL's development on narrative measures.

Conclusion

Although a great deal of work is emerging in the area of narratives as a highly viable, culturally sensitive tool for measurement, additional research is needed in this area to better understand typical trajectories of development in DLLs' narrative performance. With additional research narrative retells show promise as a viable, culturally responsive tool for language measurement that is sensitive to grammatical growth across the school year.

Acknowledgments

The authors are especially grateful to partnering schools, participating families, and support from research assistants who were essential to project completion.


Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A130460 to Florida State University. In addition the first author is supported by U.S. Office of Special Education Programs through Grant H325D140068. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

ORCID iDs

Mary Claire Wofford  <https://orcid.org/0000-0001-6243-7214>

Carla L. Wood  <https://orcid.org/0000-0002-0551-0638>

Supplemental material

Supplemental material is available online with this article.

References

- American Speech-Language-Hearing Association. (2004). *Preferred practice patterns for the profession of speech-lan-*

- guage pathology* [Preferred Practice Patterns]. Available from www.asha.org/policy
- Arias, G., & Friberg, J. (2017). Bilingual language assessment: Contemporary versus recommended practice in American schools. *Language, Speech, and Hearing Services in Schools, 48*, 1–15. doi:10.1044/2016_LSHSS-15-0090
- Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research, 41*, 1185–1192. doi:10.1044/jslhr.4105.1185
- Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism, 11*, 1–29. doi:10.2167/beb392.0
- Bedore, L. M., Pena, E. D., Gillam, R. B., & Ho, T. H. (2010). Language sample measures and language ability in Spanish-English bilingual kindergarteners. *Journal of Communication Disorders, 43*, 498–510.
- Bitetti, D., & Hammer, C. S. (2016). The home literacy environment and the English narrative development of Spanish-English bilingual children. *Journal of Speech, Language, and Hearing Research, 59*, 1159–1171. doi:10.1044/2016_JSLHR-L-15-0064
- Brooks, P. J., & Kempe, V. (Eds.). (2014). *Encyclopedia of language development*. Thousand Oaks, CA: SAGE.
- Caesar, L. G., & Kohler, P. D. (2007). The state of school-based bilingual assessment: Actual practice versus recommended guidelines. *Language, Speech, and Hearing Services in Schools, 38*, 190–200. doi:10.1044/0161-1461(2007/020)
- Cleave, P. L., Girolametto, L. E., Chen, X., & Johnson, C. J. (2010). Narrative abilities in monolingual and dual language learning children with specific language impairment. *Journal of Communication Disorders, 43*, 511–522. doi:10.1016/j.jcomdis.2010.05.005
- Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). *TVIP: Test de vocabulario en imagenes Peabody* [Test of Vocabulary of Peabody Images]. Pearson.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Minneapolis, MN: Pearson.
- Ebert, K. D., & Pham, G. (2017). Synthesizing information from language samples and standardized tests in school-age bilingual assessment. *Language, Speech, and Hearing Services in Schools, 48*, 42–55. doi:10.1044/2016_LSHSS-16-0007
- Ehrler, D. J., & McGhee, R. L. (2008). *PTONI: Primary test of nonverbal intelligence*. Austin, TX: Pro-Ed.
- Gathercole, V. C. M. (2007). Miami and North Wales, so far and yet so near: A constructivist account of morphosyntactic development in bilingual children. *International Journal of Bilingual Education and Bilingualism, 10*, 224–247. doi:10.2167/beb442.0
- Gorman, B., Bingham, G., Fiestas, C., & Terry, N. P. (2016). Assessing the narrative abilities of Spanish-speaking preschool children: A Spanish adaptation of the narrative assessment protocol. *Early Childhood Research Quarterly, 36*, 307–317. doi:10.1016/j.ecresq.2015.12.025
- Gusewski, S., & Rojas, R. (2017). Tense marking in the English narrative retells of dual language preschoolers. *Language, Speech, and Hearing Services in Schools, 48*, 183–196. doi:10.1044/2017_LSHSS-16-0093
- Gutierrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education, 13*, 175–197. doi:10.1016/S0898-5898(01)00061-4
- Gutierrez-Clellen, V. F., Restrepo, M. A., Bedore, L., Peña, E., & Anderson, R. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools, 31*, 88–98.
- Gutierrez-Clellen, V. F., & Simon-Cerejido, G. (2009). Using language sampling in clinical assessments with bilingual children: Challenges and future directions. *Seminars in Speech and Language, 30*, 234–245. doi:10.1055/s-0029-1241722
- Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology, 19*, 154–166. doi:10.1044/1058-0360(2009/08-0024)
- Heilmann, J., Nockerts, A., & Miller, J. F. (2010). Language sampling: Does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools, 41*, 393–404. doi:10.1044/0161-1461(2009/09-0023)
- Hewitt, L. E., Hammer, C. S., Yont, K. M., & Tomblin, J. B. (2005). Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. *Journal of Communication Disorders, 38*, 197–213. doi:10.1016/j.jcomdis.2004.10.002
- 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*, 1–27. doi:10.1017/S0305000910000759
- Jackson, C. W., Schatschneider, C., & Leacox, L. (2014). Longitudinal analysis of receptive vocabulary growth in young Spanish-English speaking children from migrant families. *Language, Speech, and Hearing Services in Schools, 45*, 40–51. doi:10.1044/2013_LSHSS-12-0104
- Jackson-Maldonado, D. (2004). Verbal morphology and vocabulary in monolinguals and emergent bilinguals. In B. Goldstein (Ed.), *Bilingual language development and disorders in Spanish-English speakers* (pp. 1331–1162). Baltimore, MD: Brookes.
- Jacobson, P. F., & Schwartz, R. G. (2005). English past tense use in bilingual children with language impairment. *American Journal of Speech-Language Pathology, 14*, 313–323.
- Jacobson, P. F., & Walden, P. R. (2013). Lexical diversity and omission errors as predictors of language ability in the narratives of sequential Spanish-English bilinguals: A cross-language comparison. *American Journal of Speech-Language Pathology, 22*, 554–565. doi:10.1044/1058-0360(2013/11-0055)
- Justice, L. M., Bowles, R., Pence, K., & Gosse, C. (2010). A scalable tool for assessing children's language abilities within a narrative context: The NAP (Narrative Assessment Protocol). *Early Childhood Research Quarterly, 25*, 218–234. doi:10.1016/j.ecresq.2009.11.002
- Kapantzoglou, M., Restrepo, M. A., & Thompson, M. S. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language,*

- Speech, and Hearing Services in Schools*, 43, 81–96. doi:10.1044/0161-1461(2011/10-0095)
- Kieffer, M. J., & Lesaux, N. K. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading & Writing*, 21, 783–804. doi:10.1007/s11145-007-9092-8
- La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The Elementary School Journal*, 104, 409–426. doi:10.1086/499760
- Miller, J. F., & Iglesias, A. (2010). Systematic analysis of language transcripts 2010 bilingual SE version (Computer software). Middleton, WI: SALT Software.
- Miller, L., Gillam, R. B., & Peña, E. (2001). *Dynamic assessment and intervention: Improving children's narrative abilities*. Austin, TX: PRO-ED.
- Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development & disorders* (2nd ed.). Baltimore, MD: Brookes.
- Paul, R., & Norbury, C. F. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (4th ed.). St. Louis, MO: Elsevier.
- Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2013). *Bilingual English Spanish Assessment (BESA)*. San Diego, CA: AR-Clinical Publications.
- Pence, K., Justice, L. M., & Gosse, C. (2007). *Narrative assessment protocol*. Columbus: Preschool Language and Literacy Lab, the Ohio State University.
- Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, 41, 1398–1411. doi:10.1044/jslhr.4106.1398
- Rezzonico, S., Chen, X., Cleave, P. L., Greenberg, J., Hipfner-Boucher, K., Johnson, C. J., . . . Girolametto, L. (2015). Oral narratives in monolingual and bilingual preschoolers with SLI. *International Journal of Language & Communication Disorders*, 50, 830–841. doi:10.1111/1460-6984.12179
- Squires, K. E., Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., Bohman, T. M., & Gillam, R. B. (2014). Story retelling by bilingual children with language impairments and typically developing controls. *International Journal of Language & Communication Disorders*, 49, 60–74. doi:10.1111/1460-6984.12044
- Terry, N. P., Mills, M. T., Bingham, G. E., Mansour, S., & Marencin, N. (2013). Oral narrative performance of African American prekindergartners who speak nonmainstream American English. *Language, Speech, and Hearing Services in Schools*, 44, 291–205.
- Tomasello, M. (2003). *Constructing a language: A usage based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Uccelli, P., & Páez, M. M. (2007). Narrative and vocabulary development of bilingual children from kindergarten to first grade: Developmental changes and associations among English and Spanish skills. *Language, Speech, and Hearing Services in Schools*, 38, 225–236. doi:10.1044/0161-1461(2007/024)
- Westby, C. (2005). Assessing and facilitating text comprehension problems. In H. Catts & A. Kamhi (Eds.), *Language and reading disabilities* (pp. 157–232). Boston, MA: Allyn & Bacon.
- Wood, C., Fitton, L., Petscher, Y., Rodriguez, E., Sunderman, G., & Lim, T. (in press). The effect of ebook vocabulary instruction on Spanish-English speaking children. *Journal of Speech Language Hearing Research*.
- Wood, C., Wofford, M. C., Gabas, C., & Petscher, Y. (2018). English narrative language growth across the school year: Young Spanish-English dual language learners. *Communication Disorders Quarterly*. Advance online publication. doi:10.1177/1525740118763063
- Wood, C., Wofford, M. C., & Schatschneider, C. (2018). Relationship between performance on oral narrative retells and vocabulary assessments for Spanish-English speaking children. *Communication Disorders Quarterly*, 39, 402–414. doi:10.1177/1525740117722507