Title

The Receptive–Expressive Gap in English Narratives of Spanish–English Bilingual Children With and Without Language Impairment

Journal

Journal of Speech Language and Hearing Research, 61(6)

ISSN

1092-4388

Authors

Gibson, Todd A
Peña, Elizabeth D
Bedore, Lisa M

Publication Date

2018-06-19

DOI

10.1044/2018_jslhr-l-16-0432

Peer reviewed

JSLHR

Research Article

The Receptive–Expressive Gap in English Narratives of Spanish–English Bilingual Children With and Without Language Impairment

Todd A. Gibson,[a] Elizabeth D. Peña,[b] and Lisa M. Bedore[c]

Purpose: First, we sought to extend our knowledge of second language (L2) receptive compared to expressive narrative skills in bilingual children with and without primary language impairment (PLI). Second, we sought to explore whether narrative receptive and expressive performance in bilingual children's L2 differed based on the type of contextual support.
Method: In a longitudinal group study, 20 Spanish–English bilingual children with PLI were matched by sex, age, nonverbal IQ score, and language exposure to 20 bilingual peers with typical development and administered the Test of Narrative Language (Gillam & Pearson, 2004) in English (their L2) at kindergarten and first grade.
Results: Standard scores were significantly lower for bilingual children with PLI than those without PLI. An L2 receptive–expressive gap existed for bilingual children with PLI at kindergarten but dissipated by first grade. Using single pictures during narrative generation compared to multiple pictures during narrative generation or no pictures during narrative retell appeared to minimize the presence of a receptive–expressive gap.
Conclusions: In early stages of L2 learning, bilingual children with PLI have an L2 receptive–expressive gap, but their typical development peers do not. Using a single picture during narrative generation might be advantageous for this population because it minimizes a receptive–expressive gap.

Narrative tasks are often used to assess the language of school-age children because they tap a range of functional language skills (as opposed to decontextualized language assessments that, e.g., ask children to name pictures, point to pictures, and finish sentences). Language skills engaged during narratives include but are not limited to children's ability to remember a complete story, link ideas within that story, and organize those ideas around a common theme (Gillam & Pearson, 2004). Difficulty in performing these tasks is associated with language impairment (Tsimpli, Peristeri, & Andreou, 2016). Furthermore, narrative ability develops both receptively (comprehension) and expressively (production; Bishop, 1997), which are similar but dissociable processes (Bates, 1993). A significant discrepancy between receptive and expressive ability has been a hallmark of language impairment (Gibson, Jarmulowicz, & Oller, 2018). Recent research indicates that a receptive–expressive gap occurs in vocabulary (Gibson, Oller, Jarmulowicz, & Ethington, 2012) and semantic (Gibson, Peña, & Bedore, 2014a) testing for bilingual children with typical development (TD) and is exacerbated for bilingual children with PLI (Gibson, Peña, & Bedore, 2014b). The current study seeks to expand our understanding of the receptive–expressive gap by investigating the trajectory and development of narrative receptive and expressive abilities of bilingual children with and without PLI.

Narrative Competence

Narratives develop early in life and follow a similar trajectory across languages (Berman & Slobin, 1994). In early development, children identify and link basic elements of stories, which Stein and Glenn (1979) have identified as settings, characters, actions, and events. By the age of 3 years, children produce true narratives (Applebee, 1978). By ages 5–6 years, children produce complex narratives

[a]Louisiana State University, Baton Rouge
[b]University of California, Irvine
[c]The University of Texas at Austin
Correspondence to Todd A. Gibson: toddandrewgibson@gmail.com

Editor: Sean Redmond
Associate Editor: Elizabeth Kay-Raining Bird
Received November 21, 2016
Revision received May 9, 2017
Accepted February 2, 2018
https://doi.org/10.1044/2018_JSLHR-L-16-0432

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

(Peterson & McCabe, 1983; Stein & Glenn, 1979). Similar trajectories have been observed cross-linguistically, with well-developed narratives being produced across a variety of languages by the age of 5 years (Berman & Slobin, 1994).

The development of narrative competence corresponds with and is supported by the development of short-term memory (Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009; Gathercole & Baddeley, 1993; Gathercole, Willis, Emslie, & Baddeley, 1991; Swanson, 2008). Dodwell and Bavin (2008) tested short-term memory as well as narrative comprehension and production abilities of 6-year-olds with and without primary language impairment (PLI) and identified statistically significant correlations between measures of short-term memory and measures of narrative performance. Duinmeijer, de Jong, and Scheper (2012) found similar correlations for Dutch speaking 6- to 9-year-olds with and without PLI.

Memory is important for remembering narrative events and tying those events together coherently, which is an important competency in narrative performance. Compare "He ate breakfast. He put on his clothes," to "He put on his clothes after he ate breakfast." The two short utterances include the same number of elements, but the latter version connects those elements whereas the former does not. This cohesion is accomplished by cohesive devices (Halliday & Hasan, 1976), such as the word "after" in the above example. A large body of research has shown that children with language impairment have difficulty with narrative cohesion (Boudreau & Chapman, 2000). Accomplishing cohesion requires the choice of appropriate vocabulary and grammar at the utterance level and the connection between distant aspects of a narrative, such as the initiating event and conclusion. The ability to effectively accomplish cohesion at both the utterance and story levels is what constitutes narrative competence.

Task Influence on Narrative Performance

Performance on narrative assessments varies as a function of the way narratives are elicited (Morris-Friehe & Sanger, 1992). Elicitation techniques range from minimal support (e.g., asking a child simply to tell a story with no contextual support; Spinillo & Pinto, 1994) to the highly supported task of asking a child to retell a story while viewing picture stimuli (Botting, 2002). The task demands vary considerably under these conditions. For example, because children's short-term memory was associated with scores on self-generated stories but not story retells (i.e., repeating a story that one hears), Dodwell and Bavin (2008) proposed that generating one's own narrative helped to create strong representations that were easier to remember compared to the representations created by merely listening to a story.

Narrative Retell Versus Narrative Generation

Several studies have demonstrated that outcomes on narrative retell and narrative generation tasks differ (Botting, 2002; W. M. Pearce, 2003). Merritt and Liles (1989) tested English-speaking children with and without PLI between ages 9 and 11 years. Participants retold two stories and generated three. To generate the stories, children were given story stems and asked to complete them (Merritt & Liles, 1989, p. 446). For the retell task, children watched a video of the story being told and were then asked to retell the same story. The authors found that narratives were longer and contained more story elements when they were retold compared to when they were generated based on a prompt. Similarly, Westerveld and Gillon (2010) tested children ages 7;3 to 9;3 on story retell and story generation. In the story retell task, children looked at pictures in a book while listening to a corresponding audio recording. To generate stories, children were asked to make up a story based on a single picture of a scene. Story retelling produced the most linguistically complex and longest stories. However, in the Dutch context, Duinmeijer et al. (2012) found greater complexity (e.g., embedding) associated with retelling, but length and fluency (absence of fillers and repetitions) were enhanced when generating a new narrative.

Influence of Visual Support on Narrative Performance

Even within these elicitation paradigms (retell vs. generation), task differences affect outcomes. Westerveld and Vidler (2015) asked typically developing children ages 5;3 to 8;9 to follow along with the pictures of a wordless picture book as the examiner read the story aloud. Children retold the story both with and without visual aids. Retelling a story with accompanying visual aids elicited longer and more complex narratives. However, as highlighted by Duinmeijer et al. (2012), comparisons across these elicitation techniques are misleading because they differ in more than the number of pictures provided. For example, W. M. Pearce (2003) found that children's narratives were of higher quality when based on a series of pictures compared to a single picture. Spinillo and Pinto (1994) found the opposite pattern. Duinmeijer et al. (2012) explained that this discrepancy was likely due to children in the Spinillo and Pinto (1994) study drawing their own pictures, which was a lived experience and likely enhanced significantly the children's narrative performance. This demonstrates that not only the quantity but also the quality of contextual support influences narrative outcomes.

Narrative Comprehension Versus Production

Narrative comprehension and production both require the understanding of words and the sentences in which they are embedded. However, production additionally requires strong representations that are sufficiently precise to express narrative meaning. Therefore, although children can often understand complex narratives, they may not be able to produce them at their same level of comprehension (Bates, 1993). The result might be a receptive–expressive gap.

It is unclear how differences in narrative tasks might differentially impact a receptive–expressive gap, if at all. More visual aids (vs. fewer), narrative retell (vs. narrative

generation), and narrative comprehension (vs. narrative production) all result in fewer memory demands (Baddeley, 1999; Mayer and Moreno, 2003; Paivio, 1986). Because short-term memory has a limited capacity, we might anticipate greater receptive–expressive gaps when narrative comprehension is compared to narrative generation (with greater memory demands) than when compared to narrative retell (with fewer memory demands).

Receptive–Expressive Discrepancies in Narratives

Tests of language comprehension are often considered easier than tests of language production (Gibson et al., 2014a), but direct comparisons between the two can be made by converting raw scores to standard scores (assuming a co-normed sample). This conversion mathematically controls for the differences in difficulty level. There is no reason to assume, therefore, that there would be discrepancies between receptive and expressive standard scores for groups of typically developing individuals (Gibson et al., 2014a). Indeed, such a discrepancy historically has been treated as a hallmark of expressive language disorder (Gibson et al., in press). For example, many standardized tests of omnibus language skills (e.g., Clinical Evaluation of Language Fundamentals–Fifth Edition; Wiig, Semel, & Secord, 2013; Preschool Language Scales–Fifth Edition; Zimmerman, Steiner, & Pond, 2012) provide not only separate standard scores for receptive and expressive performance but also discrepancy comparisons that allow clinicians to determine if these differences are clinically meaningful. When receptive skills exceed expressive skills by some threshold, results are often interpreted as an expressive language disorder (American Psychiatric Association, 2000, p. 58). Indeed, the World Health Organization (2005) has codified the distinction by providing separate International Classification of Diseases–Tenth Edition codes for receptive and expressive language disorders.

To our knowledge, only the Test of Narrative Language (TNL; Gillam & Pearson, 2004) provides separate comprehension and production standard scores for narrative skills, with a mean of 10 and a standard deviation of 3. However, the magnitude of discrepancies between receptive and expressive standard scores has been inconsistent across studies. Colozzo, Gillam, Wood, Schnell, and Johnston (2011) reported TNL scores for Canadian and American groups of children with and without PLI and age-matched peers with TD. Discrepancies between comprehension and production were less than 1 *SD* for both groups. Redmond (2011) found a similar discrepancy for children with PLI, but the comparison with TD had a larger discrepancy, with comprehension exceeding production by 1.05 *SD*s. Domsch et al. (2012) reported TNL comprehension and production scores for a group of late talkers and an age-matched control group of children who were not late talkers. For the late talkers, comprehension exceeded production by 1.03 *SD*s, but for the control group, this discrepancy was only 0.63 *SD*.

Another group that might present with a receptive–expressive gap in narrative performance is bilingual speakers with TD. Bohnacker (2016) administered receptive and expressive measures of narratives to 5- and 6-year-old Swedish-dominant Swedish–English bilinguals in both of their languages. Receptive measures focused on narrative elements by asking children *how* and *why* questions to determine if children could understand the goals and internal states of the characters. Expressive measures were based on narratives elicited through a four-picture series. Results were similar in both languages and showed a large gap across receptive and expressive modalities, with children more likely to understand than produce these narrative elements. This discrepancy was present for both the 5- and 6-year-old age groups, despite having TD. However, because the degrees of difficulty for receptive compared to expressive tasks were not controlled, interpreting these results is difficult.

Narrative Gap in Bilingual Children With and Without PLI

Knowledge of story elements, such as characters, settings, and events, has been referred to as macrostructure. Macrostructure appears early in narrative development and follows a similar trajectory across languages (Berman & Slobin, 1994). Fully formed narratives characterized by a problem, plot, character development, causal relationships, and resolution typically have developed by the age of 5 years (Stein & Glenn, 1979). Similar narrative developmental patterns have been identified across a variety of languages (Berman & Slobin, 1994). A distinct aspect of narrative competence is the ability to understand and produce internal narrative structures (Justice et al., 2006; Liles, Duffy, Merritt, & Purcell, 1995), and this is referred to as microstructure. At this level, speakers process the narrative within sentences by monitoring lexical and grammatical constructions and process across sentences by monitoring how sentences cohere to preceding and subsequent sentences (Liles et al., 1995). These elements are often measured in terms of productivity (e.g., how many words are produced per utterance; Paul & Smith, 1993) or complexity (e.g., the complexity of grammatical constructions; Norbury & Bishop, 2003). Indeed, researchers have found that microstructure is a multidimensional construct (Justice et al., 2006) made up of complexity and productivity (Westerveld & Gillon, 2010). Narrative competence, therefore, relies on individuals' understanding and production of both macro- and microstructure. Indeed, guidelines for best practices in the assessment of narrative ability indicate that both levels should be assessed (Hughes, McGillivray, & Schmidek, 1997).

Few studies have investigated the narratives of bilingual children with PLI, but it appears that these children's errors are similar to those of monolingual children with PLI (Gutiérrez-Clellen, Simon-Cereijido, & Sweet, 2012). Most of these comparisons, however, have been made at the level of vocabulary and syntax, for which there is general consensus that bilingual children with PLI have impoverished narratives compared to their peers with TD

(Iluz-Cohen & Walters, 2012; McCabe & Bliss, 2005). However, results comparing the macrostructural performance of bilingual children with and without PLI have been inconsistent.

Altman, Armon-Lotem, Fichman, and Walters (2016) investigated the macrostructural elements produced by Hebrew–English bilingual children with and without PLI. When comparing within subjects, bilingual children with and without PLI performed similarly in macrostructural measures in both of their languages. When comparing across groups, children with and without PLI performed similarly in their first language, but children with PLI had weaker, though statistically nonsignificant, performance in their second language (L2). In a similar study, Tsimpli et al. (2016) compared Greek-speaking bilingual children with and without PLI and found no statistically significant difference between the groups with respect to macrostructure. However, Squires et al. (2014) found that the macrostructural performance of Spanish–English bilingual children with PLI was significantly impoverished compared to their bilingual peers with TD.

Understanding the exact ways in which narratives differ between bilingual children with and without PLI is important because these differences might help identify diagnostic markers of PLI in bilingual speakers. It is not clear whether the receptive–expressive discrepancies that might be present in the narratives of monolingual children with PLI are also present in the narratives of bilingual children with PLI. Furthermore, if present, it is not clear whether they would differ from the performance of their bilingual peers with TD.

Research Questions

The purpose of the current study was twofold. First, we wished to extend our knowledge of receptive compared to expressive narrative skills in bilingual children with and without PLI. Specifically, we wished to compare the performance of bilingual children with and without PLI on the TNL (Gillam & Pearson, 2004). Most studies of narratives in bilinguals focus on production and not comprehension, and fewer still look at language impairment (however, see Altman et al., 2016; Blom & Boerma, 2016; Tsimpli et al., 2016). None, however, has used a comprehensive measure of narrative elements, and none has investigated this phenomenon among Spanish–English bilingual speakers. We focus on Spanish–English bilingual speakers because these children make up the largest proportion of bilingual speakers in the United States. In addition, we would expect to see similarities across different language combinations. Therefore, the results of the current study are potentially widely applicable. Second, we sought to explore whether narrative receptive and expressive performance differed based on contextual support. Although the issue of contextual support has been addressed for monolingual children, little is known about the effect of contextual support on narrative performance for Spanish–English bilingual children.

As part of a larger study, we had access to standardized measures of both receptive and expressive performance of narrative skills for bilingual children with and without PLI. In addition, these data were available for the same set of children at both kindergarten and first grade. Based on the above literature review and the available data, we asked the following questions.

1. Are there discrepancies between the receptive and expressive narrative performance on the TNL (Gillam & Pearson, 2004) for bilingual children's L2, and do these differ for children with and without PLI in kindergarten and first grade?
2. Are there discrepancies between receptive and expressive performance related to the degree of contextual support for children with and without PLI?

Method

Participants

Participants in the current study were reported on in Squires et al. (2014). Forty-two children from the 166 participants in a longitudinal study of bilingual diagnostic markers of PLI (Gillam, Pena, Bedore, Bohman, & Mendez-Perez, 2013) initially were included in the current study. Participants attended 12 schools from northern Utah and central Texas that served large Latino populations. Students from the targeted classrooms were invited to participate if they spoke Spanish, English, or both. The return rate for consent forms was 85%.

Children with PLI were identified using the approach developed by Tomblin, Records, and Zhang (1996) and reported previously in Gillam et al. (2013). Children's performance in vocabulary, morphosyntax, and narration in both English and Spanish was evaluated by three bilingual speech-language pathologists with experience diagnosing Spanish–English bilingual speakers with PLI. The bilingual speech-language pathologists reviewed transcriptions of narrative samples in Spanish and English, responses to standardized tests in Spanish and English, and parent and teacher reports of child proficiency in Spanish and English to make a holistic judgment of language ability based on their clinical expertise. Scores of 0, 1, 2, 3, 4, or 5 were applied to represent severe/profound impairment, moderate impairment, mild impairment, low normal performance, normal performance, or above normal performance, respectively. This scoring regime applied to both Spanish and English independently as well as overall performance. If at least two raters assigned scores of 2 or below in each language, children were identified as having PLI. Twenty-one of the 166 children were categorized as having PLI. An interrater reliability of 87% was calculated using an AC_1 statistic (Gwet, 2008), indicating high levels of agreement.

Children with PLI were matched with children with TD based on sex, age (to within 4 months of birth), nonverbal IQ score (Universal Nonverbal Intelligence Test; Bracken & McCallum, 1998), and language exposure (to

within 20% English and Spanish). Language exposure was calculated based on parent and teacher questionnaires (Bohman, Bedore, Pena, Mendez-Perez, & Gillam, 2010; Gutiérrez-Clellen & Kreiter, 2003; Restrepo, 1998) that resulted in an hour-by-hour report of language exposure. Children were also matched as closely as possible to age of first English exposure, which averaged 2.2 years (see Table 1 for demographic information). One girl from the PLI group was not administered the Dragon Story, which was a narrative generation task based on a single picture; therefore, she and her matched peer from the TD group were not included in the final analysis. Eight children (40%) from each group received dual language instruction with 16%–83% of the school day taught in Spanish according to teacher interviews. The other 12 children (60%) were in English-only classrooms.

Materials

A battery of standardized language assessments was administered to participants in kindergarten and again when those children reached first grade. Of interest to the current study was the TNL (Gillam & Pearson, 2004), which was used to test narrative abilities in English. This English language test provides standardized receptive, expressive, and overall scores with a mean of 10 and a standard deviation of 3. The TNL includes six subtests.[1] In the first, children listen to a story with 155 words about going to a McDonald's restaurant. Children then answer questions about what they heard (e.g., they are asked the name of the girl in the story). No visual support is provided. In the second subtest, children retell the McDonald's Story without visual support. For the third subtest, children listen to the Shipwreck Story, a story about a child who ruins his model boat on the way to school. The story includes 175 words and is accompanied by five wordless pictures illustrating the events in the story. This story serves as a model for the fourth subtest, the Late for School Story, in which children generate an original story based on a series of five wordless pictures illustrating a boy rising from bed and ultimately arriving late for school. In the fifth subtest, children listen to a story with 390 words about a dragon. The Dragon Story corresponds to a single picture in which appear a fire-breathing dragon, a box of treasure, a large rock, and two children. During the instructions for the Dragon Story, children are told that the next task will require them to generate their own stories. In this way, the dragon story serves as an explicit model for the subsequent subtest. In the sixth subtest, the Aliens Story, children generate their own story based on a single picture that includes a family of space aliens leaving a space ship with their pet and arriving at a park where they appear to terrify one human child and fascinate another.

[1]Descriptions of test items and scoring procedures from the Test of Narrative Language (Gillam & Pearson, 2004) used with permission. Copyright © 2004 Pro-Ed.

Receptive testing included who, what, where, how, and why questions about the stories that they heard. Children were awarded 1, 2, or 3 points for each question correctly answered (some questions could be answered with a single response, and others required up to three responses; e.g., children were asked what one of the characters ordered, which required three responses). The receptive version of the McDonald's Story had a total possible 15 points, the Shipwreck Story had a possible 11 points, and the Dragon Story had a possible 14 points. Each expressive subtest targeted aspects of narrative. In the expressive version of the McDonald's Story, children earned points by how faithfully they recalled the elements of the story that they heard. For example, they earned points for including the names of participants, including the word McDonald's, and including what they ate and what they did, for a total of 26 possible points awarded. In the Late for School Story, children received either 0, 1, or 2 points for how thoroughly they included information about the events in the story, grammar (the number of grammatical errors, whether they maintained the same tense throughout the story), and global story organization (i.e., did it make sense, was it complete?), for a total of 30 possible points. In the Aliens Story, children earned 0, 1, or 2 points for how thoroughly they included information about the setting, characters, story elements (events and temporal relationships), vocabulary and grammar, and global story organization, for a total of 34 possible points earned. A total raw score for comprehension and a separate total raw score for production were calculated by adding the total points across the three comprehension subtests and three production subtests, respectively. Separate standard scores for comprehension and production were calculated by referencing tables in the Examiner's Manual. In addition, an overall standard score can be calculated by matching the sum of standard scores to tables provided in the Examiner's Manual.

The norming sample of the TNL was chosen to match the demographic characteristics of the United States as reported by the United States Census Bureau (2001). Twelve percent of the sample was Hispanic, but L2 status was not reported. Thirteen percent of the sample included children with exceptional status, such as language/learning disorder, articulation disorder, attention-deficit disorder, or other. Because TNL norming samples did not adequately represent bilingual participants, results were not used to determine the presence or absence of language disorder.

Licensed speech-language pathologists administered the TNL and scored the tests. As per the TNL Examiner's Manual, scoring took place by listening to the digital audio recording of the testing sessions. There were some occasions when children code-switched and responded in Spanish. In the current study, we scored these occurrences as inaccurate, reflecting how the average English-speaking speech-language pathologist would likely treat these responses. To determine reliability, research assistants trained by licensed speech-language pathologists rescored a proportion of the tests. The percentage of agreement between scorers was calculated for each subtest. For 20% of the

Table 1. Means and standard deviations for demographic variables by group.

Measures	PLI, mean (*SD*)	TD, mean (*SD*)
Nonverbal IQ	88.15 (11.91)	92.75 (12.42)
Age in months	68.00 (4.58)	68.30 (3.51)
Age of first English exposure in years	2.15 (1.72)	2.45 (1.53)
% English exposure in K	55.31 (20.98)	55.55 (22.48)
% English exposure in first grade	64.47 (22.80)	60.99 (18.82)
% Girls	40	40
PLI severity level*	1.58 (.72)	4.10 (.43)
TOLD spoken language quotient at K*	63.65 (6.36)	79.89 (10.99)
BESA English semantics at K*	31.35 (13.37)	52.81 (12.90)
BESA English morphosyntax at K*	19.21 (14.30)	54.13 (24.14)
BESA Spanish semantics at K*	28.78 (15.15)	46.94 (21.54)
BESA Spanish morphosyntax at K*	26.15 (17.43)	57.30 (27.17)

Note. K = kindergarten; TD = typical development; PLI = primary language impairment. PLI severity level based on a scale of 0 to 5, with 0 = *severely impaired*, 4 = *typical*, and 5 = *above average*. TOLD = Test of Language Development (standard score with mean of 100 and *SD* of 15). BESA = Bilingual English–Spanish Assessment. BESA scores represent percent correct.

*$p < .01$ (difference between the two groups on that measure based on one-way ANOVA).

tests, the points awarded by each examiner for every item administered to a child were compared. For each subtest, the total number of scorer agreements were divided by the total number of comparisons and multiplied by 100 to form a percentage. Interrater reliability was 99.2% for narrative comprehension and 100% for oral narration.

Procedure

Testing occurred in quiet spaces at children's schools. All samples were recorded using a digital audio recorder (Sony MS-515 or ICD-P320) with an external microphone (ECM115) and then transcribed using Sony digital voice editor version 2.4.04. To ensure transcription reliability, all transcripts were transcribed by a trained research assistant and checked by a second research assistant (usually the individual who had collected the language sample data).

Statistical Analyses

In order to determine differences in the receptive–expressive gap in narrative performance, we performed repeated-measures analyses of variance (ANOVAs). Because partial eta-squared includes variance only from the target variable (Pierce, Block, & Aguinis, 2004), we report it. As published previously (Gibson et al., 2014a), we developed guidelines for the interpretation of partial eta-squared because none exists. Because of its similarity with the general linear model, we interpreted the effect sizes of 0–.10, .10–.25, .25–.50, .50–.80, and .80–1.00 as negligible, small, moderate, large, and very large, respectively. We used *t* tests to follow up with the identification of differences between individual variables. To control for multiple comparisons, we used a Bonferroni correction of .02. In order to correct for dependence among means, we used a within-group Cohen's *d* as the effect size measure (Morris & DeShon, 2002).

Although the TNL provides multiple narrative tasks, the test does not provide separate standardized scores for each subtest. However, we were able to examine each subtest by calculating standard scores based on raw scores and standard deviations shared with us by the test developers. The means and standard deviations for raw scores in the normative sample for 5-, 6-, and 7-year olds were calculated. These raw scores and standard deviations were used to create *z* scores for each of the children in the current study. The *z* scores were then transformed to standard scores based on a mean of 10 and standard deviation of 3, which is the standard score format used by the TNL.

Results

For the first analysis, we were interested in whether there were differences between receptive and expressive scores as tested by the TNL in kindergarten children with and without language impairment and whether there were changes over time. Toward that end, we performed a $2 \times 2 \times 2$ repeated-measures ANOVA with time (kindergarten vs. first grade) and modality (receptive vs. expressive) as within-subject variables and ability (PLI vs. TD) as the between-subjects variable (see Figure 1). TNL standard scores were the dependent variable. There were main effects for time, $F(1, 38) = 60.61$, $p < .01$, $\eta_p^2 = .62$, a large effect size; modality, $F(1, 38) = 4.82$, $p = .03$, $\eta_p^2 = .11$, a small effect size; and ability, $F(1, 38) = 46.59$, $p < .01$, $\eta_p^2 = .55$, a large effect size. Children increased their scores from kindergarten ($M = 4.66$, $SD = 0.28$) to first grade ($M = 6.71$, $SD = 0.31$). They scored higher in narrative receptive performance ($M = 5.97$, $SD = 0.28$) compared to narrative expressive performance ($M = 5.40$, $SD = 0.31$), and children with TD scored higher ($M = 7.52$, $SD = 0.38$) than those with PLI ($M = 3.85$, $SD = 0.38$). There was a significant interaction between time, modality, and ability, $F(1, 38) = 6.09$, $p = .02$, $\eta_p^2 = .14$,

Figure 1. Typical development (TD) and primary language impairment (PLI) total narrative performance: Modality × Time. Kinder = kindergarten.

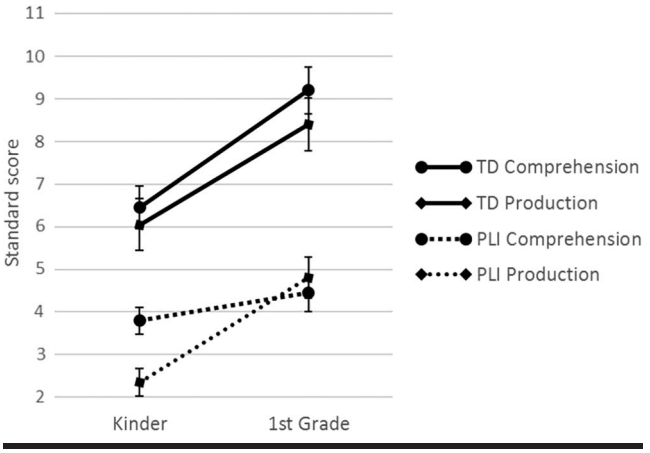
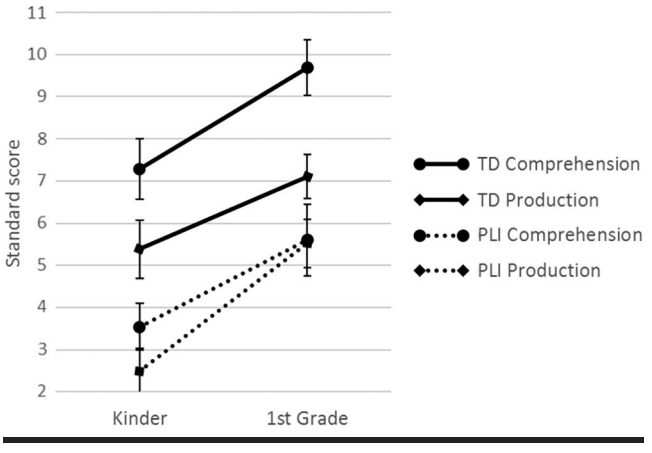


Figure 3. Typical development (TD) and primary language impairment (PLI) narrative performance in the multiple pictures + generation condition: Modality × Time. Kinder = kindergarten.



a small effect size. Follow-up analyses at the univariate level showed a receptive–expressive gap at kindergarten for children with PLI, $t(19) = 3.81$, $p = .001$, $d = 0.85$, a large effect size, but not for children with TD, $t(19) = .87$, $p = .39$, $d = 0.19$. There was no statistically significant receptive–expressive gap for either group at first grade.

We also asked whether there were discrepancies between receptive and expressive performance related to the degree of contextual support for children with and without PLI. To answer this question, we performed a 2 × 2 × 3 × 2 repeated-measures ANOVA with time (kindergarten vs. first grade), modality (receptive vs. expressive), and contextual support (no picture + retell vs. multiple pictures + generation vs. single picture + generation) as within-subject variables and ability (PLI vs. TD) as the between-subjects variable (see Figures 2, 3, and 4). Mauchly's Test of Sphericity (Mauchly, 1940) indicated a violation of sphericity for the interaction between contextual support and time, $\chi^2(2) = 8.22$, $p = .02$; therefore, we used a Greenhouse–Geisser correction for this interaction, $\varepsilon = .83$.

The comparisons of interest to answer this question were contextual support as well as its interactions. Results showed main effects for contextual support, $F(2, 76) = 11.42$, $p < .001$, $\eta_p^2 = .23$, a small effect size. Follow-up paired samples t tests with Bonferonni corrected p value of .02 showed statistically significant differences when comparing the no picture + retell condition ($M = 6.87$, $SD = 2.35$) with multiple pictures + generation ($M = 5.82$, $SD = 2.52$), $t(39) = 4.11$, $p < .001$, $d = 0.65$, a large effect size, and single pictures + generation ($M = 5.90$, $SD = 2.47$), $t(39) = 4.11$, $p < .001$, $d = 0.65$, a large effect size. There was no statistically significant difference between the single picture + generation and multiple picture + generation conditions, $t(39) = .29$, $p = .76$, $d = 0.04$. There was a statistically

Figure 2. Typical development (TD) and primary language impairment (PLI) narrative performance in the no picture + retell condition: Modality × Time. Kinder = kindergarten.

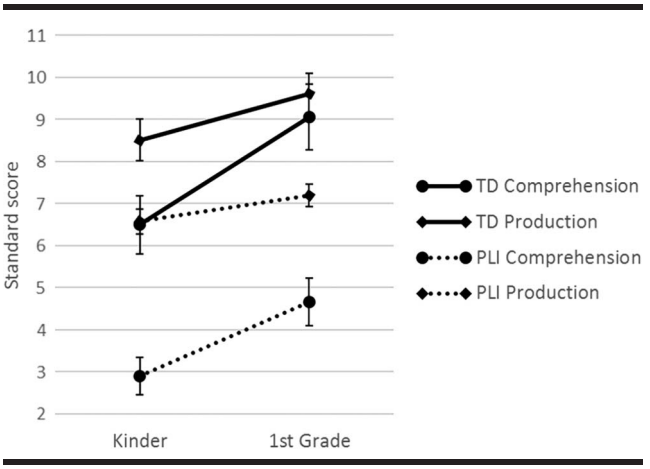
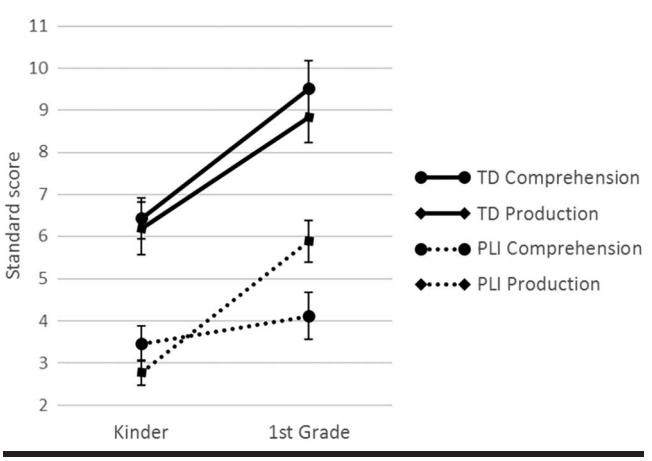


Figure 4. Typical development (TD) and primary language impairment (PLI) narrative performance in the single picture + generation condition: Modality × Time. Kinder = kindergarten.



significant three-way interaction with contextual support, which included contextual support, modality, and time, $F(2, 76) = 4.47$, $p = .02$, $\eta_p^2 = .10$, a small effect size. Scheffe's post hoc tests revealed that there were higher expressive than receptive scores for the no pictures + retell condition when compared to the single picture + generation condition at kindergarten ($p < .01$), but not at first grade. There were also higher expressive than receptive scores for no pictures + retell compared to multiple pictures + generation both at kindergarten ($p < .001$) and first grade ($p < .01$).

In addition, there was a statistically significant interaction between contextual support and modality, $F(2, 76) = 48.71$, $p < .001$, $\eta_p^2 = .56$. As a post hoc, we compared the receptive–expressive gap for each level of contextual support using paired samples t tests. There were statistically significant differences for each of the three comparisons. The greatest difference was found between the gap for no pictures + retell, $M = -2.19$, $SD = 1.87$, and multiple pictures + generation, $M = 1.40$, $SD = 2.35$, $t(39) = -10.61$, $p < .001$, $d = -1.71$, a very large effect size. The second greatest difference was between no pictures + retell and single picture + generation, $M = -0.04$, $SD = 2.21$, $t(39) = -6.01$, $p < .001$, $d = -0.95$, a very large effect size, followed by the difference between single picture + retell and multiple pictures + generation, $t(39) = -3.64$, $p = .001$, $d = -0.57$, a large effect size.

All other main effects and interactions involving contextual support were not significant. There was another significant three-way interaction, but it did not include contextual support: modality, time, and ability, $F(2, 38) = 5.02$, $p = .03$, $\eta_p^2 = .12$, a small effect size.

The absence of a statistically significant interaction between contextual support and ability, $F(1.59, 60.72) = 2.85$, $p = .07$, $\eta_p^2 = .07$, suggested that the different levels of contextual support affected bilingual children with PLI and TD similarly. There was a statistically significant interaction between contextual support and modality, $F(1.72, 65.60) = 5.85$, $p = .007$, $\eta_p^2 = .13$, a small effect size. Follow-up analyses at the univariate level using a Bonferonni-corrected p value of .02 showed that children performed better at expressive ($M = 7.97$, $SD = 1.79$) than receptive narrative skills ($M = 5.77$, $SD = 3.10$), $t(39) = -7.43$, $p < .001$, $d = -1.58$, a very large effect size, for the McDonald's story. Contextual support in the McDonald's story was provided by giving children the story first and then asking them to retell the same story, but no visual support was provided. However, when children were provided a series of pictures during a generation task, they performed better at receptive ($M = 6.53$, $SD = 3.26$) than expressive narrative skills ($M = 5.12$, $SD = 2.22$), $t(39) = 3.77$, $p = .001$, $d = 0.65$, a large effect size. There was no statistically significant difference between receptive ($M = 5.87$, $SD = 2.99$) and expressive ($M = 5.92$, $SD = 2.38$) performance when contextual support included a single picture + generation, $t(39) = -0.13$, $p = .89$, $d = -0.02$. There was no statistically significant interaction between contextual support and time, $F(2, 76) = 2.54$, $p = .08$, $\eta_p^2 = .06$.

Given the differential performance on expressive and receptive narrative testing for the bilingual children with PLI, we sought to determine if bilingual children with and without PLI had differential responses to language experience. Toward this end, we performed bivariate correlations between receptive testing, expressive testing, and language experience for each ability group at kindergarten and first grade. For bilingual children without PLI, results showed a statistically significant correlation between language experience and expressive narrative performance at kindergarten ($r = .57$, $p < .01$) and a significant correlation between language experience and receptive narrative performance at first grade ($r = .48$, $p = .03$). There were no statistically significant correlations for the bilingual children with PLI.

Discussion

We sought to determine if a receptive–expressive gap existed in the English (L2) narratives of Spanish–English bilingual children at kindergarten and first grade and, if present, whether it differed between children with and without PLI. An analysis of the overall receptive and overall expressive standard scores (i.e., the standard scores that combine the three different levels of contextual support) found that, at both kindergarten and first grade, bilingual children with PLI performed lower than their TD peers both receptively and expressively. In addition, bilingual children with PLI presented with a receptive–expressive gap at kindergarten that diminished to nonsignificance by first grade. It appeared that, for bilingual children with PLI, the greatest contributor to this diminution was improvement in productive ability. No receptive–expressive gap was present in the narratives of bilingual children with TD at either kindergarten or first grade. Although these children scored higher than their peers with PLI, they performed lower on their overall standard scores than the normative mean at kindergarten. By first grade, however, they were scoring within the normative mean for their overall narrative standard scores. It appears then that bilingual children with and without PLI were experiencing accelerated learning that was reflected in standardized language testing. Furthermore, the different types of contextual support had similar impacts on bilingual children with TD and bilingual children with PLI.

These results extend and support the findings of Altman et al. (2016) and Squires et al. (2014), who found that the narratives of bilingual children with PLI are impoverished compared to their peers with TD. Furthermore, these results suggest that a receptive–expressive gap in L2 narratives does not appear to be a phenomenon attributable to typical bilingual language development. The gap does appear, however, to be associated with bilingual PLI, at least in the earlier stages of L2 learning.

We considered the possibility that 1 year of additional practice with the language was sufficient to eliminate the receptive–expressive gap for bilingual children with PLI. However, simultaneously, these children underwent an increase in English language experience between kindergarten and first grade. Therefore, additional exposure to

English or a combination of additional exposure and additional practice may have contributed to the diminution in the receptive–expressive gap. To explore this possibility, we asked whether language experience correlated with receptive and expressive TNL performance at kindergarten and first grade for bilingual children with and without PLI. We found that there were no statistically significant correlations between language experience and TNL performance for the children with PLI. It does not appear, therefore, that additional experience with the language (in the current case, a year) plays a significant role in the diminution of the receptive–expressive gap.

We speculate that maturational development, especially in short-term memory capacity, might play a prominent role in the dissipation of the receptive–expressive gap over time. As mentioned above, both Dodwell and Bavin (2008) and Duinmeijer et al. (2012) found that measures of short-term memory correlated significantly with narrative performance. There is a substantial body of literature that has demonstrated that children's short-term memory improves with age (Barrouillet et al., 2009; Swanson, 2008), and this improvement aids in the development of language (Gathercole & Baddeley, 1993; Gathercole et al., 1991). Because narratives require children to remember large amounts of information and connect that information coherently, it might be the case that the more difficult production tasks are enhanced to a greater degree than the less difficult comprehension tasks, contributing to the diminution of the narrative receptive–expressive gap over time. Furthermore, because narrative retellings as opposed to narrative generation likely would be more affected by memory, the diminution in a receptive–expressive gap might be greatest in narrative retellings.

Additional results also suggest a role for memory. Although there was no statistically significant difference overall between the single picture + generation and multiple picture + generation conditions, their patterns of performance with respect to receptive and expressive modalities differed. In the multiple pictures + generation context, receptive was better than expressive performance at both kindergarten and first grade. The relationship between receptive and expressive performance also held at both kindergarten and first grade for the single picture + generation condition, with no statistically significant difference between receptive and expressive performance at either moment in time. One possible interpretation of these patterns is that the multiple pictures + generation context provides too many elements for these children to contend with expressively; although overall performance improves across the year for this condition, the relationship between expressive and receptive performance remains because of the heavy demands on memory to successfully perform this task. On the other hand, a single picture provides all of the elements of the story in a single frame, which perhaps taxes memory to a lesser degree and thus does not provoke a gap between receptive and expressive performance. Future studies could test this possibility by including memory measures in addition to the TNL.

Our previous research found receptive–expressive gaps for bilingual children with and without PLI during semantic tasks (Gibson et al., 2014a). We argued that a possible reason for this discrepancy was the quality of children's phonological representations. Perhaps because of limited experience with the language, L2 representations were underspecified such that they were sufficient to be successful with the easier receptive task but not sufficient to be successful with the more difficult expressive task (Gibson et al., 2014b). A similar phenomenon may have been engaged in the current generative narrative tasks. If children had difficulty in accessing individual words related to generating narratives, it would have had detrimental effects on scoring.

It might also be the case that the receptive–expressive gap for the children with PLI diminished over time because of the childhood culture in which these children lived. Corsaro and Eder (1990) identified a childhood culture in which stories play an important role. Narratives produced during pretend play help children earn peer acceptance and become a part of the social group (Hoyle, 1998). During these activities, children learn about characters and events and their organization. For example, if children pretend to take the role of mother and daughter during play, they take on the relationship characteristics of real mothers and daughters in a pretend reality, providing insight into characters, settings, and events (Goodwin, 1993; Sacks, 1992). In addition, this is a period when stories are used in the educational setting. For example, the Common Core State Standards for education in the United States provide direction on curricula from kindergarten through 12th grade (Common Core State Standards Initiative, 2010). The majority of the goals related to reading standards in kindergarten and the first grade are based on understanding and generating stories. Although we do not have information about these classroom activities for the participants in the current study, perhaps these children benefited from such educative practice. Expressive narrative abilities might have been boosted more than receptive because receptive abilities may have already been near these participants' ceiling performance. Future studies should explore this possibility.

We additionally sought to identify differences between groups based on different levels of contextual support used to elicit the narratives. Results showed that levels of contextual support had a similar impact on bilingual children with TD and bilingual children with PLI. Single picture + generation and multiple pictures + generation had a statistically similar impact on narrative performance, but performance was best in the no picture + retell context. The no picture + retell context was based on a McDonald's narrative. We suspect that, because of the popularity of McDonald's across the world, these children already had a mental model of the event of visiting a McDonald's restaurant, which they could use to support their performance. Future studies might compare performance when the narrative is based on a less frequent activity and setting. On the other hand, the McDonald's narrative was a story retell task. Because children did not have to generate

an original story in this case, it may have been easier and thus improved performance when compared to the story generation tasks. Future studies should consider this possibility.

Contextual support had a differential impact on receptive versus expressive performance. In the no picture + retell context, children performed better at expressive than receptive narrative tasks. We attribute this again to the specific story used to elicit the narrative, a visit to McDonald's, for which most children likely have a mental model to support performance. However, when answering questions about a McDonald's Story, children were required to recount a number of details that may have been missed due to weak L2 skills. Alternatively, they might have recognized the details but lacked the L2 skills to articulate them. The fact that the magnitude of the gap over time diminished most for this level of contextual support compared to the others appears to support our above proposal that maturational development in short-term memory capacity likely contributes to the diminution of the receptive–expressive gap.

When narrative testing was based on a series of pictures, receptive performance outpaced expressive performance. Gutiérrez-Clellen (2002) proposed that bilingual children attend to lexical and syntactic features of narratives in L2 to such a degree that it negatively impacts performance. We suspect that the series of pictures provided by the contextual support alleviated some of the cognitive burden imposed by memory and attention. This may have freed cognitive resources to answer questions about the narrative. Because children can succeed in the narrative expressive task without the need to recall as many specific details, the pictures appear to have aided receptive beyond expressive performance. However, we cannot separate the influence of the pictures from other support provided during testing. For example, by the time children were tested on receptive narrative performance using a series of pictures, they had already heard and retold a story about McDonald's. The preceding testing provided a model of narratives that may have contributed to children's performance. Future studies should attempt to identify the unique roles of model narratives and visual elicitation techniques.

There was no presence of a receptive–expressive gap when a single picture was provided to elicit a narrative. This suggests that single pictures might equally influence both receptive and expressive narrative performance. A single picture might support memory in a similar way as a series of pictures but to a lesser degree, thus aiding receptive performance. A single picture might also provide elements for the production of narrative, thus aiding expressive performance. The result appears to be a more-or-less balanced impact across receptive and expressive modalities.

Limitations

Although the current study compared bilingual children with and without PLI, it did not include a monolingual control group. Future studies should include monolingual comparisons. The children in the current study were only tested in English so it is not clear if results would have differed based on the language of testing. Future studies should include measures of narrative performance in both of the bilinguals' languages. Because only the TNL was administered in the current study, it is not clear how generalizable the results are; perhaps other measures of English language narrative performance would have resulted in different outcomes. Future studies should explore other measures of receptive and expressive narrative performance in English. The current study did not have information regarding speech therapy services or in-class teaching of narrative-related skills. Future studies should collect information about these data points, both of which potentially could contribute to a diminution of the gap over time.

Conclusions/Clinical Implications

Bilingual children with PLI produce L2 narratives that are impoverished compared to their TD peers. This further supports the use of narratives as a diagnostic indicator of PLI in bilingual children. In the early stages of L2 learning, children with PLI present with receptive standard score performance better than expressive standard score performance, but their TD peers do not. Future studies should explore the receptive–expressive gap in the narratives of bilingual children as a diagnostic marker of PLI. In addition, the type of task used to elicit narratives has a differential effect on narrative receptive compared to expressive performance. Using a single picture for contextual support might be advantageous for the bilingual population because it minimizes a receptive–expressive gap. This is important because a receptive–expressive gap has been treated as a hallmark of an expressive language disorder.

Acknowledgments

This work was supported by Grant R01DC007439 from the National Institute on Deafness and Other Communication Disorders, which was awarded to D. Kimbrough Oller. We thank the anonymous reviewers for their comments on this manuscript. We are grateful to the families that participated in the study. We would also like to thank Anita Méndez Pérez and Chad Bingham for their assistance with coordination of data collection, the interviewers for their assistance with collecting the data for this project, and the school districts for allowing us access to collect the data. This report does not necessarily reflect the views or policy of the National Institute on Deafness and Other Communication Disorders.

References

Altman, C., Armon-Lotem, S., Fichman, S., & Walters, J. (2016). Macrostructure, microstructure, and mental state terms in the narratives of English–Hebrew bilingual preschool children with and without specific language impairment. *Applied Psycholinguistics, 37,* 165–193.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition (Text Revision)*. Washington, DC: American Psychiatric Association.

Applebee, A. N. (1978). *The child's concept of story: Ages two to seventeen*. Chicago, IL: The University of Chicago Press.

Baddeley, A. (1999). *Essentials of human memory*. Essex, UK: Psychology Press.

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: A time-based resource-sharing model account. *Developmental Psychology, 45,* 477–490.

Bates, E. (1993). Comprehension and production in early language development. *Monographs of the Society for Research in Child Development, 58,* 222–242.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Erlbaum.

Bishop, R. (1997). Interviewing as collaborative storying. *Education Research and Perspectives, 24,* 28–47.

Blom, E., & Boerma, T. (2016). Why do children with language impairment have difficulties with narrative macrostructure? *Research in Developmental Disabilities, 55,* 301–311.

Bohman, T. M., Bedore, L. M., Pena, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: Language performance in Spanish–English bilinguals. *International Journal of Bilingual Education and Bilingualism, 13,* 325–344.

Bohnacker, U. (2016). Tell me a story in English or Swedish: Narrative production and comprehension in bilingual preschoolers and first graders. *Applied Psycholinguistics, 37,* 19–48.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy, 18,* 1–22.

Boudreau, D. M., & Chapman, R. S. (2000). The relationship between event representation and linguistic skill in narratives of children and adolescents with Down syndrome. *Journal of Speech, Language, and Hearing Research, 43,* 1146–1159.

Bracken, B., & McCallum, S. (1998). *Universal Nonverbal Intelligence Test*. Itasca, IL: Riverside.

Colozzo, P., Gillam, R., Wood, M., Schnell, R., & Johnston, J. (2011). Content and form in the narratives of children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 54,* 1609–1627.

Common Core State Standards Initiative. (2010). *Common core state standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Council of Chief State School Officers and National Governors Association. Retrieved from http://www.corestandards.org/

Corsaro, W. A., & Eder, D. (1990). Children's peer cultures. *Annual Review of Sociology, 16,* 197–220.

Dodwell, K., & Bavin, E. L. (2008). Children with specific language impairment: An investigation of their narratives and memory. *International Journal of Language & Communication Disorders, 43,* 201–218.

Domsch, C., Richels, C., Saldana, M., Coleman, C., Wimberly, C., & Maxwell, L. (2012). Narrative skill and syntactic complexity in school-age children with and without late language emergence. *International Journal of Language & Communication Disorders, 47,* 197–207.

Duinmeijer, I., de Jong, J., & Scheper, A. (2012). Narrative abilities, memory and attention in children with a specific language impairment. *International Journal of Language & Communication Disorders, 47,* 542–555.

Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education, 8,* 259–272.

Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12,* 349–367.

Gibson, T. A., Jarmulowicz, L., & Oller, D. K. (2018). Difficulties using standardized tests to identify the receptive–expressive gap in bilingual children's vocabularies. *Bilingualism: Language and Cognition, 21,* 328–339.

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

Gibson, T. A., Peña, E. D., & Bedore, L. M. (2014a). The relation between language experience and receptive–expressive semantic gaps in bilingual children. *International Journal of Bilingual Education and Bilingualism, 17,* 90–110.

Gibson, T. A., Peña, E. D., & Bedore, L. M. (2014b). The receptive–expressive gap in bilingual children with and without primary language impairment. *American Journal of Speech-Language, 23,* 655–667.

Gillam, R. B., & Pearson, N. (2004). *Test of Narrative Language*. Austin, TX: Pro-Ed.

Gillam, R. B., Pena, E. D., Bedore, L. M., Bohman, T. M., & Mendez-Perez, A. (2013). Identification of specific language impairment in bilingual children: I. Assessment in English. *Journal of Speech, Language, and Hearing Research, 56,* 1813–1823.

Goodwin, M. H. (1993). Accomplishing social organization in girls' play: Patterns of competition and cooperation in an African American working-class girls' group. In S. T. Hollis, L. Pershing, & M. J. Young (Eds.), *Feminist theory and the study of folklore* (pp. 149–165). Urbana: University of Illinois Publishing.

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education, 13,* 175–197.

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics, 24,* 267–288.

Gutiérrez-Clellen, V. F., Simon-Cereijido, G., & Sweet, M. (2012). Predictors of second language acquisition in Latino children with specific language impairment. *American Journal of Speech-Language Pathology, 21,* 64–77.

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology, 61,* 29–48.

Halliday, M. A., & Hasan, R. (1976). *Cohesion in English*. London, England: Longman.

Hoyle, S.M. (1998). Register and footing in role play. In M. S. Hoyle & C. Adger (Eds.), *Kids talk* (pp. 47–68). New York, NY: Oxford University Press.

Hughes, D. L., McGillivray, L., & Schmidek, M. (1997). *Guide to narrative language: Procedures for assessment*. Eau Claire, WI: Thinking Publications.

Iluz-Cohen, P., & Walters, J. (2012). Telling stories in two languages: Narratives of bilingual preschool children with typical and impaired language. *Bilingualism: Language and Cognition, 15,* 58–74.

Justice, L., Bowles, R., Kaderavek, J., Ukrainetz, T., Eisenberg, S., & Gillam, R. (2006). The index of narrative microstructure: A clinical tool for analyzing school-age children's narrative

performances. *American Journal of Speech-Language Pathology, 15,* 155–191.

Liles, B. Z., Duffy, R. J., Merritt, D. D., & Purcell, S. L. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech, Language, and Hearing Research, 38,* 415–425.

Mayer, R., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38,* 43–52.

Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *The Annals of Mathematical Statistics, 11*(2), 204–209.

McCabe, A., & Bliss, L. S. (2005). Narratives from Spanish-speaking children with impaired and typical language development. *Imagination, Cognition, and Personality, 24,* 331–346.

Merritt, D. D., & Liles, B. Z. (1989). Narrative analysis: Clinical applications of story generation and story retelling. *Journal of Speech and Hearing Disorders, 54,* 438–447.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125.

Morris-Friehe, M. J., & Sanger, D. D. (1992). Language samples using three story elicitation tasks and maturation effects. *Journal of Communication Disorders, 25,* 107–124.

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders, 38,* 287–313.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.

Paul, R., & Smith, R. (1993). Narrative skills in 4-year-olds with normal, impaired, and late-developing language. *Journal of Speech and Hearing Research, 36,* 592–598.

Pearce, W. M. (2003). Does the choice of stimulus affect the complexity of children's oral narratives? *Advances in Speech-Language Pathology, 5,* 95–103.

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York, NY: Plenum.

Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement, 64,* 916–924.

Redmond, S. (2011). Peer victimization among students with specific language impairment, attention-deficit/hyperactivity disorder, and typical development. *Language, Speech, and Hearing Services in Schools, 42,* 520–535.

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1398–1411.

Sacks, H. (1992). On some formal properties of children's games. In G. Jefferson (Ed.), *Lectures on Conversation* (pp. 489–506). Oxford, England: Basil Blackwell.

Spinillo, A. G., & Pinto, G. (1994). Children's narratives under different conditions: A comparative study. *British Journal of Developmental Psychology, 12,* 177–193.

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.

Stein, N. L., & Glenn, C. G. (1979). An analysis of stories comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Norwood, NJ: Able.

Swanson, H. L. (2008). Working memory and intelligence in children: What develops? *Journal of Educational Psychology, 100,* 581–602.

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research, 39,* 1284–1294.

Tsimpli, I. M., Peristeri, E., & Andreou, M. (2016). Narrative production in monolingual and bilingual children with specific language impairment. *Applied Psycholinguistics, 37,* 195–216.

United States Census Bureau. (2001). *Statistical Abstract of the United States 2001*. Washington, DC: U.S. Bureau of the Census, U.S. Government Printing Office.

Westerveld, M. F., & Gillon, G. T. (2010). Oral narrative context effects on poor readers' spoken language performance: Story retelling, story generation, and personal narratives. *International Journal of Speech-Language Pathology, 12,* 132–141.

Westerveld, M. F., & Vidler, K. (2015). The use of the Renfrew Bus Story with 5–8-year-old Australian children. *International Journal of Speech-Language Pathology, 17,* 304–313.

Wiig, E., Semel, E., & Secord, W. (2013). *Clinical Evaluation of Language Fundamentals–Fifth Edition, Spanish (CELF-5 Spanish)*. San Antonio, TX: The Psychological Corporation.

World Health Organization. (2005). *International statistical classification of diseases and related health problems* (10th revision). Geneva, Switzerland: Author.

Zimmerman, I., Steiner, V., & Pond, R. (2012). *Preschool Language Scales–Fifth Edition, Spanish (PLS-5 Spanish)*. San Antonio, TX: The Psychological Corporation.