LSHSS

Research Article

Finite Verb Morphology Composite Between Age 4 and Age 9 for the Edmonton Narrative Norms Instrument: Reference Data and Psychometric Properties

Ling-Yu Guo,[a,b] Sarita Eisenberg,[c] Phyllis Schneider,[d] and Linda Spencer[e]

Purpose: The purpose of this study was to provide reference data and evaluate the psychometric properties for the finite verb morphology composite (FVMC) measure in children between 4 and 9 years of age from the database of the Edmonton Narrative Norms Instrument (ENNI; Schneider, Dubé, & Hayward, 2005).
Method: Participants included 377 children between age 4 and age 9, including 300 children with typical language and 77 children with language impairment (LI). Narrative samples were collected using a story generation task. FVMC scores were computed from the samples. Split-half reliability, concurrent criterion validity, and diagnostic accuracy for FVMC were further evaluated.
Results: Children's performance on FVMC increased significantly between age 4 and age 9 in the typical language and LI groups. Moreover, the correlation coefficients for the split-half reliability and concurrent criterion validity of FVMC were medium to large (rs $\geq$ .429, ps < .001) at each age level. The diagnostic accuracy of FVMC was good or acceptable from age 4 to age 7, but it dropped to a poor level at age 8 and age 9.
Conclusion: With the empirical evidence, FVMC is appropriate for identifying children with LI between age 4 and age 7. The reference data of FVMC could also be used for monitoring treatment progress.
Supplemental Material: https://doi.org/10.23641/asha.10073183

A hallmark of English-speaking children with language impairment (LI) is the protracted developmental trajectory for learning to consistently use tense and agreement morphemes (Leonard, Haebig, Deevy, & Brown, 2017; Rice, Wexler, & Hershberger, 1998). Tense and agreement morphemes (hereafter, tense morphemes) in English refer to the inflections (e.g., third-person singular *–s* as in *The dog runs every day*) or function words (e.g., auxiliary *be* as in *The boys are laughing*) that give information pertaining to person, number, and time in sentences. In both spoken discourse and experimental probes, preschool children with LI were more likely to omit tense morphemes than those with typical language (TL) development (Goffman & Leonard, 2000; Leonard et al., 2017; Rice et al., 1998). The difficulty that children with LI have with using tense morphemes may persist even into the school-age years (Moyle, Karasinski, Ellis Weismer, & Gorman, 2011; Rice et al., 1998). Although there are different theoretical explanations for this tense morpheme deficit (see Leonard, 2014, for a review), there appears to be a consensus that it can be used as a clinical marker for identifying children with LI (Pawłowska, 2014; Tager-Flusberg & Cooper, 1999).

There are several different methods for calculating tense usage by English-speaking children from language

[a]Department of Communicative Disorders and Sciences, State University of New York at Buffalo
[b]Department of Audiology and Speech-Language Pathology, Asia University, Taichung, Taiwan
[c]Department of Communication Sciences and Disorders, Montclair State University, Bloomfield, NJ
[d]Department of Communication Sciences and Disorders, University of Alberta, Edmonton, Canada
[e]MSSLP Program, Rocky Mountain University of Health Professions, Provo, UT
Correspondence to Ling-Yu Guo: lingyugu@buffalo.edu

Editor-in-Chief: Holly Storkel
Editor: Marleen Westerveld
Received March 8, 2019
Revision received May 18, 2019
Accepted July 25, 2019
https://doi.org/10.1044/2019_LSHSS-19-0028

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

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

samples (Leonard et al., 2017), such as the tense and agreement productivity score (TAPS; Hadley & Short, 2005), the tense composite (Rice et al., 1998), and the finite verb morphology composite (FVMC; Bedore & Leonard, 1998; Leonard, Miller, & Gerber, 1999). TAPS documents the emergence of tense morphemes through their initial productivity. It is computed by identifying children's productive uses of five tense morphemes, including third-person singular *–s,* past tense *–ed,* copula *be,* auxiliary *be,* and auxiliary *do*. Given that productivity measures typically reveal whether a target structure is incorporated into the child's language system (Lahey, 1988; Schuele, 2013), TAPS is appropriate for documenting early development of tense usage in children (e.g., up to age 4; Gladfelter & Leonard, 2013; Guo & Eisenberg, 2014). In contrast with TAPS, the tense composite and the FVMC both evaluate the accuracy of tense usage in children. While the tense composite computes the accuracy of five tense morphemes combined (i.e., third-person singular *–s,* past tense *–ed,* copula *be,* auxiliary *be,* and auxiliary *do*), FVMC calculates the accuracy of four tense morphemes combined (i.e., third-person singular *–s,* past tense *–ed,* copula *be,* and auxiliary *be*). Because accuracy measures typically reveal the mastery of a target structure, both the tense composite and FVMC are appropriate for tracking the development of tense usage from preschool to early elementary school age (e.g., age 3 to age 8; Rice et al., 1998). It should be noted that, unlike the tense composite, FVMC does not include auxiliary *do* in the computation because English has both auxiliary *do* and main verb *do* (e.g., *He did a good job*). Including auxiliary *do* in the computation would require clinicians to differentiate auxiliary *do* and its main verb counterpart, which might not only increase the work of clinicians but also reduce the reliability in analysis (Goffman & Leonard, 2000). Given that FVMC could be used for a wider age range than the TAPS and potentially requires less work than the tense composite, we focused on FVMC in the present investigation.

Across studies, FVMC has been shown to be a promising measure for differentiating children with and without LI during their preschool and early elementary school years (Guo & Schneider, 2016; Pawłowska, 2014; Souto, Leonard, & Deevy, 2014). However, extant studies have used different language sampling protocols (e.g., free play, narrative) for calculating FVMC in preschool and school-aged children. Despite the fact that group data are available for FVMC at different ages from prior studies, the varied methodologies for data collection make it difficult for clinicians to compare across studies to obtain reference data for identifying children with LI. This issue is critical because, based on a survey by Pavelko, Owens, Ireland, and Hahs-Vaughn (2016), lack of reference data was a major reason why practicing clinicians do not regularly include language sample analysis in the assessment process. By reanalyzing the archival data of the Edmonton Narrative Norms Instrument (ENNI; Schneider, Dubé, & Hayward, 2005), this study aimed to provide reference data for FVMC in children between age 4 and age 9 who were administered a consistent language sampling protocol (i.e., a story generation task). To validate the use of FVMC computed from the ENNI protocol, we also evaluated the psychometric properties of FVMC, including split-half reliability, concurrent criterion validity, and diagnostic accuracy. In what follows, we first review the studies that investigated the performance of FVMC in English-speaking children with and without LI who were at the preschool or school ages. We then outline the scope of this study.

Performance of FVMC in Children With and Without LI

For the current report, we used the term *FVMC* to refer to the computation of tense morpheme usage based on four tense morphemes: third-person singular present *–s* (3SG *–s*), past tense *–ed,* copula *be,* and auxiliary *be* (Bedore & Leonard, 1998; Leonard et al., 1999). There have been some studies that used the term *FVMC* when auxiliary *do* was included in the computation (e.g., Gladfelter & Leonard, 2013; Leonard et al., 2017). To avoid confusion, for this report, we limited the term *FVMC* to the computation in which only the original four tense morphemes were included (Bedore & Leonard, 1998).

Guo and Eisenberg (2014) examined the performance of FVMC in 18 pairs of 3-year-old children with and without LI using conversational language samples that involved the child playing with toys with a parent. Classification of LI was based on intervention status (i.e., receiving therapy at the time of the study) and/or a standard score below 87 on the Structured Photographic Expressive Language Test–Preschool 2 (Dawson et al., 2005). The mean FVMC score was 97.44% (*SD* = 3.10%) for 3-year-old children with TL and 78.88% (*SD* = 21.83%) for those with LI. The FVMC score was significantly higher in the TL group than in the LI group, with large effect size.

In a subsequent study, Souto et al. (2014) separately examined FVMC in 14 pairs of 4-year-old and 16 pairs of 5-year-old children with and without LI using conversational language samples that involved the child playing with toys with an examiner. Classification of LI was based on a standard score below 67 on the Structured Photographic Expressive Language Test–Preschool 2 (Werner & Kresheck, 1983). The mean FVMC score was 96.97% (*SD* = 4.27%) for 4-year-old children with TL and 56.93% (*SD* = 23.91%) for those with LI. Moreover, the mean FVMC score was 97.64% (*SD* = 4.41%) for 5-year-old children with TL and 70.00% (*SD* = 18.82%) for those with LI. At both ages, the FVMC score was significantly higher in the TL group than in the LI group, with large effect sizes.

In contrast to other studies, Guo and Schneider (2016) evaluated FVMC in school-aged children with and without LI who were 6 or 8 years old using a narrative generation task. For age 6, 50 children with TL and 11 children with LI were included; for age 8, 50 children with TL and 17 children with LI were included. Classification of LI was based on intervention status and a standard score below 85 on at least one of the composite scores

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

(i.e., Receptive, Expressive, and Total Language Composites) of the Clinical Evaluation of Language Fundamentals–Third Edition (CELF-3; Semel, Wiig, & Secord, 1995). The mean FVMC score was 97% (*SD* = 3%) for 6-year-old children with TL and 79% (*SD* = 18%) for those with LI. Moreover, the mean FVMC score was 99% (*SD* = 2%) for 8-year-old children with TL and 88% (*SD* = 18%) for those with LI. The FVMC score was significantly higher in the TL group than in the LI group at both ages, with large effect sizes.

In summary, children with TL, as a group, produced FVMC at the customary level of mastery (i.e., > 90%; Brown, 1973) in conversations as early as age 3. Although prior studies have adopted different inclusionary criteria for children with LI, these studies have collectively shown that FVMC could differentiate children with and without LI from preschool to early elementary school ages. However, no studies have used one single, uniform language sampling protocol to generate reference data (e.g., mean, standard deviation) for FVMC from children with TL across the preschool and early elementary school ages to inform clinical decision making. Guo and Eisenberg (2014) used a parent-elicited conversational language sample for 3-year-old children; Souto et al. (2014) used an examiner-elicited conversational language sample for 4- and 5-year-old children; Guo and Schneider (2016) used a narrative sample for 6- and 8-year-old children. Without separate reference data for different age levels based on one single, uniform language sampling protocol, it is difficult for clinicians to determine whether a child's performance on FVMC is within the typical range or to evaluate whether a child makes significant progress on FVMC over time compared to same-aged peers.

A study by Goffman and Leonard (2000) did provide group data for FVMC based on 99 children with TL between the ages of 2;2 and 5;8 (years;months) in 5-month intervals using a consistent language sampling protocol (i.e., conversational language samples that involved children playing toys with the examiner). However, there are several issues that limit the usability of their data by clinicians for making normative comparisons. Some of the age groups had only a small number of children (e.g., nine children in the group between 3;5 and 3;9), which may not be a sufficient sample size to serve as reference data. In addition, the group data for FVMC (i.e., mean and ± 1 *SD*) were plotted in a figure, and no specific numbers were reported. Although clinicians could evaluate whether a child's FVMC falls below the −1 *SD* cutoff point, clinicians could not calculate a *z* score (i.e., number of standard deviations below the mean) given that means and standard deviations were not reported. Clinicians could not also apply a different cutoff (e.g., −1.25 *SD*) that might be required for a child to meet eligibility requirements to receive services. Furthermore, the group data in Goffman and Leonard included children up to the age of 5;8, which would not allow normative comparisons for children who are older (e.g., 6- or 7-year-old children). Thus, despite the fact that several studies have reported FVMC data from different age groups, there remains a critical need to obtain reference data for FVMC from children with TL at different ages between preschool and early school-age years using one single, uniform language sampling protocol.

This Study

Although FVMC has been shown to be a promising measure for differentiating children with and without LI, lack of adequate reference data due to the use of different language sampling protocols across studies has limited the clinical use of FVMC. To address this issue, this study aimed to provide reference data for FVMC in children between 4 and 9 years of age from the database of the ENNI (Schneider et al., 2005). The ENNI was originally designed to evaluate narrative skills in children between age 4 and age 9 using picture sequences (see Method section for details). The data in the ENNI were collected using a cross-sectional design, rather than a longitudinal design. We chose the ENNI database because a story generation task was administered consistently throughout the designated age range. The assessment protocol (e.g., pictures and instructions) and the reference data for a variety of macrostructural and microstructural measures (e.g., story grammar, syntactic complexity, total number of words) are freely available on the ENNI website. The available reference data for microstructural measures, however, do not include FVMC. This provides an opportunity for this study to generate reference data for FVMC using the archival data (i.e., narrative samples) from the ENNI database, which are publicly accessible from the website of the Child Language Data Exchange System (MacWhinney, 2000). The outcome of this study may further encourage clinicians to take advantage of the freely available ENNI protocol. It should be noted that the ENNI database is also incorporated into the normative database of the Systematic Analysis of Language Transcripts (SALT; Miller, Andriacchi, & Nockerts, 2016). Even though SALT is able to automatically generate reference data for a number of language sample measures (e.g., mean length of utterances, number of different words) using the ENNI database, it does not provide reference data for the FVMC score. Thus, despite the fact that the ENNI database is freely and commercially available in different websites/software, there remains no comprehensive reference data for evaluating children's performance on FVMC.

We are also aware of a freely available standardized test that specifically assesses children's tense usage from age 3 to age 8—the Rice Wexler Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001). To the best of our knowledge, the TEGI evaluates children's usage of five tense morphemes at the single-word or single-sentence levels, including third-person singular *–s,* past tense *–ed,* copula *be,* auxiliary *be,* and auxiliary *do*. The accuracy of using these five morphemes was combined together into a composite score. The diagnostic accuracy of the TEGI for identifying children with LI (i.e., sensitivity) and children with TL (i.e., specificity) was acceptable (80% accurate or higher;

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Plante & Vance, 1994) to good (90% accurate or higher) from age 3 to age 6, except for children between 3;0 and 3;5. Although sensitivity for the TEGI at age 7 and age 8 was acceptable, specificity was not reported. Using the ENNI protocol to compute FVMC would allow clinicians to further evaluate children's tense usage at the text level within real-life activities (i.e., narratives; Miller et al., 2016). In addition, using tense morphemes at the text level would presumably require more cognitive resources than using tense morphemes at the single-word or single-sentence levels (Charest & Johnston, 2011). If children do have deficits in tense usage, they are more likely to make errors at the text level than at the single-word or single-sentence level. This could make it relatively easier for clinicians to determine whether a child indeed has a deficit in tense usage. Moreover, because the ENNI database included children between age 4 and age 9, it allowed us to further evaluate the diagnostic accuracy (i.e., sensitivity and specificity) for children between age 7 and age 9.

This study had two goals. The first goal was to provide reference data for FVMC. To this end, we computed not only the means and standard deviations of FVMC but also 90% and 95% confidence intervals of FVMC for children with TL between age 4 and age 9 in 12-month intervals as reference data. We chose this age range as they reflected the archival data in the ENNI database. Providing reference data for language sample measures in 12-month intervals was also consistent with prior studies (e.g., Leadholm & Miller, 1992). To establish that the reference data of FVMC computed from the ENNI database was a valid clinical tool, an important initial step was to examine whether FVMC increased with age in children with and without LI and whether FVMC differed significantly between children with and without LI at each age level (i.e., construct validity; Aiken & Groth-Marnat, 2006). Thus, we also provided the means and standard deviations of FVMC from children with LI. With the group data from children with and without LI between age 4 and age 9 by age level, we were able to evaluate the initial validity for FVMC. In addition, few studies, if any, have reported confidence intervals for language sample measures. The provision of confidence intervals for FVMC based on children with TL would allow clinicians to estimate the range of a child's "true" score for FVMC, which would help clinicians interpret the results in the diagnosis process (Aiken & Groth-Marnat, 2006; McCauley & Swisher, 1984). Moreover, when tense usage is a treatment goal and FVMC is used as an outcome measure for monitoring treatment progress, the confidence intervals of FVMC would also allow clinicians to determine whether a child's performance on FVMC significantly improves over time (Hall-Mills, 2018). Specifically, the clinicians may obtain the lower and upper limits for a child's FVMC score before treatment, using 90% or 95% confidence intervals. After treatment, if the child's posttreatment FVMC score exceeds the upper limit of the pretreatment FVMC score, this would be considered as a significant improvement (Gillam et al., 2008). In contrast, if the child's posttreatment FVMC score falls below the upper limit of the pretreatment FVMC score, this would not be considered as a significant improvement even if the posttreatment FVMC score is higher than the pretreatment FVMC score.

The second goal was to further validate the use of FVMC computed from the ENNI database. To this end, we examined three psychometric properties for FVMC by age, including split-half reliability, concurrent criterion validity, and diagnostic accuracy. Split-half reliability of a measure evaluates the extent to which the results from one half of the items are consistent with those from the other half (i.e., internal consistency; Aiken & Groth-Marnat, 2006). We therefore computed the correlation between FVMC scores from one set of the stories (i.e., Story Set A) and those from the other set of the stories (i.e., Story Set B) in the ENNI protocol. Concurrent criterion validity of a measure evaluates the extent to which performance on this measure is consistent with performance on other measures designed to assess the same skill areas. We therefore computed the correlation between FVMC scores and a measure of expressive grammar—the Recalling Sentences in Contexts subtest of the Clinical Evaluation of Language Fundamentals–Preschool (CELF-P; Wiig, Secord, & Semel, 1992) or the Recalling Sentences subtest of the CELF-3 (Semel et al., 1995). Moreover, diagnostic accuracy of a measure evaluates the extent to which a measure could differentiate individuals with and without a condition (e.g., LI). For this, we examined the sensitivity, specificity, and likelihood ratios for FVMC.

It should be noted that the means, standard deviations, and diagnostic accuracy data for FVMC in 6- and 8-year-old children have been reported in Guo and Schneider (2016), whereas the rest of the data were unique to this study and have never been reported elsewhere. Thus, this study contributed to evidence-based assessment by extending the reference data for FVMC, by computing the confidence intervals for FVMC, and by examining the psychometric properties for FVMC.

In this study, we asked three questions. First, does FVMC increase significantly in children with and without LI who are between 4 and 9 years of age in the ENNI reference data? Second, do children with TL produce higher FVMC scores than those with LI in the ENNI reference data? Third, does FVMC show appropriate psychometric properties at each age level between age 4 and age 9? On the basis of prior studies (Guo & Eisenberg, 2014; Leonard et al., 2017; Rice et al., 1998; Souto et al., 2014), we predicted that the change in FVMC in the TL group would be minimal between 4 and 9 years of age because children with TL typically produced FVMC at the customary level of mastery (i.e., $\geq$ 90%) as early as age 3. In contrast, we predicted FVMC would increase significantly with age in the LI group. Therefore, children with TL may produce higher FVMC scores than those with LI between age 4 and age 9 but the magnitude of differences (i.e., effect size) may decrease with age. It was further predicted that FVMC would show appropriate psychometric properties between age 4 and age 9.

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Method

Participants

The present investigation used the data from the normative sample in the ENNI (Schneider et al., 2005). Data collection for the ENNI was approved by the institutional ethics review board at the University of Alberta. Participants in the normative sample were 377 children (300 TL, 77 LI) between 4 and 9 years of age recruited from the Edmonton area, Canada. Signed consents were obtained from the parents of all of the children who participated. There were 50 children with TL for each age level, but the number of children with LI varied (see Table 1). It should be noted again that, although the ENNI database included children between age 4 and age 9, it was a cross-sectional design, not a longitudinal design. The chronological ages were not significantly different between the TL and LI groups at any of the age levels, Fs ≤ 2.41, ps $\geq .13$, ds ≤ 0.13. The distribution of gender was also not significantly different between the TL and LI groups at any of the age levels, $\chi^2 \leq 3.03$, ps $\geq .08$. All children were from English-speaking families and spoke English at home from birth; in some cases, another language may also have been spoken in the home, as it was only specified that English must be the first language in the inclusion criteria when participants were recruited. However, the information regarding the percentage of participants who were exposed to a language other than English was not documented for the normative sample.

Children with TL in the normative sample were recruited from 13 day cares/preschools and 34 public elementary schools in the Edmonton area, all of which were randomly selected (Schneider et al., 2005). Teachers in the participating schools who had students in the target age range were asked to refer two children in the upper academic level of achievement, two children from the middle academic level, and two children in the lower academic level, with one boy and one girl at each level. This decision was made to ensure that the normative sample would consist of children with TL who had varying language skills. All children in the TL group were typically developing per teachers' reports and did not have speech or language difficulties or any other disorders such as learning disability, autism spectrum disorders, or attention-deficit/hyperactivity disorder (ADHD).

Table 1. Mean (*SD*) of demographic measures of children by language status and age.

Language status and age	*n*	Gender	Age in months	SES[a]	(Linguistic) Concepts and Directions[b]	Recalling Sentences (in Contexts)[c]	CELF-P/CELF-3[d]
Typical language							
4-year-olds	50	25 G, 25 B[e]	55.20 (2.88)	47.38 (13.58)	10.82 (3.32) 3–16[c]	9.96 (2.38) 5–18	—
5-year-olds	50	25 G, 25 B	66.12 (3.24)	46.64 (12.12)	10.74 (2.63) 3–15	9.96 (2.79) 3–16	—
6-year-olds	50	25 G, 25 B	78.94 (3.99)	48.31 (14.75)	11.58 (3.03) 6–17	11.76 (3.32) 5–17	—
7-year-olds	50	25 G, 25 B	90.48 (3.36)	45.13 (13.65)	12.24 (3.26) 4–17	11.66 (2.79) 5–17	—
8-year-olds	50	25 G, 25 B	102.92 (3.34)	45.04 (11.55)	12.16 (2.92) 4–17	10.84 (2.74) 4–16	—
9-year-olds	50	25 G, 25 B	113.88 (3.36)	48.79 (12.04)	11.84 (2.80) 6–17	11.14 (2.60) 5–16	—
Language impairment							
4-year-olds	12	3 G, 9 B	55.92 (2.76)	47.17 (10.80)	4.33 (2.6) 3–11	5.42 (1.17) 4–7	76.83 (8.62) 68–99
5-year-olds	14	6 G, 8 B	64.92 (3.12)	46.52 (12.00)	5.00 (2.88) 3–11	4.43 (1.28) 3–7	76.21 (11.35) 58–96
6-year-olds	11	5 G, 6 B	79.55 (3.17)	40.26 (13.97)	5.73 (1.79) 4–9	5.27 (2.20) 3–10	78.55 (8.18) 63–62
7-year-olds	13	3 G, 10 B	90.72 (2.76)	42.42 (13.30)	6.38 (2.36) 3–11	4.31 (1.49) 3–7	74.00 (11.05) 51–90
8-year-olds	17	7 G, 10 B	104.35 (3.12)	42.42 (7.40)	7.47 (2.37) 4–13	5.00 (1.80) 3–9	76.29 (11.94) 55–95
9-year-olds	10	5 G, 5 B	114.00 (2.52)	48.71 (9.66)	8.10 (2.55) 4–13	5.40 (1.96) 3–8	73.50 (11.27) 55–85

Note. CELF-P = Clinical Evaluation of Language Fundamentals–Preschool; CELF-3 = Clinical Evaluation of Language Fundamentals–Third Edition.

[a]SES = socioeconomic status as measured by the Blishen Scales (Blishen et al., 1987). [b]The standard score (M = 10, SD =3) was reported for the Linguistic Concepts subtest in the CELF-P (ages 4 and 5 years) and the Concepts and Directions subtest in the CELF-3 (ages 6–9 years). [c]The standard score (M = 10, SD = 3) was reported for the Recalling Sentences in Context subtest in the CELF-P (ages 4 and 5 years) and the Recalling Sentences subtest in the CELF-3 (ages 6–9 years). [d]The standard scores of Total Language Composites (M = 100, SD = 15) were reported for the LI group using the manuals of the CELF-P or the CELF-3. No Total Language Composites were reported for children with TL because they were administered only two subtests. The em dashes indicate that Total Language Composites were not available for children with TL. [e]G = girl; B = boy.

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

As part of the study protocol, two subtests of the CELF-P (Wiig et al., 1992) or the CELF-3 (Semel et al., 1995) were administered to children referred as TL, depending on the child's age. Children younger than age 6;0 were tested using the Linguistic Concepts and Recalling Sentences in Context subtests from the CELF-P. Children aged 6;0 and older were tested using the Concepts and Directions and Recalling Sentences subtests from the CELF-3. The purpose of administering these subtests was to provide a description of language skills for children with TL. Children's performance on these subtests, however, did not affect inclusion or exclusion from analysis (Schneider & Haywood, 2010).

Across all ages, 19 children with TL (1–6 per age group) had standard scores below 7 (i.e., −1 *SD*) on the Linguistic Concepts (CELF-P) or Concepts and Directions (CELF-3) subtest, 11 children (1–3 per age group) had scores below 7 on Recalling Sentences in Context (CELF-P) or Recalling Sentences (CELF-3), and six children (0–2 per age group) had scores lower than 7 on both subtests. These children were still included in the TL group for several reasons. First, the teachers did not have any concerns for these children on their speech/language skills (Bishop, Snowling, Thompson, Greenhalgh, & CATALISE Consortium, 2016; Paul, Norbury, & Gosse, 2018). Second, it is possible for a child with TL skills to score below −1 *SD* on individual subtests (Semel et al., 1995; Wiig et al., 1992). Third, only the two subtest scores were available for most of the children in the TL group (see Schneider, Hayward, & Dubé, 2006, for an explanation), which would not be adequate for identifying LI without other supporting information. This was because the CELF-P/CELF-3 manuals emphasized that, while the Composite Standard Scores (i.e., Receptive Composite, Expressive Composite, and Total Language Composite) could be used for diagnosing the presence/absence of LI, the standard scores for individual subtests cannot be used for this purpose. Finally, eliminating the children from the TD group who had the lowest CELF scores would potentially bias the sample in the direction of greater differences between the groups on the ENNI. Table 1 presents the standard scores (M = 10, SD = 3) of the two subtests in the CELF-P or CELF-3 for children with TL in the normative sample of the ENNI by age level.

The subsample of children with LI was recruited from three sites: a public school serving children with communication disorders, a rehabilitation hospital that had several programs for children with LI, and Capital Health Authority, which served preschool and school-aged children throughout the city of Edmonton. Each site was asked to refer children with a rating of 2–5 on a severity rating scale designed by Capital Health that could rate a child's LI from 1 (*mild*) to 5 (*severe*). Children could be referred even if they had a diagnosed learning disability, mild-to-moderate speech sound disorder, fine or gross motor delay, or attention-deficit disorder (ADD) or ADHD with medication. It should be noted that the participating sites were also asked not to refer children who had a diagnosis of autism, intellectual disability, hearing impairment, severe speech sound disorder, ADD/ADHD without medication, or severe visual impairment that would result in inability to see pictures even with correction. However, the information regarding whether children with LI in the normative sample also had concomitant speech disorders, motor delay, or ADD/ADHD (with medication) was not available at the time of data collection because access to children's clinical records was not obtained. Information regarding nonverbal/performance IQ was not collected.

To further confirm the language status of children who were referred as having LI, the full CELF-P or CELF-3 was administered, depending on the child's age (i.e., younger than 6;0 or not). All of the children in the LI group scored below −1 *SD* (i.e., a standard score of 85) on at least one of the composite scores (i.e., Receptive, Expressive, and Total Language Composites) of the CELF-P or CELF-3. The cutoff standard score of 85 was based on the recommendation of the CELF manuals and was consistent with the cutoff used in previous studies of children with LI (e.g., Munson, Kurtz, & Windsor, 2005). Across ages, the percentage of children with LI who scored below the cutoff was 66% (51/77) for the Receptive Composite, 95% (73/77) for the Expressive Composite, and 84% (65/77) for the Total Language Composite. Thus, all children with LI in this study were receiving language intervention at the time of data collection and scored below −1 *SD* on at least one of the composite scores on the CELF-P or CELF-3. These inclusionary criteria were consistent with prior studies (e.g., Dollaghan & Campbell, 1998; Moyle et al., 2011). Table 1 presents the Total Language Composite (M = 100, SD = 15) for children with LI in the normative sample by age.

Demographic information, including ethnicity and socioeconomic status (SES), was also collected. Ethnic composition of children in the normative sample corresponded closely to the range of ethnic diversity in the city of Edmonton according to Statistics Canada data (Statistics Canada, n.d.): Approximately 72% of the participants were of European origin, and 28% were of non-European origin. The SES of the children was estimated from parents' occupations using the Blishen Scales (Blishen, Carroll, & Moore, 1987). Based on Canadian census information, this index reflects equally weighted components of education and income level by occupation. For instance, route drivers are assigned a score of 35.73, and electrical engineers are assigned a score of 71.70 on the scale. Table 1 also presents the mean SES score by language status and age. The SES scores were not significantly different between the TL and LI groups at any of the age levels (ps ≥ .12, ds ≤ 0.42) or between the age levels regardless of children's language status (p = .39, d = 0.25).

Materials

The ENNI (Schneider et al., 2005) used a story generation task to elicit narratives from children. The task involved children telling stories about six original picture sequences with animal characters. The picture sequences were all black and white drawings drawn by a professional cartoonist based on the scripts created by the ENNI authors. The

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

picture sequences depicted stories that varied in three levels of complexity (two picture sequences for each level). To reflect the complexity of the stories, the picture sequences systematically varied in length (i.e., five, eight, and 13 pictures), number and gender of characters (i.e., two, three, and four characters), and amount of story information. The six picture sequences were equally divided into two sets (i.e., Set A and Set B) such that each set had one picture sequence from each complexity level (i.e., three picture sequences per set). For each story set, Story 1 had one episode (i.e., initiating event, attempt, consequence; Hughes, McGillivray, & Schmidek, 1997), Story 2 had two episodes, and Story 3 had three episodes. Thus, both story sets had the same total number of episodes (i.e., six episodes across stories). These picture sets may be viewed and downloaded from the ENNI website. To demonstrate the variation of complexity levels for the stories, a description of each story is provided in Supplemental Material S1.

To be particularly cautious, we examined whether the story sets had any effects on the target measure—FVMC. We found that FVMC for Set A did not differ significantly from FVMC for Set B either in the TL group, $F(1, 299) = 0.121$, $p = .728$, or in the LI group, $F(1, 76) = 0.001$, $p = .984$. Thus, we did not further consider the effect of story sets (Sets A or B) in this report.

Procedure

Each participant was seen individually by a trained examiner in the child's preschool/day care or school. Three examiners with a bachelor's degree in education or psychology were employed to evaluate the child and administer the story generation task. The task began by instructing the child that he could see all the pages first and then tell a story to the examiner. The instructions also stressed that the examiner would not be able to see the pictures so the child would have to tell a really good story in order for the examiner to understand it.

The pictures for each story were placed in page protectors in a binder. Each story was in its own binder. The examiner was required to hold the binder in such a way that she could not see the pictures as the child told the story. This meant that the child needed to be explicit in order for the examiner to understand the story. For example, the child could not legitimately use a pointing gesture to replace language when referring to a specific character in the picture. When a given story was administered, the child first went through all of the pictures to preview the story and then started to tell the story. The examiner turned the pages after the child appeared to be finished telling the story for a particular picture.

The child was first given a training story consisting of a single episode story in five pictures. Like those for Story 1 in each story set, the training story had five pictures and described one episode that contained two characters. The purpose for administering the training story was to familiarize the child with the procedure and to allow the examiner to give explicit prompts (e.g., *Once upon a time, there was a* ~) if the child had difficulties with the task. The training story, however, was not included for analysis.

After the training story, the two story sets were given. Administration of the story sets was counterbalanced across children. For Story Sets A and B, the examiner was restricted to less explicit prompts than in the training story, such as general encouragement, repetition of the child's previous utterances, and a request to tell what was happening in the story (e.g., *Tell me more about the story* or *Then what happens in the story*). Stories were audio-recorded for transcription and analysis.

Data Transcription, Coding, and Computation

The narrative samples were transcribed orthographically and coded by trained research assistants based on the conventions of SALT (Miller & Chapman, 2000). Children's narratives were segmented into communication units (C-units). A C-unit is typically an independent clause plus all of its dependent clauses (Loban, 1976). Nonclausal utterances that express complete thoughts (e.g., *A dog, a rabbit, and a sandbox*) are also counted as C-units. Only intelligible, complete, and spontaneous C-units that described the stories were included for computing the descriptive measures (e.g., mean length of C-units).

FVMC computes the percent accuracy of 3SG *–s*, regular past tense *–ed,* and contracted and uncontracted copula and auxiliary *be* forms (e.g., *am, are, is, was, were*) in obligatory contexts. An obligatory context is operationally defined as an instance in which a particular tense marker is required for the C-unit to be grammatical. For example, the C-unit "The rabbit walking on the street" has an obligatory context for auxiliary *be* but the child omits this morpheme. C-units that contained verb forms but no subjects (e.g., *Getting the airplane out of the swimming pool*) were not coded for FVMC because they did not provide obligatory contexts for tense usage. We also excluded overgeneralization of 3SG *–s* (e.g., *The elephant have**s** an airplane*) or regular past tense *–ed* (e.g., *The elephant just stand**ed** there*) from the FVMC analysis because verbs in these contexts, by definition, did not require the usage of 3SG *–s* or past tense *–ed.* It should be noted that FVMC does not include the infinitive form of *be* (e.g., *The rabbit will be sick*), present participle form of *be* (e.g., *The rabbit is being funny*), past participle form of *be* (e.g., *He has been trying to get the ball*), or gerund form of *be* (e.g., *Being* happy is easy) in the analysis. This is because these *be* forms do not mark tense. Thus, determining whether a C-unit had an obligatory context for the target morpheme of FVMC depended on the C-unit that the child produced. The number of obligatory contexts for the FVMC analysis therefore varied across children.

For each obligatory context, the target tense morpheme was coded as (a) correctly used, (b) omitted, or (c) incorrectly used (e.g., *The boy and the girl is walking*). FVMC was computed by dividing the total number of correct uses of the four target tense morphemes by the total number of obligatory contexts for these morphemes in the narrative (see Supplemental Material S2 for an example).

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

The resultant quotient was then multiplied by 100% to obtain a percentage.

Reliability of Transcription and Coding

To ensure the transcription reliability, the narratives were first transcribed by graduate research assistants majoring in speech-language pathology at the University of Alberta. The assistants were trained by the third author by practicing transcribing four stories to at least 95% accuracy before they were allowed to transcribe narratives for the ENNI. The transcripts were then checked against the recordings by the third author of this study before transcription reliability was assessed. Another graduate research assistant majoring in speech-language pathology, who did not transcribe the stories, independently transcribed one story from 24% of the participants for checking transcription reliability. The C-unit segmentation consistency was 96%, and the word-by-word consistency for transcription was 97%.

To verify the coding reliability for FVMC, we adapted a consensus procedure from Shriberg, Kwiatkowski, and Hoffman (1984). Four other graduate assistants majoring in speech-language pathology at the University at Buffalo coded the tense morphemes for FVMC for all children. They were trained by the first author and practiced transcribing a 30-min conversational language samples and coding grammatical morphemes in the sample, including tense morphemes, with at least 90% accuracy before they could start to transcribe and code language samples (e.g., conversations, narratives) in the first author's lab. At the time they coded FVMC for the current study, they all had at least 1 year of experience in transcribing language samples and coding grammatical morphemes. After the four research assistants coded FVMC for the ENNI archival data, the first author checked the coded transcripts for all children (i.e., 100% checked) without independently coding the transcripts. Discrepancies were discussed between the first and third/fourth authors. Across all children, 391 (1.41%) out of 27,715 C-units were discussed. All of the discrepancies were resolved.

Statistical Analysis

We used two-way analysis of variance (ANOVA) to evaluate the effects of language status and age on FVMC. We used the *d* value to quantify the effect size or magnitude of the differences in FVMC. Following Cohen (1988), we interpreted the effect size as small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d < 0.8$), or large ($d \geq 0.8$) whenever appropriate. Because FVMC was calculated as a percentage, it was arcsine-transformed in the ANOVAs.

To compute the 90% and 95% confidence intervals for FVMC for children with TL at each age level, we followed the steps specified in McCauley and Swisher (1984). We first obtained the split-half reliability of FVMC by computing the correlation of FVMC in Story Sets A and B for each age level, which was consistent with the current clinical practice (Dawson et al., 2005; Semel, Wiig, & Secord, 2003). The correlation coefficients (*r*) were then used to calculate the standard error of measurement (*SEM*) in the equation (i.e., $SEM = SD\sqrt{1-r}$, where *SD* is the standard deviation of FVMC for a given age level). A 90% confidence interval is computed by multiplying the *SEM* value by −1.645 (lower limit) or +1.645 (upper limit). A 95% confidence interval is computed in a similar way, except that the *SEM* is multiplied by −1.96 (lower limit) or +1.96 (upper limit).

To evaluate the split-half reliability for FVMC, we computed the correlation between FVMC scores from Story Set A and those from Story Set B for all of the children at each age level. That is, children with TL and those with LI were combined together for this analysis so that we had sufficient variability of FVMC scores among children at each age level (Kleinbaum, Kupper, Muller, & Nizam, 1998). Note that this is different from how split-half reliability was determined for calculating confidence intervals. Children with LI were not included in the computation of split-half reliability for confidence intervals because reference data were computed based only on children with TL. To determine the concurrent criterion validity of FVMC, we computed the correlation of overall FVMC scores and the raw scores of the Recalling Sentences in Context subtest of the CELF-P (4- and 5-year-olds) or the Recalling Sentences subtest of the CELF-3 (6- to 9-year-olds). Note that, although Table 1 presents the standard scores of these subtests for the ease of interpretation, raw scores, rather than standard scores, of the subtests were used for the correlation analysis. We chose these subtests as the criterion measures because sentence recall/repetition has been considered a task that taps expressive grammar and an effective measure for identifying children with LI (Pawłowska, 2014; Semel et al., 1995; Wiig et al., 1992). Although FVMC is a narrow measure of expressive grammar (Souto et al., 2014), children's performance on FVMC should be in concordance with their performance on sentence recall/repetition. Following Cohen (1988), we interpreted the correlation coefficients as small ($.1 < r < .3$), medium ($.3 < r < .5$), or large ($r > .5$) whenever appropriate.

To evaluate the diagnostic accuracy of FVMC, we computed the sensitivity, specificity, and likelihood ratios (Dollaghan, 2007). Likelihood ratios, in general, are less affected by sample size than sensitivity and specificity. Sensitivity refers to the extent to which a measure can accurately identify children with LI. It was computed as the percentage of children with LI who were also identified as LI by FVMC. In contrast, specificity refers to the extent to which a measure can accurately identify children with TL. It was computed as the percentage of children with TL who were also identified as TL by FVMC. Based on Plante and Vance (1994), sensitivity and specificity levels between 80% and 89% were considered acceptable, and sensitivity and specificity levels at or greater than 90% were considered good/preferred.

Likelihood ratios were computed from the levels of sensitivity and specificity (Dollaghan, 2007). The positive likelihood ratio (LR+) was calculated as the ratio of true

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

LI to false LI (i.e., sensitivity/[1−specificity]). A higher LR+ value for a positive test result refers to a higher likelihood that the positive result comes from a child with LI than from a child with TL. In contrast, the negative likelihood ratio (LR−) was calculated as the ratio of false TL to true TL (i.e., [1−sensitivity]/specificity). A lower LR− value for a negative result refers to a lower likelihood that the negative result comes from a child with LI than from a child with TL. According to Dollaghan (2007) and Geyman, Deyo, and Ramsey (2000), an LR+ value of ≥ 10.00 or an LR− value of ≤ 0.10 was considered as good/preferred, and an LR+ value between 5.00 and 9.99 or an LR− value between 0.11 and 0.20 was considered acceptable in this study.

To compute sensitivity, specificity, and likelihood ratios, empirically derived cutoff FVMC scores for a positive result were first determined by using the receiver operating characteristic (ROC) curve (Sackett, 1991) in the software SigmaPlot 12.0 (Systat Software, Inc., 2011). The ROC curve analysis plotted the true positive rate (i.e., sensitivity) against the false positive rate (i.e., 1−specificity) for the different possible cutoff FVMC scores (see Supplemental Material S3 for an example of the ROC curve and cutoff scores for all age groups). Thus, it automatically generated pairs of sensitivity and specificity levels for a range of cutoff FVMC scores. Following Sackett (1991), we chose the score that maximized the diagnostic accuracy, where sensitivity plus specificity divided by 2 was largest, as the cutoff.

Results

Descriptive Analyses of Narrative Samples

Table 2 presents the descriptive measures from the narratives and the number of obligatory contexts for the FVMC. Regardless of age, children with TL did not differ significantly from those with LI in number of C-units, $F(1, 375) = 0.58$, $p = .45$, $d = 0.09$. However, children with TL produced significantly longer mean lengths of C-units, more different words, and more obligatory contexts for FVMC than those with LI (Fs ≥ 15.77, ps < .001, ds ≥ 0.41). It should be noted that, although there was a significant group difference, only two children with LI produced fewer than 10 obligatory contexts for FVMC, one at age 4 and one at age 5 (see Table 2). Thus, we believe that children with LI did produce sufficient obligatory contexts in this study for the computation of FVMC. Table 2 also marks the significant differences between the TL and LI groups in these descriptive measures by age; Supplemental Material S4 further provides statistical comparisons (e.g., F values, p values) regarding group differences in total number of C-units, mean length of C-units in morphemes, number

Table 2. Mean (*SD*) and range of descriptive measures by language status and age.

Language status and age	No. of C-units	MLCUm	NDW	No. of OC for FVMC
Typical language				
4-year-olds	68.06 (19.49) 39–151	6.97 (1.08)** 4.44–8.96	130.32 (35.87)** 37–252	38.04 (14.09)** 6–89
5-year-olds	72.04 (21.23) 48–136	7.73 (0.97)** 5.51–10.22	141.40 (31.17)** 77–219	40.70 (15.74)* 5–109
6-year-olds	71.64 (19.25) 50–129	7.59 (0.97) 5.10–10.00	142.90 (29.74)* 94–225	45.64 (19.05) 19–106
7-year-olds	73.88 (18.46) 45–136	8.49 (1.15)** 5.96–10.96	154.84 (28.78) 90–228	53.24 (21.32) 19–108
8-year-olds	78.70 (21.71) 49–146	8.70 (1.07)** 6.73–11.08	172.94 (42.07)** 113–279	57.90 (22.07)* 20–117
9-year-olds	77.78 (24.08) 51–160	9.00 (1.07)** 6.80–11.18	172.06 (36.55) 112–315	62.80 (27.44) 23–166
Language impairment				
4-year-olds	58.00 (17.93) 33–106	5.18 (1.37) 3.12–6.89	84.25 (26.67) 43–140	25.25 (14.30) 6–48
5-year-olds	70.64 (22.29) 40–129	5.94 (1.35) 2.70–8.05	113.36 (30.94) 47–168	30.86 (13.12) 5–61
6-year-olds	73.63 (24.46) 52–129	7.04 (1.05) 5.50–8.50	123.45 (26.55) 89–179	41.36 (12.01) 27–60
7-year-olds	80.31 (44.07) 45–181	6.90 (1.30) 4.74–9.12	135.62 (51.35) 84–267	40.77 (33.14) 17–143
8-year-olds	73.94 (21.41) 46–124	7.20 (0.81) 5.83–8.62	141.47 (30.95) 99–202	43.18 (20.07) 21–82
9-year-olds	81.40 (38.48) 51–174	7.91 (0.91) 6.67–9.32	164.20 (45.26) 120–262	51.70 (16.67) 29–80

Note. No. of C-units = total number of C-units in the narratives; MLCUm = mean length of C-units in morphemes; NDW = number of different words; No. of OC for FVMC = number of obligatory contexts for finite verb morphology composite.

*The difference between children with and without language impairment at a given age level was significant at .05 level. **The difference was significant at the .01 level.

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

of different words, and number of obligatory contexts for FVMC by age.

Age Differences in FVMC in Children With and Without LI

Table 3 presents children's performance on FVMC by language status and age. A 2 (language status: TL vs. LI) × 6 (age level: 4–9 years) ANOVA showed that there was a main effect of language status on FVMC scores, $F(1, 365) = 243.12$, $p < .001$, $d = 1.63$. There was also a main effect of age level on FVMC scores, $F(5, 365) = 16.15$, $p < .001$, $d = 0.94$. The main effects were further qualified by a significant interaction effect, $F(5, 365) = 6.15$, $p < .001$, $d = 0.58$.

To answer our research questions, we performed the follow-up analyses in two ways. First, we evaluated the age differences in FVMC scores within the TL and LI groups separately. One-way ANOVAs showed that there was a main effect of age levels for both the TL group, $F(5, 294) = 4.48$, $p = .001$, $d = 0.55$, and the LI group, $F(5, 71) = 5.07$, $p < .001$, $d = 1.20$. Post hoc Tukey tests showed that, for children with TL, 4-year-olds produced significantly lower FVMC scores than 7-, 8-, and 9-year-olds. The difference in the mean FVMC scores between 4- and 9-year-old children who had TL was approximately 4.24%. There were no other significant age differences within the TL group. For children with LI, 4-year-olds produced significantly lower FVMC scores than 8- and 9-year-olds. Five-year-olds also produced significantly lower FVMC scores than 9-year-olds. The difference in the mean FVMC scores between the 4- and 9-year-old children who had LI was approximately 36.30%. There were no other significant age differences within the LI group.

Second, we evaluated the differences in FVMC between children with and without LI by age level. One-way ANOVAs showed that children with TL produced higher FVMC than those with LI at each age level (Fs ≥ 18.14, ps < .001), and the effect sizes were all large (ds ≥ 1.12; see Table 3). However, the effect size was larger in younger than in older children, meaning that the differences in FVMC scores between the TL and LI groups decreased with age. To ease the interpretation of this trend, we further plotted children's performance on FVMC by age and language status in Figure 1. Table 3 also presents the 90% and 95% confidence intervals for FVMC in children with TL by age. In general, the 90% and 95% confidence intervals for FVMC decreased with age.

Psychometric Properties of FVMC

Table 4 presents the correlation coefficients for the split-half reliability and concurrent criterion validity by age. The correlation coefficient for the split-half reliability of FVMC was large between age 4 and age 8 (rs > .653, ps < .001) and was medium for age 9 ($r = .429$, $p < .001$), meaning that FVMC scores computed from Set A were consistent with those computed from Set B between age 4 and age 9. Similarly, the correlation coefficient for the concurrent criterion validity of FVMC was medium or large between age 4 and age 9 (rs > .437, ps < .001), meaning that children's performance on FVMC was in concordance with their performance (i.e., raw scores) on the Recalling Sentences in Context subtest of the CELF-P (ages 4 and 5) or the Recalling Sentences subtest of the CELF-3 (ages 6–9).

Table 5 presents the indices of diagnostic accuracy for FVMC. It should be noted that, given the small sample size of children with LI per age group, the results here are considered preliminary and need to be interpreted with caution. Both sensitivity and specificity were good at age 4 and age 5 (range: 90%–100%). The sensitivity and

Table 3. Descriptive statistics of finite verb morphology composite (in percentage) by language status and age and the effect size for the between-groups difference in finite verb morphology composite by age.

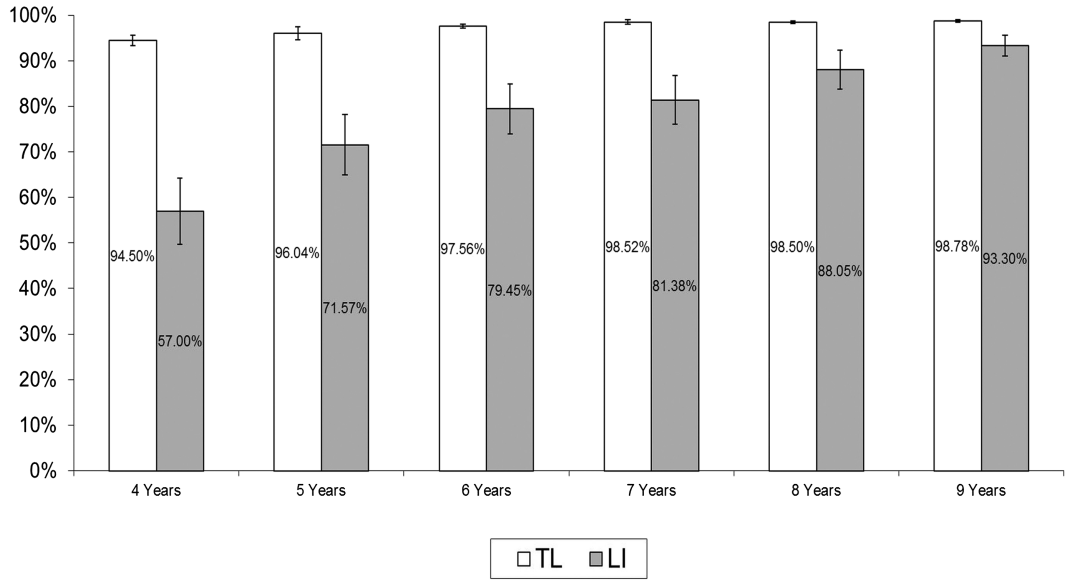
Language status and age	*M*	*SD*	Range	*SEM*[a]	90% CI[b]	95% CI	Effect size (*d*)[c]
Typical language							
4-year-olds	94.50	8.23	50.00–100	4.85	± 7.95	± 9.50	2.01
5-year-olds	97.15	5.38	72.41–100	3.49	± 5.72	± 6.83	1.83
6-year-olds	97.55	3.10	88.00–100	3.03	± 4.97	± 5.94	1.82
7-year-olds	98.48	3.60	76.60–100	1.93	± 3.17	± 3.78	1.74
8-year-olds	98.45	2.05	90.91–100	1.99	± 3.26	± 3.90	1.25
9-year-olds	98.74	2.09	87.06–100	1.92	± 3.15	± 3.77	1.12
Language impairment							
4-year-olds	56.98	25.17	22.22–100	—	—	—	—
5-year-olds	71.50	24.83	0–93.33	—	—	—	—
6-year-olds	79.58	18.29	45.16–97.92	—	—	—	—
7-year-olds	81.29	19.26	47.06–100	—	—	—	—
8-year-olds	88.09	17.78	26.09–100	—	—	—	—
9-year-olds	93.26	7.18	75.86–100	—	—	—	—

Em dashes indicate data not applicable. This is because standard errors of measurement and confidence intervals were computed only for children with typical language.

[a]*SEM* = standard error of measurement. [b]CI = confidence interval. [c]Effect size (*d*) = effect size for the difference in finite verb morphology composite between children with and without language impairment by age.

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Figure 1. The finite verb morphology composite by age and language status. Standard errors are represented in the figure by the error bars attached to each column. TL = typical language; LI = language impairment.



specificity levels were acceptable or good at age 6 and age 7 (range: 82%–90%). However, the sensitivity and specificity levels were poor or acceptable at age 8 and age 9 (range: 76%–80%). For example, at age 8, FVMC demonstrated a sensitivity level of 76%, meaning that 24% of 8-year-old children with LI in this study were not identified as LI by FVMC. Similarly, at age 9, FVMC demonstrated a specificity level of 76%, meaning that 24% of 9-year-old children with TL in this study were not identified as TL by FVMC.

The likelihood ratios showed a similar trend about the change of diagnostic accuracy across ages for FVMC. With the empirically determined cutoff values, the LR+ and LR− values were both at the good level at age 4 and age 5. For example, the LR+ value was 15.28 at age 4, meaning that 4-year-olds with LI were 15.28 times more likely to obtain a fail score (i.e., below the cutoff) for FVMC than those with TL. The LR+ and LR− values were both at the acceptable level at age 6 and age 7. However, the LR+ and LR− values were at the poor level at age 8 and age 9.

Table 4. Split-half reliability and concurrent criterion validity in correlation coefficients for finite verb morphology composite by age.

Age	Split-half reliability	Concurrent criterion validity[a]
4-Year-olds	0.821[b]	0.693
5-Year-olds	0.653	0.564
6-Year-olds	0.842	0.437
7-Year-olds	0.826	0.579
8-Year-olds	0.781	0.470
9-Year-olds	0.429	0.481

[a]Concurrent criterion validity of finite verb morphology composite was established by computing the correlation between finite verb morphology composite and the raw score from the Recalling Sentences in Context subtest of the Clinical Evaluation of Language Fundamentals–Preschool (ages 4 and 5 years) or the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals–Third Edition (ages 6–9 years). [b]All correlation coefficients were significant at the .001 level (one-tailed).

Discussion

This study provided reference data and evaluated psychometric properties for FVMC in children between 4 and 9 years of age from the ENNI database. Using one single language sampling protocol consistently across the age levels, we found that FVMC revealed age-related changes in the TL and LI groups. FVMC also demonstrated appropriate psychometric properties, except for its diagnostic accuracy. We explore these findings below.

FVMC Increased Between Age 4 and Age 9 Years in Children With and Without LI

In this study, children's FVMC score increased significantly between age 4 and age 9 in the TL and LI groups. Although significant, the magnitude of change in the TL group was much smaller (4.24%) than that in the LI group (36.30%), possibly because children with TL had reached the customary level of mastery (i.e., 90% accurate) for FVMC at age 4 in this study. Consequently, the gap in FVMC scores between children with and without LI became smaller between age 4 and age 9, which was further evidenced in the effect sizes. These findings were consistent

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Table 5. Indices of diagnostic accuracy for finite verb morphology composite by age using the cutoff score from the receiver operating characteristic curve analysis.

Age level	Cutoff	Sensitivity[a]	Specificity	Overall accuracy	LR+[b]	LR–
4-Year-olds	83.77%	92%** (11/12)	94%** (47/50)	94% (58/62)	15.28**	0.09**
5-Year-olds	93.46%	100%** (14/14)	90%** (45/50)	92% (59/64)	10.00**	< 0.01**
6-Year-olds	93.50%	82%* (9/11)	90%** (45/50)	89% (54/61)	8.18*	0.20*
7-Year-olds	96.64%	85%* (11/13)	86%* (43/50)	86% (54/63)	6.14*	0.18*
8-Year-olds	97.50%	76% (13/17)	80%* (40/50)	79% (53/67)	3.82	0.29
9-Year-olds	97.85%	80%* (8/10)	76% (38/50)	77% (46/60)	3.33	0.26

[a]For the columns of sensitivity/specificity, a single asterisk indicates that sensitivity/specificity of a given measure reaches the acceptable level of accuracy, that is, 80% accuracy (Plante & Vance, 1994). Double asterisks indicate that sensitivity/specificity of a given measure reaches a good or preferred level of accuracy, that is, 90% accuracy. The numbers within the parentheses indicate the number of children who are correctly classified; for example, nine of eleven 6-year-olds with language impairment were correctly classified by the finite verb morphology composite. [b]LR+ = positive likelihood ratio; LR– = negative likelihood ratio. For the columns of LR+/LR–, a single asterisk indicates that the LR+/LR– of a given measure reaches the acceptable level, that is, an LR+ value between 5.00 and 9.99 or an LR– value between 0.11 and 0.20 (Dollaghan, 2007; Geyman et al., 2000). Double asterisks indicate that the LR+/LR– of a given measure reaches the good level, that is, an LR+ value at or above 10.00 or an LR– value at or below 0.10.

with our predictions and prior studies (Rice et al., 1998; Souto et al., 2014).

Although this study was a cross-sectional study and the age-related changes could not be directly used to infer the development of FVMC over time, the data from children with TL in this study can be compared to those in Rice et al. (1998), a longitudinal study. Note that Rice et al. used a different measure of tense usage and combined conversational language samples and experimental probes to evaluate children's performance on tense usage. Despite the discrepancy in methods between this study and Rice et al., children with TL in both studies showed comparable levels of accuracy in using tense morphemes. For example, by age 4, the mean FVMC score was 94.50% in this study, and the mean tense composite score was 94.90% in Rice et al. Similar findings were observed in Goffman and Leonard (2000), as well. This study also showed that the FVMC scores were not significantly different between age 5 and age 9 in children with TL, indicating that those children's performance on FVMC was stable at this age range. Together, the present and prior studies (Rice et al., 1998; Souto et al., 2014) suggest that children with TL, as a group, reach the customary level of mastery for tense usage around age 4 and that their performance on FVMC is stable after age 5.

Also similar to the children with LI in Rice et al. (1998), children with LI in this study, as a group, performed below the customary level of mastery for FVMC at age 8. The mean FVMC score for 8-year-olds with LI in this study was 88.05%, and the mean tense composite score for those in Rice et al. was 89.20%. This study expanded on the previous studies by showing that children with LI, as a group, produced FVMC at the customary level of mastery (93.26%) by age 9. This shows that tense usage in children with LI does not plateau at age 8 years but rather continues to increase. However, we are not claiming that children with LI tended to grow out of the tense morpheme deficits by age 9. After all, children with LI still produced significantly lower FVMC scores than those with TL (i.e., below age expectation) at age 9, which further supports the hypothesis that tense morpheme deficits are a phenotypic marker for English-speaking children with LI (Rice et al., 1998). We do not have evidence indicating whether or when children with LI will close the gap with those with TL on FVMC computed from a narrative generation task. Even if children with LI do close the gap at older ages, this might only mean that FVMC computed from this type of language sampling protocol is not sufficiently sensitive for documenting continued tense morpheme deficits in individuals with LI at older ages (Poll, Betz, & Miller, 2010). That is, although tense morpheme deficits could be a phenotypic marker for English-speaking children with LI, we should not expect that a single measure for tense morphemes (e.g., FVMC) would be able to identify children with LI at all ages. Different tasks tapping knowledge of tense morphemes would be needed in order to identify older children with LI (e.g., Poll et al., 2010).

FVMC Exhibited Appropriate Psychometric Properties Between Age 4 and Age 9 Years, Except for Its Diagnostic Accuracy

In this study, we evaluated three psychometric properties for FVMC computed from the ENNI database. First, we found that FVMC showed appropriate split-half reliability (i.e., internal consistency). This was evidenced in the result that FVMC scores from Story Set A were significantly correlated with those from Story Set B between age 4 and age 9 with medium or large correlation coefficients. Second, we found that FVMC showed appropriate concurrent criterion validity. This was supported by the result that FVMC scores were significantly correlated with raw scores for the Sentence Recall subtest of the CELF-P/CELF-3, a measure of expressive grammar, between age 4 and age 9 with medium or large correlation coefficients. The finding further suggests that FVMC and Sentence Recall tap related but distinct aspects of expressive grammar and are not redundant to each other. Third, we found that FVMC yielded good diagnostic accuracy for 4- and 5-year-olds and

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

acceptable diagnostic accuracy for 6- and 7-year-olds. The current findings were comparable to those in Souto et al. (2014), which found that FVMC showed good diagnostic accuracy for both 4- and 5-year-olds. Despite the discrepancies in methodologies, our findings were also consistent with those in Rice and Wexler (2001), which found that the TEGI showed acceptable to good diagnostic accuracy from 3;6 to 6;11. In concert with this study, Rice and Wexler also found that the TEGI showed acceptable sensitivity for age 7, although specificity was not reported. Thus, this study extended the previous study by demonstrating that FVMC, a tense measure, had acceptable specificity for age 7 when sensitivity was considered at the same time. On the other hand, we found that the diagnostic accuracy of FVMC was unacceptably low for 8- and 9-year-olds. For example, the sensitivity of FVMC was below the acceptable level for age 8, a finding at odds with that in Rice and Wexler (2001). The gradient downward change of diagnostic accuracy for FVMC between 4 and 9 years of age possibly resulted from the decreasing group differences and hence increasing overlap in FVMC scores between children with and without LI over time. Together, the results from the age-related changes and the psychometric properties collectively suggest that FVMC computed from the ENNI protocol is reliable and valid between age 4 and age 9 and could be used for identifying children with LI between age 4 and age 7.

Given the poor diagnostic accuracy for FVMC at age 8 and age 9 in this study, one could argue that the ENNI protocol may not be appropriate for diagnostic purposes at age 8 or age 9. It is possible that the ENNI protocol is not challenging enough for children at age 8 or age 9. Thus, clinicians may need to use a different narrative protocol with attested diagnostic accuracy for identifying children with LI at age 8 or older (e.g., Test of Narrative Language–Second Edition; Gillam & Pearson, 2017). However, one could also argue that the ENNI protocol may still be used for diagnostic purposes at age 8 and age 9 years but that a different measure should be computed because this measure might be more sensitive to LI than FVMC at age 8 or age 9. In fact, in our other study (Guo, Eisenberg, Schneider, & Spencer, 2019), we examined the psychometric properties of another measure computed from the ENNI database—percent grammatical utterances (PGUs). We found that PGU demonstrated acceptable to good diagnostic accuracy from age 4 to age 9. Thus, we believe that the ENNI protocol can still be used for diagnostic purposes at age 8 and age 9 if a different measure such as PGU is used.

Limitations

By taking advantage of the existing normative sample from the ENNI, we face limitations that must be considered. First, we had a small number of children with LI at each age level because the ENNI authors attempted to avoid overrepresenting children with LI in the normative sample (see Schneider et al., 2006, for further discussion). One problem of having a small number of children with LI is that the differences in sensitivity between age levels could have been overinterpreted, although we sought to address this issue by including likelihood ratios as additional indices for evaluating the diagnostic accuracy of FVMC because these were less affected by sample size. For example, we had 10 children with LI in the 9-year-old group. Misclassification of one child with LI could lead to a 10% difference in the sensitivity level. In addition, although FVMC misclassified two children with LI at both age 6 and age 9, the sensitivity was 85% at age 6 and 80% at age 9 simply because we had more children with LI at age 6. A related issue is that this study was a two-gate, rather than one-gate, design because we preselected children with TL and those with LI for evaluating the diagnostic accuracy of FVMC instead of testing a large, unselected group of children (Dollaghan & Horner, 2011). One potential problem of a two-gate design is that diagnostic accuracy may have been inflated relative to a one-gate design. Because of the nature of the design and the small number of children with LI, the present investigation could only be viewed as an early-phase study for the diagnostic accuracy of FVMC (Dollaghan & Horner, 2011). Therefore, the diagnostic accuracy for FVMC in this study was preliminary. Future one-gate studies that include a large number of participants are needed in order to verify the current findings.

Second, information about children's nonverbal intelligence was not available in this study because it was not collected for the normative sample of ENNI. However, our findings were comparable to those in the studies that screened children for their nonverbal intelligence (e.g., Rice et al., 1998). Similarly, we were not able to obtain the information about whether children with LI also had concomitant speech disorders, motor delays, or ADD/ADHD (with medication). The information regarding the proportion of children who were exposed to a language other than English was also not documented in the normative sample of ENNI. The lack of information regarding the presence/absence of concomitant disorders and children's language background could limit the generalizability of the current findings to those who clearly do not have any of these disorders or those who are exposed only to English. On the other hand, one could argue that the current findings may still be applicable to the clinical population who are likely to have similar profiles to children with LI in this study.

Clinical Implications and Applications

Given the current findings, clinicians may use FVMC computed from the ENNI protocol as part of a diagnostic battery for children between age 4 and age 7 who are suspected of LI. For example, some work settings mandate the use of prescriptive cutoffs (e.g., −1.25 *SD,* −1.5 *SD*) to determine the eligibility for speech-language services. In such cases, the means and standard deviations for children with TL from this study can be used for calculating the *z* scores (i.e., number of standard deviations from the mean) needed to make diagnostic decisions. Supplemental

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Material S5 provides a calculation template for clinicians to convert FVMC raw scores to *z* scores. To demonstrate, consider a 6-year-old child from a state that uses −1.5 *SD* as the cutoff to determine the eligibility for speech-language service. If the child produces an FVMC score of 72.12% in the ENNI protocol, we can use the reference data to compute his *z* score for FVMC, which would be −8.20 (i.e., [72.12%−97.55%]/3.10% = −8.20). This *z* score could be considered as one piece of evidence suggesting that this child is eligible for therapy. Other pieces of evidence may come from parent/teacher rating/interview (i.e., whether parents or teachers have concerns about the child's language skills) and standardized tests because the diagnosis of LI must consider both environmental and norm-referenced expectations (Bishop et al., 2016; Paul et al., 2018).

At this point of time, there is no trustworthy evidence indicating that FVMC computed from narratives samples can be used for identifying LI in children who are 8 years old or older. However, this does not necessarily mean that clinicians should not evaluate FVMC for children who are 8 years old or older. Given that tense deficits are a hallmark of children with LI, some children with LI who are 8 years old or older may still show difficulty using tense morphemes in spoken discourse. If production of tense morphemes is the treatment goal, FVMC could be computed to monitor treatment progress for children aged 8 years and older as well as for younger children. When FVMC is used for this purpose, the upper limits of the 90% or 95% confidence intervals from this study could be used to determine whether a child makes clinically significant progress in FVMC (McCauley & Swisher, 1984).

As an example, consider a 4-year-old with LI who has an FVMC score of 68.24% from the ENNI protocol before treatment. Using the value of 95% confidence interval from this study (i.e., ± 9.50%), the clinician estimates that the child's true FVMC score falls between 58.74% (lower limit) and 77.74% (upper limit; see Supplemental Material S5 for a calculation template for all ages). After the treatment, this child achieves an FVMC score of 74.59%. Although the FVMC score improves from pre- to posttreatment, the posttreatment FVMC score does not exceed the upper limit of the pretreatment FVMC score (i.e., 77.74%), and therefore, the improvement would not be considered clinically significant (Gillam et al., 2008).

Concluding Thoughts

In a recent survey, Finestack and Satterlund (2018) found that more than 25% of the clinicians in their study reported to use the type–token ratio, a vocabulary measure from language samples, to monitor progress for grammatical treatment. The use of type–token ratio for this purpose, indeed, goes against evidence-based practice. This finding suggests that valid grammatical measures are critically needed for clinicians to document treatment progress. This study addressed this need by providing the reference data for FVMC, a grammatical measure evaluating children's skills in producing tense and agreement morphemes. In addition to documenting treatment progress, we also demonstrated that FVMC was reliable and valid and could be used for identifying children with LI from preschool to early elementary school ages (i.e., up to age 7). FVMC was, however, not an appropriate for identifying children with LI who were 8 years old or older due to limited psychometric properties (i.e., diagnostic accuracy). With the reference data and psychometric properties for FVMC, we hope this study would shed light on the clinical value for language sample analysis, in general, and for FVMC, in specific.

Acknowledgments

The development of the ENNI was supported by a grant from the Children's Health Foundation of Northern Alberta. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Children's Health Foundation of Northern Alberta.

We are grateful to the children who participated and the teachers and clinicians who referred children to this study. We also thank Amy Briggs, Katelynn Imagna, Kayla Kuehlewind, and Sanjana Nair for coding the data and Gwyneth Rost for her comments on this article.

References

Aiken, L. R., & Groth-Marnat, G. (2006). *Psychological testing and assessment* (12th ed.). Boston, MA: Allyn & Bacon.

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*(5), 1185–1192. https://doi.org/10.1044/jslhr.4105.1185

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE Consortium. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLOS ONE, 11*(7), e0158753.

Blishen, B. R., Carroll, W. K., & Moore, C. (1987). The 1981 socioeconomic index for occupations in Canada. *Canadian Review of Sociology and Anthropology, 24,* 465–488.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Charest, M., & Johnston, J. R. (2011). Processing load in children's language production: A clinically oriented review of research. *Canadian Journal of Speech-Language Pathology & Audiology, 35*(1), 18–35.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dawson, J., Stout, C., Eyer, J., Tattersall, P., Fonkalsrud, J., & Croley, K. (2005). *Structured Photographic Expressive Language Test–Preschool 2 (SPELT-P 2)*. DeKalb, IL: Janelle Publication.

Dollaghan, C. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore, MD: Brookes.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*(5), 1136–1146.

Dollaghan, C., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research, 54*(4), 1077–1088.

Finestack, L. H., & Satterlund, K. E. (2018). Current practice of child grammar intervention: A survey of speech-language

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

pathologists. *American Journal of Speech-Language Pathology, 27*(4), 1329–1351. https://doi.org/10.1044/2018_AJSLP-17-0168

Geyman, J., Deyo, R., & Ramsey, S. (2000). *Evidence-based clinical practice: Concepts and approaches*. Boston, MA: Butterworth-Heinemann.

Gillam, R. B., Loeb, D. F., Hoffman, L. M., Bohman, T., Champlin, C. A., Thibodeau, L., . . . Friel-Patti, S. (Eds.). (2008). The efficacy of Fast ForWord language intervention in school-age children with language impairment: A randomized controlled trial. *Journal of Speech, Language, and Hearing Research, 51*(1), 97–119. https://doi.org/10.1044/1092-4388(2008/007)

Gillam, R. B., & Pearson, N. A. (2017). *Test of Narrative Language–Second Edition (TNL-2)*. Austin, TX: Pro-Ed.

Gladfelter, A., & Leonard, L. B. (2013). Alternative tense and agreement morpheme measures for assessing grammatical deficits during the preschool period. *Journal of Speech, Language, and Hearing Research, 56*(2), 542–552. https://doi.org/10.1044/1092-4388(2012/12-0100)

Goffman, L., & Leonard, J. (2000). Growth of language skills in preschool children with specific language impairment: Implication for assessment and intervention. *American Journal of Speech-Language Pathology, 9,* 151–161. https://doi.org/10.1044/1058-0360.0902.151

Guo, L.-Y., & Eisenberg, S. (2014). The diagnostic accuracy of two tense measures for identifying 3-year-olds with language impairment. *American Journal of Speech-Language Pathology, 23*(2), 203–212. https://doi.org/10.1044/2013_AJSLP-13-0007

Guo, L.-Y., Eisenberg, S., Schneider, P., & Spencer, L. (2019). Percent grammatical utterances between 4 and 9 years of age for the Edmonton Narrative Norms Instrument: Reference data and psychometric properties. *American Journal of Speech-Language Pathology*. Advance online publication. https://doi.org/10.1044/2019_AJSLP-18-0228

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

Hadley, P., & Short, H. (2005). The onset of tense marking in children at risk for specific language impairment. *Journal of Speech, Language, and Hearing Research, 48,* 1344–1362.

Hall-Mills, S. (2018). Language progress monitoring for elementary students. *Perspectives of the ASHA Special Interest Groups, 3*(1), 170–179. https://doi.org/10.1044/persp3.SIG1.170

Hughes, D., McGillivray, L., & Schmidek, M. (1997). *Guide to narrative language*. Austin, TX: Pro-Ed.

Kleinbaum, D., Kupper, L., Muller, K., & Nizam, A. (1988). Applied regression analysis and other multivariable methods (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Lahey, M. (1988). *Language disorders and language development*. Needham, MA: Macmillan.

Leadholm, B. J., & Miller, J. F. (1992). *Language sample analysis: The Wisconsin guide*. Madison, WI: Wisconsin Department of Public Instruction.

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.

Leonard, L. B., Haebig, E., Deevy, P., & Brown, B. (2017). Tracking the growth of tense and agreement in children with specific language impairment: Differences between measures of accuracy, diversity, and productivity. *Journal of Speech, Language, and Hearing Research, 60*(12), 3590–3600. https://doi.org/10.1044/2017_jslhr-l-16-0427

Leonard, L. B., Miller, C., & Gerber, E. (1999). Grammatical morphology and the lexicon in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*(3), 678–689.

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL: National Council of Teachers of English.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). New York, NY: Psychology Press.

McCauley, R. J., & Swisher, L. (1984). Use and misuse of norm-referenced test in clinical assessment: A hypothetical case. *Journal of Speech and Hearing Disorders, 49*(4), 338–348.

Miller, J., Andriacchi, K., & Nockerts, A. (2016). *Assessing language production using SALT software: A clinician's guide to language sample analysis* (2nd ed.). Madison, WI: SALT Software.

Miller, J., & Chapman, R. S. (2000). Systematic Analysis of Language Transcripts [Computer software]. Madison: University of Wisconsin.

Moyle, M., Karasinski, C., Ellis Weismer, S., & Gorman, B. (2011). Grammatical morphology in school-age children with and without language impairment: A discriminant function analysis. *Language, Speech, and Hearing Services in Schools, 42,* 550–560.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*(5), 1033–1047. https://doi.org/10.1044/1092-4388(2005/072)

Pavelko, S. L., Owens, J. R. E., Ireland, M., & Hahs-Vaughn, D. L. (2016). Use of language sample analysis by school-based SLPs: Results of a nationwide survey. *Language, Speech, and Hearing Services in Schools, 47*(3), 246–258. https://doi.org/10.1044/2016_LSHSS-15-0044

Paul, R., Norbury, C. F., & Gosse, C. (2018). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (5th ed.). Atlanta, GA: Elsevier.

Pawłowska, M. (2014). Evaluation of three proposed markers for language impairment in English: A meta-analysis of diagnostic accuracy studies. *Journal of Speech, Language, and Hearing Research, 57*(6), 2261–2273. https://doi.org/10.1044/2014_JSLHR-L-13-0189

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25*(1), 15–24.

Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of clinical markers of specific language impairment in adults. *Journal of Speech, Language, and Hearing Research, 53*(2), 414–429. https://doi.org/10.1044/1092-4388(2009/08-0016)

Rice, M., & Wexler, K. (2001). Test for Early Grammatical Impairment. San Antonio, TX: The Psychological Corporation.

Rice, M., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1412–1431. https://doi.org/10.1044/jslhr.4106.1412

Sackett, D. (1991). *Clinical epidemiology: A basic science for clinical medicine* (2nd ed.). Boston, MA: Little Brown.

Schneider, P., Dubé, R. V., & Hayward, D. (2005). *The Edmonton Narrative Norms Instrument*. Retrieved from http://www.rehabmed.ualberta.ca/spa/enni/

Schneider, P., & Hayward, D. (2010). Who does what to whom: Introduction of referents in children's storytelling from pictures. *Language, Speech, and Hearing Services in Schools, 41*(4), 459–473. https://doi.org/10.1044/0161-1461(2010/09-0040)

Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton narrative norms instrument.

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions

Journal of Speech-Language Pathology and Audiology, 30(4), 224–238.

Schuele, C. M. (2013). Beyond 14 grammatical morphemes toward a broader view of grammatical development. *Topics in Language Disorders, 33*(2), 118–124. https://doi.org/10.1097/TLD.0b013e3182928dc2

Semel, E., Wiig, E. H., & Secord, W. A. (1995). *Clinical Evaluation of Language Fundamentals–Third Edition (CELF-3)*. San Antonio, TX: The Psychological Corporation.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4)*. San Antonio, TX: Pearson.

Shriberg, L. D., Kwiatkowski, J., & Hoffman, K. (1984). A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Research, 27,* 456–465.

Souto, S. M., Leonard, L. B., & Deevy, P. (2014). Identifying risk for specific language impairment with narrow and global measures of grammar. *Clinical Linguistics & Phonetics, 28*(10), 741–756. https://doi.org/10.3109/02699206.2014.893372

Statistics Canada. (n.d.). Canada dimensions: The people [online]. Retrieved from http://www.statcan.ca

Systat Software, Inc. (2011). *SigmaPlot® 12.0*. Point Richmond, CA: Author.

Tager-Flusberg, H., & Cooper, J. (1999). Present and future possibilities for defining a phenotype for specific language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 1275–1278. https://doi.org/10.1044/jslhr.4205.1275

Werner, E. O., & Kresheck, J. D. (1983). *Structured Photographic Expressive Language Test–2*. DeKalb, IL: Janelle Publications.

Wiig, E., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals–Preschool (CELF-P)*. San Antonio, TX: The Psychological Corporation.

Downloaded from: https://pubs.asha.org Sarita Eisenberg on 11/08/2019, Terms of Use: https://pubs.asha.org/pubs/rights_and_permissions