Research Article

The Finite Verb Morphology Composite: Values From a Community Sample

Johanna M. Rudolph,^a Christine A. Dollaghan,^a and Simone Crotteau^{a,b}

Purpose: We investigated the finite verb morphology composite (FVMC), a measure associated with developmental language disorder (DLD), in a large community sample to provide evidence on its distribution, its association with other variables, and its sensitivity and specificity.

Method: We coded percent occurrence of FVMC morphemes in obligatory contexts in archival language samples from 676 six-year-olds conversing with an adult family member. Using multiple regression, we quantified the variance in FVMC scores accounted for by language sample characteristics (number of utterances and obligatory contexts) and child/family variables (performance IQ, family history, maternal education, and adult FVMC). We calculated sensitivity and specificity of low (< 85%) FVMC scores for predicting low (–1 SD)

scores on measures of utterance length, receptive vocabulary, and nonword repetition.

Results: FVMC scores ranged from 33% to 100% ($M = 93\%$). Number of obligatory contexts ($R^2 = 3\%$), performance IQ ($R^2 = 11\%$), and adult FVMC ($R^2 = 18\%$) were significant covariates. Sensitivity ranged from 25% to 35%; specificity was from 91% to 92%.

Conclusion: FVMC scores at the age of 6 years were generally high, but 12% of the participants had FVMCs below 85%, and scores were significantly associated with characteristics of the language samples, children, and adult interlocutors. Sensitivity of the FVMC was considerably lower in this community sample than in previous studies comparing groups of children who met criteria for DLD or typical language. Evidence from large representative samples is important when developing and validating potential clinical markers of DLD.

Deficits in morphology are characteristic of English-speaking children with specific language impairment or developmental language disorder (DLD; the label used here cf. Bishop, Snowling, Thompson, Greenhalgh, & the CATALISE-2 Consortium, 2017). A number of investigations have shown that tense and agreement (T/A) morphemes represent an area of particular weakness for this population (see Gladfelter & Leonard, 2013). Descriptions, definitions, and theoretical accounts vary somewhat, but it is generally agreed that the T/A system includes third-person singular present tense *-s* (3S), past tense *-ed*, copula and auxiliary forms of BE (*am*, *is*, *are*, *was*, *were*), and auxiliary forms of DO (*do*, *does*, *did*). Considerable evidence shows that, as a group, children with DLD perform at lower levels than age- or mean length of

utterance (MLU)–matched peers, both in producing T/A morphemes spontaneously and in responding to elicited production and grammaticality judgment probes, leading several investigators to examine the potential value of T/A morphemes as clinical markers of DLD (e.g., Hadley & Short, 2005; Marchman, Wulfeck, & Ellis Weismer, 1999; Rice & Wexler, 2001; Simkin & Conti-Ramsden, 2001).

One of the earliest (Bedore & Leonard, 1998) and most studied measures of T/A morphemes is the finite verb morphology composite (FVMC; Leonard, Miller, & Gerber, 1999). The FVMC is derived from a spontaneous language sample, in either a free-play or elicited narrative scenario, and reflects the percent occurrence in obligatory contexts of eight T/A morphemes: regular past tense *-ed*, 3S, and present tense uncontracted and contracted copula and auxiliary BE forms (*am*, *is*, *are*). Traditionally, such forms have been defined as “mastered” when they are produced in more than 85% of at least three or four obligatory contexts, which, for children with typical language (TL) development, occurs by the age of 6 years (Rice, 2012; Rice & Wexler, 2001); however, little is known about the distribution of FVMC scores at that age or other ages. In addition, the extent to which characteristics

^aCallier Center for Communication Disorders, University of Texas at Dallas

^bReliant Rehabilitation, Plano, TX

Correspondence to Johanna M. Rudolph:

Johanna.Rudolph@utdallas.edu

Editor-in-Chief: Sean Redmond

Editor: Jan de Jong

Received October 31, 2018

Revision received December 17, 2018

Accepted January 18, 2019

https://doi.org/10.1044/2019_JSLHR-L-18-0437

Disclosure: Christine A. Dollaghan has authored a book that touches on many of the topics discussed in this article and receives royalties for book sales. The remaining authors report no relevant conflicts of interest related to the article.

of the language sample, such as the number of utterances or the number of obligatory contexts, might influence FVMC scores is unknown. Similarly, although some investigations have controlled for socioeconomic, dialectical, and cognitive factors in studying the FVMC (e.g., Gladfelter & Leonard, 2013; Guo & Eisenberg, 2014), the influence of variables such as family history, maternal education, caregiver input, and performance IQ, which have been shown or are assumed to be associated with language performance (e.g., Dollaghan et al., 1999; Rice, Smith, & Gayan, 2009; Weisleder & Fernald, 2013), has not been investigated. Knowledge of these associations could provide valuable information about when and for whom the FVMC is an appropriate and unbiased clinical measure.

Finally, the available evidence concerning the FVMC's potential as a clinical marker for distinguishing between children with DLD and children with TL suffers from several limitations. Table 1 summarizes information on a set of representative studies, including sample sizes, point estimates, and 95% confidence intervals (CIs) for the FVMC's sensitivity (the percentage of children with DLD it correctly identified) and specificity (the percentage of children with TL it correctly identified). Table 1 shows that sample sizes were generally small, particularly for children with DLD, and that 95% CIs were wide. As a result, although most point estimates for sensitivity and specificity exceeded 80% (the criterion for acceptable diagnostic accuracy proposed by Plante & Vance, 1994), none of the surrounding CIs had lower bounds this high and some fell far below this target. Furthermore, criteria for recruiting participants in these studies have varied and, in some cases, may have maximized group separation a priori, for example, by requiring severe language deficits in the DLD group but allowing language skills as high as the 99th percentile in the TL group (Gladfelter & Leonard, 2013; Souto, Leonard, & Deevy, 2014). Research on the FVMC's accuracy in more representative samples of children with DLD and TL is an

important prerequisite for evaluating its potential as a clinical marker.

In this study, we examined the FVMC in 676 six-year-old children from whom language samples had been obtained in the course of a previous investigation. This archival data set allowed us to examine the distribution of the FVMC in a representative group of children and provided the opportunity to determine the extent to which the FVMC is affected by certain child, family, and language sample characteristics. We also examined the FVMC's accuracy in identifying children who had significantly low scores on three other measures that have been used to identify DLD in previous investigations. Six-year-olds were of particular interest for this portion of the study because children with TL are expected to master the FVMC and its constituent morphemes by the age of 6 years (e.g., Rice & Wexler, 2001), whereas some older children with DLD still use these morphemes inconsistently (Guo & Schneider, 2016; see also Moyle, Karasinski, Ellis Weismer, & Gorman, 2011). Therefore, we reasoned that a pronounced disparity between typical and atypical FVMC use might be especially apparent at this age.

We had four specific aims:

1. to describe the distribution of the FVMC;
2. to determine the amount of FVMC variance accounted for by two language sample characteristics: number of utterances and number of obligatory contexts;
3. to determine the amount of FVMC variance accounted for by four child/family characteristics: child's performance IQ, family history of developmental communication disorders, maternal education, and caregiver use of the target morphemes; and
4. to quantify the sensitivity and specificity of low FVMC scores for identifying children with low scores on measures of utterance length, receptive vocabulary, and nonword repetition.

Table 1. Ages, numbers (*n*) of children with language impairment (LI) and typical language (TL), finite verb morphology composite (FVMC) means (*M*), standard deviations (*SD*), cutoff values, point estimates, and 95% confidence intervals (CI) for sensitivity and specificity from previous studies.

Study	Age (years)	LI		TL		Cutoff	Sensitivity [95% CI]	Specificity [95% CI]
		<i>n</i>	<i>M</i> (<i>SD</i>)	<i>n</i>	<i>M</i> (<i>SD</i>)			
Bedore & Leonard (1998)	4–5	6	7% (9%)	6	95% (6%)	—	100 [4, 100]	100 [54, 100]
Gladfelter & Leonard (2013)	4	12	56% (21%)	16	96% (4%)	84%	100 [74, 100]	100 [78, 100]
	5	13	68% (19%)	15	97% (5%)	85%	92 [64, 100]	93 [68, 100]
Guo & Eisenberg (2014)	3	18	78% (21%)	18	97% (3%)	95%	83 [59, 96]	89 [65, 99]
Souto et al. (2014)	4 ^a	14	57% (24%)	16	97% (4%)	—	93 [66, 100]	94 [70, 100]
	4 ^b	14	62% (21%)	16	99% (2%)	—	93 [66, 100]	100 [79, 100]
	5 ^a	11	70% (19%)	15	97% (4%)	—	91 [59, 100]	93 [68, 100]
	5 ^b	11	64% (22%)	15	97% (3%)	—	82 [48, 98]	93 [68, 100]
Guo & Schneider (2016)	6	11	79% (18%)	50	97% (3%)	94%	82 [48, 98]	90 [78, 97]
	8	17	88% (18%)	50	99% (2%)	98%	77 [50, 93]	80 [6, 90]

Note. Em dashes indicate that these values were not reported in the studies.

^aRandomly assigned original group. ^bRandomly assigned replication group.

Method

Participants

This research was performed with approval from the University of Texas at Dallas Institutional Review Board (IRB 14-62); no consent forms were obtained as the study focused on de-identified archival data. Participants were 6-year-old children (age range: 71–75 months) followed longitudinally in the course of a series of investigations into otitis media and child development; detailed descriptions of the methods employed in the original study are available in previous publications (e.g., Paradise et al., 2005). In brief, participants in the original study were recruited between the ages of 2 days and 2 months from urban, rural, and suburban sites in Pittsburgh, Pennsylvania, and surrounding areas; infants were excluded based on medical risks (e.g., birth weight < 5 lb, multiple birth, neonatal asphyxia) and other demographic factors (e.g., multilingual household, foster care or adopted, mother < 18 years of age). From 6,350 healthy infants originally enrolled, 748 met criteria for inclusion in the longitudinal portion of that study and participated in one or more annual testing sessions scheduled to occur within 2 months of their third, fourth, and sixth birthdays. At the 3-year-old testing session, the interviewer estimated the child's race or ethnic group as Black, White, or other/indeterminate and parents completed questionnaires on the number of first-degree family members with a history of developmental communication disorders (Tomblin, 1989), and the highest level of education completed by the child's mother.

Of the 748 eligible children, 729 (97.5%) participated in an individual testing session at the age of 6 years. During the session, a number of norm-referenced tests and criterion-referenced procedures were administered, and a language sample was audio-recorded as children and adult family members conversed for approximately 15 min; adults were instructed to "play and talk with [child's name] as you would at home." Language samples from 53 of the 729 participants were excluded from this study either because of recording problems or because the adult interlocutor was someone other than the child's mother, father, grandmother, grandfather, aunt, or uncle. Language samples from the remaining 676 dyads were the starting point for the analyses of the FVMC we report below. Demographic characteristics of this subset were similar to those of the entire sample and representative of the surrounding community (Paradise et al., 2005). Specifically, 45% of the 676 participants were girls; 30% were Black, 67% were White, and 3% were classified as other/indeterminate.

Measures

FVMC

The FVMC was the primary measure of interest for all analyses in this study. To calculate it, three trained and blinded coders began by reviewing the 676 electronic files that had been transcribed during the original study by trained research assistants using Systematic Analysis of Language

Transcripts (SALT; Miller & Chapman, 2000) software. Coders independently examined both child and adult utterances and entered electronic codes to indicate obligatory contexts, occurrences, and absences of the eight grammatical morphemes included in the original FVMC (Bedore & Leonard, 1998; Leonard et al., 1999): 3S, regular past tense *-ed*, and present tense contracted and uncontracted BE forms *am*, *is*, and *are*. We specified two additional rules for forms that are common in language samples from diverse populations but, to our knowledge, have not been addressed explicitly in the FVMC literature. First, because deletion of fronted BE from yes/no questions (e.g., *You hungry?*) is an optional feature of both dialect (inner-city English) and stylistic (informal English) variations (e.g., Akmajian, Farmer, Bickmore, Demers, & Harnish, 2017), we only coded yes/no questions as obligatory contexts for BE if a BE form was present (*Are you hungry?*), not if it was absent (*You hungry?*). Second, because zero subject–verb agreement or concord (e.g., *They is baking*) is an optional feature of African American English (Oetting & McDonald, 2001; Washington & Craig, 1994), we coded BE forms as present in utterances irrespective of concord.

After coding, we calculated the percentage of occurrence in obligatory contexts for each morpheme individually in all complete and intelligible utterances. Finally, we calculated the FVMC by dividing the total number of target morphemes present in obligatory contexts by the total number of obligatory contexts; we did not require a minimum number of obligatory contexts.

Covariates/Predictors

The two language sample covariates were (a) the number of complete and intelligible utterances as a measure of how talkative the child was and (b) the number of obligatory contexts as a measure of how many opportunities the child had to produce the FVMC morphemes. SALT software yielded values for these two covariates for all 676 transcripts.

The four child/family covariates included (a) performance IQ, (b) family history of developmental communication disorders, (c) maternal education level, and (d) caregiver use of the T/A morphemes. We used standard scores on the Performance IQ subtest of the Wechsler Intelligence Scale for Children–Third Edition (WISC-III; Wechsler, 1991) as the first of these covariates; scores were available for 674 of the 676 participants. The covariate for family history was the number of first-degree relatives (father, mother, siblings) reported to have had developmental communication disorders (Tomblin, 1989); this information was available for 670 of the 676 participants. The covariate for maternal education had five levels: (a) less than a high school degree, (b) high school graduate, (c) technical school/some college/associate degree, (d) college graduate, and (e) postgraduate (cf. Patra, Greene, Patel, & Meier, 2016); this information was available for 674 of the 676 participants. The final covariate was caregiver use of the target morphemes as measured by FVMCs from adult interlocutors calculated from the language

samples using the same procedures as those used to determine children's FVMCs; this information was available for all 676 participants.

Reference Measures

To determine the accuracy of a potential clinical marker, each individual's binary classification on the index measure (affected/positive or unaffected/negative) is compared to that individual's classification on a reference measure that is assumed to provide the child's "true" status. To explore the FVMC as a possible marker for DLD, we began by dichotomizing children's FVMC scores using the 85% cutoff identified by Gladfelter and Leonard (2013), defining FVMC scores below 85% as positive and scores 85% or above as negative. Unfortunately, there is no uniform reference standard for identifying children with DLD, and various research groups have employed a variety of reference measures for this purpose, including test batteries (e.g., the EpiSLI criteria; Tomblin, Records, & Zhang, 1996), omnibus tests (e.g., Bedore & Leonard, 1998), and tests of specific language skills (e.g., Gladfelter & Leonard, 2013). To explore the FVMC's potential as a marker of DLD in the current study, we compared its classifications to those of three measures of specific language skills. We selected these reference measures based on the following criteria: (a) used to identify deficits in different language domains (e.g., vocabulary, syntax, phonology) and modalities (e.g., receptive, expressive), (b) used or recommended for use to identify children with DLD in a clinical context, and (c) suggested to be a clinical marker of DLD or used as a reference standard in prior research investigating T/A morpheme use as a clinical marker for DLD.

The first reference measure, MLU, was the SALT value for MLU in morphemes from all complete and intelligible utterances. This measure of expressive syntactic length was employed as a reference standard in seminal research on T/A use as a clinical marker for DLD (Rice & Wexler, 1996) and was found to be most effective at differentiating clinical groups at a cutoff of 1 *SD* below the mean. For the current study, we defined positive MLU values as those falling 1 *SD* or more below the mean of our sample, which appears to be the largest and most representative data set on this measure in children at this age.

The second reference measure, the Peabody Picture Vocabulary Test–Revised (PPVT-R), Form M (Dunn & Dunn, 1981), was selected following the precedent of Oetting, Cantrell, and Horohov (1999), who suggested that receptive vocabulary measures are useful for identifying DLD, particularly if dialect differences might confound scores on expressive language measures. This measure was also found to most effectively differentiate between clinical groups at a cutoff of 1 *SD* below the mean (Rice & Wexler, 1996). We therefore defined positive PPVT-R standard scores as those falling 1 *SD* or more below the mean standard score of our sample.

The third reference measure was nonword repetition, which has been suggested and referred to (Pawłowska,

2014) as a clinical marker of DLD. We defined this variable as the total percentage of phonemes repeated correctly (TOTPPC) on a 16-item nonword repetition task (Dollaghan & Campbell, 1998). Fifty-two children failed to attempt at least one nonword at all four nonword lengths (one, two, three, and four syllables) and were excluded from analyses involving this measure. Because no cut-off precedent exists for TOTPPC, we again defined positive scores as those that fell 1 *SD* or more below the mean of the 624 children for whom TOTPPC scores were calculated.

Statistical Analysis

To address Aim 1, we calculated descriptive statistics for the FVMC and plotted a histogram to investigate the distribution of scores in this sample. To address Aims 2 and 3, we conducted multiple regression analyses to determine the amount of variance in FVMC scores accounted for by (a) number of utterances and number of obligatory contexts and (b) performance IQ, family history, maternal education, and FVMC scores of the adult interlocutors. To address Aim 4, we calculated sensitivity, specificity, and associated 95% CIs for the FVMC with reference to MLU, PPVT-R, and TOTPPC in turn.

Results

Reliability

The second author (C. A. D.) independently coded a random selection of 10% of the language samples for the FVMC variables; the average percentage of agreement for coding target morphemes as present or absent in obligatory contexts was 97.2% for child samples and 96.7% for adult samples. Interrater reliability details for the transcribed language samples and reference standards are detailed in Paradise et al. (2005). In brief, 10%–20% of the audio-recorded language samples and nonword repetition tasks in the original study were randomly selected for independent transcription; interrater reliability (percentages of agreement for utterance segmentation, bound morpheme coding, and TOTPPC scores) ranged from 91% to 94%. All WISC-III and PPVT-R scores were verified independently in the original study with discrepancies resolved by discussion.

Aim 1: Distribution of the FVMC and Constituent Morphemes

Consistent with previous reports suggesting mastery of finite verb morphemes at approximately 6 years of age (e.g., Rice & Wexler, 2001), the average FVMC score was high ($M = 93\%$, $SD = 10$, range: 33%–100%). As shown in Figure 1, the distribution was unimodal and left-skewed, indicating that most of the children had scores of 85% or above. Most scores below 85% fell in the 70%–84% range; scores below 70% were distributed widely and almost equally among the lower intervals.

Figure 1. Histogram of finite verb morphology composite scores in increments of 5 percentage points from < 50% to 100%.

Aim 2: Association Between FVMC and Language Sample Factors

On average, children's language samples contained 34 obligatory contexts and 138 complete and intelligible utterances (see Table 2). It does not appear that a criterion has been set for a minimum number of obligatory contexts required to calculate the FVMC, and to our knowledge, only one study of the FVMC reported the average number

Table 2. Descriptive statistics for covariates.

Continuous covariates	<i>M</i>	<i>SD</i>	Range
FVMC OC	34	15	1–96
CIVU	138	41	29–276
WISC-PIQ	100	15	59–141
Adult FVMC	96%	7%	33%–100%
Ordinal covariates	<i>n</i>	%	
Maternal education			
< High school	86	13	
High school graduate	244	36	
Technical school, etc.	256	38	
College graduate	72	11	
Postgraduate	16	2	
Family history (parents/siblings)			
0	459	69	
1	140	21	
2	48	7	
3	15	2	
4	7	1	
5	1	< 1	

Note. FVMC = finite verb morphology composite; OC = obligatory context; CIVU = complete intelligible verbal utterances; WISC-PIQ = Wechsler Intelligence Scale for Children–Performance IQ.

of obligatory contexts produced by the sample (Guo & Schneider, 2016). Similarly, we are not aware of guidelines concerning the minimum number of utterances required to calculate the FVMC; the average number of utterances analyzed in previous studies has ranged from 50 to 1,000 utterances, although this information has not always been reported. The results of the multiple regression analysis (see Table 3) indicate that the number of complete and intelligible utterances and the number of obligatory contexts together accounted for only 3% of the variance in FVMC scores ($F = 9.80$, $p < .001$). Between these two variables, the number of obligatory contexts was statistically significant, with children who produced more obligatory contexts having higher scores. A series of simple linear regressions between FVMC scores and number of obligatory contexts was performed in which children who had less than a specified number of obligatory contexts were systematically eliminated. The association remained statistically significant until children who had less than 25 obligatory contexts were removed (see Table 4).

Aim 3: Association Between FVMC and Child/Family Factors

Our sample followed a normal distribution with regard to scores on the WISC-III Performance IQ subtest (see Table 2). Average FVMC score for the adult interlocutors was high (96%), but individual scores ranged widely. In terms of maternal education, 13% of the participants had mothers with less than a high school education, 74% had mothers who were high school graduates, and 13% had mothers who were college graduates. This distribution is representative of the surrounding community at the time of the original

Table 3. Multiple regression results for covariates.

Language sample covariates	Estimate	SE	t	Significance
Intercept	0.894	0.013	69.53	< .001
FVMC OC	0.001	0.000	3.30	.001*
CIVU	0.000	0.000	-0.14	.888
Child and family covariates				
Intercept	0.279	0.050	5.60	< .001
WISC-PIQ	0.001	0.000	5.16	< .001*
Maternal education	0.004	0.004	1.04	.298
Family history	0.000	0.004	-0.16	.875
Adult FVMC	0.536	0.054	9.85	< .001*

Note. Sample size was 676 for the language sample factor analysis; sample size was 668 for the child and family factor analysis. FVMC = finite verb morphology composite; OC = obligatory context; CIVU = complete intelligible verbal utterances; WISC-PIQ = Wechsler Intelligence Scale for Children-Performance IQ.

* $p < .01$.

study (Paradise et al., 2005). The majority of children had no family history of developmental communication disorders. The results of the multiple regression analysis for these covariates are reported in Table 3. Performance IQ, family history,

Table 4. Simple linear regression series for obligatory contexts, performance IQ, and adult finite verb morphology composite (FVMC).

Factor	Threshold	n Included	R ²	Significance
OC	N/A	676	3%	< .001*
	10	661	2%	< .001*
	15	637	2%	< .001*
	20	569	1%	.005*
	25	483	1%	.025
WISC-PIQ	N/A	676	11%	< .001*
	70	667	11%	< .001*
	75	647	8%	< .001*
	80	631	7%	< .001*
	85	569	6%	< .001*
	90	524	6%	< .001*
	95	438	2%	.002*
	100	345	2%	.002*
	105	248	2%	.009*
	106	248	2%	.009*
	107	228	2%	.054
Adult FVMC	N/A	676	18%	< .001*
	70%	666	15%	< .001*
	75%	662	15%	< .001*
	80%	649	17%	< .001*
	85%	629	16%	< .001*
	90%	549	12%	< .001*
	95%	513	1%	.008*
	96%	479	2%	.003*
	97%	446	3%	< .001*
	98%	389	1%	.104

Note. Bolded values identify thresholds for nonsignificance. OC = obligatory context; N/A = no threshold was used (i.e., all of the participants were included in the regression analysis); WISC-PIQ = Wechsler Intelligence Scale for Children-Performance IQ.

* $p < .01$.

maternal education, and adult FVMC together accounted for 22% of the variance in FVMC scores ($F = 47.85$, $p < .001$). WISC-III performance and adult FVMC scores were statistically significant in the model. Children with higher WISC-III scores and whose adult family members had higher FVMC scores also had higher FVMC scores. A series of simple linear regressions between FVMC scores and WISC-III scores was performed in which children who scored below a specified standard score on the WISC-III were systematically eliminated. The association remained statistically significant until all children who scored below 107 were removed (see Table 4). Similarly, a series of simple linear regressions between child FVMC scores and adult FVMC scores was performed in which children whose adult interlocutors scored below a specified threshold were systematically eliminated. The association remained statistically significant until all children whose adult family members scored less than 98% on the FVMC were removed (see Table 4).

Aim 4: Accuracy of the FVMC

Descriptive information for FVMC and the three reference measures (MLU, PPVT-R, and TOTPPC) is reported in Table 5, which also shows the cutoff value (-1.0 SD) that defined scores as positive and the resulting number and percentage of positive and negative scores on each measure. Table 6 shows exact values for true positives, false positives, false negatives, and true negatives as well as point estimates and 95% CIs for the FVMC's sensitivity and specificity relative to each reference measure. Sensitivity of the FVMC was low: All point estimates were below 40%, and lower bounds were below 30%, indicating that a positive score on the FVMC did not accurately predict a positive score on the other measures. Point estimates and lower bounds for the FVMC's specificity, on the other hand, all exceeded 85%, indicating that a negative FVMC score was predictive of negative scores on the other measures.

Discussion

We calculated the FVMC in a large community sample of 6-year-old children who conversed with an adult family member for approximately 15 min in a controlled setting. On average, child FVMCs were higher than 90%, but scores ranged widely, and 12% of the children fell below the threshold (85%) used to define mastery on this composite (Gladfelter & Leonard, 2013). Children's FVMC scores were significantly associated with the number of obligatory contexts they produced, their performance IQ scores, and T/A morpheme production by adult interlocutors, as measured by adult FVMC values. Three series of simple linear regressions were performed to identify the limits of those significant associations. Sensitivity of the FVMC for predicting poor performance on other language measures was low (point estimates < 40%, lower bounds < 30%), and specificity was uniformly high (point estimates > 90%, lower bounds > 85%). We discuss each aim in turn.

Table 5. Sample size, mean score, cutoff value, and number and percentage of positive and negative scores for each measure of children's language.

Measure	N	M	SD	Cutoff score	Positives, n (%)	Negatives, n (%)
FVMC	676	93%	10%	85%	81 (12)	595 (88)
MLU	676	3.87	0.73	3.13	98 (14)	578 (86)
PPVT-R	674	97	15	82	95 (14)	579 (86)
TOTPPC	624	76%	11%	65%	89 (14)	535 (86)

Note. FVMC = finite verb morphology composite; MLU = mean length of utterance in morphemes; PPVT-R = Peabody Picture Vocabulary Test–Revised (standard score); TOTPPC = total percentage of phonemes correct on nonword repetition task.

Aim 1: Distribution of the FVMC

The histogram plot of FVMC scores indicated that the distribution in this sample was unimodal with no clear evidence for different peaks to indicate a unique mode for children with DLD or other possible subgroups, such as dialect users, in the sample. Our findings concur with those of Rice (2012) and Rice and Wexler (2001) in that majority of the participants (88%) had mastered the constituent morphemes by the age of 6 years, but the 12% prevalence of those who fell below this cutoff was higher than might be expected given the prevalence rates previously and recently reported for children with DLD (e.g., Norbury et al., 2016; Tomblin et al., 1997). There are several possible reasons for the higher prevalence rates observed in this study, some of which are discussed below.

Aims 2 and 3: FVMC Covariates

The results of Aims 2 and 3 indicate that children's FVMC performance was influenced by three of six factors examined in this study. The first of these was the number of obligatory contexts the children produced: Children with fewer contexts had lower FVMC scores. Previous studies of the FVMC have generally reported the number of complete utterances produced by the participants, but this factor is not significantly associated with FVMC performance. The number of obligatory contexts has only rarely been mentioned, and no requirement for a minimum number of contexts has been set. Our results suggest that the association between the number of contexts and FVMC scores disappears if children have produced at least 25 obligatory contexts. This could be considered a

conservative guideline for the number of obligatory contexts required to achieve a robust sample for calculating this measure.

The second factor was performance IQ: Children with lower performance IQ scores also had lower FVMC scores. Our study differed from some previous studies of the FVMC in that we did not exclude participants who fell below a specified IQ threshold. Note, however, that the association between performance IQ and FVMC scores remained significant until all children with IQs below 107 were excluded—a value well above the “normal range” cutoffs employed in previous investigations. It is generally assumed that a standard score of 85 is sufficient to eliminate potential associations between verbal and nonverbal skills and to reduce sample heterogeneity associated with cognitive abilities, but the present results indicate that this threshold is not sufficiently stringent in the case of the FVMC. Similar results may be observed for other language measures. It does not seem reasonable to suggest that the FVMC be used exclusively with children who have performance IQs at or above 107 or that future studies of children with DLD focus exclusively on children who score above this cutoff. This finding echoes recent debates concerning cutoff values that should be used to define children with DLD and, indeed, whether nonverbal cognition should be an exclusionary factor in studies of such children (e.g., Reilly et al., 2014).

The third significant covariate was use of T/A morphemes by adult interlocutors, as reflected by adult FVMC scores: Children whose adult interlocutors had lower FVMC scores, themselves, had lower FVMC scores. This association, the strongest of any examined in this study, showed adult FVMCs accounted for 18% of the

Table 6. Accuracy values and metrics for the finite verb morphology composite relative to each reference measure.

Reference measure	TP	FP	FN	TN	Sensitivity [95% CI]	Specificity [95%CI]
MLU	25	56	73	522	26 [18, 35]	90 [88, 93]
PPVT-R	33	48	62	531	35 [26, 45]	92 [89, 94]
TOTPPC	27	47	62	488	30 [22, 41]	91 [89, 93]

Note. TP = true positives; FP = false positives; FN = false negatives; TN = true negatives; CI = confidence interval; MLU = mean length of utterance in morphemes; PPVT-R = Peabody Picture Vocabulary Test–Revised; TOTPPC = total percentage of phonemes correct on nonword repetition task.

variance in child FVMCs. This association remained significant until all children whose adult family members scored below 98% on the FVMC were eliminated. This finding could reflect the influence of factors such as dialect exposure, although these brief language samples may not provide an adequate assessment of the level of T/A morpheme exposure that these participants regularly experienced. Conservatively, it is reasonable to recommend that the FVMC should only be used with children who are regularly exposed to mainstream American English and that adult interlocutors collecting their language samples use the target T/A morphemes consistently. This finding supports the decision to exclude children who are dialect speakers from investigations of the FVMC.

Aim 4: Accuracy of the FVMC

The sensitivity values observed in the current study were below 40%, much lower than the values of 80% or higher that have been reported in some previous studies of the FVMC. There are several factors that might have contributed to this discrepancy, including the relatively short duration of the language samples we examined, the slight variation in FVMC coding procedures, our focus on a diverse community sample in which children were included irrespective of cognitive abilities and dialect background, the older age of the sample as compared to the samples in some previous investigations of the FVMC, and the more inclusive cutoff criteria and relatively narrow focus of the three reference measures we explored. Accordingly, we recharacterized the sample seven times to account for these differences and then recalculated sensitivity and specificity values for the reference measures as appropriate. First, we adjusted participant FVMC scores to account for departures from traditional criteria; that is, deletion of fronted BE in yes/no questions and lack of subject–verb BE form agreement were counted as omissions in obligatory contexts. Second, we eliminated all participants with performance IQ scores less than 85. Third, we eliminated all participants whose adult interlocutor scored less than 98% on the FVMC. Fourth, children who did not attempt a nonword at every syllable length were classified as positive rather than excluded. Fifth, children who scored more than 1.5 *SDs* below the mean on the reference measures were characterized as positive. Sixth, those children who scored more than 2.0 *SDs* below the mean were characterized as positive. Finally, those who scored below the cutoff on two or more reference measures were characterized as positive to replicate circumstances in which omnibus tests are used as the reference standard. Sensitivity was subsequently recalculated 17 times, and no significant differences from the values reported in Table 6 were found in any instance.

In our judgment, the most important contributor to the low levels of sensitivity observed in this study is the nature and composition of the present sample as compared to the samples employed in the studies listed in Table 1. As a community sample, participants in this study represented the full spectrum of performance on the measures we

examined, with many children, as expected, scoring near the cutoffs. By contrast, in the clinically identified samples recruited for some previous studies, participants with DLD have had language deficits that would characterize them as severely affected, whereas participants with TL have had scores well above the normal range (e.g., Gladfelter & Leonard, 2013; Souto et al., 2014). It is plausible that the FVMC's sensitivity would be higher under these conditions than in samples where the group differences are less extreme. Furthermore, the base rates (prevalence) of positive participants in our study, approximately 12%–15%, were much closer to population expectations and, accordingly, were smaller than in previous studies, where children with DLD have been oversampled to yield base rates as high as 50%. The impact of base rate on various accuracy metrics (Babchishin & Helmus, 2016; Robinson, Boissoneault, Sevel, Letzen, & Staud, 2016) has been discussed in the diagnostic literature as the “needle in a haystack” problem (Straus, Richardson, Glasziou, & Haynes, 2005). In the present context, it is plausible that the higher sensitivity values found in previous studies are related to their much higher base rates, in which there are relatively more positive children to be identified.

In summary, our study investigated aspects of the FVMC that have not been examined in previous research—its distribution, potential covariates that may influence children's performance on this measure, and the thresholds of significance for these covariates—as well as its accuracy in identifying children with low scores on other language measures in a representative sample. The covariate results suggest that this measure could provide a relatively unbiased assessment of T/A morpheme use in 6-year-old children if certain conditions are met (i.e., children with mainstream American English background and exposure provide at least 25 obligatory contexts for the target morphemes). However, whether the FVMC is a clinical marker of DLD is an open question. The most useful clinical markers are those for which the results from affected individuals are clearly distinct from the results from unaffected individuals (Sackett, Haynes, Guyatt, & Tugwell, 1991). An ideal clinical marker should therefore be distributed in such a way that there is minimal score overlap between those with the condition and those without the condition (Stokes, Wong, Fletcher, & Leonard, 2006). The results of the distribution analysis indicate that the FVMC is unimodally distributed, and the results of the accuracy analyses indicate that the FVMC does not adequately identify children with deficits in expressive syntax, receptive vocabulary, or nonword repetition. The findings of this study suggest a need for further investigation of the FVMC and other proposed clinical markers for DLD in studies with large samples that are representative of the conditions in which such markers will be employed.

Acknowledgments

The original study on which these analyses are based was supported by funding from the National Institutes of Health

(National Institute of Child Health and Human Development and National Institute on Deafness and Other Communication Disorders), the Agency for Health Care Policy and Research, SmithKline Beecham Laboratories, and Pfizer, Inc. We thank the principal investigator of the original study, Jack L. Paradise, and members of that research team: Thomas F. Campbell, Heidi M. Feldman, Janine E. Janosky, Marcia Kurs-Lasky, Dayna N. Pitcairn, Howard E. Rockette, Clyde G. Smith, and Diane L. Sabo. We also thank the pediatricians, research assistants, parents, and children who participated in that study.

References

- Akmajian, A., Farmer, A., Bickmore, L., Demers, R., & Harnish, R. M. (2017). *Linguistics: An introduction to language and communication* (7th ed.). Cambridge, MA: The MIT Press.
- Babchishin, K. M., & Helmus, L.-M. (2016). The influence of base rates on correlations: An evaluation of proposed alternative effect sizes with real-world data. *Behavior Research Methods*, 48(3), 1021–1031. <https://doi.org/10.3758/s13428-015-0627-7>
- Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research*, 41, 1185–1192. <https://doi.org/10.1044/jslhr.4105.1185>
- Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the CATALISE-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *The Journal of Child Psychology and Psychiatry*, 58(10), 1068–1080. <https://doi.org/10.1111/jcpp.12721>
- Dollaghan, C. A., & Campbell, T. F. (1998). Nonword repetition and language impairment. *Journal of Speech, Language, and Hearing Research*, 41, 1136–1146. <https://doi.org/10.1044/jslhr.4105.1136>
- Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research*, 42, 1432–1443. <https://doi.org/10.1044/jslhr.4206.1432>
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised: Manual for Forms L and M (PPVT-R)*. Circle Pines, MN: AGS.
- 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, 542–552.
- 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, 203–212. https://doi.org/10.1044/2013_AJSLP-13-0007
- 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, 317–329. https://doi.org/10.1044/2015_JSLHR-L-15-0066
- Hadley, P. A., & 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. [https://doi.org/10.1044/1092-4388\(2005\)094](https://doi.org/10.1044/1092-4388(2005)094)
- 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, 678–689. <https://doi.org/10.1044/jslhr.4203.678>
- Marchman, V. A., Wulfeck, B., & Ellis Weismer, S. (1999). Morphological productivity in children with normal language and SLI: A study of English past tense. *Journal of Speech, Language, and Hearing Research*, 42, 206–219. <https://doi.org/10.1044/jslhr.4201.206>
- Miller, J., & Chapman, R. (2000). SALT: Systematic Analysis of Language Transcripts (Version 7.0) [Computer software]. Madison: University of Wisconsin-Madison.
- Moyle, M. J., Karasinski, C., Ellis Weismer, S., & Gorman, B. K. (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. [https://doi.org/10.1044/0161-1461\(2011\)10-0029](https://doi.org/10.1044/0161-1461(2011)10-0029)
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., ... Pickels, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *The Journal of Child Psychology and Psychiatry*, 57(11), 1247–1257. <https://doi.org/10.1111/jcpp.12573>
- Oetting, J. B., Cantrell, J. P., & Horohov, J. E. (1999). A study of specific language impairment (SLI) in the context of non-standard dialect. *Clinical Linguistics & Phonetics*, 13(1), 25–44. <https://doi.org/10.1080/026992099299220>
- Oetting, J. B., & McDonald, J. L. (2001). Nonmainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research*, 44, 207–223. [https://doi.org/10.1044/1092-4388\(2001\)018](https://doi.org/10.1044/1092-4388(2001)018)
- Paradise, J. L., Campbell, T. F., Dollaghan, C. A., Feldman, H. M., Bernard, B. S., Colborn, K., ... Smith, C. G. (2005). Developmental outcomes after early or delayed insertion of tympanostomy tubes. *New England Journal of Medicine*, 353, 576–586. <https://doi.org/10.1056/NEJMoa050406>
- Patra, K., Greene, M. M., Patel, A. L., & Meier, P. (2016). Maternal education level predicts cognitive, language, and motor outcome in preterm infants in the second year of life. *American Journal of Perinatology*, 33(8), 738–744. <https://doi.org/10.1055/s-0036-1572532>
- 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, 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, 15–24. <https://doi.org/10.1044/0161-1461.2501.15>
- Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., ... Wake, M. (2014). Specific language impairment: A convenient label for whom? *International Journal of Language & Communication Disorders*, 49(4), 416–451. <https://doi.org/10.1111/1460-6984.12102>
- Rice, M. L. (2012). Towards epigenetic and gene regulation models of specific language impairment: Looking for links among growth, genes, and impairment. *Journal of Neurodevelopmental Disorders*, 1–14. <https://doi.org/10.1186/1866-1955-4-27>
- Rice, M. L., Smith, S. D., & Gayán, J. (2009). Convergent genetic linkage and associations to language, speech and reading measures in families of probands with specific language impairment. *Journal of Neurodevelopmental Disorders*, 1, 264–282. <https://doi.org/10.1007/s11689-009-9031-x>
- Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research*, 39, 1239–1257. <https://doi.org/10.1044/jslhr.3906.1239>
- Rice, M. L., & Wexler, K. (2001). *Rice/Wexler Test of Early Grammatical Impairment*. San Antonio, TX: The Psychological Corporation.

-
- Robinson, M., Boissoneault, J., Sevel, L., Letzen, J., & Staud, R.** (2016). The effect of base rate on the predictive value of brain biomarkers. *The Journal of Pain*, 17(6), 637–641. <https://doi.org/10.1016/j.jpain.2016.01.476>
- Sackett, D., Haynes, R., Guyatt, G., & Tugwell, P.** (1991). *Clinical epidemiology: A basic science for clinical medicine*. Boston, MA: Little, Brown and Company.
- Simkin, Z., & Conti-Ramsden, G.** (2001). Non/word repetition and grammatical morphology: Normative data for children in their final year of primary school. *International Journal of Language & Communication Disorders*, 36, 395–404.
- 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, 741–756. <https://doi.org/10.3109/02699206.2014.893372>
- Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Leonard, L. B.** (2006). Nonword repetition and sentence repetition as clinical markers of specific language impairment: The case of Cantonese. *Journal of Speech, Language, and Hearing Research*, 49, 219–236. [https://doi.org/10.1044/1092-4388\(2006/019\)](https://doi.org/10.1044/1092-4388(2006/019))
- Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, R. B.** (2005). *Evidence-based medicine: How to practice and teach EBM* (3rd ed.). Edinburgh, United Kingdom: Elsevier.
- Tomblin, J. B.** (1989). Familial concentration of developmental language impairment. *Journal of Speech and Hearing Disorders*, 54, 287–295. <https://doi.org/10.1044/jshd.5402.287>
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M.** (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research*, 40, 1245–1260. <https://doi.org/10.1044/jslhr.4006.1245>
- 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. <https://doi.org/10.1044/jshr.3906.1284>
- Washington, J. A., & Craig, H. K.** (1994). Dialectal forms during discourse of poor, urban, African American preschoolers. *Journal of Speech and Hearing Research*, 37, 816–823. <https://doi.org/10.1044/jshr.3704.816>
- Wechsler, D.** (1991). *Wechsler Intelligence Scale for Children—Third Edition (WISC-III)*. San Antonio, TX: The Psychological Corporation.
- Weisleder, A., & Fernald, A.** (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>