Using Language Sample Analysis in Clinical Practice: Measures of Grammatical Accuracy for Identifying Language Impairment in Preschool and School-Aged Children

Sarita Eisenberg, Ph.D.[1] and Ling-Yu Guo, Ph.D.[2]

ABSTRACT

This article reviews the existing literature on the diagnostic accuracy of two grammatical accuracy measures for differentiating children with and without language impairment (LI) at preschool and early school age based on language samples. The first measure, the finite verb morphology composite (FVMC), is a narrow grammatical measure that computes children's overall accuracy of four verb tense morphemes. The second measure, percent grammatical utterances (PGU), is a broader grammatical measure that computes children's accuracy in producing grammatical utterances. The extant studies show that FVMC demonstrates acceptable (i.e., 80 to 89% accurate) to good (i.e., 90% accurate or higher) diagnostic accuracy for children between 4;0 (years;months) and 6;11 in conversational or narrative samples. In contrast, PGU yields acceptable to good diagnostic accuracy for children between 3;0 and 8;11 regardless of sample types. Given the diagnostic accuracy shown in the literature, we suggest that FVMC and PGU can be used as one piece of evidence for identifying children with LI in assessment when appropriate. However, FVMC or PGU should not be used as therapy goals directly. Instead, when children are low in FVMC or PGU, we suggest that follow-up analyses should be conducted to determine the verb tense morphemes or grammatical structures that children have difficulty with.

KEYWORDS: Finite verb morphology composite, percent grammatical utterances, diagnostic accuracy, language sample analysis

[1]Department of Communication Sciences and Disorders, Montclair State University, Montclair, New Jersey; [2]University at Buffalo–The State University of New York, Buffalo, New York.

Address for correspondence: Sarita Eisenberg, Ph.D., Department of Communication Sciences and Disorders, Montclair State University, 1515 Broad Street, Bloomfield, NJ 07003 (e-mail: eisenbergs@mail.montclair.edu).

Automating Child Speech, Language and Fluency Analysis; Guest Editor, Brian MacWhinney, Ph.D.

Semin Speech Lang 2016;37:106–116.

DOI: http://dx.doi.org/10.1055/s-0036-1580740.
ISSN 0734-0478.

Learning Outcomes: As a result of this activity, the reader will be able to (1) analyze a language sample for finite verb morphology composite; (2) analyze a language sample for percent grammatical utterances; and (3) comment on the relevance and validity of these measures for identifying language impairment, selecting goals, and measuring treatment outcome.

Identifying the presence of a language disorder can be challenging, particularly for children with a primary language impairment (LI) who have milder degrees of impairment. Although this clinical decision is typically based on norm-referenced standardized tests,[1–3] several studies have concluded that quantitative language sample analysis (LSA) measures may be more accurate for identifying LI than tests.[4,5]

Not all LSA measures, however, can be used for identifying LI. To be valid for this purpose, LSA measures must have certain psychometric properties.[6] One such property is reliability, which measures whether the same performance would be obtained under different conditions (e.g., different sample types, different sample lengths) and with repeated administrations. Another property is diagnostic accuracy, defined as the ability of the assessment to correctly identify both children with LI (i.e., sensitivity) and children with typical language (TL; i.e., specificity). According to Plante and Vance,[7] a 90% level for both sensitivity and specificity is considered good and an 80 to 89% level is considered acceptable diagnostic accuracy for differentiating children with and without LI. An additional property important for clinical application is the availability of normative data and an empirically determined cutoff score for interpreting performance.

In this article, we focus on two measures of grammatical accuracy that are computed from language samples—a finite verb morphology composite (FVMC) and percent grammatical utterances (PCU). We will first describe each measure and how it is calculated. We will then review studies about diagnostic accuracy, reliability, and interpretation.

FINITE VERB MORPHOLOGY COMPOSITE

FVMC measures usage of grammatical morphemes that mark verb tense and agreement (henceforth referred to as verb tense morphemes). The motivation for this measure comes out of studies documenting that children with LI tend to have particular and persisting difficulty learning and consistently using verb tense morphemes.[8–12]

Computing Finite Verb Morphology Composite

Rather than separately calculating usage frequency for individual verb tense morphemes, FVMC computes the overall percentage of correct use in obligatory contexts of third-person singular present *-s*, regular past tense *-ed*, and copula and auxiliary *be* (i.e., *am, are, is, was, were*) with a single measure. It should be noted that FVMC does not include the infinitive form of *be* (e.g., I will *be* happy; she wants to *be* a teacher), present participle form of *be* (e.g., she is *being* difficult), past participle form of *be* (e.g., she has *been* very helpful), or gerund form of *be* (e.g., *being* a student is fun) in the analysis. This is because those forms, by definition, do not mark tense in English.[7,13,14]

FVMC does not count all morphemes that mark tense in English. For example, FVMC does not include auxiliary *do* (e.g., what *do* you like) even though this morpheme is included in another measure of tense usage (i.e., Rice and colleagues' tense composite[11]). FVMC also does not include auxiliary *have* (e.g., he *has* studied for 3 hours), irregular past tense (e.g., he *ran* away), or irregular third-person singular present tense (e.g., he *has* a book). Finally, FVMC does not include present tense verbs with first-person, second-person, or plural subjects (e.g., *I/you/we/they/the dogs* like *cakes*) in the analysis. In those contexts, the verbs are marked with tense covertly and do not require the use of third-person singular present *-s*, regular past tense *-ed*, copula *be*, or auxiliary *be*.

There are several steps for computing FVMC. Step 1 involves identifying the obligatory contexts for use of the four target verb tense morphemes. Step 2 involves coding uses and nonuses of the morphemes within those obligatory contexts. The number of correct uses of the four verb tense morphemes is divided by the total number of obligatory contexts and the resultant quotient is multiplied by 100 to obtain a percentage.

IDENTIFYING OBLIGATORY CONTEXTS

Obligatory contexts are identified by looking for instances in which verb tense morphemes are required for the utterance to be grammatical and match the situational context. Uses of verb tense morphemes that are not required are not counted as obligatory contexts and are, therefore, excluded from the FVMC analysis. This means that overgeneralizations of regular past tense *-ed* or third-person singular present *-s* with irregular verbs (e.g., *eated*, *dranked*, *haves*) are excluded from the FVMC analysis. This also means that overuses of third-person singular present *-s* with plural subjects (e.g., the boys *plays*) are excluded from the FVMC analysis.

Verb forms that occur without a preceding subject are also excluded from the FVMC analysis because verb tense marking is only obligatory when there is a sentence subject. The utterances in Table 1 are all missing subjects. Note that sentence subjects are required in English. However, it is pragmatically allowable to omit the sentence subject when the referent for that subject is recoverable from the previous utterance and therefore redundant in the discourse,[15] as shown in utterance 1. In this utterance, the subject *I* was omitted and therefore the auxiliary *am* was not required. Thus, this child utterance would be excluded from the FVMC analysis as it does not provide an obligatory context for use of a verb tense morpheme.

In utterances 2 and 3, nonproduction of the sentence subject is ungrammatical because the examiner's prior question obligated a full sentence response. However, the lack of a subject in both utterances means that use of the past tense morpheme *-ed* is not required and both utterances would be excluded from the FVMC analysis. Note that utterance 3 also demonstrates an overuse of the past tense morpheme (i.e., use in a context that does not make the morpheme obligatory), which provides another reason for excluding this utterance from the FVMC analysis.

Table 1 Subjectless Utterances

1. Examiner: What are you doing? Child: *Making* a tower.
2. Examiner: What happened? Child: *Walk* there.
3. Examiner: What happened? Child: *Walked* there.

CODING USES AND NONUSES

For each obligatory context, morphemes are coded as accurate use (e.g., the boys *are* sleeping, to talk about an ongoing current action), omitted (e.g., the boys Ø sleeping [note: the symbol Ø indicates the omission of a morpheme]), or inaccurate use (e.g., the boys *is* sleeping, involving an agreement error for number; the boys *are* sleeping yesterday, to talk about a past action and thus involving an error in time reference). FVMC is computed by dividing the number of accurate uses by the number of obligatory contexts and then multiplying by 100. (See Appendix 1 for an example.) Note that overgeneralizations and overuses of regular third-person and past tense and nonuse of verb tense morphemes in subjectless utterances do not get counted as errors in the FVMC calculation because they are excluded from the FVMC analysis from the outset.

Diagnostic Accuracy of Finite Verb Morphology Composite

Across the studies we reviewed, the mean FVMC reported for children with TL was 90% or higher with little variability (standard deviation [SD] < 5%) so, for each study, we will report FVMC only for the LI group. Although existing studies consistently found a significant difference in FVMC between children with LI and children with TL, specificity and sensitivity were not consistently in the good (i.e., 90 to 100%) or even acceptable (i.e., 80 to 89%) range across the studies.

The earliest study on the diagnostic accuracy of FVMC was conducted by Bedore and Leonard.[13] These authors investigated FVMC based on play samples elicited from children aged 3;7 (years;months) to 5;9. LI status was based on low (i.e., below age level) mean length of utterance and a score lower than 1 SD below the mean on an omnibus language test. Results were reported separately in two experiments. In experiment 1, mean FVMC for the LI group was 49% (SD = 26%). Specificity was acceptable (84%) and sensitivity was good (100%). In experiment 2, mean FVMC for the LI group was considerably lower (mean = 7%, SD = 9%) than that in experiment 1, and both specificity and sensitivity were good (100%). The cutoff score used for group differentiation was not specified. It should be noted that children with LI in this study were pooled together into one group and were not separated by age.

Two other studies by Leonard and his colleagues both focused on 4- and 5-year-old children separately using conversational samples.[14,16] One of these studies used a broader measure that also included auxiliary *do* in addition to the four verb tense morphemes in FVMC.[16] In both studies, classification as LI was based on scoring lower than 2.2 SD below the mean on a standardized test of expressive grammar. Across the two studies, mean FVMC in the LI group was 53 to 56% (SD = 21 to 24%) for 4-year-olds and 65 to 68% (SD = 19 to 22%) for 5-year-olds. Specificity and sensitivity were acceptable to good (82 to 100%) for 4- and 5-year-olds across the two studies. Only Gladfelter and Leonard reported cutoff scores: 84% for 4-year-olds and 85% for 5-year-olds.[16]

A study by Moyle et al looked at FVMC for school-aged children between the ages of 5;5 and 9;8 without separating children by age.[17] Classification as LI was based on intervention status (i.e., receiving therapy at the time of the study) and a score lower than 1 SD below the mean on a standardized test of receptive vocabulary or receptive grammar. Language samples were elicited by asking children to describe a fictional or personal experience and to explain how to do something. Mean FVMC for the LI group was 94% (SD = 6%). Specificity was acceptable (86%) but sensitivity was poor (only 50%). The cutoff score was not provided and children were not separated by age.

Guo and Schneider also investigated FVMC in school-aged children but provided separate data for children who were 6 and 8 years of age.[18] LI status was based on referral from a speech-language pathologist and a score lower than 1 SD below the mean on at least one composite score of an omnibus standardized language test. These authors used narrative samples based on picture sequences. Mean FVMC for children with LI was 79% (SD = 18%) for 6-year-olds and 88% (SD = 18%) for 8-year-olds. For the 6-year-old group, specificity was good (90%) and sensitivity was acceptable (82%) at a cutoff score of 93%. For the 8-year-old group, specificity was acceptable (80%) but sensitivity was only fair (76%) at a cutoff score of 97%.

There was also one study focusing only on preschool children. Guo and Eisenberg investigated FVMC in 3-year-olds (3;0 to 3;11) based on play samples.[19] All of the children in the LI group were receiving therapy at the time of the study or were subsequently enrolled in therapy. Mean FVMC for the LI group was 79% (SD = 23%). At a cutoff of 95%, both specificity (89%) and sensitivity (83%) were acceptable. However, when children were separated into younger (3;0 to 3;5) and older (3;6 to 3;11) age groups, specificity dropped to fair (75%) for the younger 3-year-olds and specificity dropped to fair for the older 3-year-olds.

In sum, FVMC showed at least acceptable diagnostic accuracy for children aged 4 to 6 years both for play (i.e., conversational) samples and for narrative samples (see Table 2 for a summary across studies). Above this age range, children with LI have achieved a sufficiently high usage level for verb tense morphemes so that FVMC is insufficiently sensitive. Below this age range, usage of tense morphemes is much more variable across children, thus reducing diagnostic accuracy for FVMC.

Reliability of Finite Verb Morphology Composite

We know of only one study that looked at reliability of FVMC as a function of sample size. Guo and Eisenberg compared FVMC for 3-year-old children based on 100-utterance and 50-utterance samples elicited during play.[19] Mean FVMC was not affected by sample size for either group. That is, mean FVMC was 97%

Table 2 Specificity, Sensitivity, and Cutoff Scores for FVMC

Study	Age‡	Sample Type	Cutoff (%)	Specificity (%)	Sensitivity (%)
Guo and Eisenberg (2014)[19,*]	3;0–3;11	Conversation	95	89	83
	3;0–3;5	Conversation	95	75	100
	3;6–3;11	Conversation	94	100	70
Gladfelter and Leonard (2013)†	4;0–4;6	Conversation	84	100	100
	5;0–5;6	Conversation	85	93	92
Souto et al[14,16,*]	4;0–4;6	Conversation	NA	94–100	93
	5;0–5;6	Conversation	NA	93	82–91
Guo and Schneider (in press)[18,*]	6;0–6;11	Narrative	93	90	82
	8;0–8;11	Narrative	97	80	76

Abbreviations: FVMC, finite verb morphology composite; NA, not applicable.
*FVMC calculation included auxiliary *be*, copula *be*, regular past *-ed*, and regular third-person singular *-s*.
†FVMC calculation also included auxiliary *do*.
‡Year;month.

for the TL children and 79% for the LI children for both sample sizes. This suggests that FVMC could be calculated based on samples of 50 utterances. However, sample size did affect diagnostic accuracy. For 100-utterance samples using a cutoff of 95%, specificity and sensitivity were both acceptable (89 and 83%). For 50 utterance samples using a cutoff of 91%, specificity was good (100%) but sensitivity was poor (67%).

We found no studies that compared FVMC across sample types. However, comparing FVMC across studies shows no difference in FVMC between conversation and narration, suggesting that FVMC may be impervious to sample type. We found no studies that looked at temporal stability for FVMC.

Interpreting Finite Verb Morphology Composite

As we noted previously, the reported FVMC for children with TL was 90% or better even for 3-year-olds at the group level. These data may seem surprising in light of the growth trajectory reported for verb tense marking in a longitudinal study by Rice et al.[11] These authors observed a sharp increase in verb morpheme usage after the age of 3, starting around 50% usage and only reaching a 90% usage level at age 4. There were, however, several significant differences between how verb tense usage was computed in Rice et al[11] and how FVMC was computed in other studies.[14,16,18,19] First, Rice et al included auxiliary *do* in their calculation as well as the four morphemes included in FVMC.[11] Second, they included data from probes as well as from conversational speech. Third, they averaged seven separate calculations of usage frequency for individual morphemes and for different data types rather than calculating a single usage frequency based on obligatory contexts for all morphemes combined.

As noted previously, only some of the studies about FVMC reported cutoffs. Those that did suggested high usage cutoffs of at least 84%.[16,18,19] Given the large variability in FVMC by children with LI in those studies (i.e., SDs $\geq$ 18%), there will be some children who score in the clinical range for verb tense marking (i.e., who fall below the cutoff) who nonetheless show sufficiently high usage of verb tense morphemes. Verb tense usage would thus not be prioritized as a therapy goal for these children.

PERCENT GRAMMATICAL UTTERANCES

PGU is a more comprehensive measure of grammatical accuracy than FVMC. The motivation for this measure comes from studies that reported errors by children with LI on aspects of grammar other than verb tense marking. These include argument omissions,[20,21] production of fragments when a complete sentence is required,[22] pronominal form error,[23,24] and omissions and incorrect uses of other grammatical morphemes.[25,26]

Another motivation for this measure is the persisting higher rates of grammatical errors produced by school-aged children with LI.[27,28]

Computing Percent Grammatical Utterances

There are several steps for computing PGU. Step 1 involves identifying the analysis set, that is, the set of utterances to be included in the PGU analysis. Step 2 involves coding the utterances that are ungrammatical. PGU is computed by first subtracting the number of ungrammatical utterances from the total number of utterances that are included for the PGU analysis to obtain the number of grammatical utterances. The number of grammatical utterances is then divided by the total number of utterances that are included for the PGU analysis and multiplied by 100 to obtain a percentage. (See Appendix 2 for an example.)

UTTERANCE INCLUSION CRITERIA

Only utterances that require both a subject and a verb are included in the PGU analysis. Utterances with pragmatically omitted subjects (e.g., utterance 1 in Table 1) are excluded. However, utterances for which subject omission would be ungrammatical (e.g., utterances 2 and 3 in Table 1) are included.

CODING FOR GRAMMATICALITY

Computation of PGU involves judgments about utterance grammaticality. Any utterance that has even one grammatical error is coded as ungrammatical. Although this judgment can be made holistically, it can be helpful to first code for specific types of errors. Eisenberg and colleagues coded for the following error types: (1) production of fragments when full clauses are obligated; (2) argument errors; (3) pronoun errors; (4) verb form errors; (5) errors on other grammatical morphemes; and (6) other errors.[22,29] Examples of these error types are provided in Table 3.

Diagnostic Accuracy of Grammaticality Measures

Eisenberg and Guo investigated the diagnostic accuracy of PGU for 3-year-olds based on a picture description sample.[22] All of the children in the LI group were receiving therapy at the time of the study or were subsequently enrolled in therapy. Mean PGU for the LI children was lower than the mean for children with TL (mean = 38%, SD = 12% versus mean = 72%, SD = 12%). Specificity was acceptable (88%) and sensitivity was good (100%) at a cutoff score of 58%.

Guo and Schneider further investigated the diagnostic accuracy of PGU for 6- and 8-year-olds based on narrative samples using picture sequences.[18] Mean PGU for the LI group was lower than the mean for the TL group (6-year-olds: mean = 64%, SD = 19% versus mean = 91%, SD = 7%; 8-year-olds: mean = 78%, SD = 15% versus mean = 95%, SD = 3%). For 6-year-olds, specificity was good (90%) and sensitivity was acceptable (82%) at cutoff score of 83%. For the 8-year-olds, both specificity and sensitivity were acceptable (84 and 88%) at a cutoff score of 91%.

There is also diagnostic accuracy information for another measure of grammaticality, percent sentence point (PSP) based on *Developmental Sentence Scoring*.[30] This measure is similar to PGU but includes only utterances that have both a subject and a main verb. That means that ungrammatical sentences with an omitted subject and/or an omitted main verb (usually the copula; e.g., he Ø at home) are excluded from the PSP analysis in step 1 before utterances are judged for grammaticality in step 2. PSP scores, consequently, will be somewhat higher than PGU. In addition to PGU, Eisenberg and Guo also examined the diagnostic accuracy of PSP.[22] Mean PSP was 75% (SD = 11%) for children with TL and 45% (SD = 11%) for children with LI (compared with 72 and 38% for PGU).[18] At a cutoff of 68%, specificity for PSP was acceptable (82%), which was slightly lower than the specificity for PGU. Sensitivity for PSP was good (100%), which was equal to that for PGU.

Souto et al looked at diagnostic accuracy for PSP in 4- and 5-year-olds based on conversational samples.[14] Mean PSP for 4-year-olds was 57 to 60% (SD = 14 to 24%) in the LI group versus 91 to 92% (SD = 4 to 8%) in the TL group. Mean PSP for 5-year-olds was 59 to 70% (SD = 9 to 18%) in the LI group versus 92 to 93% (SD = 3 to 5%) in the TL group. Both specificity and sensitivity

Table 3 Error Types

Error Type	Error Subtype	Examples*
Fragments		E: What is happening in the picture?
		C: A boy and two girls.
Argument errors	Subject omission	_ coming home from school.
	Object omission	The boy broke _.
	Indirect object omission	They're giving _ a donut.
	Locative phrase omission	She's putting the boy _.
	Prepositional object omission	Him knock over _.
Pronoun errors	Subject case error	*Her* has dad's hat on.
	Possessive case error	A mom lost *she* seatbelt.
	Reflexive case error	He can't get down by *heself*.
	Gender error	He got wet on her shoe (referring to same girl).
Verb form errors	Auxiliary *be* omission	The mom _ pouring juice.
	Copula *be* omission	The cat _ stuck in the tree.
	Modal *do* error	He *don't* get hurt.
	Bare verb	The dog *splash* the water.
	Overgeneralization	He *taked* it out.
	Overuse (agreement) error	The boys *wants* get the kitty down.
Morpheme errors	Plural omission or overgeneralization	The dad is raking up the leaf_.
	Article omission or overuse	The *firemans* are coming out.
	Participle omission or error	_ cat is on _ tree. A water is coming out.
	Preposition omission or error	Papa is come_ to get her doll.
	Infinitive marker omission	They're playing _ the leaf pile. The boy fall on the stool.
	Conjunction omission or error	He wants _ get kitty down.
		The mom is making breakfast _ put some juice in the cup.
Other errors	Lexical error	They're *brooming*.
	Word order error	She's washing the water with her hands.

*Underscore (_) represents an omission; italic represents an incorrect usage.

were good (93 to 100%) at each age level. The cutoff score was not provided.

In sum, PGU showed at least acceptable diagnostic accuracy for children aged 3 to 8 years both for picture description samples and for narrative samples. An alternative measure, PSP, also showed at least acceptable diagnostic accuracy for children aged 3 to 5 years for picture description and conversational samples (see Table 4 for a summary across studies). Whether PSP would show at least acceptable diagnostic accuracy beyond age 5 has yet to be determined.

Reliability of Percent Grammatical Utterances

We know of only one study that looked at reliability of PGU as a function of sample size. Eisenberg and Guo compared PGU for longer and shorter samples for 3-year-old children based on picture description samples.[31] Sample size was based on the number of pictures used to elicit the sample. PGU was not affected by sample size. Mean PGU was 57% for both the longer samples (15 pictures, ~70 utterances) and shorter samples (7 pictures, ~30 utterances), suggesting that PGU can be calculated using smaller samples. However, sample size affected the consistency of clinical decision. Using a 58% cutoff, overall agreement between longer and shorter samples for clinical decisions (i.e., making pass and fail decisions based on scores falling above or below the cutoff) was good for pass decisions (90%) but only fair for fail decisions (77%).

Table 4 Specificity, Sensitivity, and Cutoff Scores for PGU and PSP

Measure/Study	Age	Sample Type	Cutoff (%)	Specificity (%)	Sensitivity (%)
PGU					
Eisenberg and Guo (2013)[22]	3;0–3;11	Picture description	58	88	100
Guo and Schneider (in press)[18]	6;0–6;11	Narrative	83	90	82
Guo and Schneider (in press)[18]	8;0–8;11	Narrative	91	84	88
PSP					
Guo and Eisenberg (2013)[22]	3;0–3;11	Picture description	67	82	100
Souto et al (2014)[14]	4;0–4;6	Conversation	N/A	94–100	93–100
Souto et al (2014)[14]	5;0–5;6	Conversation	N/A	100	100

Abbreviations: N/A, not applicable; PGU, percent grammatical utterances ; PSP, percent sentence point.
*Year;month.

We found no studies that compared PGU across sample types. We cannot make comparisons across studies because different sample types were used with children at different ages. That is, in the extant literature, PGU was based on picture description for preschool children and narration for school-aged children.[18,22,29] We know of no studies that looked at temporal stability for PGU.

Interpreting Percent Grammatical Utterances

Across studies, PGU and the suggested cutoff scores for PGU varied as a function of both age and sample type. This means that different PGU cutoffs need to be used for older and younger children and for different sample types.

In Eisenberg and Guo,[31] all of the disagreements for pass/fail decisions were for children who scored between 42 and 62% for PGU. This suggests that we should be cautious in applying a single point cutoff score to make binary decisions. We use confidence intervals for interpreting norm-referenced standardized tests to take into account measurement error.[32] Similarly, we want to construct score intervals for interpreting criterion referenced assessment, below which we can be certain about failing decisions and above which we can be certain about passing decisions.

USING LANGUAGE SAMPLE ANALYSIS FOR OTHER PURPOSES

In our review, we have focused on use of FVMC and PGU for the purpose of identifying children with LI. LSA measures are often suggested for two other purposes—goal selection and measuring treatment outcome.[2,3,28] However, establishing the validity of these measures in identifying LI does not establish their validity for these two purposes. Rather, validity needs to be separately evaluated for each purpose.[6]

Goal Selection

Children with LI show lower usage of verb tense morphemes and lower grammaticality than do children with TL. These would each, therefore, seem to be reasonable goal areas for language intervention. However, this does not mean that the intervention goal would be an increase in FVMC or PGU. FVMC is only one possible measure of verb tense morpheme usage. Similarly, PGU is only one possible measure of grammaticality. To state these as therapy goals would be to confuse a linguistic property or construct with the means for measuring that property.[33]

A low score on FVMC might suggest that further assessment of verb tense markers is warranted. This assessment would involve activities that would provide contexts for a variety of verb tense morphemes. Note that this follow-up assessment could include verb tense morphemes that were not included in the FVMC analysis, such as auxiliary *do* or irregular past tense.

A low score on PGU could be followed up with activities to evaluate whether a child produces various types of grammatical errors, such as those listed in Table 3. Recall that we suggest coding for error type as part of the PGU analysis because this makes it easier to judge utterances for grammaticality. This should not, however, constrain potential goals to just those errors produced during the PGU analysis. Follow-up assessment should consider error types that are not manifest within the sample used for PGU as well as providing additional information about error types that are demonstrated on this sample.

Measuring Treatment Outcome

When assessing performance on treatment goals, we want to know that any change is real and important.[34] To establish that change is important, we need to go beyond measures of performance on the specific therapy targets and use general language performance measures that reflect a wider range of linguistic features.[28] Measures such as FVMC and PGU are useful for this. However, we have to choose a general language performance measure that is relevant to the treatment goals. For instance, FVMC might not be the best choice of outcome measure if therapy focused on modal *do* or irregular verb use, two aspects of verb tense usage that are not included in this measure.

To establish that change is real requires measures that are reliable. However, as noted previously, there are no studies that have looked at the temporal stability of FVMC or PGU. Without such studies, we cannot know whether increases in these measures over time reflect real changes in knowledge and use of grammatical features rather than random fluctuations in performance. Without this evidence and in light of the large SDs reported for children with LI, clinicians should be cautious in interpreting small differences in FVMC and PGU.

CLINICAL IMPLICATIONS

Speech-language pathologists assess language for a variety of purposes. The common clinical assumption is that decisions about the presence or absence of an LI must be based on norm-referenced standardized tests and that LSA should be used only for goal selection and measuring outcome. We question this assumption. Evidence suggests that LSA measures—specifically quantitative measures of grammatical accuracy such as FVMC and PGU—are useful for identifying LI in preschool and early school-aged children. We caution against interpreting quantitative measures as therapy goals in themselves. Rather, low performance on these measures should be followed up with descriptive analyses and probes to determine therapy goals. Finally, although we believe it is important that clinicians go beyond measuring performance on individual therapy goals, we also urge caution in interpreting these measures when assessing treatment progress.

REFERENCES

1. Huang R, Hopkins J, Nippold MA. Satisfaction with standardized language testing: a survey of speech-language pathologists. Lang Speech Hear Serv Sch 1997;28:12–29
2. McCauley RJ. Assessment of Language Disorders in Children. Mahwah, NJ: Erlbaum; 2001
3. Paul R, Norbury CF. Language Disorders from Infancy through Adolescence, 4th ed. St. Louis, MO: Elsevier; 2012
4. Aram DM, Morris R, Hall NE. Clinical and research congruence in identifying children with specific language impairment. J Speech Hear Res 1993;36(3):580–591
5. Dunn M, Flax J, Sliwinski M, Aram D. The use of spontaneous language measures as criteria for identifying children with specific language impairment: an attempt to reconcile clinical and research incongruence. J Speech Hear Res 1996;39(3):643–654
6. Eisenberg SL, Fersko McGovern T, Lundgren C. The use of MLU for identifying language impairment in preschool children. Am J Speech Lang Pathol 2001;10:323–342
7. Plante E, Vance R. Selection of preschool language tests: a data-based approach. Lang Speech Hear Serv Sch 1994;25:15–24
8. Goffman L, Leonard J. Growth of language skills in preschool children with specific language impairment: implications for assessment and intervention. Am J Speech Lang Pathol 2000;9:151–161
9. Leonard LB. Children with Specific Language Impairment, 2nd ed. Cambridge, MA: The MIT Press; 2014
10. Poll GH, Betz SK, Miller CA. Identification of clinical markers of specific language impairment in adults. J Speech Lang Hear Res 2010;53(2):414–429
11. Rice ML, Wexler K, Hershberger S. Tense over time: the longitudinal course of tense acquisition in children with specific language impairment. J Speech Lang Hear Res 1998;41(6):1412–1431
12. Windsor J, Scott CM, Street CK. Verb and noun morphology in the spoken and written language of

children with language learning disabilities. J Speech Lang Hear Res 2000;43(6):1322–1336
13. Bedore LM, Leonard LB. Specific language impairment and grammatical morphology: a discriminant function analysis. J Speech Lang Hear Res 1998;41(5):1185–1192
14. Souto SM, Leonard LB, Deevy P. Identifying risk for specific language impairment with narrow and global measures of grammar. Clin Linguist Phon 2014;28(10):741–756
15. Biber D, Johansson S, Leech G, Conrad S, Finegan E. Longman Grammar of Spoken and Written English. Essex, UK: Pearson Education Limited; 1999
16. Gladfelter A, Leonard LB. Alternative tense and agreement morpheme measures for assessing grammatical deficits during the preschool period. J Speech Lang Hear Res 2013;56(2):542–552
17. Moyle MJ, Karasinski C, Ellis Weismer S, Gorman BK. Grammatical morphology in school-age children with and without language impairment: a discriminant function analysis. Lang Speech Hear Serv Sch 2011;42(4):550–560
18. Guo LY, Schneider P. Differentiating school-aged children with and without language impairment using tense and grammaticality measures from a narrative task. J Speech Lang Hear Res, In press
19. Guo LY, Eisenberg S. The diagnostic accuracy of two tense measures for identifying 3-year-olds with language impairment. Am J Speech Lang Pathol 2014;23(2):203–212
20. Ebbels SH, van der Lely HK, Dockrell JE. Intervention for verb argument structure in children with persistent SLI: a randomized control trial. J Speech Lang Hear Res 2007;50(5):1330–1349
21. Grela B, Leonard L. The use of subject arguments by children with specific language impairment. Clin Linguist Phon 1997;11(6):443–453
22. Eisenberg SL, Guo LY. Differentiating children with and without language impairment based on grammaticality. Lang Speech Hear Serv Sch 2013;44(1):20–31
23. Loeb DF, Leonard LB. Subject case marking and verb morphology in normally developing and specifically language-impaired children. J Speech Hear Res 1991;34(2):340–346
24. Moore ME. Third person pronoun errors by children with and without language impairment. J Commun Disord 2001;34(3):207–228
25. Leonard LB, Eyer JA, Bedore LM, Grela BG. Three accounts of the grammatical morpheme difficulties of English-speaking children with specific language impairment. J Speech Lang Hear Res 1997;40(4):741–753
26. Watkins RV, Rice ML. Verb particle and preposition acquisition in language-impaired preschoolers. J Speech Hear Res 1991;34(5):1130–1141
27. Fey ME, Catts HW, Proctor-Williams K, Tomblin JB, Zhang X. Oral and written story composition skills of children with language impairment. J Speech Lang Hear Res 2004; 47(6):1301–1318
28. Scott CM, Windsor J. General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. J Speech Lang Hear Res 2000;43(2):324–339
29. Eisenberg SL, Guo LY, Germezia M. How grammatical are 3-year-olds? Lang Speech Hear Serv Sch 2012;43(1):36–52
30. Lee LL. Developmental Sentence Analysis. Evanston, IL: Northwestern University Press; 1973
31. Eisenberg SL, Guo LY. Sample size for measuring grammaticality in preschool children from picture-elicited language samples. Lang Speech Hear Serv Sch 2015;46(2):81–93
32. Hutchinson TA. What to look for in the technical manual: twenty questions for users. Lang Speech Hear Serv Sch 1996;27:109–121
33. Sabers DL. By their tests we will know them. Lang Speech Hear Serv Sch 1996;27:102–108
34. Bain BA, Dollaghan CA. Treatment efficacy: the notion of clinically significant change. Lang Speech Hear Serv Sch 1991;22:264–270

Appendix 1 Sample FVMC Analysis

Utterance	Obligatory context	Accurate	Not accurate
1. This _ for dinner.	Yes		Omitted copula
2. You knock _ the food down.	Yes		Omitted past tense
3. It's still hot.	Yes	Copula *'s*	
4. Where _ I put this?	No (auxiliary *do* not included in FVMC)		
5. And this is the crib.	Yes	Copula *is*	
6. I stays in here cause	No (overuse third-person singular)	Copula *is*	
this is the crib room.	Yes		
7. And this _ the highchair.	Yes		Omitted copula
8. They have to watch TV.	No (plural subject does not obligate third-person *-s*)		
9. The baby sits in her chair.	Yes	Third-person singular *-s*	
10. And brother sit _ here.	Yes		Omitted 3rd person
11. The sister need/s a horse.	Yes	Third-person singular *-s*	
12. Now her _ riding.	Yes		Omitted auxiliary

Abbreviation: FVMC, finite verb morphology composite.
Note: Number of obligatory contexts for verb tense morphemes included on FVMC: 10, FVMC = 50%.

Appendix 2 Sample PGU Analysis

Utterance	Included	Grammatical	Error Type
1. The mom is so angry.	Yes	Yes	
2. He go under the couch.	Yes	No	Verb form
3. Make footprints.*	Yes	No	Argument (subject omission)
4. Her give a donut to him.	Yes	No	Pronoun, verb form
5. The mom driving.	Yes	No	Verb form
6. The mom ask the sister "no more donuts."	Yes	No	Other (*ask* instead of *tell*); verb form
7. The mom washing dishes.	Yes	No	Verb form
8. Him getting new cookie.	Yes	No	Pronoun, verb form, morpheme (article)
9. The water's gonna spill on the floor.	Yes	Yes	
10. Being quiet.	Yes	No	Argument (subject omission)

Abbreviation: PGU, percent grammatical utterances.
*In response to examiner prompt, "The dog ate some of the cake and what's going to happen next?"
Note: Number of utterances included on PGU: 10, PGU = 50%.