Top Lang Disorders
Vol. 40, No. 2, pp. 135–148

Using General Language Performance Measures to Assess Grammar Learning

Sarita L. Eisenberg

Grammar is an important goal area for children with language impairment. The current article considers the use of General Language Performance Measures (GLPMs) to assess outcomes for this basic goal area and for 3 intermediate grammar goals that contribute to children's developing ability to construct increasingly longer and more complex sentences: production of word combinations, production of sentences with required and optional constituents, and production of complex sentences. These goals are very important because they address a significant deficit area for children with language impairment and enable these children to increase the informativeness of their utterances. Using GLPMs in addition to session data and probes enables speech–language pathologists to determine whether these forms are being incorporated into conversational speech. **Key words:** *child language*, *language sample analysis*, *language therapy*, *outcome measurement*, *syntax*

OUTCOME MEASUREMENT is an essential component of clinical practice when working with children with language impairments. The collection of such data enables speech–language pathologists (SLPs) to determine whether or not children are showing sufficient progress as a result of language intervention and to make changes when children are not adequately responding to intervention. Outcome measurement also enables SLPs to make decisions about dismissing children from intervention services.

In a recent survey by Finestack and Sutterland (2018), only 30% of SLPs in early intervention (i.e., serving children aged 0–5 years) and only 21% of SLPs serving elementary school-aged children reported frequently using language sample analysis (LSA) to measure therapy outcomes. This contrasted with larger numbers of SLPs who reported frequently using observation (83% of SLPs in early intervention settings; 66% of SLPs in elementary schools), informal language sampling (64% of SLPs in early intervention settings; 53% of SLPs in elementary schools), and probes (57% of SLPs in early intervention settings; 64% of SLPs in elementary schools) to measure therapy outcomes[1]. A review of therapeutic intervention studies shows a similarly limited use of LSA for measuring treatment outcomes. For instance, of the studies

Author Affiliation: *Montclair State University, Montclair, New Jersey.*

The author has indicated that she has no financial and no nonfinancial relationships to disclose.

Corresponding Author: *Sarita L. Eisenberg, PhD, Department of Communication Sciences and Disorders, Montclair State University, 1515 Broad St, Bloomfield, NJ 07003 (eisenbergs@montclair.edu).*

DOI: 10.1097/TLD.0000000000000214

[1]For both *language sample analysis* and *informal language sampling*, the SLP elicits conversational or narrative language samples. For language sample analysis, the clinician transcribes and systematically analyzes the language sample, whereas informal language sampling is more impressionistic, with the SLP noting observations about the child's language usage. This is similar to *observation* except that, for observation, someone other than the SLP interacts with the child. *Probes* involve a more structured assessment in which the SLP provides discrete opportunities for the child to attempt the targeted grammatical form (Finestack, personal communication, October 2019).

included in a review of language therapy by Ebbels (2014), only 18% of the studies used LSA to measure outcomes, compared with the use of probes in 88% of the studies.

The limited use of LSA to measure outcomes is unfortunate because LSA offers a means to assess broader learning and language growth beyond performance on therapy goals during treatment session activities. In this article, LSA will be considered for this purpose. More specifically, I will explore use of LSA for assessing achievements in grammar at the emerging and developing language stages. These stages span the ages between 18 months and 5 years for typical children and extend into the early school years for children with language impairment. During these stages, children first learn to combine words and then to combine phrasal constituents and clauses to build increasingly complex sentences. Children with language impairment show a late onset for these combinations and continue to produce less elaborated sentences than their peers with typical language throughout the school years (King & Fletcher, 1993; Leonard, 2014).

INTERVENTION GOALS

Therapy goals can be conceptualized within a hierarchical organization. McCauley and Fey (2006), for instance, suggested a hierarchy of four levels. The most general level is the *basic goal* area, which encompasses the domain and modality that is the focus of therapy (e.g., expressive spoken grammar). *Intermediate goals* reflect categories of related grammatical forms within a basic goal area (e.g., complex sentences). *Specific goals* include specific grammatical forms or exemplars within an intermediate goal category (e.g., propositional complements such as *I don't know what that is*). *Subgoals* involve a limited set of exemplars (e.g., propositional complements with five main verbs—*know*, *think*, *pretend*, *tell*, and *ask*) and specify the conditions during which those exemplars will be presented and practiced (e.g., production during play when provided with a high density of models). The procedures for measuring outcomes should account for each of these levels.

Performance on subgoals can be measured with session data and probes. Speech–language pathologists also can use probes to measure responses to changes in the conditions for subgoal performance, such as reduction in scaffolding or increases in utterance complexity. At the specific goal level, SLPs can use probes to check for generalization of targeted grammatical forms to untrained contexts—with different vocabulary, in different sentence positions and different sentence types, and for different pragmatic functions. Probes also can assess achievement of intermediate level goals by assessing usage of untrained category members. Such probes are commonly characterized as measuring *response generalization*. Probes, however, do not measure performance within communicative interactions (commonly characterized as measuring *stimulus generalization*). This requires use of LSA.

There are several types of LSA that can be used for measuring conversational usage. One type of LSA involves a fine-grained structural analysis of usage and errors on individual grammatical forms (e.g., percent usage of subject noun phrases). This fine-grained LSA can measure attainment of specific goals (i.e., the exemplars targeted in therapy) or of intermediate goals (i.e., by examining use of untrained category members). There are also broader structural analyses, such as Developmental Sentence Scoring (Lee, 1974) and the Index of Productive Syntax (Altenberg, Roberts, & Scarborough, 2018; Scarborough, 1990), that award points based on usage and/or errors on a variety of grammatical features from different intermediate goal areas. These broad LSA protocols can be useful for measuring outcomes at the basic goal level as well.

Another type of LSA, and the focus of this article, is General Language Performance Measures (GLPMs). General Language Performance Measures characterize an entire language sample with a single score that represents productivity, fluency, vocabulary, utterance length, grammatical complexity, or grammatical accuracy (Scott & Windsor, 2000). Some GLPMs are limited in scope to a single category of related features and are appropriate for measuring outcomes on the corresponding intermediate goal area. Other GLPMs that are broader in scope, including features from different categories, are appropriate for measuring broader basic goal–level outcomes.

In this article, I will focus on the basic goal area of expressive grammar within spoken discourse. Within that basic goal area, I will focus on three intermediate goals for children at the emerging and developing language levels that are critical for sentence construction: combining words, combining phrasal constituents to form sentences, and combining clauses to produce complex sentences. I will propose GLPMs that can be used to measure outcomes for each of these intermediate goal areas.

GENERAL CONSIDERATIONS FOR USING GLPMS

General Language Performance Measures should be computed before the start of therapy to determine the baseline level and then periodically throughout intervention to measure progress. A minimum of three data points is recommended to establish that the baseline level is stable (i.e., varies by no more than 10%; Bain & Dollaghan, 1991; McCauley, 2001). If preintervention performance fluctuates more than 10%, then larger changes in performance would be needed to conclude that any change was due to therapeutic intervention.

To allow for comparison of performance over time, it is important that the same sample type (e.g., conversation, narration), sample size (i.e., number of utterances, duration), transcription conventions (e.g., segmentation into morphological units or C-units [clausal units]), utterance inclusion criteria, and implementation guidelines be followed. Note that because the GLPMs are not being used to compare a child's score to normative data, these variables can be individualized for each child as long as there is consistency for that particular child across sampling times.

Identifying the analysis set for computing GLPMs

Before calculating a GLPM, the *analysis set* must be selected. This is the set of utterances that will be used for the analysis. Only utterances that are complete and that can be fully transcribed are included (Brown, 1973). To that end, utterances are excluded if they were interrupted or abandoned before being completed or if they contain unintelligible segments. Studies have documented that LSA measures are influenced by discourse factors (Johnston, 2001; Oosthuizen & Southwood, 2009) that can confound the ability of those measures to determine change in performance over time as a result of therapy. To reduce this possible confound, the following types of utterances also should be excluded: exact self-repetitions (Johnston, 2001; Miller, 1981), exact repetitions of the adult's immediately preceding utterance (Johnston, 2001; Lund & Duchan, 1993), routines such as reciting nursery rhymes or songs (Lund & Duchan, 1993; Miller, 1981), enumerations such as counting, reciting the alphabet, or successive labeling of objects (Lund & Duchan, 1993; Miller, 1981), utterances with long strings of conjoined words or phrases (Lee, 1974; Miller & Chapman, 1981), and single-word yes/no acknowledgments and responses to questions (Johnston, 2001; Lund & Duchan, 1993). Within each utterance, mazes (e.g., interjections, nonword fillers, false starts, repetitions) are separated from the main body of the utterance (Brown, 1973) and excluded from the analysis. Some GLPMs may have other utterance inclusion criteria and these will be noted later in the description of those particular GLPMs.

Types of GLPMs

There are several types of GLPM scores: (1) *Frequency counts* are a tally of the total number of occurrences of a grammatical form in the analysis set. To allow comparison between samples over time, sample length must be standardized on the basis of either some number of utterances or some amount of time. (2) *Mean usage rate* for a grammatical form is calculated by summing the frequency count (i.e., total number of occurrences of the form) and dividing that total by the total number of utterances in the analysis set. (3) *Percent usage* is calculated by multiplying the mean usage rate by 100%. Percent usage adjusts for sample length so that samples of different lengths can be compared. In Table 1, each GLPM is classified by type and the intermediate goal area measured by that GLPM is indicated.

MLU AS A BASIC GOAL GLPM

In the survey by Finestack and Sutterland (2018), mean length of utterance (MLU) was the most frequently used LSA measure by SLPs who used LSA to measure outcomes, with use reported by 94% of SLPs in early intervention settings and 86% of SLPs in elementary schools. This was followed by type-token ratio (a measure of vocabulary diversity) used, respectively, by 25% and 33% of respondents. No other LSA measures were used by more than 15% of respondents. Therefore, I will first discuss MLU before considering other GLPMs.

It is important to distinguish between outcome measurement and therapy goals. It is not uncommon to see "increasing MLU" stated as an intervention goal (see, for instance, Loeb & Armstrong, 2001; Yoder, Spruytenberg, Edwards, & Davies, 1995). However, this confounds *goal*—a skill to be learned—with *outcome measurement*—the means for measuring acquisition of that skill (Sabers, 1996). Note that an increase in utterance length does not indicate what specific aspects of grammar were targeted to achieve that increase. Rather, changes in utterance length can result from intervening with any one of a number of different aspects of grammar, such as adding content words, sentence constituents, clauses, or grammatical

Table 1. General Language Performance Measures

GLPM	Acronym	Goal Area to Be Measured	Type of GLPM
Mean Length of Utterance in Words	MLUw	Word combinations, sentence constituents, complex sentences	Mean usage rate
Number of Word Combinations	NWC	Word combinations	Frequency count
Percent Word Combinations	PWC	Word combinations	Percent usage
Unique Syntactic Types	UST	Word combinations	Frequency count
Percent Verb Combinations	PVT	Word combinations with verbs	Percent usage
Mean Sentence Constituents per Utterance	MSCU	Sentence constituents	Mean usage rate
Mean Arguments per Utterance	MAU	Sentence constituents	Mean usage rate
Percent Complex Sentences	PCS	Complex sentences	Percent usage
Mean Clauses per Utterance	MCU	Complex sentences	Mean usage rate
Mean Finite Clauses per Utterance	MFCU	Complex sentences	Mean usage rate

morphemes. Such aspects of grammar are addressed in therapy (i.e., the goals of therapy), whereas change in utterance length is an outcome of attaining those goals.

Mean length of utterance is a measure of the mean usage rate in a sample of either morphemes or words (Rice et al., 2010). In clinical practice by SLPs, MLU is more commonly measured in morphemes (i.e., mean length of utterance in morphemes, MLUm). MLUm is computed by counting all of the morphemes (i.e., word roots, bound inflectional morphemes, and free grammatical morphemes) in each utterance, summing the total morphemes in all of the utterances, and then dividing total morphemes by the total number of utterances. In contrast, mean length of utterance in words (MLUw) is computed by counting only the words (i.e., content words and free grammatical morphemes) in each utterance, summing the total words in all of the utterances, and then dividing total words by the total number of utterances.

Mean length of utterance can be used as a basic goal-level assessment measure because the addition of any constituent type—morphemes, words, phrases, and clauses—will affect utterance length, However, it is important to recognize that MLUw and MLUm reflect different aspects of language learning, with MLUw providing information about content and MLUm providing information about grammatical morphology (Wieczorek, 2010). As goals for word combinations, sentence constituents, and complex sentences all add content to the child's utterances, the appropriate utterance length measure when targeting those forms is one that measures utterance length in words, MLUw.

When transcribing samples for calculating MLUw, decisions must be made about what constitutes one or more than one word. Following Brown (1973), the following are counted as one word: compounds (e.g., ice_cream), ritualized reduplications (e.g., night_night), proper names (Mr_Smith), catenatives (e.g., wanna), and contracted forms (e.g., isn't, he'll). Selection of the analysis set follows the general guidelines for GLPMs. To make MLUw more sensitive to small changes for more advanced children, one could exclude single-word utterances once MLUw reaches three words or more.

The suggested sample size recommended for calculating MLU in clinical practice ranges from 50 utterances (Miller & Chapman, 1981) to 100 utterances (Guo & Eisenberg, 2015; Leadholm & Miller, 1992). However, studies investigating reliability of MLU as a function of sample length suggest that samples of at least 80–90 utterances are necessary to achieve adequate reliability. Guo and Eisenberg (2015) reported acceptable internal consistency reliability for MLUm for shorter samples—0.88 for 7-min samples (mean of 63 and range of 32–107 utterances) and 0.93 for 10-min samples (mean of 91 and range of 40–152 utterances). In a study by Gavin and Giles (1996), however, temporal reliability of MLUm was only 0.61 for samples of 50 utterances and 0.82 for samples of 100 utterances, not reaching at least 0.90 until the sample size was 175 utterances. Note, however, that the suggested sample sizes here are from studies focused only on MLUm. We lack comparable information for MLUw.

GLPMS TO MEASURE INTERMEDIATE GOAL OUTCOMES

The GLPMs in this section can each be used for measuring particular intermediate goal areas. Because of their narrower scope, they are likely to be more sensitive to changes in achievement of those intermediate goals than MLUw. For each intermediate goal area, I first briefly summarize the difficulty often shown by children with language impairment and how that goal area could be addressed in therapy. I then discuss one or more GLPMs to measure outcomes for that intermediate goal area. Whenever possible, I have provided a reference for each of the suggested GLPMs from a study investigating language development, language deficits in children with impairment, and/or therapy efficacy. However, I did not find a reference for some of the suggested GLPMs. Note also that the labels that I use for the GLPMs were designed to reflect the type of GLPM—frequency counts, mean usage rate, or percent usage—as well as the goal area, and may not be the same terms used in the cited references.

Measuring production of word combinations

Children with language impairment often show a delay in achieving longer and more varied word combinations (Trauner, Wulfeck, Tallal, & Hesselink, 1995, as cited in Leonard, 2014). Intervention studies for this goal area have targeted specific word combinations (e.g., Long, Olswang, Brian, & Dale, 1997; Scherer & Olswang, 1989) or provided general language stimulation by training parents to model a variety of word combinations (e.g., Girolametto, Pearce, & Weitzman, 1996). Suggested specific goals for word combinations are listed in Table 2.

Girolametto et al. (1996) used a frequency count measure, Number of Word Combinations (NWC), to measure outcomes in their treatment study. Number of Word Combinations was calculated by adding up all utterances longer than one word produced in 20 min of language sampling. The authors reported a significant difference in NWC after 4 months between children who had and had not received the intervention. Number of Word Combinations can be converted to a percent usage measure (Percent Word Combinations, PWC; Paul, Norbury, & Gosse, 2018) by dividing NWC by the total number of utterances in the sample and then multiplying by 100.

A limitation of both NWC and PWC is that they treat all utterances of more than one word as the same. That is, they do not distinguish between utterances that include nonsyntactic elements (e.g., *Mommy* in the utterance *Mommy, doggie* to get the mother's attention; *no* in the utterance *No, cookies* in response to the question "Do you want cake?") and true syntactic combinations such as *my mommy* or *no cookies* (said in response to seeing an empty cookie package). They also do not take into account multiple instances of the same word combination. Thus, both NCW and PCW could potentially overestimate a child's ability to combine words.

Hadley (1999) proposed Unique Syntactic Types (UST) as an alternative means to measure changes in word combining that countered these limitations. An important difference from NWC is that only word combinations composed of syntactic elements are counted. Unique Syntactic Types distinguishes between syntactic words (i.e., nouns,

Table 2. Specific goals for word combinations

Phrasal constructions	• Demonstrative + entity (e.g., *that ball*) • Recurrence + entity (e.g., *more juice*; *another cookie*) • Negative + entity (e.g., *no cookies*) • Possessor + entity (e.g., *my dolly*) • Attribute + entity (e.g., *big doggy*)
Action (verb) constructions	• Agent + action (e.g., *baby sleeping*) • Action + patient (e.g., *push car*) • Action + recipient (e.g., *give mommy*) • Action + preposition/location (e.g., *put in*)
Locative constructions	• Entity + preposition (e.g., *ball in*) • Preposition + location (e.g., *in car*) • Action + preposition/location (e.g., *put in*)
State constructions	• State + entity (e.g., *want cookie*)

verbs, adjectives, pronouns, prepositions, and other free grammatical morphemes) and nonsyntactic elements that attach to the start or end of utterances (e.g., *hi*, *uh oh*, and *please*). A second difference from NWC is that UST distinguishes between types (unique instances of a word combination) and tokens. To be credited as unique, a word combination must involve a combination of different words (e.g., *Mommy go* and *Doggie go*) or morphological variations of one of the words (e.g., *Doggie go* and *Doggie going*). Hadley reported high interexaminer agreement (98%) for differentiating between syntactic and nonsyntactic word combinations. Unique Syntactic Types showed moderate to high temporal stability reliability between scores obtained on different days, with the magnitude of the correlation increasing over time as UST increased (0.49 at Time 1; 0.79 at Time 2; 0.87 at Time 3). Unique Syntactic Types showed a significant increase after therapy for each of three 3-month time intervals.

The UST analysis is based on a fairly brief sample of 12 min of conversation. It can be calculated cumulatively over more than 1 day and it can be implemented without transcribing the sample, by jotting down word combinations as they occur and subsequently eliminating duplications. This outcome measure can be used until the child is producing so many word combinations that these can no longer be written down in real time. When that occurs, it may be time to move on to another more advanced intermediate grammar goal.

First, however, it may be necessary to target word combinations with verbs. Children typically first produce phrasal utterances (i.e., word combinations without verbs) before starting to produce combinations with verbs (Brown, 1973). Combinations with verbs are particularly important as verbs are the foundation for building sentences. I propose an alternative measure that would specifically capture progress in achieving word combinations with verbs, Percent Verb Combinations. The analysis set for this calculation would be limited to syntactic word combinations (i.e., utterances with two or more words that are syntactically linked). The percentage of verb combinations is calculated by summing the NWC that include verbs and dividing by the total number of syntactic word combinations produced.

Measuring production of sentence constituents

Children with language impairment produce a more limited range of sentence constituents and produce fewer sentence constituents per utterance (King & Fletcher, 1993), omit more required constituents (termed *verb arguments*; Grela & Leonard, 1997; Thordardottir & Ellis Weismer, 2002), and use fewer optional constituents (termed *adjuncts*; Ingham, Fletcher, Schletter, & Sinka, 1998; King, 2000). Therapeutic intervention studies for this goal area have focused on production of specific sentence constituents (Bolderson, Dosanjh, Milligan, Oring, & Chiat, 2011) or on production of sentence frames involving basic subject–verb–object English sentence structure (Loeb & Armstrong, 2001; Robertson & Ellis Weismer, 1999) and other combinations of constituents (Bolderson et al., 2011; Ebbels, van der Lely, & Dockrell, 2007; Spooner, 2002). Suggested specific goals for sentence frames and adjuncts are listed in Table 3.

One way to measure outcomes for this goal area is by calculating the mean usage rate of sentence constituents per utterance. Note that this is similar to MLU, but with length measured in number of sentence constituents rather than in words or morphemes. There are two ways to calculate mean usage rate of sentence constituents: one that includes all sentence constituents, required and optional (Mean Sentence Constituents per Utterance, MSCU) and another that counts only required constituents (Mean Arguments per Utterance, MAU; see Table 4). The analysis set for both measures is limited to single-clause utterances with verbs.

Mean Sentence Constituents per Utterance is computed by counting the number of sentence constituents in each utterance,

Table 3. Specific goals for sentence constituents

1. Sentence frames with action verbs
   a. Intransitive verb frame: agent/subject + action (e.g., *The boy is jumping*)
   b. Transitive verb frame: agent/subject + action + patient/direct object (e.g., *The girl ate breakfast)*
   c. Ditransitive verb frame: agent/subject + action + patient/direct object + recipient/indirect object (e.g., *The girl gave her mother a present; The girl gave a present to her mother*)
   d. Complex transitive verb frame: agent/subject + action + patient + location (e.g., *The boy put the book on the shelf*)
2. Sentence frames with intensive verbs
   a. Subject + verb + noun phrase complement (e.g., *The boy is a good student*)
   b. Subject + verb + prepositional phrase (locative) complement (e.g., *The book is on the shelf*)
   c. Subject + verb + adjective phrase (descriptive) complement (e.g., *The girl felt happy*)
3. Adjuncts (optional constituents that can be added to verb frames)
   a. Location (prepositional phrase or *here/there*) (e.g., *The boy is jumping on the bed*)
   b. Time (prepositional phrase or adverb) (e.g., *The girl ate breakfast before school)*

Table 4. An example of calculating Mean Arguments per Utterance (MAU) and Mean Sentence Constituents per Utterance (MSCU)

Constituents	Verb Type	No. of Arguments	No. of Constituents
(T) S N 1. [Now] [here]'s [some cookies].	Intensive	2	3
S 2. [I]'m gonna cook.	Transitive	1	1
S P 3. [We]'ll use [this].	Transitive	2	2
S P 4. [We] spilled [it].	Transitive	2	2
S P 5. [I] have [a spatula].	Transitive	2	2
S P 6. Should [we] cook [the bacon]?	Transitive	2	2
S A 7. [It]'s [okay].	Intensive	2	2
S P 8. [You] can use [this].	Transitive	2	2
S P (T) 9. [I] like [the cherry one] [sometimes].	Transitive	2	3
	Total	17	19

Note. Arguments: S = subject; P = patient; N = noun phrase complement; A = adjective phrase complement; Adjuncts: (T) = time. Calculations: MAU = No. of arguments ÷ No. of utterances = 17 ÷ 9 = 1.89; MSCU = No. of constituents ÷ No. of utterances = 19 ÷ 9 = 2.11.

summing the total number of sentence constituents in all of the utterances, and then dividing that total by the total number of single-clause utterances with a main verb. Mean Arguments per Utterance is computed by counting the number of verb arguments in each utterance, summing the total number of verb arguments in all of the utterances, and then dividing that total by the total number of single- clause utterances with a main verb. Mean Sentence Constituents per Utterance could be used when targeting either required constituents such as subjects, patients (also referred to as direct objects, e.g., *The girl ate* ***the apple***), and datives (also referred to as indirect objects, e.g., *The girl baked* ***her friend*** *a cake*) or optional constituents that express location and time. Mean Arguments per Utterance would only be appropriate when targeting required constituents. In addition, because MAU counts only required constituents, it might be more sensitive to increases in production of those constituents than MSCU.

Measuring production of complex sentences

Complex and coordinated sentences (hereafter grouped together as complex sentences) are sentences with more than one clause. This includes sentences with two independent clauses and sentences with a main clause and one or more dependent clauses. Dependent clauses include nonfinite clauses (i.e., clauses that include verb forms that do not show tense or agreement, such as infinitives, e.g., *I need* ***to pick up more animals***, and gerunds, e.g., *Keep* ***it moving***) and finite clauses (i.e., clauses that include a subject and a verb form that can be marked for tense and agreement such as propositional complements, e.g., *I think* ***she's gonna eat her dinner***, subordinate clauses, e.g., *You have to blow on them* ***cause they'll be really hot***, and relative clauses, e.g., *This is the person* ***that rides***). Suggested specific goals for complex sentences are listed in Table 5.

Children with language impairment show a late onset and slower rate of development for complex sentences (Schuele & Dykes, 2005) and produce fewer complex sentences than children with typical language skills, and their problems persist through the school years (Fletcher, 1991; Marinellie, 2004; Nippold, Mansfield, Billow, & Tomblin, 2009). Studies also have shown that these children display difficulty with specific types of complex sentences including infinitival complements (Eisenberg, 2003, 2004), propositional complements (Owens Van Horne & Lin, 2011; Steel, Rose & Eadie, 2016), and subordinate clauses (Marinellie, 2004). I found only one intervention study with preschool children that focused exclusively on complex sentences (Tyack, 1981). Other studies have included specific types of complex sentences as therapeutic goals for some children (e.g., Camarata & Nelson, 1992; Camarata, Nelson, & Camarata, 1994; Nelson, Camarata, Welsh, Butkovsky, & Camarata, 1996). Included in Table 5 is a list of complex sentence types that would be appropriate specific goals for young children.

Percent Complex Sentences (PCS; Tyack & Gottlesben, 1986) is computed by adding up all utterances with more than one clause, dividing this number by the total number of utterances in the sample, and then multiplying by 100. For this analysis, both finite and nonfinite clauses are counted. The analysis set for this measure is limited to utterances with at least one clause. Thus, all utterances that do not include at least one verb would be excluded. Percent Complex Sentences can be computed without transcribing the sample by tallying utterances with one clause and utterances with more than one clause.

Another type of measure is *clausal density*, the mean usage rate of the number of clauses per utterance. There are two ways to calculate clausal density, one that counts all clauses (i.e., finite and nonfinite, Mean Clauses per Utterance, MCU; Kemper, Rice, & Chen, 1995) and one that counts only finite clauses (Mean Finite Clauses per Utterance, MFCU; Scott & Windsor, 2000, also called the Subordination Index; see Table 6). Mean Clauses per Utterance is computed by counting

Table 5. Specific goals for complex sentences

1. Coordinate clauses with conjunctions *and*, *but*, *or*
   a. Full clause coordination (e.g., *You play with that one and I play with this*)
   b. Predicate coordination (e.g., *I went to the aquarium and saw the fish*)
2. Complement clauses functioning as the object of the main verb (also called nominal clauses)
   a. Infinitival complements
      1) Simple infinitives (e.g., *I try to do it*)
      2) Simple infinitives with verbs that require a patient noun phrase (e.g., *Mommy told me to clean my room*)
      3) Infinitives with a different subject (e.g., *I want Bill to have it*)
   b. Propositional complements
      1) With mental verbs (e.g., know, think) (e.g., *I guess she's sick*)
      2) With perception verbs (e.g., see, hear) (e.g., *I heard that the teacher is really mean*)
      3) With communication verbs (e.g., say, tell) (e.g., *My mother says I should eat breakfast*)
3. Adverbial clauses
   a. Subordinate clauses with temporal conjunctions (e.g., before, after, when) (e.g., *I went to the movies after I finished my homework*)
   b. Subordinate clauses with causal conjunctions (e.g., because) (e.g., *I want this doll because she's big*)
   c. Subordinate clauses with conditional conjunctions (e.g., if) (e.g., *It must be mine if it has sprinkles*)
4. Relative clauses (also called adjectival clauses)
   a. Object relative clauses (e.g., *I like the one he has*)
   b. Subject relative clauses (e.g., *The one I wanted is gone*)

the number of clauses and dividing by the total number of utterances with at least one clause. Mean Finite Clauses per Utterance is computed by counting the number of finite clauses and dividing by the total number of utterances with at least one clause. Note that single-clause utterances and main clauses are counted as finite clauses. As for PCS, the analysis set for these clausal density measures is limited to utterances with at least one clause, so that utterances without any verbs would be excluded. Kemper et al. (1995) reported 94% interrater agreement for identifying embedded clauses for the MCU analysis.

Although PCS could be used for any child who is working on complex sentences, it may be particularly useful when working on complex sentences with children who produce few or no complex sentences. Mean Clauses per Utterance could be used when targeting complex sentences with either nonfinite or finite clauses, whereas MFCU would be appropriate only when targeting complex sentences with finite clauses. In addition, because MFCU counts only finite clauses, it might be more sensitive to increases in production of finite clauses than MCU.

EVIDENCE FOR GLPMS

There are several psychometric properties that are desirable for quantitative LSA measures such as GLPMs that are used for measuring performance over time. These include evidence of both reliability and validity (Eisenberg, Fersko, & Lundgren, 2001; McCauley, 1996). To evaluate reliability, we would want to know about temporal stability, consistency, and interexaminer agreement. Temporal stability means that, in the absence of treatment, performance on the measure will not significantly change over a short period of time. However, data about temporal reliability are available only for UST (Hadley, 1999). Consistency means that small differences in the sampling procedures (e.g., different

Table 6. An example of calculating Mean Clauses per Utterance (MCU) and Mean Finite Clauses per Utterance (MFCU)

Clause Type	No. of Clauses	No. of Finite Clauses
M 1. [I can't open it].	1	1
M F 2. [I wonder] [what that is].	2	2
M N 3. [They have] [to fix it].	2	1
M N F 4. [We'll have] [to do it] [so the car can see].	3	2
M 5. [You love it].	1	1
M F N 6. [I think] [we have] [to get you aboard]	3	2
M F 7. [I know] [who can help].	2	2
M 8. [He can lift you up].	1	1
M N F 9. [You have] [to blow on them] [cause they'll be really hot].	3	2
Total	18	14

Note. Clause type: M = main clause; F = finite clause; N = nonfinite clause. Calculations: MCU = No. of clauses ÷ No. of utterances = 18 ÷ 9 = 2.00; MFCU = No. of finite clauses ÷ No. of utterances = 14 ÷ 9 = 1.56.

toys or pictures) will not significantly affect performance. There have been no studies that have investigated consistency for any of the suggested GLPMs. Interexaminer agreement means that independent examiners agree on the coding of utterances. This information is available only for UST (Hadley, 1999) and MCU (Kemper et al., 1995).

To evaluate the validity of using a GLPM to measure outcomes, we want evidence that the GLPM is sufficiently sensitive to show changes for individual children over time. Note that longitudinal studies reporting only group data are not sufficient. What is needed are studies that track growth trajectories for individual children and show whether there are fluctuations in the GLPM over time. This information would allow us to establish the best time interval for GLPM measurements and a criterion for the amount of change needed to demonstrate meaningful progress.

We also want evidence that the GLPM changes as a result of intervention. Of the suggested GLPMs, only NCW (Girolametto et al., 1996) and UST (Hadley, 1999) were actually used for outcome measurement in therapeutic treatment studies and both showed significant changes as a result of treatment. However, those data are not sufficient to establish that a GLPM reflects real changes in the grammatical forms targeted by intervention that are indeed due to the treatment rather than simply to maturation. What is needed is evidence from discriminant analysis comparing change in performance on therapeutic goals, for which change in performance is expected on the outcome measure, and control goals, for which no change in performance is expected.

CONCLUDING THOUGHTS

In this article, I considered the use of GLPMs for measuring outcomes on oral expressive grammar goals. Grammatical morphemes and assessments to measure morpheme acquisition and use are a common focus of therapy for young children. However, I chose to focus instead on three intermediate goal areas that I consider very important because they enable children to increase the informativeness of their utterance—production of word combinations, production of sentences with required and optional constituents, and production of complex sentences. This is consistent with the principle that therapy should focus on elaborations that increase a child's ability to convey information before focusing on grammatical correctness (Nelson, 2013). It also is consistent with how children learn language (Barako Arndt & Schuele, 2013; Limber, 1973; Paul, 1981; Tyack & Gottsleben, 1986) and addresses significant deficit areas for children with language impairment (Eisenberg, 2013; Leonard, 2014).

The GLPMs could be administered every 3–4 months to supplement session data, probes, and structural analyses for trained and untrained exemplars. This would determine whether broad-based learning was occurring for that intermediate goal area. In addition, I suggest less frequent administration of MLUw—perhaps every 6–8 months—to measure children's overall ability to construct increasingly longer, more complex, and more informative sentences throughout the emerging and developing language stages.

To date, there are limited data about the reliability of GLPMs and the validity of using the GLPMs to measure treatment outcomes. Research is needed to investigate these issues. Because of this, although I believe that it is important that SLPs go beyond session data and probes to measure progress, I also urge caution in interpreting GLPMs when assessing treatment progress.

REFERENCES

Altenberg, E. P., Roberts, J. A., & Scarborough, H. S. (2018). Young children's structure production: A revision of the index of productive syntax. *Language, Speech, and Hearing Services in Schools*, *49*, 1–14.

Bain, B. A., & Dollaghan, C. A. (1991). Treatment efficacy: The notion of clinically significant change. *Language, Speech, and Hearing Services in Schools*, *22*, 264–270.

Barako Arndt, K., & Schuele, C. M. (2013). Multiclausal utterances aren't just for big kids: A framework for analysis of complex syntax production in spoken language of preschool- and early school-age children. *Topics in Language Disorders*, *33*, 125–139.

Bolderson, S., Dosanjh, C., Milligan, C., Oring, T., & Chiat, S. (2011). Colorful semantics: A clinical investigation. *Child Language Teaching and Therapy*, *27*, 344–353.

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

Camarata, S. M., & Nelson, K. E. (1992). Treatment efficiency as a function of target selection in the remediation of child language disorders. *Clinical Linguistics & Phonetics*, *6*, 167–178.

Camarata, S. M., Nelson, K. E., & Camarata, M. N. (1994). Comparison of conversational-recasting procedures and imitative procedures for training grammatical structures in children with specific language impairment. *Journal of Speech and Hearing Research*, *37*, 1414–1423.

Ebbels, S. H. (2014). Effectiveness of intervention for grammar in school-aged children with primary language impairments: A review of the evidence. *Child Language Teaching and Therapy*, *30*, 7–40.

Ebbels, S. H., van der Lely, H. K. J., & Dockrell, J. E. (2007). Intervention for verb argument structure in children with persistent SLI: A randomized control trial. *Journal of Speech Language and Hearing Research*, *50*, 1330–1349.

Eisenberg, S. L. (2003). Production of infinitival object complements in the conversational speech of 5-year-old children with language impairment. *First Language*, *23*, 327–341.

Eisenberg, S. L. (2004). Production of infinitives by 5-year-old language-impaired children on an elicited production task. *First Language*, *24*, 305–321.

Eisenberg, S. L. (2013). Grammar intervention: Content and procedures for facilitating children's language. *Topics in Language Disorders*, *33*, 165–178.

Eisenberg, S., Fersko, T.M., & Lundgren, C. (2001). Use of MLU for identifying language impairment in preschool children: A review. *American Journal of Speech-Language Pathology*, *10*, 323–342.

Finestack, L. H., & Satterlund, K. E. (2018). Current practice of child grammar intervention: A survey of speech-language pathologists. *American Journal of Speech-Language Pathology*, *27*, 1329–1351.

Fletcher, P. (1991). Evidence from syntax for language impairment. In J. Miller (Ed.), *Research on child language disorder: A decade of progress* (pp. 177–194). East Sussex, England: Psychology Press.

Gavin, W. J., & Giles, L. (1996). Temporal reliability of language sample measures. *Journal of Speech and Hearing Research*, *39*, 1258–1262.

Girolametto, L., Pearce, P. S., & Weitzman, E. (1996). Interactive focused stimulation for toddlers with expressive vocabulary delays. *Journal of Speech and Hearing Research*, *39*, 1274–1283.

Grela, B. G., & Leonard, L. B. (1997). The use of subject arguments by children with specific language impairment. *Clinical Linguistics & Phonetics*, *11*(6), 443–453.

Guo, L., & Eisenberg, S. (2015). Sample length affects the reliability of language sample measures in three-year-olds: Evidence from parent-elicited conversational samples. *Language, Speech, and Hearing Services in Schools*, *46*, 81–93.

Hadley, P. A. (1999). Validating a rate-based measure of early grammatical abilities: Unique syntactic types. *American Journal of Speech-Language Pathology*, *8*, 261–272.

Ingham, R., Fletcher, P., Schletter, C., & Sinka, L. (1998). Resultative VPs and specific language impairment. *Language Acquisition*, *7*, 87–111.

Johnston, J. R. (2001). An alternate MLU calculation: Magnitude and variability effects. *Journal of Speech Language and Hearing Research*, *44*, 156–164.

Kemper, S., Rice, K., & Chen, Y-J. (1995). Complexity metrics and growth curves for measuring grammatical development from five to ten. *First Language*, *15*, 151–166.

King, G. (2000). Verb complementation in language impaired school aged children. In M. Aldridge (Ed.), *Child language* (pp. 84–91). Cleveden, England: Multilingual Matters.

King, G., & Fletcher, P. (1993). Grammatical problems in school-aged children with specific language impairment. *Clinical Linguistics & Phonetics*, 7, 339–352.

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

Lee, L. L. (1974). *Developmental sentence analysis*. Evanston, IL: Northwestern University Press.

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

Limber, J. (1973). The genesis of complex sentences. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 169–185). New York, NY: Academic Press.

Loeb, D. F., & Armstrong, N. (2001). Case studies on the efficacy of expansions and subject-verb-object models in early language acquisition. *Child Language Teaching and Therapy*, *17*, 35–53.

Long, S. H., Olswang, L. B., Brian, J., & Dale, P. S. (1997). Productivity of emerging word combinations in toddlers with specific expressive language impairment. *American Journal of Speech-Language Pathology*, *6*, 35–47.

Lund, N. J., & Duchan, J. F. (1993). *Assessing children's language in naturalistic contexts* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Marinellie, S. A. (2004). Complex syntax used by school-age children with specific language impairment (SLI) in child-adult conversation. *Journal of Communication Disorders*, *37*, 517–533.

McCauley, R. J. (1996). Familiar strangers: Criterion-referenced measures in communication disorders. *Language, Speech, and Hearing Services in Schools*, *27*, 122–131.

McCauley, R. J. (2001). *Assessment of language disorders in children*. Mahwah, NJ: Lawrence Erlbaum.

McCauley, R. J., & Fey, M. E. (2006). Introduction. In R. J. McCauley & M. E. Fey (Eds.), *Treatment of language disorders in children* (pp. 1–17). Baltimore, MD: Brookes.

Miller, J. F. (1981). *Assessing language production in children: Experimental procedures*. Baltimore, MD: University Park Press.

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, *24*, 154–161.

Nelson, K. E., Camarata, S. M., Welsh, J., Butkovsky, L., & Camarata, M. (1996). Effects of imitative and conversational recasting treatment on the acquisition of grammar in children with specific language impairment and younger language-normal children. *Journal of Speech Language and Hearing Research*, *39*, 850–859.

Nelson, N. W. (2013). Simple to complex: Targeting elaboration before correctness. *Topics in Language Disorders*, *33*, 115–117.

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2009). Syntactic development in adolescents with language impairments: A follow up investigation. *American Journal of Speech-Language Pathology*, *18*, 241–251.

Oosthuizen, H., & Southwood, F. (2009). Methodological issues in the calculation of mean length of utterance (MLU). *South African Journal of Communication Disorders*, *56*(1), 76–87.

Owen van Horne, A. J., & Lin, S. (2011). Cognitive state verbs and complement clauses in children with SLI and their typically developing peers. *Clinical Linguistics & Phonetics*, *25*, 881–898.

Paul, R. (1981). Analyzing complex sentence development. In J. F. Miller (Ed.), *Assessing language production in children: Experimental procedures* (pp. 36–40). Baltimore, MD: University Park Press.

Paul, R., Norbury, C. F., & Gosse, C. (2018). *Language disorders from infancy through adolescence* (5th ed.). St. Louis, MI: Elsevier.

Rice, M., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language and Hearing Research*, *53*, 333–349.

Robertson, S. B., & Ellis Weismer, S. (1999). Effects of treatment on linguistic and social skills in toddlers with delayed language development. *Journal of Speech Language, and Hearing Research*, *42*, 1234–1248.

Sabers, D. L. (1996). By their tests we will know them. *Language, Speech, and Hearing Services in Schools*, *27*, 102–109.

Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics*, *11*, 1–22.

Scherer, N. J., & Olswang, L. B. (1989). Using structured discourse as a language intervention technique with autistic children. *Journal of Speech and Hearing Disorders*, *54*, 383–394.

Schuele, C. M., & Dykes, J. C. (2005). Complex sentence acquisition: A longitudinal case study of a child with specific language impairment. *Clinical Linguistics & Phonetics*, *19*, 295–318.

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech Language and Hearing Research*, *43*, 324–339.

Spooner, L. (2002). Addressing expressive language disorder in children who also have severe receptive language disorder: A psycholinguistic approach. *Child Language Teaching and Therapy*, *18*, 289–313.

Steel, G., Rose, M., & Eadie, P. (2016). The production of complement clauses in children with language impairment. *Journal of Speech, Language, and Hearing Research*, *59*, 330–341.

Thordardottir, E. T., & Ellis Weismer, S. (2002). Verb argument structure weaknesses in specific language impairment in relation to age and utterance length. *Clinical Linguistics & Phonetics*, *16*, 233–250.

Tyack, D. L. (1981). Teaching complex sentences. *Language, Speech, and Hearing Services in Schools*, *12*, 49–56.

Tyack, D. L., & Gottsleben, R. H. (1986). Acquisition of complex sentences. *Language, Speech, and Hearing Services in Schools*, *17*, 160–174.

Wieczorek, R. (2010). Using MLU to study early language development in English. *Psychology of Language and Communication*, *14*, 59–69.

Yoder, P. J., Spruytenberg, H., Edwards, A., & Davies, B. (1995). Effect of verbal routine contexts on gains in the mean length of utterance in children with language delays. *Language, Speech, and Hearing Services in Schools*, *26*, 21–32.