

Tutorial

Using Computer Programs for Language Sample Analysis

Mollee J. Pezold,^a Caitlin M. Imgrund,^b and Holly L. Storkel^a

Purpose: Although language sample analysis is widely recommended for assessing children's expressive language, few school-based speech-language pathologists routinely use it, citing a lack of time, resources, and training (Pavelko, Owens, Ireland, & Hahs-Vaughn, 2016). The purpose of this clinical tutorial is (a) to describe options for language sample analysis using computer programs and (b) to demonstrate a process of using language sample analysis focused on the assessment of 2 preschool children as case studies.

Method: We provide an overview of collecting and analyzing child language samples and describe 3 programs for language sample analysis: 2 dedicated software programs (Computerized Language Analysis [MacWhinney, 2000] and Systematic Analysis of Language Transcripts [Miller & Iglesias, 2015]) and 1 protocol for using word processing

software to analyze language samples (Sampling Utterances and Grammatical Analysis Revised; Pavelko & Owens, 2017). We also present analysis results from each program for play-based language samples from 2 preschool children and detailed analysis of the samples with potential treatment goals.

Results: Each program offers different analyses, comparison databases, and sampling contexts. We present options for additional analysis, clinical interpretations, and potential treatment goals based on the 2 preschool cases.

Conclusion: Clinicians can use computer programs for language sample analysis as part of a process to make naturalistic language assessment more feasible.

Supplemental Material: <https://doi.org/10.23641/asha.10093403>

School-based speech-language pathologists' (SLPs') primary responsibilities include determining whether a child presents with a language disorder and how his or her language weaknesses impact educational progress (American Speech-Language-Hearing Association, 2010). One way to assess children's language is by using language sample analysis—procedures for eliciting, transcribing, and analyzing children's use of language in different contexts. The primary benefit of language sample analysis is that it captures a child's language use in naturalistic settings that mirror the communication demands of everyday social and academic situations (Costanza-Smith, 2010). Language samples can be analyzed in depth and with descriptive detail, and SLPs can repeatedly collect language samples to monitor progress over time (Rojas & Iglesias, 2009).

Despite the benefits of using language sample analysis, its use in clinical practice is limited. Pavelko, Owens, Ireland, and Hahs-Vaughn (2016) found that only around two thirds of school SLPs had used language sample analysis in the past year, and of those, most collected 10 or fewer samples. More than half of SLPs transcribed language samples while talking to children. Few SLPs used formal methods for language sample analysis, and of those who did, most reported using self-designed methods. A large majority (78%) of SLPs reported that language sample analysis took too much time. Other commonly reported barriers included a lack of resources and training. Fulcher-Rood, Castilla-Earls, and Higginbotham (2018) also described the prevalence of real-time and informal sampling: 40% of school SLPs in their sample transcribed language samples as they talked to children; 21% made broad informal judgments about language skill while speaking with children; and 21% collected, transcribed, and analyzed language samples based on traditional guidelines. These studies suggest few SLPs are taking advantage of formal language sample analysis protocols.

Computer programs can automate many language sample analyses, making language sample analysis faster and more accurate. However, each program has different requirements for coding samples and completing analyses. In addition, SLPs using comparison databases to describe

^aSpeech-Language-Hearing: Sciences and Disorders, University of Kansas, Lawrence

^bCommunication Sciences & Disorders, Florida Atlantic University, Boca Raton

Correspondence to Mollee J. Pezold: mollee@ku.edu

Editor-in-Chief: Shelley Gray

Editor: Kerry Ebert

Received December 20, 2018

Revision received April 23, 2019

Accepted July 29, 2019

https://doi.org/10.1044/2019_LSHSS-18-0148

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

children’s performance relative to others must consider that not all programs have data for the same age ranges or contexts. Using a language sample analysis program can save time, but SLPs must spend time learning how to use the program and how to code samples for analysis with the program.

To help SLPs choose a language sample analysis program, this tutorial outlines the differences between three currently available language sample analysis programs: Computerized Language Analysis (CLAN; MacWhinney, 2000), Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2015), and Sampling Utterances and Grammatical Analysis Revised (SUGAR; Pavelko & Owens, 2017). We focus on preschool children because all three programs include database samples from preschool-age children. We begin with an overview of considerations for collecting language samples, provide a general description of the three programs, and present two cases in which the same language sample was analyzed in each program along with estimates of the time needed to complete the analysis and the results for each program. Finally, we demonstrate a process of analyzing a language sample in depth, choosing additional measures based on the results, and identifying possible treatment goals for our two preschool cases. Supplemental Material S1 (referenced throughout) contains detailed information for further study.

General Considerations for Language Sample Analysis

Sampling Contexts

It is important to sample children’s language in contexts that are age appropriate and challenging in order to best capture their language strengths and weaknesses. The appropriate context for a language sample depends on the child’s age and language level as well as the information sought. Figure 1 provides an overview of recommended sampling contexts at different ages.

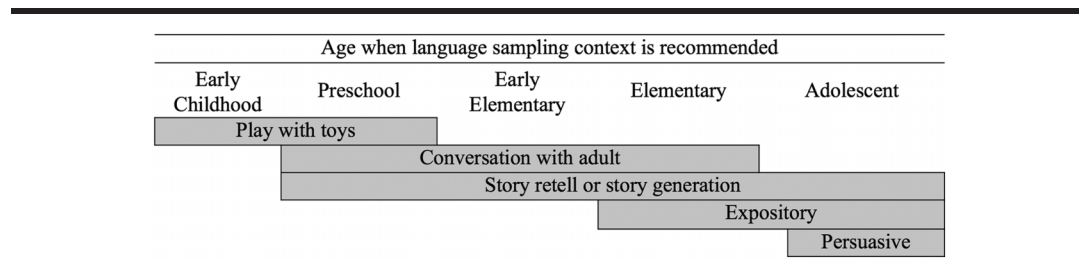
Miller, Andriacchi, and Nockerts (2015) recommend collecting language samples during adult–child play with toys for children younger than 4–5 years old. For older children (i.e., children ages 5–6 years and older), conversation with an adult is typically more appropriate. In the later preschool period and beyond (i.e., age of 4 years and older), children can also complete narration tasks by telling personal stories, generating stories from pictures, or retelling stories

they have just heard. However, younger children may produce more utterances and be more easily understood in play or conversation (Southwood & Russell, 2004; Wagner, Nettelbladt, Sahlen, & Nilholm, 2000), so SLPs may want to collect samples from both conversation and narrative tasks for a more complete sample. Narrative tasks can also be used with school-age children (i.e., aged 5 years and older), who tend to produce more complex utterances and provide more information in narrative tasks than in conversation (Thordardottir, 2008). For older school-age children (i.e., ages 8–10 years and older), expository tasks such as explaining how to play a sport or game are likely to elicit more complex language than conversation (Nippold, Hesketh, Duthie, & Mansfield, 2005) or narrative (Berman & Verhoeven, 2002) samples. Adolescents (i.e., aged 14 years and older) can also complete persuasive tasks such as arguing for changes in a rule. Adolescents may produce more complex language in persuasive than expository contexts (Brimo & Hall-Mills, 2019). SLPs have some flexibility when choosing a sampling context for a specific child. However, if SLPs use comparison databases to evaluate a child’s performance relative to peers, they should make sure that the context is the same, as analysis results can vary across contexts. Also note that children’s performance can differ with specific tasks even within a single context (narrative generation vs. narrative retell, expository summary vs. process description, etc.). We provided more guidance for collecting language samples in Supplemental Material S1 (Section 1).

Sample Length

In terms of length, 50 utterances is often considered an adequate sample of children’s language, with acceptable reliability demonstrated for mean length of utterance (MLU) and vocabulary measures derived from samples of 50 utterances or fewer (Casby, 2011; Heilmann, Nockerts, & Miller, 2010). However, Guo and Eisenberg (2014) found that 100-utterance samples were more diagnostically accurate and reliable than 50-utterance samples when computing overall tense and agreement accuracy measures for 3-year-old children. Balason and Dollaghan (2002) found that even samples of 100 or more utterances did not provide adequate opportunities for 4-year-old children to produce individual grammatical morphemes. In summary, the research on the effect of transcript length on language sample analysis results is mixed. While 50-utterance samples are likely

Figure 1. Recommended language sample contexts by age.



adequate for computing broad measures of utterance length and lexical diversity, 100-utterance samples may be needed for more detailed coding of grammatical structures, and even then, the child may not have many opportunities to produce every target. Longer samples are also more time-consuming to collect and analyze. On balance, we recommend that SLPs planning to compute language sample analysis measures strive to collect samples of at least 50 utterances in most cases. SLPs should ensure that their samples are similar in length to those in any comparison database being used, because some measures (vocabulary measures in particular) vary across transcript length (Owen & Leonard, 2002).

Analyses

The benefit of computerized language sample analysis is the calculation of quantitative measures of children's language production. These most often include measures of utterance length, productivity, vocabulary diversity, and grammatical accuracy. Here, we describe quantitative measures relevant to the preschool children described in this tutorial that can be calculated from 50-utterance language samples. We describe measures of grammar that require hand-coding in more detail and provide references for measures relevant to older children in Supplemental Material S1 (Sections 2 and 3).

MLU

MLU indicates children's average utterance length in words or morphemes across a sample. For example, for the following two utterances—"Look out" and "He's running"—the MLU in morphemes value would be three (six morphemes divided by two utterances). Although both sentences contain the same number of words, in the second sentence, both words contain a bound morpheme (contracted auxiliary "is" in "he's" and present-progressive "-ing" in "running"). MLU therefore broadly reflects a child's morphosyntactic ability, or the number of morphemes that he or she is able to combine in an utterance. A low MLU relative to same-age peers may indicate that a child's morphosyntax skills are a concern, but MLU alone does not indicate which skills are lacking. As with all language measures, MLU alone should not be used to diagnose a child with language impairment (Bedore & Leonard, 1998; Eisenberg, Fersko, & Lundgren, 2001). MLU is also not recommended for indicating language impairment after the age of 7 years (Rice et al., 2010). In summary, MLU is a measure of utterance length in morphemes that indexes overall morphosyntactic skill and can support diagnostic decisions and track progress over time in younger children.

Total Number of Words

Total number of words (TNW) counts all words a child uses in a language sample. In samples not matched for number of utterances, TNW reflects a child's speaking rate (in time-based samples) and talkativeness (in narrative samples; Miller, 1981). TNW increases as children get older (Klee, 1992) but has low test-retest reliability (Gavin & Giles, 1996) and low correlation with other language sample

analysis measures (Dethorne, Johnson, & Loeb, 2005). A downside of TNW is that it does not reflect the diversity of vocabulary used in a sample. For example, the sentences "We did things and got stuff" and "My class hiked and found treasures" both have a word count of six, but a language sample with many sentences such as the former is likely to contain a smaller range of words than a sample with sentences such as the latter even if TNW is equal. Because of its low reliability and unclear implications, TNW results are difficult to interpret.

Number of Different Words

Number of different words (NDW) counts the number of unique words in a language sample, roughly indicating the diversity of vocabulary used (Miller, 1981). Like TNW, NDW increases with age (Klee, 1992), but it has higher test-retest reliability, at least in longer samples (Gavin & Giles, 1996). Ukrainetz and Blomquist (2002) found that NDW results were significantly correlated with standardized tests of vocabulary. Watkins, Kelly, Harbers, and Hollis (1995) found that preschoolers with language impairment had significantly lower NDW than typical peers, whereas Thordardottir and Ellis Weismer (2001) found no significant difference in NDW in school-age children with and without language impairment. The nature of NDW makes it susceptible to changes in sample length in words. Owen and Leonard (2002) explained that NDW comparisons matched for length of utterances are influenced by MLU because children with higher MLUs will produce more words in a sample. Even when limiting samples by the number of words, NDW will level off as sample length increases, because words are likely to be repeated. While low NDW alone should not be used as the only evidence of language impairment, SLPs can have more confidence in NDW than TNW values for indicating potential vocabulary weaknesses, particularly when comparing same-length samples from younger children.

Type-Token Ratio

Type-token ratio (TTR) is calculated by dividing NDW by TNW to find the ratio of different words to total words in a sample. TTR is intended to account for language samples of varying length. However, TTR is not sensitive to changes in age (Klee, 1992; Miller, 1981) and does not differentiate between children with and without language disorders (Watkins et al., 1995). As with NDW, TTR is affected by sample length, because function words repeated in longer samples reduce lexical diversity (Owen & Leonard, 2002).

Error Coding

Children with language disorders have particular difficulty with syntax and morphology (Leonard, 2014). Recent research has highlighted the usefulness of measuring the presence and type of grammatical errors in language samples (Guo & Eisenberg, 2014; Guo & Schneider, 2016). Computer programs can facilitate error analysis by automatically tallying the number of errors or types of errors. For example, SLPs might differentiate morphemes that are omitted ("he like it") versus overgeneralized ("they likes it")

using different codes. Measures such as percent grammatical utterances (PGU) can be used to compute a sentence-level grammaticality score that can help differentiate children with and without language disorders (Eisenberg & Guo, 2016). We describe PGU further in Supplemental Material S1 (Section 2).

Using Database Comparisons

A benefit of calculating language sample measures such as MLU and NDW is that they allow SLPs to compare a child's performance to the performance of children his or her age. SLPs computing language sample measures by hand can compare a child's results to values reported in research articles (e.g., Rice et al., 2010). The three language sample analysis programs described in the next section also provide data from a set of comparison database samples. We use the term *database samples* here rather than *normative database* because none of the programs described below meets the criteria for normative databases. For example, Andersson (2005) calls for normative samples with at least 100 children per age group and demographic characteristics representative of the population, among other requirements. None of the database comparisons built into the three language sample analysis programs we describe here meets both of these conditions. In addition, the diagnostic accuracy for any one language sample analysis measure is unclear (Ebert & Scott, 2014; Eisenberg et al., 2001; Heilmann, Miller, & Nockerts, 2010; but see preliminary findings from Pavelko & Owens, 2019). Although language sample analysis results may support other diagnostic information, no single piece of information in isolation should be used to determine service eligibility or diagnostic status. When used carefully, results from comparison databases may indicate overall areas of weakness relative to peers, which can be confirmed by qualitatively analyzing a language sample to determine appropriate treatment targets.

Comparison of Language Sample Analysis Programs

This tutorial describes three currently available programs for language sample analysis. Two are dedicated software, and a third employs generic word processing features. A fourth program, Computerized Profiling (Long, 2012), has also been described. Computerized Profiling is not currently available online, but access to the download is available from its author (S. Long, personal communication, February 6, 2019). Table 1 provides a visual overview of the features of each program. Each program is briefly described below, with further details and resources included in Supplemental Material S1 (Sections 4–6).

CLAN

CLAN (MacWhinney, 2000; available at <http://dali.talkbank.org/clan/>) is a powerful, customizable program that can compute language sample analysis measures across 49 languages. A unique feature of CLAN is that it does not

Table 1. Comparison database features across CLAN, SALT, and SUGAR.

Feature	Category	CLAN	SALT	SUGAR
Age	Early childhood	✓	✓	
	Preschool (3–5)	✓	✓	✓
	Early school age (up to 7)		✓	✓
	Later school age (8+)		✓	
Type of sample	Play	✓	✓	
	Conversation		✓	✓
	Narrative		✓	
	Expository		✓	
Languages	Persuasion		✓	
	English	✓	✓	✓
	Bilingual English/Spanish		✓	
Cost	Spanish		✓	
	Free	✓		✓

Note. CLAN = Computerized Language Analysis; SALT = Systematic Analysis of Language Transcripts; SUGAR = Sampling Utterances and Grammatical Analysis Revised.

require users to manually code morphemes. Rather than coding morphemes by hand, users can run a command that makes CLAN identify the morphemes. Morpheme coding by CLAN is 94% accurate on average, and human coding of morphemes is unlikely to be 100% accurate (Bernstein Ratner & MacWhinney, 2016). Users do add codes to indicate intelligible segments, repetitions, and fillers (which are excluded from analyses) and abandoned utterances (which are included in analyses). Users can indicate errors by using the [*] code. CLAN also allows users to link audio or video files to a transcript and “walk” through media files using keyboard commands that are similar to the features of pedal transcribers (described further in Supplemental Material S1, Section 7). CLAN users may compare play samples for children under 6 years old to Child Language Data Exchange System (CHILDES) database samples using the KIDEVAL analysis option, although these comparisons are preliminary. The clinician manual for CLAN also includes published values for MLU, NDW, and other measures to which SLPs may compare language sample analysis results as appropriate (Bernstein Ratner & Brundage, 2018). Although CLAN can complete analyses in 49 languages, the comparison database and the published values listed in the manual are for English-speaking children only.

In terms of database comparisons, CLAN's KIDEVAL feature compares samples to a database drawn from the CHILDES corpora (MacWhinney, 2000). CHILDES includes language samples from various studies. The specific corpora that KIDEVAL uses are listed in the CLAN manual (MacWhinney, 2018, p. 110) but are not individually described. An informal tally by the first author indicated that the corpora include roughly 450 typically developing children, although some studies were longitudinal, so children are included in the database multiple times. For the age band used in this tutorial (4;6–4;11 [years;months]), KIDEVAL returned results from 141 samples. It is difficult to easily summarize socioeconomic data for the CHILDES corpora used because samples are drawn from 30 different studies. Regarding use

of the comparison database, the CLAN manual states that “this method for using KIDEVAL is still under development and the results should not yet be used for serious diagnostic purposes” (MacWhinney, 2018, p. 109). SLPs can also use CLAN to compute measures and then compare them to published values.

SALT

SALT (Miller & Iglesias, 2015; available at <http://saltsoftware.com/products/software>) includes options for quantifying language skills in children and adolescents across several sampling contexts in both English and Spanish. SALT users must individually code morphemes in transcripts in order to perform analyses such as MLU. Users also indicate the presence of fillers, repetitions, abandoned utterances, and intelligible segments, which are automatically excluded from analyses. Codes can be used to mark different errors (e.g., [EO:___] for overgeneralization errors or [EU] for utterance-level errors). Users can automatically compare the performance of their clients to a built-in comparison database for similar-age children (ages 2–18 years) in the same context (play, conversation, narrative, expository, or persuasive). The SALT comparison database samples also feature additional coding to capture the overall structure of the sample separated by sample type: narrative (ages 4–13 years), expository (ages 10–18 years), and persuasive (ages 14–18 years). SLPs can score client samples using rubrics and compare results to the comparison database. SALT offers paid transcription services in English and Spanish, allowing SLPs to securely send audio files to be transcribed and coded by SALT staff. SALT is unique in its inclusion of samples in Spanish and samples from Spanish–English bilingual children in the comparison database. We included resources for bilingual language sample analysis in Supplemental Material S1 (Section 8).

For the context used here (adult–child play), the SALT database contains 69 samples across ages 2;8–5;8. Children all attended public preschool or kindergarten in Madison, Wisconsin, and had a primary language of English. Samples were diverse in terms of economic background and ability levels, but specific data on race or parent education were not available (Miller et al., 2015). For the age band used in this tutorial (4;3–5;3), SALT returned results from 25 fifty-utterance samples. Note that this age band differs from the other two programs. We began with SALT’s default value of plus or minus 6 months. We attempted to compare samples to the 6-month range used by CLAN and SUGAR (4;6–4;11), but this returned only four 50-utterance samples, so we continued with the wider age range. Compared to the other two programs, SALT has fewer preschool database samples. Unlike the other two programs, SALT also has narrative and expository database samples from older children.

SUGAR

SUGAR (Pavelko & Owens, 2017; details at <https://www.sugarlanguage.org/>) is not a standalone software but rather a protocol for using Microsoft Word or other word

processing software to complete several modified language sample analysis procedures. In contrast to CLAN and SALT, SUGAR coding is minimal and requires only that SLPs add spaces between specified morphemes and line breaks before clauses. Instead of transcribing an utterance exactly as a child says it, transcribers omit filler words, repetitions, and reformulations, meaning that these extraneous words do not need to be specially coded to be excluded from analyses. Although SUGAR does not include procedures for error coding, users could mark errors with an asterisk without affecting analysis results. SLPs can compare their client’s performance on SUGAR analyses to values from adult–child conversation samples from children ages 3–7 years (Pavelko & Owens, 2017). The SUGAR protocol for sample elicitation encourages the use of process questions and narrative elicitations, which are intended to increase the language complexity children use during conversation.

Of the three options, SUGAR describes its comparison database in the most detail. Pavelko and Owens (2017) reported race and parent education data for each age band in their sample, which includes 270 fifty-utterance samples from children ages 3;0–7;11 across a single sampling context (adult–child interview) and 55 children in the age band for our cases (4;6–4;11).

Case Demonstrations of Preschool Language Sample Analysis

We present two preschool cases to demonstrate the potential use of language sample analysis in the clinical setting. Both children discussed here received a comprehensive speech and language evaluation as part of a larger study of language development in children who were born preterm (Imgrund, Loeb, & Barlow, 2019). Testing included the completion of the Clinical Evaluation of Language Fundamentals Preschool–Second Edition (CELF Preschool-2; Semel, Wiig, & Secord, 2004) and the collection of a 100-utterance, play-based language sample. We used pseudonyms here to protect the children’s privacy.

Sam

Sam is age 4;9 and a Caucasian boy. On the CELF Preschool-2, Sam achieved an Expressive Language Index standard score of 83, a Receptive Language Index standard score of 86, and a Core Language standard score of 84. At the time of the evaluation, Sam’s mother expressed concerns about his overall readiness for kindergarten.

Julia

Julia is age 4;9 and a Caucasian girl. Julia was receiving services from an SLP at the time of her evaluation. Julia’s mother stated that Julia demonstrated particular difficulty with expressive language. Julia achieved an Expressive Language Index standard score of 98, a Receptive Language Index standard score of 109, and a Core Language standard score of 100.

Language Samples

The language samples were obtained during examiner-child play and were video-recorded for transcription. Utterances were equally divided across three play contexts (farm set, puppets, and baby doll set), which were presented in random order. For this tutorial, we selected 50-utterance samples beginning with the farm set context and continuing to the puppets context. For Sam, these were the first and second contexts of the play session, and for Julia, these were the second and third contexts. Note that while the CLAN and SALT comparison databases contain adult-child play samples, SUGAR provides values only for adult-child conversation. The SUGAR conversational protocol (Pavelko & Owens, 2017) encourages SLPs to use narrative elicitations (e.g., “Tell me about [a recent event]”) and process questions (e.g., “How does [X] work?”), which were not used in our samples. Although we recommend that SLPs collecting samples to be analyzed with SUGAR follow its sample collection guidelines as closely as possible, we analyzed these samples in SUGAR to illustrate the differences across the language analysis methods on the samples that were available. When interpreting the results, it is important to keep in mind that the SUGAR comparisons may not be valid because our samples do not match those in the comparison database.

Transcription and Analysis Procedure

In the previous study (Imgrund et al., 2019), research assistants transcribed and coded child and examiner utterances using SALT software. Although the samples were already transcribed, we transcribed them again to evaluate the use of speech-to-text software as an alternative and potentially faster way to complete transcription that may be of interest to SLPs. The second author transcribed child and examiner utterances from the 50 child utterance segment for each language sample using Google’s voice typing feature. The second author listened to the recording using headphones while repeating the child and examiner utterances aloud so that Google voice typing could convert her speech to text. The second author verified the accuracy of this transcription as it occurred and while listening to the recording a second time to verify the complete transcript. It took 22 min to transcribe an 8-min segment (Sam) and 18 min to transcribe a 6.5-min segment (Julia) using this method, including two passes through the transcript. We included further directions for using speech-to-text to aid in transcription in Supplemental Material S1 (Section 7).

For this tutorial, we removed all coding from the existing SALT transcripts and created three versions of each 50-utterance transcript to analyze in CLAN, SALT, and SUGAR. For the SUGAR transcript, this included removing examiner utterances and filler words or repeated segments (e.g., “um” or “he went he went outside”). Each version of the transcript contained slightly different utterances because of differing rules for utterance inclusion. SUGAR analyses include abandoned utterances and utterances with unintelligible segments, CLAN analyses include abandoned utterances

but not utterances with unintelligible segments, and SALT analyses include neither abandoned utterances nor utterances with unintelligible segments. The discrepancy in utterance inclusion rules meant that we were not comparing the programs’ analyses on the same set of utterances. In this tutorial, we describe the results based on these unequal samples, because these are the results that SLPs following each program’s guidelines would get. We also completed the analyses separately on a matched set of utterances to show specific differences between the program, with results in Supplemental Material S2.

Next, a research assistant coded the six bare transcripts in semirandom order according to instructions for each program. The research assistant was a second-year master’s degree student in speech-language pathology with some familiarity with SALT from an undergraduate language analysis course. The first author also coded duplicate copies of the six bare transcripts, compared the two newly coded transcripts for disagreements, and resolved any discrepancies by consulting the respective programs’ manuals. An example of the same utterance transcribed and coded in each program is listed in Table 2. Both coders tracked the time they spent coding each sample (reported in Table 3).

Transcription

Details on coding across the three programs are listed in Table 3. We attempted to identify the standard set of codes for each program that allowed us to complete similar analyses. SALT and SUGAR both require coding morphemes, while CLAN codes morphemes automatically. CLAN and SALT both call for coding fillers, repetitions, errors, and omissions, while SUGAR omits them from the transcript to save time and facilitate analysis in Microsoft Word. In SUGAR, we did not code words per sentence or clauses per sentence because we considered this a supplemental analysis that we could not easily compare to CLAN and SALT results. Overall, SUGAR coding was the quickest to complete and the most reliable between the two coders (see Table 3). Coders were likely less reliable with CLAN and SALT coding because they were coding for more features (i.e., retraces, mazes, errors).

Analyses

We computed MLU (in all three programs) and NDW (in CLAN and SALT) and compared children’s results to program database means. We did not include TNW because of its questionable clinical interpretation but did include it in Supplemental Material S2 for comparison between programs. We did not formally track the time needed to conduct analyses on coded transcripts. Completing analyses is relatively instant. SALT and CLAN can automatically compare multiple language sample analysis measures to their comparison databases, while SUGAR requires that the user compare numbers to a table and determine standard deviation differences by hand. Table 4 contains analysis results.

Table 2. Coded transcription of the same child utterance across CLAN, SALT, and SUGAR.

Original utterance	Child: And sometime um whenever it's time for us to put them in bed we put the horse in here.
CLAN	CHI ^a : and <sometime> [//] &-um whenever it's time for us to put them in bed we put the horse in here.
SALT	C ^a And (sometime um) whenever it's time for us to put them in bed we put the horse in here.
SUGAR	And whenever it's time for us to put them in bed we put the horse in here.

Note. CLAN = Computerized Language Analysis; SALT = Systematic Analysis of Language Transcripts; SUGAR = Sampling Utterances and Grammatical Analysis Revised.

^aCHI and C are the speaker codes for the child speaker in CLAN and SALT, respectively.

Differences in Results Across Programs

Raw Values

As reported in Table 4, for both children, raw scores for MLU were similar across all three programs (i.e., within .16 morphemes of one another). Rice et al. (2010) found that, for children aged 4;6–4;11, the standard deviation for MLU in morphemes was .70 for children with language disorders and .79 for children without language disorders. In other words, the differences between MLU values of .16 or less across programs were small in terms of the average distribution of MLU scores for same-age children reported by Rice et al. NDW raw scores differed by only one word between CLAN and SALT for both children. The slight differences reflect differences in coding across programs (see Supplemental Material S2 for details).

Database Comparisons

While children's raw scores were similar, the distance of their raw scores from the database mean in standard deviation units was different across programs. For Sam, MLU and NDW standard deviation results were low across all programs, although to varying degrees. Sam's results fit generally with his borderline scores on the CELF Preschool-2 and his mother's concerns about kindergarten readiness. For Julia, MLU standard deviation results were low according to SALT and SUGAR but not CLAN. By comparing Julia's MLU to published values listed in the clinician manual for

CLAN (Bernstein Ratner & Brundage, 2018), SLPs might draw different conclusions about Julia's CLAN results. SLPs may choose to verify MLU values with the Rice data because the samples and children are described in more detail than the KIDEVAL samples. Rice et al. (2010) reported a 1-SD range of 3.96–5.54 for MLU in morphemes for 74 typical children ages 4;6–4;11. Julia's result of 3.92 falls just below this range. Julia's NDW standard deviation results were also low according to CLAN and SALT. Julia's results are in contrast to her average scores on the CELF Preschool-2, but fit with reported parent concern and ongoing SLP services.

Database comparisons varied slightly across programs because of differences in the database samples for each program, none of which meets the requirements for normative samples. SLPs using database comparisons should consider how similar their sample and client are to those from comparison databases, as well as the number of samples being compared. Our approach violated these recommendations for SUGAR. We used play-based samples, while the SUGAR samples were collected through a specific conversational protocol (Pavelko & Owens, 2017). Likewise, our samples contained 50 utterances, which we were able to match to database samples in SALT and SUGAR but not CLAN. These differences likely contributed to the observed variability.

In cases where results are borderline or unclear, like for Julia's, SLPs can complete further assessment—by analyzing the language sample in depth, completing additional

Table 3. Transcription and coding procedures in CLAN, SALT, and SUGAR.

Measure	CLAN	SALT	SUGAR
Coding completed	Trailing off, self-interruption, retracing, filled and unfilled pauses, omissions, and errors	Trailed off or abandoned utterances, bound morphemes, mazes, omissions, and errors	Bound morphemes
Time to code	5–11 min	5–15 min	2–4 min
Percentage of utterances with coding disagreements	5%–10%	11%–17%	2%–6%
Summary of disagreements	Most involved coding omissions (e.g., "so it be [i:] [i] soft." vs. "so it 0will be soft.")	Error and omission coding, morpheme coding (e.g., let's vs. let/us)	Morpheme coding
Impact of disagreements	No impact on our analyses	Error and omission coding does not impact our analyses; morpheme coding does impact analyses	Impacts MLU results

Note. CLAN = Computerized Language Analysis; SALT = Systematic Analysis of Language Transcripts; SUGAR = Sampling Utterances and Grammatical Analysis Revised; MLU = mean length of utterance.

Table 4. Analysis results across each program for Sam and Julia.

Measure	Sam			Julia		
	CLAN ^a	SALT ^b	SUGAR ^c	CLAN ^a	SALT ^b	SUGAR ^c
MLU raw	3.08	2.96	3.00 ^d	3.90	3.92	4.06 ^d
MLU <i>SD</i> from program database mean	-1.13	-1.93	-2.41	-0.39	-1.08	-1.61
NDW raw	68	69		79	78	
NDW <i>SD</i> from program database mean	-1.11	-1.95		-1.07	-1.45	

Note. CLAN = Computerized Language Analysis; SALT = Systematic Analysis of Language Transcripts; SUGAR = Sampling Utterances and Grammatical Analysis Revised; MLU = mean length of utterance in morphemes; NDW = number of different words.

^aCompared to free-play samples from 141 children ages 4;6–4;11 using KIDEVAL. ^bCompared to 50-utterance play samples from 25 children ages 4;3–5;3 using SALT database. ^cCompared to conversation samples from 55 children ages 4;6–4;11 reported by Pavelko and Owens (2017). ^dValues for MLU SUGAR.

observation, or administering probe tasks—to determine whether goals are needed for a certain language domain. Cross-referencing published reference values for specific analyses may also be helpful. Because databases vary, SLPs should use one database across samples from the same context to provide consistent information over time when possible. The goal of our database comparisons was not to make a diagnosis but to quantify language functioning in naturalistic contexts that are relevant to children’s everyday language use. Importantly, results for our cases seemed to match parents’ concerns. The language samples also allow us to investigate potential language weaknesses in greater depth to identify treatment goals and track progress over time.

Determining Treatment Goals

A primary benefit of collecting and analyzing a language sample is that it helps SLPs identify treatment goals (Costanza-Smith, 2010). We depicted a sample process that SLPs might employ for using language sample analysis to guide treatment planning and monitor progress in Figure 2. The process might begin by collecting a 50-utterance language sample and computing measures with the help of a language sample analysis program. However, the automatic analyses described here (MLU, NDW) do not translate directly into meaningful and easily measurable treatment goals once children are combining a variety of words into sentences. Instead, these measures point SLPs to overall areas of potential deficit. This streamlines the process of choosing methods of further hand-coding or choosing additional measures or probe tasks. SLPs can then use these results to determine specific treatment targets. Once this process is completed, shorter probe tasks can be administered more frequently to monitor generalization on treatment targets. Finally, SLPs can repeat language sampling to assess a child’s generalization of skills and identify new goals as needed. This can be done at the end of a monitoring period or when significant progress or a lack of progress on probe tasks indicates that new goals may be warranted.

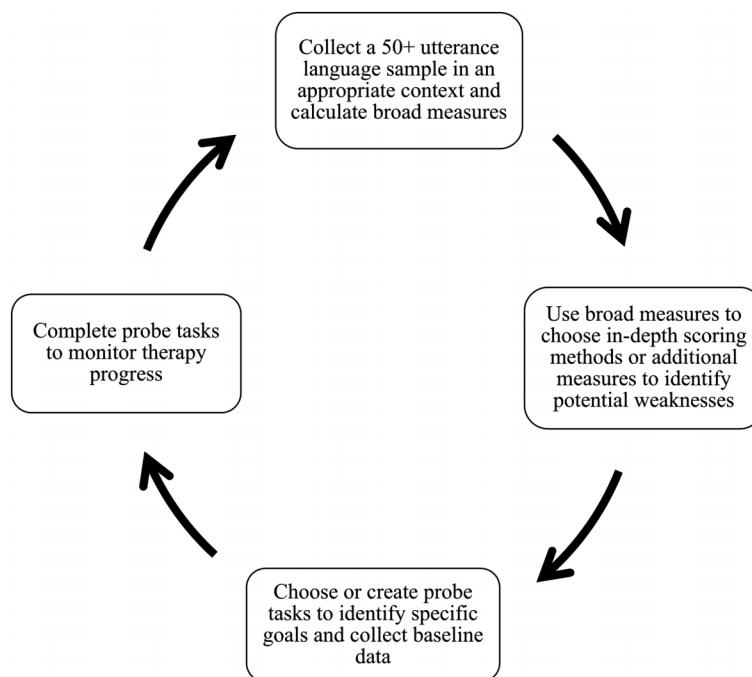
Methods of Hand-Coding

A variety of options for detailed analysis of children’s language use within a sample are available. We used the

language sample analysis results presented above to identify broad areas of potential need for both children and then analyzed the transcripts by hand to determine possible treatment goals. Clinicians who are particularly comfortable with language analysis or who have a specific goal or grammatical structure in mind may accomplish this analysis informally. For those who want a more formal method to analyze samples in depth, several procedures are available to systematically tally children’s use of language structures. We describe Developmental Sentence Scoring (DSS; Lee, 1974; Lee & Canter, 1971) and the Index of Productive Syntax (Scarborough, 1990) in Supplemental Material S1 (Section 2). Also available are Miller’s (1981) Assigning Structural Stage Procedure, which scores obligatory use of Brown’s grammatical morphemes, and a process designed to go with SUGAR samples that identifies children’s use of phrases and morphemes (Owens, Pavelko, & Babinelli, 2018). For children with many errors, an analysis of all errored sentences may be adequate. PGU (Eisenberg & Guo, 2016), described in Supplemental Material S1 (Section 2), could accomplish this. All of these methods are viable options for cataloging the structures children are using and showing where there are weaknesses or what next steps in assessment might be.

For our cases, we chose to complete DSS (Lee, 1974). We chose DSS here because both children had low MLU results, indicating potential morphosyntactic weaknesses. When we reviewed the children’s samples, they had few consistent grammatical errors, so we did not think that a measure such as PGU would lead directly to treatment goals. DSS is also sensitive to changes in grammar skills following treatment (Fey, Cleave, Long, & Hughes, 1993). The DSS procedure evaluates grammatical complexity. It includes eight grammatical categories: indefinite pronouns, personal pronouns, main verbs, secondary verbs, negatives, conjunctions, interrogative reversals, and *wh*-questions. For every utterance, the child is given a point value between 1 and 8 for each of the syntactic categories used within the utterance. Lower point values are assigned to simple, early-acquired grammatical forms, and higher point values are assigned for more complex, later acquired grammatical forms. For example, the coordinating conjunction *and* is awarded 3 points, whereas the more complex subordinating

Figure 2. Sample process for goal selection and progress monitoring using language sample analysis.



conjunction *when* is awarded 8 points. Additionally, each utterance that meets adult standards (from a semantic and syntactic standpoint) earns one extra “sentence point.” An average DSS is then calculated for the utterances in the sample. Only utterances that are complete (i.e., the utterance contains a subject and a verb), unique, and intelligible are scored. However, calculating a total score requires 50 eligible utterances. For the two children here, even their 100-utterance samples from the larger study (Imgrund et al., 2019) did not contain enough eligible utterances to calculate a total score. DSS is still valuable, however, because it allows us to look at which structures are in error and which children are attempting. Hughes, Fey, and Long (1992) recommend identifying treatment goals by examining the types of errors, the forms that children attempt frequently, and the variety of low- versus high-scoring forms present. Using these suggestions, we evaluated the DSS results for potential intervention targets.

Sam’s language sample results showed low MLU, indicating a potential weakness in morphosyntax skills. For the 50-utterance sample used here, 22 of Sam’s utterances were eligible for DSS. We used this sample to catalog the utterances Sam was producing, attempting, or not attempting. Looking at Sam’s DSS results (reported in Supplemental Material S3), he often attempted main verbs but omitted “do” in questions (e.g., “What *does this one look/3s like?”). This indicated that the ability to ask *wh*-questions with inverted auxiliaries might be an appropriate goal. Other structures that were attempted only once or not at all include negatives and conjunctions. Because of the open-endedness of language samples, a 50-utterance (or

even 100- or 200-utterance) language sample is not guaranteed to have an adequate number of opportunities for a child to attempt a certain structure (Balason & Dollaghan, 2002). Qualitatively analyzing a language sample with a procedure such as DSS tells us which structures we might want to probe. SLPs might create probes to elicit specific targets directly (e.g., Finestack, Bangert, & Huang, 2017) or use or adapt existing probes (e.g., the free Test of Early Grammatical Impairment; Rice & Wexler, 2001). Probes also have the benefit of being quicker for therapy progress monitoring. For Sam, we would recommend probing auxiliary verbs (particularly in questions), negatives, and conjunctions. Based on the results of probe tasks, we could choose to target one to three specific structures for which Sam’s accuracy was low. We would target those goals using general procedures of grammar intervention such as modeling or recasts (e.g., Smith-Lock, Leitao, Lambert, & Nickels, 2013).

Julia’s MLU was also low according to SALT, SUGAR, and published reference values. Julia’s sample, like Sam’s, did not have 50 utterances with a subject and a verb, so we could not compute an overall DSS score. Looking at the 23 utterances that could be scored (listed in Supplemental Material S3), Julia had many low scores on indefinite pronouns or noun modifiers. We could probe Julia’s ability to use more advanced forms such as *nothing*, *any*, or *everything* (which may also tie into vocabulary goal setting, described in the following section). Although Julia frequently scored points for main verbs, she also had six unsuccessful attempts (denoted by slashes). Broadly looking at these sentences, it seems that copula and auxiliary verbs may warrant attention. Julia also showed some errors on

past tense verbs. For Julia, we would create probes for indefinite pronouns, copula and auxiliary verbs, and regular past tense verbs. Depending on Julia's performance on probe tasks, we might select any or all of these structures for treatment targets. Treatment might include procedures such as variable recasts (Plante et al., 2014) or "hard" past tense verbs (Owen Van Horne, Curran, Larson, & Fey, 2018).

Choosing Other Assessments

Both children's vocabulary diversity results (i.e., NDW) were also low. The above examples of hand-coding are mostly grammatical, and we are not aware of a method for detailed coding of language samples to identify vocabulary goals. SLPs may be able to do this in samples where children are obviously using mazes or substituting general words (e.g., *stuff, things*) or when monitoring for the use of specific words or classes of words. For our cases, we reviewed the samples in detail and did not notice any obvious vocabulary needs. The children's potential vocabulary weakness could instead be supported by classroom observations or by parent or teacher surveys of language ability (e.g., the Children's Communication Checklist [Bishop, 2003] or the Teacher Rating of Oral Language and Literacy [Dickinson, McCabe, & Sprague, 2001]). If we were concerned about specific words (e.g., quantifiers, comparatives), we might develop a probe task to assess this skill. We might also use expressive and receptive vocabulary subtests from standardized tests to confirm that overall vocabulary skills are a concern. The combination of observations, interviews, and tests might indicate that vocabulary treatment is warranted and provide direction for specific targets and treatment approaches (e.g., Hadley, Dickinson, Hirsh-Pasek, & Golinkoff, 2019).

Conclusion

Based on the cases presented here, SLPs can expect to spend 30–45 min in total collecting, transcribing, coding, and analyzing 50-utterance conversational language samples. In these preschool samples, raw values for MLU and NDW were similar regardless of the program used. Programs did differ when comparing children's performance to database sample means, which reiterates the importance of carefully matching contexts for sample collection. Carefully used database comparisons can broadly identify areas of need and lead to further in-depth coding to identify potential treatment targets. Taken together, computer-assisted language sample analysis is a relatively quick way to gain quantitative information about children's everyday language functioning within the realities of clinical practice, complementing other more commonly used approaches.

Of course, SLPs may analyze standardized test items to complete a similar process of goal selection. We argue that analyzing language samples in depth is a better way to identify potential treatment goals. A child may demonstrate a skill in structured tasks but not in natural contexts, or the reverse could be true. The naturalistic and academically relevant contexts used for language sampling provide the best picture of the impact of children's communication

weaknesses on everyday functioning (Costanza-Smith, 2010). Language samples also provide the opportunity to broadly quantify multiple skills at once (and to compare those results to those of peers, if program database or published reference values are used). Language samples can also be analyzed in greater detail to identify absent or attempted language structures. While we acknowledge that many SLPs consider language sample analysis too time-consuming (Pavelko et al., 2016), we hope that we have demonstrated an efficient method of using language samples to complement and guide the language assessment process within the realities of clinical practice.

Acknowledgments

This study was supported by the Council of Academic Programs in Communication Sciences and Disorders PhD Scholarship and the ASHA Foundation Student Research Grant in Early Childhood Language Development, both to Caitlin Imgrund. We thank Katharine Kesler for her assistance with this project. We also thank members of the Clinical Research for SLPs Facebook group for their comments and suggestions on this topic. Open access was funded by the University of Kansas College of Liberal Arts and Sciences.

References

- American Speech-Language-Hearing Association.** (2010). *Roles and responsibilities of speech-language pathologists in schools* [Professional issues statement]. Retrieved from <https://www.asha.org/policy/PI2010-00317/>
- Andersson, L.** (2005). Determining the adequacy of tests of children's language. *Communication Disorders Quarterly*, 26(4), 207–225.
- Balason, D. V., & Dollaghan, C. A.** (2002). Grammatical morpheme production in 4-year-old children. *Journal of Speech, Language, and Hearing Research*, 45(5), 961–969. [https://doi.org/10.1044/1092-4388\(2002\)078](https://doi.org/10.1044/1092-4388(2002)078)
- Bedore, L. M., & Leonard, L. B.** (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research*, 41(5), 1185–1192. <https://doi.org/10.1044/jslhr.4105.1185>
- Berman, R. A., & Verhoeven, L.** (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language and Literacy*, 5(1), 1–43. <https://doi.org/10.1075/wll.5.1.02ber>
- Bernstein Ratner, N., & Brundage, S. B.** (2018). *A clinician's complete guide to CLAN and PRAAT*. Retrieved from <https://talkbank.org/manuals/Clin-CLAN.pdf>
- Bernstein Ratner, N., & MacWhinney, B.** (2016). Your laptop to the rescue: Using the Child Language Data Exchange System archive and CLAN utilities to improve child language sample analysis. *Seminars in Speech and Language*, 37(2), 74–84. <https://doi.org/10.1055/s-0036-1580742>
- Bishop, D. V. M.** (2003). *The Children's Communication Checklist—Second Edition*. London, United Kingdom: The Psychological Corporation.
- Brimo, D., & Hall-Mills, S.** (2019). Adolescents' production of complex syntax in spoken and written expository and persuasive genres. *Clinical Linguistics & Phonetics*, 33, 237–255. <https://doi.org/10.1080/02699206.2018.1504987>
- Casby, M. W.** (2011). An examination of the relationship of sample size and mean length of utterance for children with developmental

- language impairment. *Child Language Teaching and Therapy*, 27(3), 286–293. <https://doi.org/10.1177/0265659010394387>
- Costanza-Smith, A.** (2010). The clinical utility of language samples. *Perspectives on Language Learning and Education*, 17(1), 9–15. <https://doi.org/10.1044/lle17.1.9>
- Dethorne, L. S., Johnson, B. W., & Loeb, J. W.** (2005). A closer look at MLU: What does it really measure? *Clinical Linguistics & Phonetics*, 19(8), 635–648. <https://doi.org/10.1080/02699200410001716165>
- Dickinson, D. K., McCabe, A., & Sprague, K.** (2001). *Teacher Rating of Oral Language and Literacy (TROLL): A research-based tool*. Ann Arbor: Center for the Improvement of Early Reading Achievement, University of Michigan.
- Ebert, K. D., & Scott, C. M.** (2014). Relationships between narrative language samples and norm-referenced test scores in language assessments of school-age children. *Language, Speech, and Hearing Services in Schools*, 45(4), 337–350. https://doi.org/10.1044/2014_LSHSS-14-0034
- Eisenberg, S. L., Fersko, T. M., & Lundgren, C.** (2001). The use of MLU for identifying language impairment in preschool children: A review. *American Journal of Speech-Language Pathology*, 10(4), 323–342. [https://doi.org/10.1044/1058-0360\(2001\)028](https://doi.org/10.1044/1058-0360(2001)028)
- Eisenberg, S., & Guo, L.-Y.** (2016). Using language sample analysis in clinical practice: Measures of grammatical accuracy for identifying language impairment in preschool and school-aged children. *Seminars in Speech and Language*, 37(2), 106–116. <https://doi.org/10.1055/s-0036-1580740>
- Fey, M. E., Cleave, P. L., Long, S. H., & Hughes, D. L.** (1993). Two approaches to the facilitation of grammar in children with language impairment. *Journal of Speech and Hearing Research*, 36(1), 141–157. <https://doi.org/10.1044/jshr.3601.141>
- Finestack, L. H., Bangert, K., & Huang, T.** (2017, November). *Using language samples to develop grammatical intervention goals*. Paper presented at the annual conference of the American Speech-Language-Hearing Association, Los Angeles, CA. Retrieved from http://www.finestackclil.com/wp-content/uploads/2017/11/Finestack-Bangert-Huang-ASHA-2017_final.pdf
- Fulcher-Rood, K., Castilla-Earls, A. P., & Higginbotham, J.** (2018). School-based speech-language pathologists' perspectives on diagnostic decision making. *American Journal of Speech-Language Pathology*, 27(2), 796–812. https://doi.org/10.1044/2018_AJSLP-16-0121
- Gavin, W. J., & Giles, L.** (1996). Sample size effects on temporal reliability of language sample measures of preschool children. *Journal of Speech and Hearing Research*, 39(6), 1258–1262. <https://doi.org/10.1044/jshr.3906.1258>
- Guo, L.-Y., & Eisenberg, S.** (2014). The diagnostic accuracy of two tense measures for identifying 3-year-olds with language impairment. *American Journal of Speech-Language Pathology*, 23(2), 203–212. https://doi.org/10.1044/2013_AJSLP-13-0007
- Guo, L.-Y., & Schneider, P.** (2016). Differentiating school-aged children with and without language impairment using tense and grammaticality measures from a narrative task. *Journal of Speech, Language, and Hearing Research*, 59(2), 317–329. https://doi.org/10.1044/2015_JSLHR-L-15-0066
- Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., & Golinkoff, R. M.** (2019). Building semantic networks: The impact of a vocabulary intervention on preschoolers' depth of word knowledge. *Reading Research Quarterly*, 54(1), 41–61. <https://doi.org/10.1002/rrq.225>
- Heilmann, J., Miller, J. F., & Nockerts, A.** (2010). Using language sample databases. *Language, Speech, and Hearing Services in Schools*, 41(1), 84–95. [https://doi.org/10.1044/0161-1461\(2009\)08-0075](https://doi.org/10.1044/0161-1461(2009)08-0075)
- Heilmann, J., Nockerts, A., & Miller, J. F.** (2010). Language sampling: Does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools*, 41(4), 393–404. [https://doi.org/10.1044/0161-1461\(2009\)09-0023](https://doi.org/10.1044/0161-1461(2009)09-0023)
- Hughes, D. L., Fey, M. E., & Long, S. H.** (1992). Developmental Sentence Scoring: Still useful after all these years. *Topics in Language Disorders*, 12(2), 1–12.
- Ingroud, C. M., Loeb, D. F., & Barlow, S. M.** (2019). Expressive language in preschoolers born preterm: Results of language sample analysis and standardized assessment. *Journal of Speech, Language, and Hearing Research*, 62(4), 884–895. https://doi.org/10.1044/2018_JSLHR-L-18-0224
- Klee, T.** (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders*, 12(2), 28–41.
- Lee, L. L.** (1974). *Developmental sentence analysis: A grammatical assessment procedure for speech and language clinicians*. Evanston, IL: Northwestern University Press.
- Lee, L. L., & Canter, S. M.** (1971). Developmental Sentence Scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders*, 36(3), 315–340.
- Leonard, L. B.** (2014). *Children with specific language impairment*. Cambridge, MA: MIT Press.
- Long, S.** (2012). 'Computerized Profiling' of clinical language samples and the issue of time. In M. Ball, D. Crystal, & P. Fletcher (Eds.), *Assessing grammar: The languages of LARSP* (pp. 29–42). Bristol, United Kingdom: Multilingual Matters.
- MacWhinney, B.** (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.
- MacWhinney, B.** (2018). *Tools for analyzing talk part 2: The CLAN program*. Mahwah, NJ: Lawrence Erlbaum Associates. <https://doi.org/10.21415/T5G10R>
- Miller, J. F.** (1981). *Assessing language production in children: Experimental procedures*. Baltimore, MD: University Park Press.
- Miller, J. F., Andriacchi, K., & Nockerts, A.** (2015). *Assessing language production using SALT software: A clinician's guide to language sample analysis* (2nd ed.). Middleton, WI: SALT Software.
- Miller, J. F., & Iglesias, A.** (2015). *Systematic Analysis of Language Transcripts* [Computer software]. Middleton, WI: Salt Software, LLC.
- Nippold, M. A., Hesketh, L. J., Duthie, J. K., & Mansfield, T. C.** (2005). Conversational versus expository discourse: A study of syntactic development in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research*, 48(5), 1048–1064. [https://doi.org/10.1044/1092-4388\(2005\)073](https://doi.org/10.1044/1092-4388(2005)073)
- Owen, A. J., & Leonard, L. B.** (2002). Lexical diversity in the spontaneous speech of children with specific language impairment: Application of D. *Journal of Speech, Language, and Hearing Research*, 45(5), 927–937. [https://doi.org/10.1044/1092-4388\(2002\)075](https://doi.org/10.1044/1092-4388(2002)075)
- Owens, R. E., Pavelko, S. L., & Babinelli, D.** (2018). Moving beyond mean length of utterance: Analyzing language samples to identify intervention targets. *Perspectives of the ASHA Special Interest Groups*, 3(1), 5–22. <https://doi.org/10.1044/persp3.SIG1.5>
- Owen Van Horne, A. J., Curran, M., Larson, C., & Fey, M. E.** (2018). Effects of a complexity-based approach on generalization of past tense -ed and related morphemes. *Language, Speech, and Hearing Services in Schools*, 49(3S), 681–693. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0142
- Pavelko, S. L., & Owens, R. E.** (2017). Sampling Utterances and Grammatical Analysis Revised (SUGAR): New normative values for language sample analysis measures. *Language, Speech, and*

- Hearing Services in Schools*, 48(3), 197–215. https://doi.org/10.1044/2017_LSHSS-17-0022
- Pavelko, S. L., & Owens, R. E.** (2019). Diagnostic accuracy of the Sampling Utterances and Grammatical Analysis Revised (SUGAR) measures for identifying children with language impairment. *Language, Speech, and Hearing Services in Schools*, 50, 211–223. https://doi.org/10.1044/2018_LSHSS-18-0050
- Pavelko, S. L., Owens, R. E., Ireland, M., & Hahs-Vaughn, D. L.** (2016). Use of language sample analysis by school-based SLPs: Results of a nationwide survey. *Language, Speech, and Hearing Services in Schools*, 47(3), 246–258.
- Plante, E., Ogilvie, T., Vance, R., Aguilar, J. M., Dailey, N. S., Meyers, C., . . . Burton, R.** (2014). Variability in the language input to children enhances learning in a treatment context. *American Journal of Speech-Language Pathology*, 23(4), 530–545. https://doi.org/10.1044/2014_AJSLP-13-0038
- Rice, M. L., 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(2), 333–349. [https://doi.org/10.1044/1092-4388\(2009/08-0183\)](https://doi.org/10.1044/1092-4388(2009/08-0183))
- Rice, M. L., & Wexler, K.** (2001). *Rice/Wexler Test of Early Grammatical Impairment*. San Antonio, TX: Pearson.
- Rojas, R., & Iglesias, A.** (2009). Making a case for language sampling: Assessment and intervention with (Spanish–English) second language learners. *The ASHA Leader*, 14(3), 10–13. <https://doi.org/10.1044/leader.FTR1.14032009.10>
- Scarborough, H. S.** (1990). Index of Productive Syntax. *Applied Psycholinguistics*, 11(1), 1–22.
- Semel, E., Wiig, E. H., & Secord, W. A.** (2004). *Clinical Evaluation of Language Fundamentals Preschool—Second Edition*. San Antonio, TX: Harcourt Assessment.
- Smith-Lock, K. M., Leitao, S., Lambert, L., & Nickels, L.** (2013). Effective intervention for expressive grammar in children with specific language impairment. *International Journal of Language & Communication Disorders*, 48(3), 265–282. <https://doi.org/10.1111/1460-6984.12003>
- Southwood, F., & Russell, A. F.** (2004). Comparison of conversation, freeplay, and story generation as methods of language sample elicitation. *Journal of Speech, Language, and Hearing Research*, 47(2), 366–376. [https://doi.org/10.1044/1092-4388\(2004/030\)](https://doi.org/10.1044/1092-4388(2004/030))
- Thordardottir, E. T.** (2008). Language-specific effects of task demands on the manifestation of specific language impairment: A comparison of English and Icelandic. *Journal of Speech, Language, and Hearing Research*, 51(4), 922–937. [https://doi.org/10.1044/1092-4388\(2008/068\)](https://doi.org/10.1044/1092-4388(2008/068))
- Thordardottir, E. T., & Ellis Weismer, S.** (2001). High-frequency verbs and verb diversity in the spontaneous speech of school-age children with specific language impairment. *International Journal of Language & Communication Disorders*, 36(2), 221–244. <https://doi.org/10.1080/13682820118239>
- Ukrainetz, T. A., & Blomquist, C.** (2002). The criterion validity of four vocabulary tests compared with a language sample. *Child Language Teaching and Therapy*, 18(1), 59–78. <https://doi.org/10.1191/0265659002ct2270a>
- Wagner, C. R., Nettelbladt, U., Sahlen, B., & Nilholm, C.** (2000). Conversation versus narration in pre-school children with language impairment. *International Journal of Language & Communication Disorders*, 35(1), 83–93. <https://doi.org/10.1080/136828200247269>
- Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W.** (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech and Hearing Research*, 38(6), 1349–1355. <https://doi.org/10.1044/jshr.3806.1349>