

Letter to the Editor

Use of Computerized Language Analysis to Assess Child Language

Julianne Garbarino,^a  Nan Bernstein Ratner,^a  and Brian MacWhinney^b 

We want to thank Pezold et al. (2020) for their excellent tutorial on using computer programs for language sample analysis (LSA) and for their fair and detailed comparison of the three programs reviewed. We are pleased to see language sampling programs receiving such serious attention in the clinical literature.

Readers will be happy to know that there have been many important improvements and updates to the Computerized Language ANalysis (CLAN; MacWhinney, 2000) software reviewed by Pezold et al. (2020). This new work is supported by grants from the National Institute on Deafness and Other Communication Disorders (NIDCD) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) designed specifically to facilitate clinical applications. Here, we will discuss these updates and provide clarifications about how CLAN computes many measures automatically using the KIDEVAL program.

When recommending a set sample length for LSA, it is important to think about the requirements of profiling instruments that have assumed a certain minimum number of eligible utterances. As Pezold et al. (2020) note, CLAN's KIDEVAL program can provide automatic computation of both the Index of Productive Syntax (IPSyn; Scarborough, 1990) and the Developmental Sentence Score (DSS; Lee & Canter, 1971). However, having just 50 utterances, as used in Sampling Utterances and Grammatical Analysis Revised (SUGAR; Pavelko & Owens, 2017, 2019) analyses, is not enough for computing these two important and valid indices. Not every utterance produced by a young child is

sufficiently complete to be included in these analyses. IPSyn requires 100 eligible utterances, as does traditional advisement on calculation of mean length of utterance (MLU), and DSS requires 50. However, once ineligible utterances are excluded, a corpus with only 50 utterances would likely not be large enough for DSS, as Pezold et al. note, let alone IPSyn.

Regarding computation of DSS, we would like to emphasize that KIDEVAL computes this complex measure fully automatically, providing both the total score and standard deviation scores measured against a large set of comparison transcripts. The authors appear to have computed DSS by hand. However, all of this can be done automatically. The overall KIDEVAL package outputs just the DSS and the standard deviation. However, if the clinician is interested in further detail, DSS can also be run by itself, outside of the KIDEVAL command, by simply typing the command "DSS <filename>". This will provide the full traditional DSS grid displaying observed elements in the child's sample, similar to the grids the authors filled out by hand and provided in their Supplemental Material S3 (Pezold et al., 2020). There is no need for practicing speech-language pathologists (SLPs) to conduct this analysis by hand. CLAN will provide both the total and computed utterance-by-utterance analysis quickly and automatically.

CLAN also performs an IPSyn analysis as part of the KIDEVAL report and can generate an IPSyn computation grid using the separate IPSyn command "IPSYN <filename>". For IPSyn, in order to compute the child's total score, SLPs must collect 100 eligible utterances. In practice, this will typically require more than 100 utterances, since strict inclusion criteria apply.

In regard to the analysis of lexical diversity, KIDEVAL standardizes number of different words (NDW) over a sample of 100 words. We should note that researchers have computed NDW in various ways, sometimes over a set number of utterances (Ukrainetz & Blomquist, 2002; Watkins

^aUniversity of Maryland, College Park

^bCarnegie Mellon University, Pittsburgh, PA

Correspondence to Julianne Garbarino: jgarbari@umd.edu

Editor-in-Chief: Shelley Gray

Editor: Sherrie Hill

Received December 10, 2019

Accepted January 2, 2020

https://doi.org/10.1044/2020_LSHSS-19-00118

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

et al., 1995) and sometimes over a set number of tokens (Thordardottir & Ellis Weismer, 2001). We agree with the authors when they note that it is also important to consider methods for standardizing computation of total number of words (TNW). This measure must absolutely be standardized over a time frame—one cannot usefully compare TNW gathered over a 30-min sample with one gathered over 10 min—or within a specific task or number of utterances. In addition, we note that computing this measure becomes trickier when the child is engaged in conversation, since part of the time frame is now subject to the behavior of the adult interlocutor. This may be one reason why TNW, as the authors note, tends to show low reliability. Finally, we also note the availability of a third measure of lexical diversity not discussed in the article, called *D* or *VocD*, that is available only in CLAN. This measure shows excellent growth profile over childhood while avoiding some of the pitfalls facing NDW and type–token ratio (Bernstein Ratner & MacWhinney, 2016).

When comparing alternative measures computed during LSA, important information presented in our article cited by the authors (Bernstein Ratner & MacWhinney, 2016) should be noted. In this article, we compared the apparent sensitivity of many LSA measures to growth in child age. Briefly, we found the following, using more than 600 Child Language Data Exchange System (CHILDES) samples:

- MLU fails to display a differential growth function after roughly 3 and a half years of age, making it potentially less useful for evaluation purposes; Brown (1973) warned of this many years ago, and he appears to have been correct. In addition, MLU on its own provides limited treatment guidance.
- Type–token ratio shows no apparent age function, whereas VocD shows a robust age function. Of the three LSA programs discussed, only CLAN provides VocD.

In regard to transcription, the authors appear to have first transcribed the samples in Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2015) and then imported them into CHAT (the format used by CLAN). While we understand why this was done for the purposes of comparing transcription programs, this is not encouraged for clinical purposes because it will slow users down. CLAN users can and should transcribe using CLAN. We suggest use of CLAN's Walker Controller function, which the authors describe in their Supplemental Material S1, to speed transcription time.

In regard to reference values for LSA analyses, the authors state that the corpora used in the KIDEVAL reference database are not individually described and that this makes it difficult to know whether they include children with similar demographic characteristics to the child whose language is being assessed. We agree that consideration of demographic factors of the comparison sample is extremely important, and we point out that descriptions of the individual corpora are available at <https://childes.talkbank.org/access/>. Children are compared to others of the same sex,

within a 6-month age band, in more detail than provided by many other sources. We fully acknowledge the authors' point that it is, at the moment, difficult for a user to easily obtain socioeconomic information about children in the comparison database. By the end of the current grant project, we plan to consolidate this information so that it is easy for clinicians to assess whether the normative data fairly apply to their client's profile. SLPs may also be interested to know that the federal support for CLAN KIDEVAL is also being used to develop additional LSA procedures more appropriate to evaluation of children who use African-American English.

Pezold et al. (2020) note,

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. (p. 107)

We would like to make two observations here. First, in CLAN, the SLP is not required to identify targeted morphemes because of the very accurate MOR parser that the authors describe. Secondly, we'd like to offer the opinion that omitting mazes in SUGAR may miss valuable information that is captured by both KIDEVAL and SALT outputs. There is growing concern that disfluency, or "maze," behaviors may provide information about the child's language abilities (Boscolo et al., 2002; Miller et al., 2015). With that said, while we encourage careful transcription of mazes and parent or examiner speech, neither CLAN software nor the KIDEVAL program requires this information. If an SLP wants to study only the basic aspects of a session, a CHAT transcription can be done using the same type of minimalist transcription used in SUGAR. (And, in fact, even more minimal transcription can be used because, as we and the authors have noted, CLAN's parser obviates the need for the minimal morphological parsing that must be done by hand in SUGAR and in SALT.) In general, practicing clinicians can and should use CLAN flexibly as a tool to accomplish their goals, rather than as a set of stringent guidelines.

We now turn to reviewing the updates that have been made to CLAN since the authors wrote their article. Some of these updates are critically relevant to concerns noted by the authors.

First, KIDEVAL, the CLAN analysis program described, has undergone many changes. One is the reference database: The comparison data set was larger, even in 2016, when the authors cite our publication discussing KIDEVAL—it had over 600 children then, and it has over 940 children now.

Second, KIDEVAL output no longer warns about using its comparison values for assessment purposes, thanks to improvements enabled by the federal grant referenced above. Each of our 6-month intervals in the comparison

database now have well over 100 transcripts, the size of normative sample that the authors and Andersson (2005) recommend. Moreover, database comparisons can also be retrieved separately for male or female children within a selected 6-month age range. Although the comparison databases are smaller, reference values are now also available for Mandarin Chinese and Japanese.

SLPs will be happy to know that, once a sample is transcribed, the complete SUGAR analysis can now be performed automatically in CLAN, using the command "SUGAR <filename>". As we have noted elsewhere (Guo et al., 2018), the morphemes counted in SUGAR analyses are different from those used in traditional analyses, such as those counted for the MLU measures found in KIDEVAL.

Finally, although the current KIDEVAL comparison database relies entirely on conversational samples, there is also work underway to include the numerous narrative samples in CHILDES for the purposes of LSA with this genre, which, as the authors note, is likely to differ substantially in profile from that of conversational or play samples.

In conclusion, we greatly appreciate Pezold et al.'s (2020) favorable discussion of the benefits of using programs such as CLAN's KIDEVAL to make language assessment quicker, more accurate, more in-depth, and more usable for establishing therapeutic goals. However, we also wanted to call attention to the clarifications and updates described above for transcribing, analyzing, and interpreting CLAN KIDEVAL results. Because we are only in the second year of our current federal funding, we anticipate making even greater improvements to the program, and we look forward to SLP feedback in this process.

Acknowledgments

Nan Bernstein Ratner and Brian MacWhinney are the principal investigators of the Child Language Assessment Project (funded by NIDCD R01-015494) and the CHILDES Project (funded by NICHD R01-082736), respectively. Julianne Garbarino is funded by NIDCD R01-015494 to Nan Bernstein Ratner.

References

- Andersson, L. (2005). Determining the adequacy of tests of children's language. *Communication Disorders Quarterly*, 26(4), 207–225. <https://doi.org/10.1177/15257401050260040301>
- 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>
- Boscolo, B., Bernstein Ratner, N., & Rescorla, L. (2002). Fluency of school-aged children with a history of specific expressive language impairment: An exploratory study. *American Journal of Speech-Language Pathology*, 11(1), 41–49. [https://doi.org/10.1044/1058-0360\(2002\)005](https://doi.org/10.1044/1058-0360(2002)005)
- Brown, R. (1973). *A first language: The early stages*. Harvard University Press.
- Guo, L.-Y., Eisenberg, S., Bernstein Ratner, N., & MacWhinney, B. (2018). Is putting SUGAR (Sampling Utterances of Grammatical Analysis Revised) into language sample analysis a good thing? A response to Pavelko and Owens (2017). *Language, Speech, and Hearing Services in Schools*, 49(3), 622–627. https://doi.org/10.1044/2018_LSHSS-17-0084
- 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. <https://doi.org/10.1044/jshd.3603.315>
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Erlbaum.
- Miller, J. F., Andriacchi, K., & Nockerts, A. (2015). *Assessing language production using SALT software: A clinician's guide to language sample analysis* (2nd ed.). LLC: SALT Software.
- Miller, J. F., & Iglesias, A. (2015). *Systematic Analysis of Language Transcripts* [Computer software]. SALT Software, LLC.
- 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(2), 211–223. https://doi.org/10.1044/2018_LSHSS-18-0050
- Pezold, M. J., Imgrund, C. M., & Storkel, H. L. (2020). Using computer programs for language sample analysis. *Language, Speech, and Hearing Services in Schools*, 51(1), 103–114. https://doi.org/10.1044/2019_LSHSS-18-0148
- Scarborough, H. (1990). Index of Productive Syntax. *Applied Psycholinguistics*, 11(1), 1–22. <https://doi.org/10.1017/S0142716400008262>
- 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/10.1191/0265659002ct227oa>
- 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>