**Title**: Reliability of Reduced Language Sample Length for the IPSyn: Research and Clinical Considerations

**Authors**: Valtierra, A.M.1, Byrd, E.2, Dietrich, M.S.1,2, Hall, M.L.3, Walker, A.2, Bustos, J.4,& McDaniel, J.5

**Introduction**: Language sample analysis (LSA) can provide information about children’s spontaneous, expressive language for research and clinical purposes. Using LSA to measure deaf and hard of hearing (DHH) children’s expressive language captures areas of weakness that are often overlooked by omnibus language assessments[1] and serves as an avenue for evidence-based language assessment and intervention. Currently, speech-language pathologists often use self-designed methods rather than standardized analyses for LSA[2]. Increasing the use of standardized LSA procedures, such as the Index of Productive Syntax (IPSyn)[3], is expected to benefit DHH children by systematically identifying and monitoring their language needs. However, a barrier for using the IPSyn, particularly for clinicians, is the time it takes to collect and analyze the required 100 child utterances[2,4-6]). Yang and colleagues[7] found that their shortened IPSyn (IPSyn-C) improved reliability over the original IPSyn when scoring from reduced samples of 50 rather than 100 utterances.

We extend this work to our sample of DHH children with the following research questions: Is the IPSyn scored from 50 utterances from play and conversation contexts sufficiently reliable for (RQ1) research purposes and (RQ2) clinical purposes? We addressed these questions via: (a) correlations between IPSyn scores from reduced 50-utterance and full 100-utterance samples, (b) correlations between IPSyn scores from the samples of the first 50 and second 50 utterances, (c) internal consistency of items for reduced and full samples, and (d) comparison of IPSyn and IPSyn-C reliability. Efficient LSA analyses are needed for research and clinical purposes. Yet, the relative importance of saving time versus maximizing precision may be different. Researchers may prioritize measurement precision over time to minimize error and maximize power[8]. On the other hand, clinically, time may be weighted more heavily than precision.

**Method**: We transcribed English language samples from DHH children aged 2 to 6 years in two contexts: play (*n* = 32) and conversation (*n* = 19), each 12-20 minutes long. We scored the IPSyn and IPSyn-C using the first 50, second 50, and first 100 eligible utterances to derive total scores and scores for each of the four subscales.

**Results**: To evaluate the correlation between scores from full and reduced samples, we calculated Spearman’s rho in play and conversation contexts between IPSyn scores derived from the first and second 50 utterances of the sample and IPSyn scores derived from the full sample of 100 utterances. As shown in Table 1, Spearman’s rho for all comparisons exceeded our criterion of .80, as did all subscale scores derived from the first 50 utterances from a play context. To test whether the IPSyn-C improved reliability compared to the IPSyn, we compared these same pairs of scores for the IPSyn-C. Spearman’s rho exceeded our criterion of .80 for three of the four comparisons on the IPSyn-C. However, as shown in Table 1, the IPSyn-C did not increase reliability compared to the IPSyn.

As a more stringent test of reliability, we calculated Spearman’s rho between IPSyn scores derived from the first 50 and second 50 utterances. As shown in Table 1, Spearman’s rho exceeded our reliability criterion (.80) for the IPSyn only in the play context for the total score with some subscales falling below .80. The same pattern held for the IPSyn-C.

Finally, to test internal consistency, Cronbach’s alpha for all items on the IPSyn was .93 for each of the three sample lengths (i.e., full sample, first 50, and second 50 utterances) combined across play and conversation contexts. When considering items by subscales, only the first 50 utterance sample exceeded the .70 criterion for all four subscales.

**Discussion:** For research purposes (RQ1), measures with high precision are desired to minimize measurement error. Given this relatively high value placed on precision, the IPSyn scores from 50 utterances did not demonstrate sufficient reliability, especially if subscale scores are of interest. Therefore, continued use of 100 utterance samples for the IPSyn is recommended for research purposes. Unlike the findings of Yang and colleagues (2022), the IPSyn-C did not increase any measures of reliability compared to the IPSyn for our sample. These results support the continued use of the IPSyn, rather than the IPSyn-C, which will also permit comparison of results with the extant literature.

Clinically (RQ2), reducing the required number of utterances from 100 to 50 to score the IPSyn would enable clinicians to use this tool for a wider range of DHH children and could save time. These benefits and the increased use of standardized LSA procedures over current practices may outweigh the reduction in precision that occurs when reducing the sample length. Our results support the use of 50 utterances to score the IPSyn for total score, with caution for interpreting subscale scores, especially if the alternative is to omit standardized LSA procedures. To save additional time, clinicians could score the IPSyn-C from 50 utterances. However, clinicians should consider available trainings and resources when deciding between the IPSyn and the IPSyn-C, as there are currently more resources for the IPSyn. By saving time on LSA, clinicians could collect language samples more frequently for progress monitoring and/or for more children on their caseload.

Table 1. *Spearman’s Rho for Total Scores and Subscale Scores*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sample | IPSyn | |  | IPSyn-C | |
| Play | Conversation |  | Play | Conversation |
| First 50 vs 100 Utterances – Total Score | .93 | .87 |  | .91 | .80 |
| Subscale Ranges | .85 – .92 | .73 – .91 |  | .69 – .93 | .69 – .85 |
| Second 50 vs 100 Utterances – Total Score | .93 | .84 |  | .94 | .75 |
| Subscale Ranges | .77 – .94 | .45 – .86 |  | .76 – .93 | -.27 – .86 |
| First 50 vs Second 50 Utterances – Total Score | .85 | .54 |  | .84 | .66 |
| Subscale Ranges | .67 – .81 | .09 – .87 |  | .66 – .81 | -.39 – .87 |

**References:**

[1]Werfel, K. L., & Douglas, M. (2017). Are we slipping them through the cracks? The insufficiency of norm-referenced assessments for identifying language weaknesses in children with hearing loss. *Perspectives of the ASHA Special Interest Groups*, *2*(9), 43-53. https://doi.org/10.1044/persp2.SIG9.43

[2]Wilder, A., & Redmond, S. M. (2024). Updates on clinical language sampling practices: A survey of speech-language pathologists practicing in the United States. *Language, Speech, and Hearing Services in Schools*, *55*(4), 1151-1166. https://doi.org/doi:10.1044/2024\_LSHSS-24-00035

[3]Scarborough, H. S. (1990). Index of productive syntax. *Applied psycholinguistics*, *11*(1), 1-22. https://doi.org/https://doi.org/10.1017/S0142716400008262

[4]Finestack, L. H., Rohwer, B., Hilliard, L., & Abbeduto, L. (2020). Using computerized language analysis to evaluate grammatical skills. *Language, Speech, and Hearing Services in Schools*, *51*(2), 184-204. https://doi.org/10.1044/2019\_LSHSS-19-00032

[5]Roberts, J. A., Altenberg, E. P., Ferrugio, H. R., & Rosenberg, J. E. (2022). How to use the Index of Productive Syntax to select goals and monitor progress in preschool children. *Language, Speech, and Hearing Services in Schools*, *53*(3), 803-824. https://doi.org/doi:10.1044/2021\_LSHSS-21-00096

[6]Scarborough, H. S., Rescorla, L., Tager-Flusberg, H., Fowler, A. E., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics*, *12*(1), 23-46. https://doi.org/10.1017/S014271640000936X

[7]Yang, J. S., MacWhinney, B., & Ratner, N. B. (2022). The Index of Productive Syntax: Psychometric properties and suggested modifications. *American Journal of Speech-Language Pathology*, *31*(1), 239-256. https://doi.org/doi:10.1044/2021\_AJSLP-21-00084

[8]Yoder, P. J., Symons, F. J., & Lloyd, B. (2018). *Observational measurement of behavior* (2nd ed.). Brookes Publishing.

Vanderbilt University

2 Vanderbilt University Medical Center

3 Temple University

4 Kansas State School for the Deaf

5 Vanderbilt University School of Medicine