

Alternatives to Traditional Language Sample Measures With Emergent Bilingual Preschoolers

Mark Guiberson

The aim of this article was to evaluate the use of alternatives to traditional language sample analysis (LSA) measures with bilingual preschool-aged children. A database of 184 bilingual preschool children, including children with developmental language disorder (DLD), was used in this study. The goals of the study were to establish the convergent validity between traditional language sample measures and alternative measures of clinician-observed and parent-reported longest utterance, compare groups across traditional and alternative measures, and establish to what extent alternative measures predict language status (typical or DLD). Results indicated that alternative measures may be useful in describing the language of preschool bilingual children, but none of the measures alone would provide adequate diagnostic information to make a determination of DLD status.

Key words: *alternative measures, bilingual, language sample analysis, preschoolers*

LANGUAGE SAMPLE ANALYSIS (LSA) can be a naturalistic and unbiased indicator of linguistic development in preschool-aged bilingual children (for a review, see Gutie Rrez-Clellen, Restrepo, Bedore, Peña, & Anderson, 2000; Rojas & Iglesias, 2006). However, transcription of 50–100 utterances is time-consuming and many speech-language pathologists (SLPs) report that their use of formal LSA is limited because of time constraints (Fulcher-Rood, Castilla-Earls, & Higginbotham, 2018; Pavelko, Owens,

Ireland, & Hahs-Vaughn, 2016; Westerveld & Claessen, 2014). Calculation of mean length of utterance (MLU) is a common LSA measure. The MLU is a global measure of syntactic complexity that is often used with preschool-aged children (Brown, 1973; Fenson et al., 1994). A child's longest utterance also has been suggested to be a good indicator of overall child language development (Brown, 1973). A parallel and alternate measure is parent report on a child's longest utterances, which has been found to be strongly associated with other measures of linguistic development in English- and Spanish-speaking children (Fenson et al., 1994; Guiberson, Rodríguez, & Dale, 2011). Given how labor-intensive LSA is, and the shortage of bilingual SLPs qualified to conduct LSA with bilingual children, alternative measures should be considered. Alternative measures based on longest utterance(s) observed or the longest utterance(s) reported by parents may be viable, practical, and efficient ways to describe the language development of bilingual preschool-aged children. The goal of this study is to complete exploratory analysis with

Author Affiliation: *Division of Communication Disorders, University of Wyoming, Laramie.*

This project was supported by grants from the National Center for Research Resources (5P20RR016474) and the National Institute of General Medical Sciences (8P20 GM103432) from the National Institutes of Health.

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

Corresponding Author: Mark Guiberson, PhD, Division of Communication Disorders, University of Wyoming, Dept 3311, 1000 E. University Ave, Laramie WY 82071 (mguibers@uwyo.edu).

DOI: 10.1097/TLD.0000000000000208

an existing corpus of data to establish the potential of alternative LSA measures when used with bilingual preschoolers. The author sought to (1) establish the convergent validity between traditional LSA measures, alternative LSA measures, and a standardized language measure; (2) compare traditional and alternative measures of a typically developing (TD) group with that of a group with developmental language disorder (DLD); and (3) complete exploratory analysis describing to what extent alternative measures predict language status.

METHOD

Participants

One hundred eighty-four bilingual preschool-aged children (aged 3;0–5;10 years) participated in this study. Children who participated in the study were emergent bilingual, predominately Spanish-speaking (spoke Spanish 80% of the time or more according to parent report), had normal hearing, and had no known neurodevelopmental disorders, cognitive disability, or other sensory impairments. Children were categorized as having DLD or as TD; DLD was established using triangulation of three sources of information: (a) identification by a bilingual SLP; (b) report of parent concerns about the child's language development; and (c) expressive language scores on the Spanish edition of the Preschool Language Scales–Fourth Edition (PLS-4 Spanish) of 77 or less (1.5 *SD* below the mean). The DLD group included 59 children (27 girls and 32 boys) and the TD group included 125 children (58 girls and 67 boys). There were no significant group differences in terms of the children's age ($t = -0.41$), children's percent Spanish use ($t = 1.08$), or parent's percent Spanish use ($t = 0.08$).

Measures, procedures, and reliability

Language sample analysis

The board book *Pato Está Sucio* by Kitamura (1998) was used to elicit language

samples from children. This book comprises seven parts that are illustrated across two pages each. The language in the book is very simple; the Appendix presents an English translation of the text in the book. The story is about a duck, the primary character, who gets dirty and faces a number of obstacles as he goes on a walk, until finally he washes off in a pond and is happy. Parents were asked to look at the book with their children as they normally would and then ask the children to tell them the story presented. Parents were not instructed to read the story, but some parents did opt to do so. All utterances produced by the child during this interaction were recorded and included in transcripts. The Systematic Analysis of Language Transcripts program (SALT; Miller & Iglesias, 2008) was used to obtain traditional LSA measures. The child's language was transcribed and segmented into clausal units (*C*-units).

Traditional language sample measures

The recommended traditional LSA measures for use with Spanish-speaking children were selected and included number of different words (NDW), total number of words (TNW), and mean length of utterance in words (MLU-W; Gutie Rrez-Clellen et al., 2000; Rojas & Iglesias, 2006).

Alternative language sample measures

Two alternative LSA measures were obtained from complete transcripts; these included length of longest utterance produced in words (LU-W) and average of three longest utterances in words (L3U-W). The L3U-W measure was calculated by adding the number of words produced for the three longest utterances provided and then dividing by three.

Reported longest utterance measures

Parents were asked to report the three longest utterances that they had heard their children say recently. From this two separate measures were obtained: longest reported utterance in words (rLU-W) and mean length of the three longest reported utterances in

words (r3LU-W). The r3LU-W measure was calculated by adding the number of words for each of the three utterances provided and then dividing by three.

Preschool Language Scales—Fourth Edition, Spanish

The PLS-4 Spanish is an assessment that includes receptive and expressive language subtests (Zimmerman, Steiner, & Pond, 2002). Using Plante and Vance's (1994) interpretation of sensitivity and specificity values, the expressive subtest of the PLS-4 Spanish has good sensitivity (0.92) and less than adequate specificity (0.68). To strengthen the less than adequate specificity, triangulation was used that included diagnosis of DLD by a bilingual SLP and parent report of concern of language development. The PLS-4 Spanish expressive subtest was administered and standard scores were calculated.

Procedure

The research team scheduled study visits with families at collaborating preschool centers in the Mountain West region of the United States to collect language samples and standardized language measures (PLS-4 Spanish). During these visits, parent report on utterances was collected as part of intake paperwork, and if left incomplete, this information was gathered by a member of the research team. Also during these visits, a Spanish-English bilingual SLP administered the PLS-4 Spanish in Spanish. Parents then showed their children the book *Pato Está Sucio* to elicit language samples. These study visits generally lasted between 30 and 45 min.

Reliability

A total of 13 bilingual graduate student coders were involved in language transcription using SALT. Coders received 8 hr of training in language transcription and completed transcription of three training videos. Before independently coding, they achieved 90% or higher point-by-point interrater agreement for word-for-word agreement and C-unit segmentation agreement. Interrater reliability

checks were completed with 20% ($n = 37$) of the language sample data. Interrater reliability for word-for-word agreement was 93% and interrater reliability for C-unit segmentation agreement was 97%. Seven of these graduate students were also involved in the hand calculation of the alternate LSA variables: L3U-W and r3LU-W. These students were trained by the author. In addition, reliability checks were completed with 20% ($n = 37$) of these calculations. Exact agreement for L3U-W calculation was 97%, and r3LU-W calculation was 100%.

RESULTS

Convergent validity between measures

One of the aims of this study was to establish the convergent validity between traditional LSA measures, alternative LSA measures, and a standardized language measure. Partial correlations (controlling for age) were completed with these measures in order to establish whether the alternative measures were evaluating similar constructs as traditional and standardized measures. Cohen's (2013) guidelines were used to describe effect size based on correlation magnitude. Table 1 presents the coefficients obtained. The p values were adjusted for multiple comparisons using a Benjamini-Hochberg correction. Of the traditional LSA measures, NDW was significantly associated with PLS-4 Spanish scores ($r = .33$, $p \leq .01$), with medium effect size observed, whereas TNW ($r = .23$, $p \leq .01$) and MLU-W ($r = .26$, $p \leq .01$) were significantly associated with PLS-4 Spanish scores, but with small effect sizes observed.

The alternative LSA measures, LU-W and L3U-W, were significantly associated with the traditional LSA measures of NDW and MLU-W ($r = .57-.81$, $p \leq .01$), with large effect sizes observed. Both LU-W and L3U-W had smaller but significant associations with parent report measures ($r = .23-.32$, $p \leq .01$), with small to medium effect sizes observed. Both parent report measures (rLU-W and r3LU-W) had

Table 1. Partial correlations, controlling for age, between language sample measures, reported linguistic measures, and standardized language scores

Measures	1	2	3	4	5	6	7	8
<i>Language sample measures</i>								
1. PLS-4 Spanish	-							
2. TNW	.23*	-						
3. NDW	.33*	.93*	-					
4. MLU-W	.26*	.34*	.30*	-				
5. LU-W	.32*	.57*	.57*	.76*	-			
6. L3U-W	.33*	.62*	.60*	.81*	.93*	-		
<i>Parent report measures</i>								
7. rLU-W	.47*	.14	.20**	.29*	.29*	.32*	-	
8. r3LU-W	.45*	.1	.17	.22**	.23*	.26*	.94*	-

Note. LU-W = longest utterance produced in words; L3U-W = three longest utterances in words; MLU-W = mean length of utterance in words; NDW = number of different words; PLS-4 Spanish = Preschool Language Scales-Fourth Edition, Spanish; rLU-W = longest reported utterance in words; r3LU-W = three longest reported utterances in words; TNW = total number of words. *p* values were adjusted for multiple comparisons using a Benjamini-Hochberg correction.

p* ≤ .01. *p* ≤ .05.

significant associations with the PLS-4 Spanish scores ($r = .45-.47$, $p \leq .01$), with medium effect sizes observed. However, both parent report measures had weaker associations with traditional LSA measures ($r = .10-.29$), with only half of these associations reaching significance.

Group comparisons of traditional and alternative measures

A second aim of this study was to compare traditional and alternative LSA measures used with TD children versus children with DLD. As a first step, means and standard deviations were examined; these are presented in Table 2. The TD group had higher traditional and alternative LSA values than the DLD group. Next, independent-samples *t* tests were performed to examine group differences across these variables. To control for Type I errors, a Bonferroni adjustment was calculated, and the level of significance was adjusted to $p \leq .01$. Mean difference effect sizes were estimated using Hedges' *g* (Hedges & Olkin, 2014) and Cohen's (2013) suggestions for interpretations. Of the traditional measures, significant group differences were detected for the TNW ($t = 2.97$, $p \leq .01$) and

NDW measures ($t = 4.40$, $p \leq .01$) but not for MLU-W ($t = 2.20$). A medium effect size was observed for the NDW group differences, and a small (approaching medium) effect size was observed for TNW. Of the alternative measures, LU-W ($t = 3.81$, $p \leq .01$) and L3U-W ($t = 3.75$, $p \leq .01$) values were significantly different, with medium effect sizes observed. The parent report measures, rLU-W ($t = 5.37$, $p \leq .01$) and r3LU-W ($t = 5.03$, $p \leq .01$), were significantly different for the two groups, with large effect sizes observed.

Exploratory analysis predicting language status

An exploratory logistic regression model was estimated to identify which variables may account for the most variability in language status. The model was developed specifically to identify how much variance traditional and alternative LSA measures accounted for when combined. Because of multicollinearity issues, several variables that were strongly intercorrelated had to be dropped from the model, including MLU-W, TNW, L3U-W, and r3LU-W. For the remaining variables (LU-W, NDW, and rLU-W), variance inflation factor values were acceptable (Leech, Barrett, &

Table 2. Descriptive and group comparisons for language sample analysis and parent report measures

	TD Group (<i>n</i> = 125)		DLD Group (<i>n</i> = 59)		<i>t</i> Tests	Effect Size, Hedge's <i>g</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
<i>Language sample measures</i>						
TNW	50.82	24.59	38.29	27.69	2.97*	0.49
NDW	29.14	12.47	20.66	11.59	4.40*	0.70
MLU-W	3.12	1.36	2.65	1.36	2.20	0.35
LU-W	6.74	2.86	5.08	2.47	3.81*	0.61
L3U-W	5.59	2.37	4.23	2.10	3.75*	0.59
<i>Parent report measures</i>						
rLU-W	7.42	2.78	5.07	2.82	5.37*	0.84
r3LU-W	6.05	2.32	4.13	2.38	5.03*	0.82

Note. DLD = developmental language disorder; LU-W = longest utterance produced in words; L3U-W = three longest utterances in words; MLU-W = mean length of utterance in words; NDW = number of different words; rLU-W = longest reported utterance in words; r3LU-W = three longest reported utterances in words; TD = typically developing; TNW = total number of words.

**p* ≤ .01.

Morgan, 2008). When LU-W, NDW, and rLU-W were considered together, they significantly predicted language status ($\chi^2 = 43.87$, *df* = 3, *N* = 183, *p* ≤ .001), accounting for 30% of the variability in language status. The model classified 90% of TD children correctly and 48% of the DLD children correctly.

DISCUSSION

The current study provided information about the potential of alternative LSA measures to describe the language development of emergent bilingual preschoolers. First, LU-W and L3U-W were significantly associated with traditional LSA measures, and significant group differences with medium effect sizes were observed for these measures. Studies of English-speaking toddlers and preschool-aged children have found that longest utterance measures parallel MLU and are a good predictor of future MLU values and language status (Smith & Jakins, 2014). Measures such as LU-W and L3U-W may appeal to clinicians who use real-time language sampling. There is a growing body of research describing SLPs' language sampling practices that have found

that SLPs frequently collect language samples in real-time while interacting with a child, guided by their own methods and clinical judgments (Fulcher-Rood et al., 2018; Pavelko et al., 2016; Westerveld & Claessen, 2014). To calculate LU-W and L3U-W through real-time transcriptions, SLPs could transcribe only the longer utterances they hear during their interactions with children. However, further research is needed to establish procedures and rules to obtain these measures, to evaluate the reliability of real-time sampling, and to establish whether LU-W and L3U-W are equally as informative for developmental levels beyond those typically seen in preschool-aged children.

A second finding was that parent report measures of utterance length appeared to provide descriptive developmental information. The measures rLU-W and r3LU-W were highly associated with PLS-4 Spanish scores, and significant group differences with large effect sizes were observed with these two measures. However, like the transcription-derived longest utterance measures, these parent report measures do not have the classification accuracy to be used alone to

identify DLD in young Spanish-speaking children. Nonetheless, best practices in identifying DLD in young bilingual children should involve multiple sources of converging information rather than overreliance on a single test or measure (Guiberson & Banerjee, 2012). Longest utterance observed and longest utterances reported appear to provide important *descriptive* information that could be clinically useful for assessment and progress monitoring purposes, especially if combined with other more robust measures.

LIMITATIONS

The current study included a sample of Spanish-speaking children living in the Moun-

tain West region of the United States. Further research is needed with other samples of young Spanish-speaking children. Future research also should evaluate whether the alternative LSA measures used in the current study, when combined with standardized assessment data, improve diagnostic accuracy.

Appendix

English translation of Satoshi Kitamura's *Pato Está Sucio*

Duck is going for a walk.

Uh-oh, it's raining.

Uh-oh, lots of mud.

Uh-oh, lots of wind.

Oops.

Splash.

That's better.

REFERENCES

- Brown, R. W. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), 1-185.
- 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.
- Guiberson, M., & Banerjee, R. (2012). Using questionnaires to screen young dual language learners for language disorders. In *14th Young Exceptional Children Monograph: Supporting young children who are dual language learners with or at-risk for disabilities* (pp. 75-93). Missoula, MT: Council for Exceptional Children Division for Early Childhood.
- Guiberson, M., Rodríguez, B. L., & Dale, P. S. (2011). Classification accuracy of brief parent report measures of language development in Spanish-speaking toddlers. *Language, Speech, and Hearing Services in Schools*, 42(4), 536-549.
- Gutierrez-Rez-Clellen, V. F., Restrepo, M. A., Bedore, L., Peña, E., & Anderson, R. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools*, 31(1), 88-98.
- Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Cambridge, MA: Academic Press.
- Kitamura, S. (1998). *Pato está sucio*. Mexico City, Mexico: Fondo de Cultura Económica.
- Leech, N. L., Barrett, K. C., & Morgan, G. A. (2008). *SPSS for intermediate statistics: Use and interpretation* (3rd ed.). Mahwah, NJ: Erlbaum.
- Miller, J., & Iglesias, A. (2008). Systematic Analysis of Language Transcripts (Research Version 9.1) [Computer software]. Madison, WI: Language Analysis Lab.
- Pavelko, S. L., Owens, R. E. Jr., 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., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, 25(1), 15-24.
- Rojas, R., & Iglesias, A. (2006). Bilingual (Spanish-English) narrative language analyses: Why and how? *Perspectives on Communication Disorders and Sciences in Culturally and Linguistically Diverse Populations*, 13(1), 3-8.
- Smith, A. B., & Jackins, M. (2014). Relationship between longest utterances and later MLU in late talkers. *Clinical Linguistics & Phonetics*, 28(3), 143-152.
- Westerveld, M. F., & Claessen, M. (2014). Clinician survey of language sampling practices in Australia. *International Journal of Speech-Language Pathology*, 16(3), 242-249.
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool language scale* (4th ed. Spanish). San Antonio, TX: Harcourt Assessment.

The CE test for this course is available online only at <https://alliedhealth.ceconnection.com/browse/professions/10>. Test takers can search for this course by title. Remember to login before taking the test.