Original Article

http://e-cacd.org/ eISSN: 2508-5948

Clinical Archives of Communication Disorders / Vol. 2, No. 1:43-50 / April 2017

Open Access

https://doi.org/10.21849/cacd.2017.00052

Relationships Among Conversational Language Samples and Norm-Referenced Test Scores

Robert E. Owens[1]*, Stacey L. Pavelko[2]

[1]Department of Communication Sciences and Disorders, The College of Saint Rose, Albany, NY, USA; [2]Department of Communication Sciences and Disorders, James Madison University, Harrisonburg, VA, USA

Purpose: Research demonstrates that many speech-language pathologists (SLPs) do not routinely include language sample analysis (LSA) in their clinical practice because LSA has limited recognition as a valid assessment measure. Limited research suggests that some LSA values obtained from narrative samples correlate with the results of standardized language tests. This research examined the relationship among values obtained from conversational language samples and the results of standardized testing.

Methods: This study investigated whether LSA values obtained from conversational language samples shared a relationship with the results of standardized language testing. A total of 16 children ages 43-90 months (M=61.5 months) completed three subtests of a standardized language test and a 15-minute conversational language sample. Fifty-utterance language samples were analyzed for four LSA values including mean length of utterance (MLU_S), total number of words (TNW), clauses per sentence (CPS) and words per sentence (WPS).

Results: Results revealed that three of the four LSA values (MLU_S, TNW, and WPS) demonstrated statistically significant ($ps < .006$) strong correlations ($rs > .65$) with the results of norm-referenced language testing. The partial correlations and the zero-order correlations were significant, suggesting age had little influence in controlling for the relationships.

Conclusions: Conversational language samples complement norm-referenced tests well. Results support further exploration of the relationships among LSA measures obtained from conversational samples and the results of standardized language testing.

Keywords: Language sample analysis, Language disorders, Language assessment

INTRODUCTION

Language sample analysis (LSA) has long been recognized as an integral part of a comprehensive assessment of language functioning in children [1]. Although LSA is considered one of the most ecologically valid methods of evaluating language production in children [2], many SLPs do not routinely include it as part of their clinical practice and rely primarily on standardized tests [3-7]. Pavelko, Owens, Ireland, and Hahs-Vaughn [7] surveyed 1,399 school-based SLPs regarding their use of LSA. Results indicated that one-third (33%) of the SLPs they surveyed did not use LSA at all in the previous school year. Of the SLPs who reported using LSA (67%), most (55%) reported analyzing less than ten samples per year despite having much larger caseloads.

Received: March 1, 2017
Revision: April 21, 2017
Accepted: April 26, 2017

Correspondence:
Robert E. Owens

Department of Communication Sciences and Disorders, The College of Saint Rose, 432 Western Avenue, Albany, NY, USA
Tel: +518-454-5258
Fax: +518-454-2083
E-mail: owensr@mail.strose.edu

Limited use of LSA has been similarly reported in both the U.S. and Australia [6,8]. In a survey of SLPs' perspectives and experiences of multilingualism in Australia, Williams and McLeod [6] reported that 20% (n=22) of respondents indicated "sometimes" using informal procedures to assess the language skills of children from multilingual backgrounds. Similarly, Westerveld and Claessen [8] reported that, although 90.8% of their respondents reported collecting spontaneous language samples, 11% reported never or rarely listening to the language samples they collected and 11% reported never or rarely transcribing the samples.

In contrast, most SLPs reported routinely using standardized assessments. Wilson, Blackmon, Hall, and Elcholtz [3] reported that nearly all SLPs they surveyed (265 of 266) indicated using standardized assessments during a language evaluation. Eickhoff, Betz and Ristow [5] reported that nearly 100% of the SLPs they surveyed rated standardized tests as one of the five most important assessment measures and 50% indicated standardized tests were the most important assessment measure. Similar results have also been reported for SLPs who work in Australia. Williams and McLeod [6] reported that 81.9% of their respondents indicated "always" or "sometimes" using English-only standardized tests when assessing the language skills of multilingual children. Westerveld and Claessen [8] reported 97.3% of the SLPs they surveyed reported using standardized language assessments when assessing children suspected of having a language impairment. Taken together, these results suggest that although LSA is recognized as an ecologically valid way to assess children's language, many SLPs continue to rely primarily on standardized tests when assessing children suspected of having language impairments and do not routinely use LSA.

Limitations to standardized testing

Researchers have cautioned SLPs to use standardized tests carefully due to concerns regarding diagnostic accuracy (sensitivity and specificity), the potential for cultural/linguistic bias and the possibility of over identifying children who speak nonstandard dialects or are English Learners [9-12]. Many tests do not report sensitivity and specificity levels [9]. Further, even in tests that report sensitivity and specificity, many report values below 80% [9]. Plante and Vance [10] suggested that 90% should be considered good accuracy, 80-89% fair accuracy, and tests with accuracy below 80% should not be used because misidentifications "occur at unacceptably high rates" (p. 21).

Spaulding, Plante, and Farinella [11] reviewed 43 standardized language tests and reported that only four of the test manuals provided sufficient information to allow clinicians to determine sensitivity and specificity values. Similarly, Betz and colleagues [9] surveyed 364 SLPs regarding their use of standardized tests. Results indicated that, of the ten most frequently used standardized tests, only two had acceptable levels of sensitivity and specificity.

With respect to issues of cultural/linguistic bias, McCabe and Champion's [12] study found that African American children from low-income backgrounds scored significantly lower than the normative sample on tests of vocabulary. The authors suggested that standardized tests should be chosen carefully because language varies "as a function of geography, socioeconomic status, and family practices" (p.168).

Because of these limitations, standardized tests may help define impairment, but are not sufficient to determine a disability [13]. Comprehensive language assessment does not rely solely, or even primarily, on norm-referenced assessment instruments to determine a student's communication abilities [14]. Researchers, test authors, and state education agencies have all advocated using a variety of assessments, such as LSA, dynamic assessment, criterion-referenced measures, and language sampling in combination with standardized testing when assessing the language abilities of children suspected of having language impairments [13-17].

Challenges to LSA use

Correlations with Standardized Testing. In contrast to standardized language testing, administered outside the normal contexts in which a child communicates that do not capture the complexities or the subtle nuances of the communication process, language sampling is elicited in natural and functional communication contexts and provides the type of naturalistic information required to plan relevant intervention [2,18]. Further, LSA has been called the gold standard in language assessment with bilingual children [19]. Despite this, SLPs report not using LSA because it has limited recognition as a valid assessment measure [7]. In an effort to validate the results of LSA, several researchers have correlated the results of standardized testing with various measures obtained from language samples. For example, Bishop and Donlan [20] reported positive correlations among mean length of utterance (MLU_S) in words and standardized test measures; however, the results did not reach significance.

Bedore, Pena, Gillam, and Ho [21] analyzed the relation-

ships among four measures taken from narrative language samples and children's language abilities as measured by the Bilingual English Spanish Assessment (BESA), a standardized test of language ability for Spanish-English bilingual children. Results indicated that both English MLU_S and Spanish MLU_S shared statistically significant correlations with BESA results ($ps < .05$).

Ebert and Scott [22] examined the relationships among eight measures taken from narrative language samples and 11 subtest scores from four different norm-referenced tests. Of the 73 participants, 70% had language impairment. Results indicated that MLU_S was not significantly correlated with one of the standardized test measures for older children (ages 9;1-12;8). In younger children (ages 6;0-8;11), MLU_S significantly correlated with seven of the 11 norm-referenced subtest scores (all $ps < .05$). In older children, total number of words (TNW) significantly correlated one of the norm-referenced language measures.

Finally, Ebert and Pham [23] analyzed the relationships among five language measures taken from narrative language samples and the scores from three different norm-referenced tests. The 51 bilingual children who participated in the study had been diagnosed with language impairment. In the younger group of children (ages 5;6-8;11), MLU_S significantly correlated with two of the five test scores ($ps < .05$). For the older children, MLU_S significantly correlated with one of the test scores ($p < .05$).

By demonstrating that at least some measures obtained from narrative language samples share significant relationships with the results of standardized testing, these studies are an important step in reducing the perception of LSA having limited recognition as a valid assessment measure. One limitation of these studies is they all used a narrative language sampling task. Because conversation is the most frequently used LSA task [7], it is important to document how LSA measures from conversational language samples intersect with the results of standardized testing.

Current study

Recently, Pavelko and Owens [24] collected 50-utterance conversational language samples from 270 typically developing children ages 3;0-7;11. Results indicated statistically significant age-related increases in MLU_S, TNW, clauses per sentence (CPS), and words per sentence (WPS). Additionally, results indicated that language samples could be collected, transcribed, and analyzed in approximately 20 minutes. Although these results suggest that LSA is an efficient and effective evaluation tool for clinical use, they are limited in that they do not address whether there might be associations between LSA measures and norm-referenced tests.

The associations between norm-referenced tests and measures obtained from narrative language samples have ranged substantially [22]. Emerging evidence suggests that these relationships are stronger in younger children, but are less robust with older children [22,23]. Notably absent from the literature is research examining the relationships between conversational language sample measures and norm-referenced tests. Because a comprehensive evaluation of language includes both standardized and non-standardized data, and conversation is the most frequently used LSA task [7], understanding how conversational language sample measures and norm-referenced tests interact is important in determining a pattern of strengths and weaknesses for a particular child.

The purpose of this study was to explore whether LSA values obtained from conversational language samples correlated with standardized test results. We extend prior work in this area in several ways. First, we examine conversational, rather than narrative, language samples. Second, we consider the role of age in the relationship between these two types of measures. Finally, we include subtests from a norm-referenced test measuring syntax, comprehension, and pragmatics. Because the study is exploratory, a range of outcomes is possible; however, we propose the main hypotheses as follows:

Hypothesis 1. Associations between norm-referenced tests and conversational measures will be strongest on expressive measures of syntax, followed by measures of paragraph comprehension, then measures of pragmatic language. Previous researchers have demonstrated that MLU_S correlates with some measures of standardized testing when using a narrative task [22,23]. Therefore, we expect the relationship between MLU_S and a measure of syntactic construction to be stronger than the relationship between MLU_S and comprehension or pragmatics.

Hypothesis 2. We expect TNW to be more strongly related to measure of syntactic construction than comprehension or pragmatics. Because previous researchers have indicated that the relationship between TNW from narrative language samples and norm-referenced language measures is only significant for older children [22,23], it is possible that, because the children in this study are younger, there will not be a significant relationship between the two.

Hypothesis 3. We expect the relationship between CPS and a measure of syntactic construction to be stronger than the relationship between CPS and comprehension or pragmatics. Researchers have indicated that a similar measure, subordination index, shared a significant relationship with norm-referenced test measures in younger children [22].

Hypothesis 4. We expect the relationship between WPS and a measure of syntactic construction to be stronger than the relationship between WPS and comprehension or pragmatics. Previous researchers have demonstrated that MLU_S in words, a similar but more general measure, correlates with some measures of standardized testing when using a narrative task [22,23].

METHODS

Participants

A total of 16 children (7 boys, 9 girls), ages 43-90 months (M=61.5 months), participated in this study. All participants were middle class, monolingual English speakers attending regular public schools. Of the 16 children, nine (56%) had previously been identified as having moderate specific language impairment (SLI) by a licensed and American Speech-Language Hearing Association (ASHA) certified SLP and were receiving intervention services in a college-based speech and hearing clinic. All participants were determined to have no hearing loss or co-occurring disorders, such as a learning disability.

Procedures and materials

Participants completed one testing session that consisted of a 15-minute digitally recorded language sample and administration of three subtest of the *Comprehensive Assessment of Spoken Language* (CASL) [25]. The order of presentation was randomized across participants. Language samples were collected using the Conversational Protocol proposed by Owens and Pavelko [24]. Briefly, the protocol offers children the opportunity to produce their most complex language by having examiners use a communicative interactional style, reducing children's one-word or minimal responses, and encouraging complex language though narrative elicitations and expository explanations. Samples were recorded on a digital recorder. Table 1 lists the participants' language sample results and scores on the CASL subtests.

Participants completed three subtests of the CASL, Pragmatic Judgement (PJ), Syntax Construction (SC), and Para-

Table 1. Participant performance on language and norm-referenced measures

Age	MLU_S	TNW	CPS	WPS	SC	PC	PJ
53	3.48	160	1	3.63	73	88	70
52	2.74	118	1	2.75	89	84	93
56	4.3	177	1	4.42	96	105	96
63	5.32	227	1.22	5.25	129	115	124
56	2.06	96	1	3.08	71	103	74
49	4.18	199	1.05	4.63	91	91	93
49	4.52	145	1.13	4.83	91	100	102
59	5.04	220	1.33	6.11	93	103	94
60	2.6	127	1	3.59	83	89	80
72	7.12	364	1.28	7.32	135	117	122
47	5.96	224	1.06	4.7	102	118	124
90	10.74	487	1.9	11.45	75	110	94
73	8.68	396	1.45	9.1	115	110	94
88	10.48	505	1.7	11.18	121	117	116
74	9.4	436	1.39	8.74	123	126	108
43	5.12	227	1.17	5.56	90	107	93

Table displays sample characteristics for all participants in terms of age, norm-referenced test scores, and language sample measures. Age is reported in months. Test scores are reported as standard scores (i.e., M=100, SD=15) for CASL subtests. SC=Syntax Construction; PC=Paragraph Comprehension; PJ=Pragmatic Judgement; MLU=mean Length of Utterance in morphemes; TNW=Total Number of Words; CPS=Clauses per Sentence; WPS=Words per Sentence.

graph Comprehension of Syntax (PC). The CASL was chosen because it is a frequently used language test [9]. These specific subtests were chosen because they are age-appropriate for all of the participants; included measures of both expressive and receptive language; allowed for expressive responses longer than a single word; and assessed differing aspects of language, rather than only language form. The PJ subtest of the CASL measures knowledge and use of pragmatic rules. In completing the subtest, children listen to a series of vignettes and then either judge the appropriateness of the language used in the vignettes or supply appropriate language for the situation. The SC subtest examines use of morphological rules to formulate and express sentences. For younger children, the subtest requires them to imitate, complete sentences using phrases, answer questions, and formulate sentences. The PC subtest measures comprehension of spoken syntax. In this subtest, children listen to a paragraph and then answer questions about the paragraph. Test responses for the CASL were recorded digitally and on test forms.

Training procedures

All researchers had previous experience collecting and analyzing language samples, completed the National Institutes of Health Protection of Research Participants tutorial, and completed two training sessions. Student researchers had participated in at least two semesters of in-house practicum in a college speech and hearing clinic or were finishing their second semester and were supervised by licensed and ASHA-certified SLPs. The first training session consisted of one hour on the administration and scoring of three CASL subtests. The second training session consisted of two hours on collecting, transcribing and analyzing language samples using the methods outlined in Pavelko and Owens [24].

Sampling transcription and analysis

Using the transcription and analysis procedures outlined in Pavelko and Owens [24], a trained researcher transcribed and analyzed each child's language sample in Microsoft Word® on a personal computer. Only the first 50 child utterances were transcribed. If three or more words in and utterance were unintelligible, the utterance was excluded and numbering continued with the next utterance. Otherwise, unintelligible words were included in the transcript and were indicated with "XX". Of the sixteen 50-utterance samples collected, only a total of six words from four different children were determined to be unintelligible, accounting for approximately 0.15% of the total words collected and less than 1% of the four children's utterances.

Samples were analyzed for four LSA metrics including TNW, MLU_S, WPS, and CPS. TNW was calculated by determining the total number of words the child produced, including any unintelligible words. MLU_S was calculated using the following rules, adapted from Pavelko and Owens [24]:

- One morpheme: All free morphemes, Brown's [26] 5 grammatical morphemes, 15 inflectional morphemes identified by Pavelko and Owens [24] (-ing, -s, -ed, dis-, -er, -est, -ful, -ish, -ly, -ment, re-, -sion, -tion, un-, and -y), and each word in a proper name (i.e., Aunt Sally);
- Two morphemes: all contractions (e.g., "don't", "won't", "I'm") and the words "hafta", "wanna", and "gotta";
- Three morphemes: the word "gonna".

Words per sentence was calculated by determining the number of sentences and then dividing the total number of words in those sentences by the number of sentences. Similarly, clauses per sentence were calculated by determining the number of clauses and dividing that number by the number of sentences. Although no sentence could have more than two clauses joined by the conjunction "and", multiple clauses could be joined by other morphemes [24].

Inter-rater reliability

The first author rescored approximately 30% of each child's recorded responses from the CASL. Inter-judge reliability was 99.6%. All disagreements were discussed and resolved. The resolved scores were used in the study.

The first author transcribed approximately 25% of each child's sample. Inter-judge reliability for utterance boundaries was 98.4%. Word-by-word inter-judge reliability was 98.5%. The first author also rated approximately 25% of the utterances in each sample for reliability. Inter-judge reliability for each value is as follows: morpheme count, 99.4%; word count, 99.7%; element count in the noun phrase, 98.2%; word count in the verb phrase, 98.6%; clause count, 96.4%; sentence count, 99.4%; and noun phrase count, 99.1%. Disagreements were discussed and resolved. The resolved samples were used for analysis.

Analyses

The distributions of all variables were examined prior to conducting any analyses. TNW and CPS showed evidence of non-normal distribution. Therefore, those measures were subjected to LOG10 transformations to increase normality.

Partial correlation analyses, controlling for the effects of age, were conducted to examine the association between language sample measures and norm-referenced tests. Age was controlled in these analyses because language sample measures would be expected to improve with development.

RESULTS

The results of the partial correlations between MLUs, WPS, and TNW, and each of the norm-referenced language measures while controlling for age were all significant (all *p*s < .014). However, zero-order correlations also revealed statistically significant, large correlations, indicating that age had very little influence in controlling for the relationships between MLUs, WPS, and TNW, and each of the norm-referenced language measures. Therefore, only the Pearson Product Moment Correlation (PPMC) results are reported (see Table 2). Specifically, MLUs was significantly correlated with SC (r (14) = .719, $p < .02$), PC (r (14) = .778, $p < .000$), and PJ (r (14) = .719, $p < .002$). TNW was significantly correlated with SC (r (14) = .765, $p < .001$), PC (r (14) = .767, $p < .001$), and PJ (r (14) = .722, $p < .002$). WPS was significantly correlated with SC (r (14) = .650, $p < .006$), PC (r (14) = .738, $p < .001$), and PJ (r (14) = .639, $p < .008$).

The results of the partial correlations between CPS and each of the norm-referenced language measures while controlling for age were indicated that only the relationship between CPS and PC was statistically significant (r (13) = .585, $p < .022$). However, zero-order correlations also revealed statistically significant, large correlations indicating that age had very little influence in controlling for the relationships correlations between CPS and PC (see Table 2).

DISCUSSION

Results discussed here should be considered exploratory due to the absence of tight controls on participant eligibility or assessment procedures. They may, however, reflect the realities of clinical practice. We consider our results below in terms of the four main hypotheses posed in the introductory section.

The relationship of MLUs to norm-referenced testing

Results indicated that MLUs significantly correlated with standardized test scores. These findings are similar to results reported by Ebert and Scott [22] and Ebert and Pham [23], who found significant correlations of MLUs calculated from narrative language samples with subtests of standardized language tests. Unlike previous research [22,23], the results from this study indicate that age had very little influence on the relationships between MLUs and each of the subtests because both the partial and the zero-order correlations indicated statistically significant relationships. One reason for this discrepancy could be differences in sampling context (narrative verses conversation), as well as differences in calculating MLUs. In both the Ebert and Scott [22] and Ebert and Pham [23] studies, MLUs was calculated in words, whereas the current study calculated MLUs in morphemes.

The relationship of TNW to norm-referenced testing

Results indicated that TNW significantly correlated with standardized test scores. These results are different from Ebert and Scott [22], who only found significant correlations of TNW calculated from narrative language samples with subtests of standardized language tests for older, but not younger, children. It is likely that differences in sampling context could ac-

Table 2. Bivariate Correlations among the Four LSA metrics and Three CASL Subtests

Measure	1	2	3	4	5	6
MLU	--					
TNW	.971*	--				
CPS	.836*	.764*	--			
WPS	.976*	.945*	.904*	--		
SC	.719*	.765*	.495	.650*	--	
PC	.778*	.767*	.692*	.738*	.862*	--
PJ	.719*	.722*	.527	.639*	.930*	.920*

MLU = mean length of utterance; TNW = total number of words; CPS = clauses per sentence; WPS = words per sentence; SC = Syntactic Construction; PC = Paragraph Comprehension; PJ = Pragmatic Judgement.
*$p < .006$.

count for the disparate findings. This study used a conversational protocol of 50-utterances, whereas Ebert and Scott [22] used a narrative protocol. Because other researchers have documented that narrative language samples from typically developing children can range from 9 to 14 C-units (Tilstra and McMaster (2007), it is possible that children completed the narrative in less than 50 utterances, thereby resulting in lower TNW scores.

The relationship of CPS to norm-referenced testing

Results indicated that CPS only shared a significant relationship with PC. This result is somewhat perplexing because previous work has indicted that a similar measure, subordination index (SI), shared significant correlations with the results of norm-referenced assessments [22]. It is possible that the lack of significant findings in the current study are due to differences in the norm-referenced assessments and/or the ages of the participants. Ebert and Scott [22] used the Clinical Evaluation of Language Fundamentals, whereas the current study used the CASL. Additionally, the participants' mean age of the younger participants in the Ebert and Scott [22] study was 7;5, whereas the mean age of the participants in the current study was 5;1.

The relationship of WPS to norm-referenced testing

Results indicated that WPS significantly correlated with standardized test scores. WPS is essentially the same measure as MLU_S in words. Ebert and Scott [22] and Ebert and Pham [23], both reported significant correlations of MLU_S in words with subtests of standardized language tests. Unlike previous research [22,23], however, the results from this study indicate that age had very little influence on the relationships between TNW and each of the subtests because both the partial and the zero-order correlations indicated statistically significant relationships. Perhaps the conversational sampling protocol used in this study encouraged children to produce longer utterances.

Clinical implications

The results reported in this paper offer some important clinical implications for SLPs. One reason SLPs cite for not using sampling is that LSA has limited recognition as a valid assessment measure [7]. This study identified three LSA measures (TNW, MLU_S, WPS) that significantly correlated with the results of standardized testing. Therefore, SLPs should have both increased confidence in the validity of these LSA measures and increase their use of LSA. The identification of valid LSA measures may enable SLPs to focus on these specific measures, thereby shortening the amount of time needed to analyze a language sample. Shortening the time involved in analysis has the potential to encourage more widespread use of LSA.

Second, many SLPs report having limited training or expertise in LSA [7]. SLPs could use the relatively easy techniques in the Pavelko and Owens [24] sampling protocol to collect a language sample. The sampling protocol offers assistance in structuring conversations that afford children a better opportunity than found in typical conversation to produce more complex language while being flexible enough to be easily adapted for use with many different clients. After collecting a language sample, SLPs could use the transcription rules and methods of calculation to quickly determine LSA values (see Pavelko and Owens [24]).

Limitations and future directions

Because this study is a first step in identifying clinically significant LSA values, there are several limitations. First, the authors acknowledge that the these LSA values are superficial in that they alone cannot adequately describe a child's language skills. Additional data, such as results of standardized tests, other non-standardized data, dynamic assessment, parent/teacher interviews, and language sample sub-analysis will be needed to fully describe a child's language and identify all appropriate intervention goals. We echo previous researchers in advocating that clinicians consider both tools when assessing children suspected of having language impairments.

This exploratory study suggests a number of directions for future work. The lack of a significant relationship between CPS and a measure of syntactic construction was unexpected and future research could further explore reasons accounting for this lack of finding. Exploring these relationships in older children (i.e., those older than 7;5) to determine whether stronger relationships exist under these circumstances could be helpful to clinicians when assessing school age children. Additionally, the lack of significant results for CPS could be due to the small number of participants included in the study. Finally, continued exploration of the relationship between conversational language samples and norm-referenced assessments is important in assisting clinicians in understanding how these measures intersect.

REFERENCES

1. Evans J. Plotting the complexities of language sample analysis: Linear and non-linear dynamical models of assessment. In: Cole K, Dale P, Thal D, editors. Assessment of communication and language. Baltimore: Paul H. Brookes; 1996.
2. Nippold MA. Language sampling with adolescents: Implications for intervention. 2nd ed. San Diego, CA: Plural Publishing; 2014.
3. Wilson KS, Blackmon RC, Hall RE, Elcholtz GE. Methods of language assessment: A survey of California public school clinicians. Lang Speech Hear Serv Sch. 1991;22:236-241.
4. Caesar LG, Kohler PD. The state of school-based bilingual assessment: Actual practice versus recommended guidelines. Lang Speech Hear Serv Sch. 2007;38:190-200.
5. Eickhoff JR, Betz SK, Ristow J. Clinical procedures used by speech-language pathologists to diagnose SLI. Symposium on Research in Child Language Disorders. 2010 June; Madison, WI.
6. Williams CJ, McLeod S. Speech-language pathologists' assessment and intervention practices with multilingual children. Int J Speech Lang Pathol. 2012;14:292-305.
7. Pavelko SL, Owens R, Ireland M, Hahs-Vaughn DL. Use of language sample analysis by school based SLPs: Results of a nationwide survey. Lang Speech Hear Serv Sch. 2016;47:246-258.
8. Westerveld MF, Claessen M. Clinician survey of language sampling practices in Australia. Int J Speech Lang Pathol. 2014;16:242-249.
9. Betz SK, Eickhoff JR, Sullivan SF. Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. Lang Speech Hear Serv Sch. 2013;44(2):133-146.
10. Plante E, Vance R. Selection of preschool language tests: A data-based approach. Lang Speech Hear Serv Sch. 1994;25:15-24.
11. Spaulding TJ, Plante E, Farinella KA. Eligibility criteria for language impairment: Is the low end of normal always appropriate? Lang Speech Hear Serv Sch. 2006;37(1):61-72.
12. McCabe A, Champion T. A matter of vocabulary II: Low-income African American children's performance on the expressive vocabulary test. Commun Disord Q. 2010;31(3):162-169.
13. Connecticut State Department of Education. Guidelines for speech and language programs: Determining eligibility for special education speech and language services under IDEA. Hartford: Author; 2008.
14. Virginia Department of Education. Speech-language pathology services in schools: Guidelines for best practice. Richmond, VA: Author; 2011.
15. Florida Department of Education. Exceptional student education eligibility for students with language impairments. Tallahassee: Author; 2010.
16. Zimmerman IL, Steiner VG, Pond RE. Preschool Language Scales. 5th ed. Bloomington, MN: Pearson; 2011.
17. Wiig E, Semel E, Secord W. Clinical Evaluation of Language Fundamentals. 5th ed. Bloomington, MN: Pearson; 2013.
18. Miller JF, Andriacchi K, Nockerts A, editors. Assessing language production using SALT software: A clinician's guide to language sample analysis. Middleton, WI: SALT Software LLC; 2015.
19. Heilmann JJ, Rojas R, Iglesias A, Miller JF. Clinical impact of wordless picture storybooks on bilingual narrative language production: A comparison of the "Frog" stories. Int J Lang Commun Disord. 2016;51:339-345.
20. Bishop D, Donlan C. The role of syntax in encoding and recall of pictorial narratives: Evidence from specific language impairment. Br J Dev Psychol. 2005;23:25-46.
21. Bedore LM, Peña ED, Gillam RB, Ho T-H. Language sample measures and language ability in Spanish-English bilingual kindergarteners. J Commun Disord. 2010;43:498-510.
22. Ebert KD, Scott CM. Relationships between narrative language samples and norm-referenced test scores in language assessments of school-age children. Lang Speech Hear Serv Sch. 2014; 45:337-350.
23. Ebert KD, Pham G. Synthesizing information from language samples and standardized tests in school-age bilingual assessment. Lang Speech Hear Serv Sch. 2017;48:42-55.
24. Pavelko SL, Owens RE. (2017). Sampling utterances and grammatical analysis revisited (SUGAR): New normative values for language sample analysis measures. Submitted for publication.
25. Carrow-Woolfolk E. Comprehensive assessment of spoken language (CASL). Torrance, CA: WPS; 1999.
26. Brown, R. A first language: The early stages. Cambridge, MA: Harvard University Press; 1973.
27. Cohen, J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
28. Kroecker J, Lyle K, Allen K, Filippini E, Galvin M, Johnson M, et al. Effects of student training on child language sample quality. Contemp Issues Commun Sci Disord. 2010;37:4-13.
29. Rice ML, Smolik F, Perpich D, Thompson T, Rytting N, Blossom M. Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. J Speech Lang Hear Res. 2010;53:333-349.