

Short-term Stability of Phonological Measures in a Sample of Two-year-old Late Talkers

Shari Leigh DeVeney, Kelsey O. Sheridan

University of Nebraska at Omaha, Omaha, Nebraska, United States

Purpose: When assessing phonological development, speech-language pathologists (SLPs) may use connected-speech samples and informal analyses. If the analyses' results are inconsistent, SLPs may misidentify children's abilities leading to inappropriate treatment decisions. Thus, use of reliable measures in the diagnostic process is paramount. The present exploratory investigation of short-term test/re-test stability included several informal phonological analyses with a clinically-significant population, late talkers.

Methods: Three male participants (24- to 31-months-of-age) identified as late talkers were video-recorded twice engaging in play-based parent-child interactions one week apart under near-identical circumstances. The samples were transcribed using broad transcription of the International Phonetic Alphabet. Following transcription, four informal analyses were completed and compared across data collection sessions: phonetic inventory and word shape analysis (independent analyses) and place-manner-voice analysis and percent consonant correct-revised (relational analyses).

Results: Results showed at least one inconsistent outcome across participants for each analysis. More particularly, one participant demonstrated inconsistent outcomes for phonetic inventory in initial and final position consonants, two participants produced inconsistent results for the word shape analysis, two participants indicated substantive differences across sessions on the articulatory substitution error categories (e.g., place, manner, and/or voice errors) on the place-manner-voice analysis, and variations in accurate consonant productions so substantive that it resulted in changes in severity rating for two participants as calculated using the percent consonant correct-revised.

Conclusions: These single case study findings, although limited, indicate potentially unreliable information associated with informal phonological measures. The present study findings provide important information warranting continued examination of this area of study.

Keywords: late talkers, phonology, phonological analysis, communication impairment



Received: October 24, 2017

Accepted: December 27, 2017

Correspondence:

Shari Leigh DeVeney

University of Nebraska at Omaha,
6005 Dodge Street, Omaha, Nebraska,
United States

Tel: +4025542993

Fax: +4025543572

E-mail: sdeveney@unomaha.edu

© 2017 The Korean Association of Speech-Language Pathologists

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

For speech-language pathologists (SLPs) conducting assessments with young children to determine the extent of their communication delays, part of that process typically involves collecting and analyzing a sample of the child's spontaneous conversational speech [1]. Although size, format, and topic of communication samples may vary [1-4], analysis of the sample for speech sound use through informal analyses is customary. These analyses often includes descriptive measures not compared to a normative database like those reported for formal standardized assessments such as the *Goldman-*

Fristoe Test of Articulation - 3 [5], *Structured Photographic Articulation Test* [6], and *Toddler Phonology Test* [7]. Rather, these descriptive measures involve independent and relational analyses indicating functional, naturalistic speech sound use [8] in order to obtain a holistic representation of a child's phonological system [9]. Even for children not yet 36-months-of-age, appropriate identification and treatment of phonological disorders may predict and prevent later negative consequences associated with speech sound deficits [10, 7, 11], showing a need for evaluation of these skills through formal and/or informal means. Regardless of the type of assessment tools used, SLPs need to use measures that are consistent and accurate because these instruments "serve as gateways to services" [1] (Crais, 342).

One informal instrument, an independent analysis of small speech samples, distinguishes sounds a child produces without comparison to typical adult productions. Independent measures are able to provide essential information about the speech sounds produced by children who present with many sounds in error, those who may not have a large amount of speech sounds, and/or those who are dual language learners [8, 1, 12]. For example, if the child produced "gog" for "dog," the child's production of the /g/ sounds would be recorded and analyzed without reference to the target word form. Conversely, relational analyses are used to determine the extent of speech-sound differences young children produce compared with adult production standards [13]. For example, comparing a child's production of "gog" to the adult form, "dog" and determining the severity of the child's speech sound delay based on the discrepancy between the target adult form and the child's production. These analyses are used by SLPs to determine treatment needs so reliability is essential. If SLPs are unable to accurately measure and assess the presence of a potential phonological delay, it is unlikely that they will be able to provide the most effective intervention course.

Few attempts have been made to study the effectiveness and reliability of informal measures. Limbrick, McCormack, and McLeod (2013) conducted a systematic review summarizing informal measures used for speech sound assessment in terms of the measures' conceptualization and operationalization [14]. Although the researchers found few informal measures addressed operational criteria and most lacked evidence of effectiveness for clinical use, they excluded measures used with words from spontaneous communication samples and only included measures conducted with predetermined word lists. Morris (2009) and Wittler and DeVeney (2016)

studied the reliable use of informal phonological measures with young children who were typically developing [12, 15]. Morris (2009) focused on children between the ages of 18-and-22 months who were typically developing. The children and their mothers were recorded during two 20-minute play samples which occurred one week apart under near identical circumstances. Recordings of these samples were analyzed with independent phonological measures including phonetic inventory, word shape analysis, syllable structure level, and index of phonetic complexity. The results showed high reliability for syllable structure level and the index of phonetic complexity; however, only moderate reliability was found with word-final phonetic inventory and word shape analysis, and word-initial phonetic inventory was not found to be a reliable measure. Discrepancy of informal independent analyses indicated they may not be representative of the child's speech sound range. Implications could be inaccuracy in baseline measurements used to determine and implement intervention such that perceptions of therapeutic progress may not truly represent progress, but instead may be "an artifact of an unstable measure." [12, p. 46]. Similar findings were noted in a pilot study conducted by Wittler and DeVeney (2016) [15] with three children without language delay. Participants were 25-to 33-months of age and the investigators used the same procedure of two 20-minute play samples recorded a week apart as Morris (2009) [12]. The findings indicated support for Morris' (2009) [12] results in regards to phonetic inventories. Consistency was obtained between two out of three participants phonetic inventories for word-initial sound productions, but inconsistencies were documented in two out of three participants in regards to word-final sound productions and consonant cluster production (i.e., production of two adjacent consonant sounds such as "sn" in "snake" or "pl" in "plate") over the two data collection sessions.

Although these findings indicate the potential for unreliable results for young children with typical development, SLPs primarily assessment and intervene with young children who show delayed speech and language use. For instance, informal analyses may be utilized when working with children who are 'late talkers.' Late talkers are young children identified with a language delay in the absence of any causal developmental disorders (e.g., autism spectrum disorder, intellectual deficits) [16]. For example, a 24-month-old child who is saying fewer than 50 words in his/her expressive vocabulary may be considered a late talker if other aspects of development are typical. Not only do late talkers exhibit language delays, but

they also exhibit delays in speech sound development. Late talkers have been found to produce fewer complex syllable structures in words [17], vocalize less often [18, 19], and use more limited speech sound repertoires [20, 21] compared to typically developing peers.

These two developmental systems - speech and language - influence each other in early development. Stoel-Gammon (1989) found a link between the size of a child's phonological repertoire and the size of his/her expressive vocabulary, such that children with smaller expressive vocabularies used relatively fewer speech sounds and children with more robust vocabularies used more [22]. Storkel and Morrisette (2002) theorized that children who use more words are able to produce more sounds, whereas children who use fewer words are not able to produce a large number of different sounds [23]. Several experimental studies have supported this theory that vocabulary and phonological development influence each other. For example, Girolametto, Pearce, and Weitzman (1997) found that when focusing treatment on expanding a child's expressive vocabulary, improvements in phonological diversity, the use of a wide array of different speech sounds, were also evident [24].

Though the relationship between an early language delay and delays in phonological development is well documented, little empirical evidence regarding the reliability of informal analyses of speech samples from young children presenting with language delays is available. The aim of the present study was to determine the test/re-test stability of selected independent or relational analyses of small speech samples for a clinically-relevant population of young children. Additionally, even for typically developing young children, little is known regarding the temporal stability of informal relational analyses. The purpose of the present study was to extend the work of Morris (2009) [12] and Wittler and DeVeney (2016) [15] in two ways: (1) replicate study procedures with a clinically-significant population who have similar language proficiency skills compared with the children who participated in the previous studies and (2) include informal relational analyses. The goal is to determine the test/re-test stability of four informal analyses of speech samples currently used by SLPs, two independent measures (phonetic inventory and word shape analysis) and two relational (percent consonants correct-revised and place-manner-voice analysis). In this exploratory pilot study, the investigators determined if the instability noted by Morris (2009) [12] and Wittler and DeVeney (2016) [15] was also indicated with participants who were late talkers. The fol-

lowing research questions was addressed:

1. What is the short-term test-retest reliability (over a one-week time period) of independent informal analyses speech samples calculated using intelligible words produced during a 20-minute speech sample for young children identified as late talkers?
2. What is the short-term test-retest reliability (over a one-week time period) of relational informal analyses of speech samples calculated using intelligible words produced during a 20-minute speech sample for young children identified as late talkers?

METHODS

All participant interactions, recruitment, and project procedures were conducted in accordance with the ethical standards of the University of Nebraska at Omaha and University of Nebraska Medical Center Institutional Review Board. The original research was approved by this governing body prior to the beginning of data collection (IRB #140-15-EP).

Participants

Participants were three males aged 24- to 31-months ($M=26.67$, $SD=3.79$). Participants were recruited through educational and childcare centers in a large Midwest-metropolitan area. Each participant was identified as a late talker based on their assessment performance and parental report using four measurement tools. (1) The *MacArthur Bates Communicative Development Inventory* (CDI) [25], a 680-word standardized and norm-referenced parent checklist used to measure expressive vocabulary, was administered along with (2) the *Preschool Language Scale - Fifth Edition* (PLS-5) [26]. The PLS-5 is a standardized, norm-referenced assessment instrument used to evaluate the receptive and expressive language skills of young children. (3) The *Modified Checklist for Autism in Toddlers* (M-CHAT) [27] a standardized screening tool to identify possible autism spectrum disorder (ASD) characteristics in toddlers was administered to determine risk for ASD and (4) the *Ages and Stages Questionnaire - Third Edition* (ASQ-3) [28], a standardized, norm-referenced assessment tool was administered to determine if young children are "at risk" for a variety of developmental skill sets (communication, gross motor, fine motor, problem solving, and personal-social).

In order to participate in the study, participants scored

within the 10th percentile on the CDI. Additionally, participants obtained a standard score ($M=100$, $SD=15$) of 85 or below on the expressive communication subtest of the PLS-5. All participants achieved a passing score on the M-CHAT and passed at least two subsections of the ASQ-3. All were monolingual native English speakers whose parents reported no concerns for hearing or vision abilities. See Table 1 for descriptive information about the participants. Additional intake information was collected through parent interview. Parents were asked about birth and developmental history, potential presence of sensory deficits, and unusual family circumstances that may influence their child's performance. All parents reported typical birth and developmental histories with no recent unusual family circumstances to report. All indicated concerns for delayed speech-language development; however, the first participant (P1) was the only child currently receiving speech-language intervention at the time of the study. Two parents reported some second language exposure (Japanese, Spanish) although all parents indicated English as the primary language spoken in the home.

Setting and procedures

Using procedures consistent with those of Morris (2009) [12] and Wittler and DeVeney (2016) [15], two 20-minute play-based parent-child interactions were recorded one week apart in a university clinic setting with the same age-appropriate toy sets available each time (i.e., grocery items, cars with garage, blocks and tools, and farm). Each room included an adjustable table, three to four chairs, a free-standing cabinet, a wall-mounted white board, and a video camera attached to a tripod. Communication samples were obtained during play interactions with the parent. Prior to each play session, parents were provided the same standard instructions from a graduate student research assistant. The instructions were, "I want to see what kind of activities ____ enjoys. I'd like to see how

____ communicates when s/he enjoys what s/he is doing. So, play and have fun. Help ____ enjoy what s/he's doing." The child and parent were instructed that they could utilize the toys if they would like either simultaneously or one at a time based on the child's interest. All the toy sets were initially located on top of the cabinet within the room so that the child needed to request a toy set in order to play with it. During each play-based session, parents and child were seated on the floor.

All sessions were video recorded for later review using a Cannon HD R500 camcorder with a mounted external microphone supported by an adjustable tripod. The camera and tripod were moved as needed during the sample to provide maximal view of the child's face. Because the sound quality from the camera with external microphone was sufficient for transcription, additional microphones were not employed during the data collection.

All child utterances from the two play-based samples were independently transcribed from video recordings by two graduate student research assistants trained in phonetic transcription using broad transcription techniques and the International Phonetic Alphabet (IPA). After each transcribed the entire data set ($n=6$ play-based sessions), inter-rater reliability was calculated at 81.83% (range = 72-91%). Only agreed upon consonant phones interpreted as belonging to the same word by both transcribers were included in the analysis corpus of the present study.

Using these agreed-upon IPA transcriptions, one graduate student research assistant calculated the independent measures of phonetic inventory (PI) and word shape (WS) analyses for each participant session. The other calculated the relational analyses place-manner-voice (PMV) analysis and percent consonants correct-revised (PCC-R). Inter-rater reliability for independent and relational analyses was established with the first author who re-analyzed 20% of the measures. In-

Table 1. Descriptive participant information

Descriptor	Participant		
	One (P1)	Two (P2)	Three (P3)
Age	25 months	31 months	24 months
Gender	Male	Male	Male
PLS-5 Exp. (SS %ile)	85 (21)	77 (6)	82 (12)
PLS-5 Aud. (%ile)	82 (16)	81 (10)	94 (34)
M-CHAT	PASS	PASS	PASS
CDI/CDI III (%ile)	<5%	<5%	<5%
ASQ-3	PASS 2/5 subsections	PASS 5/5 subsections	PASS 5/5 subsections

ter-rater reliability for PI range from 89-100% and was 100% for WS. For PMV, inter-rater reliability ranged from 84-93% and 89-95% for PCC-R.

Phonetic inventory

To obtain data for the PI analysis, the researchers documented the manner of production (i.e., stops, nasal fricative, affricate, glide, and liquids), word position of the phone (i.e., initial, medial, final word positions) and presence of consonant clusters in the sample. Phones produced twice in two different words within a given word position were noted as “productive” phones and differentiated from those phones produced fewer than twice across two different words within a given word position. These phone productions were classified as “emerging.” For example, if a child produced the /g/ sound in initial word position in both “go” and “game,” the phone would be classified as productive; however, if a child only produced an initial /g/ in “go” then the phone would be classified as an emerging sound. Productive and emerging sounds were combined for a total number of phones produced across word positions and clusters.

Word shape

WS analysis is a measure of word complexity and documentation of consonants (C) and vowels (V) present in a given word production. For example, the word “nut” /nʌt/ has a CVC structure while “peanut” /pinʌt/ has a more complex CVCVC structure. If a child produced “nu” /nʌ/, his/her individual production would be classified as CV regardless of the intended target word shape. There are a number of different conventions for measuring word shape complexity, for this study, the authors used the same procedures as Morris (2009) [12] and Wittler and DeVeney (2016) [15] which consisted of determining the consistent presence of eight different target word shapes: V, CV, CVCV, VC, CVC, CCVC, CVCC, and CVCVC.

Place-manner-voice

A PMV analysis is used to categorize and describe patterns of sound substitution errors. A child’s productions are compared to typical adult-standard productions in terms of place produced within the vocal tract, manner of air restriction during production, and presence of voicing [13]. For example, if a child produced “fog” /fɔg/ for “dog” /dɔg/, this relatively uncommon error pattern would be described as representing a difference in place (more anterior than target), manner (fricative

rather than a stop-plosive), and voice (/f/ is voiceless and /d/ is voiced) such that one errored production may ‘count’ in all three different production categories depending on the extent of differentiation from the target phone. For the present study, each participant’s speech sound productions were analyzed and described in terms of substitution patterns that resulted in deviation from adult-standard productions based on the procedural description from Williams (2003) [29].

Percent consonants correct – revised

To complete a PCC-R analysis, as with PMV, a child’s word production is compared to the adult-standard form. The number of consonants in a target word are recorded and compared to the number of consonants a child accurately generated in their own production of the word. For example, if the adult target is “cat” /kæt/ and a child produced “ca” /kæ/, he/she accurately produced one out of two consonants present in the adult form. Based on this example, a child would have 50% consonants correct. According to Shriberg, Austin, Lewis, McSweeney, and Wilson (1997) [30] and Tattersall and Dawson (2016) [6], use of PCC-R rather than PCC is more appropriate for analyzing the speech productions of young children because distortions of sound productions (e.g., lateral or frontal /s/ production distortion) are scored as correct. A Severity Rating Scale developed by Shriberg and Kwiatkowski (1982) [31] is associated with PCC and PCC-R when the calculated percentage is obtained from a conversational speech sample. The severity rating is as follows: mild = PCC of 85-100%; mild-moderate = 65-85%; moderate-severe = 50-65%; and severe = less than 50% (Shriberg & Kwiatkowski, 1982). To calculate PCC-R, Shriberg et al. (1997) [30] recommend use of a 5- to 10-minute continuous speech sample which many have interpreted as a sample consisting of 50 to 100 utterances [32]. It is worth noting that Shriberg (1982) developed the PCC measure for speech sound disorders in general, not phonological disorder specifically, and his findings were based on a study of children aged 4;1 to 8;6 [33]. Consequently, the severity increments (mild, moderate, etc.) were originally intended to be applied to children within this age range and it remains unclear if these categorical severities are appropriate for children outside this age range [34]. However, Shriberg and colleagues developed a reference database on hundreds of 3- to 17-year-olds with typical speech productions [35] available for use by SLPs for comparative norms within this age range. Because this measure is being extended to younger children than originally intended, the authors de-

terminated its inclusion in the current study was fitting in order to continue scientific exploration of the measure's viable use with young children who may be presenting with speech sound delays.

RESULTS

The present study represents a single case exploratory study using within-subject comparison [36] and, as such, results are displayed through visual representation of descriptive indicators.

Independent analysis reliability: Phonetic inventory

The visual analysis of the results for productive, emerging, and total consonants used in initial and final word positions is shown in Table 2 and Figure 1. Inconsistencies across session one (S1) and session two (S2) were indicated if a difference of three or more consonant productions were present in a target word position. This distinction was determined because it represented a difference of just over one standard deviation for both initial (2.75 for session one and 2.91 session two) and final consonants (1.48 for session one and 2.05 for session two) in the Morris (2009) [12] study findings and was used as a cutoff point for Wittler and DeVney (2016) [15].

Initial consonant productions

For productive initial consonants, the first and second participants, P1 and P2 respectively, were consistent in the number of consonant productions across the two sessions. P1 had six productive initial consonants during S1 and seven in S2. Similarly, P2 had six in S1 and six in S2. However, the third participant (P3) showed inconsistency in productive initial consonants with two in S1 and seven in S2. For emerging initial consonants, all participants were consistent. P1 had two during S1 and three in S2, P2 had two in S1 and one in S2, and P3 had two in both S1 and S2. In regards to total initial consonants produced, P1 and P2 were consistent with P1 who had eight in S1 and ten in S2 and P2 with eight in S1 and seven in S2. P3 was inconsistent in total initial consonants produced with four and nine, respectively.

Final consonant productions

For productive final consonants, P1 was consistent with four in S1 and three in S2. P3 was inconsistent with two in S1 and four in S2. For emerging final consonants, P3 was consistent with zero in S1 and one in S2 and P1 was inconsistent with

Table 2. Participant Descriptive Data: Sample-based Measures

Descriptor	Participants		
	P1	P2	P3
MLU^a			
Session 1	1.08	1.00	1.00
Session 2	1.00	1.22	1.08
Total Number of Different Words Used			
Session 1	22	10	7
Session 2	25	12	13
Total Number of Words Used			
Session 1	86	23	20
Session 2	55	23	40
PI^b			
Session 1			
Initial Consonants	6	6	2
Final Consonants	4	0	2
Session 2			
Initial Consonants	7	6	7
Final Consonants	3	0	4
Word Shape^c			
Session 1	2/6	2/4	4/5
Session 2	2/6	2/4	4/5
Consistency (%)	33%	50%	80%
PMV			
Session 1			
Place (# of errors)	3	3	2
Manner (# of errors)	0	3	2
Voice (# of errors)	2	1	1
Session 2			
Place (# of errors)	10	1	5
Manner (# of errors)	10	1	2
Voice (# of errors)	7	2	5
PCC-R			
Session 1 (% Severity)	47%	53%	71%
	(Severe)	(Mod-Severe)	(Mild-Mod)
Session 2 (% Severity)	34%	65%	50%
	(Severe)	(Mild-Mod)	(Mod-Severe)

^aMean Length of Utterance (MLU); ^bPhonetic Inventory (PI) for productive consonants, those produced in a particular position across at least two different words during the sample; ^cWord Shape Analysis of eight different target word shapes: V, CV, CVCV, VC, CVC, CCVC, CVCC, and CVCVC.

zero in S1 and three in S2. P2 consistently produced no final consonants in either S1 or S2. With total final consonants generated, P1 and P2 were consistent with P1 having four in S1

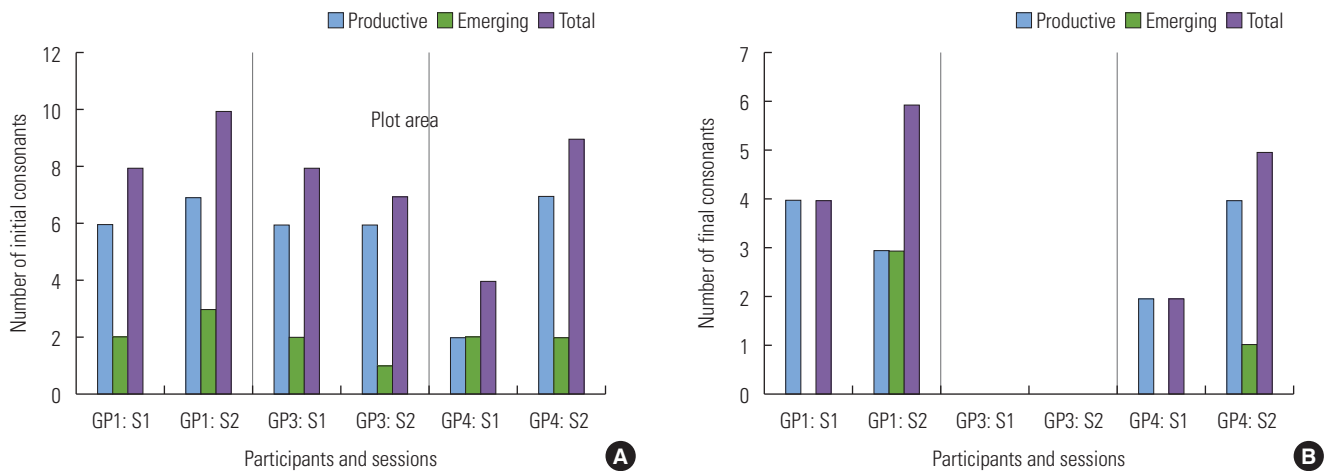


Figure 1. Phonetic inventories of initial (A) and final (B) consonants by session.

and six in S2 and P2 with no final consonants produced in S1 and S2. P3 was inconsistent with two in S1 and five in S2.

Independent analysis reliability: Word shape

When calculating target word shapes, participants were credited with use of a word shape if the sequence of sounds were produced in at least two different words across the communication sample. Specifically, the researchers targeted the following eight different word shapes: V, CV, CVCV, VC, CVC, CCVC, CVCC, and CVCVC. Inconsistencies were noted as a difference of at least 50% across the two data collection periods. Findings indicated inconsistent results for two of the three participants (see Table 2). For P1, two of six word shapes were consistently produced across the two data collection sessions indicating 33% consistency across the two sessions. For P2, two of four word shapes were consistently produced across the two data collection sessions indicated 50% consistency. P3 noted consistent productions of four of five different target words shapes resulting in 80% consistency.

Relational analysis reliability: Place manner voice analysis

Findings from the PMV analyses are shown in Table 2. Discrepancy criteria was met when the number of phone substitutions between the two communication samples differed by at least two noted substitutions across place, manner, or voicing within a particular word position. For P1, substantive differences were noted across sessions for place, manner, and voice errors. P2 did not exhibit substantial differences across the two sessions for any of the articulatory substitution error categories. P3 showed substantive differences across the sessions regarding place and voice errors, but not manner.

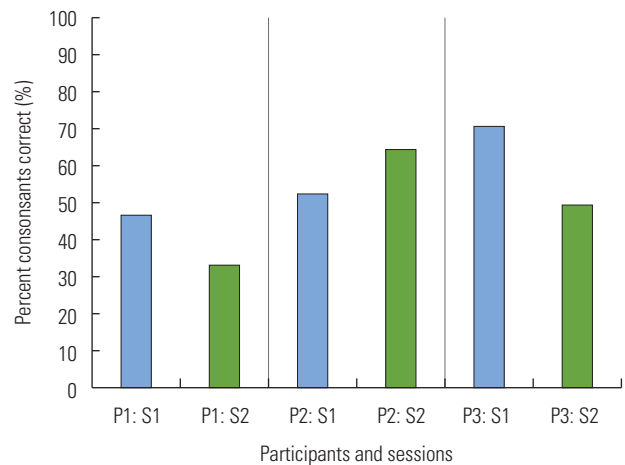


Figure 2. Percent Consonants Correct - Revised (PCC-R) by session.

Relational analysis reliability: Percent consonants correct-revised

Findings from the PCC-R are displayed in Table 2 and Figure 2. For the purposes of this study, discrepancy in PCC-R measures were noted when the calculated percentage resulted in a difference in severity rating. Differences in PCC-R noted across sessions 1 and 2 resulted in changes of severity rating for two of three participants. For P2, the session 1 percentage was 53%, corresponding with a moderate-severe severity rating; however, for session 2 the percentage was 65%, which corresponded with a mild-moderate severity rating. For P3, the session 1 calculation was 71% (mild-moderate) and 50% (moderate-severe) for session 2. For P1, the severity rating did not change across data collection sessions as both were within the severe rating at 47% and 34% respectively.

DISCUSSION

The present study aimed to investigate the test/retest stability of four informal analyses of small speech samples: phonetic inventory, word shape analysis, place manner voice analysis, and percent consonant correct-revised with late talking 2-year-olds. This is valuable information in regard to diagnostic data, goal development for therapeutic interactions, and intervention approach selection. Although single case study findings, such as the findings associated with the present study, have limited generalization, they provide important information regarding areas warranting further investigation (Kazdin, 2011). This function aligns with the purpose of the present exploratory pilot study, in which the investigators sought to determine if the instability noted by Morris (2009) [12] and Wittler and DeVeney (2016) [15] was also indicated with participants who were late talking, warranting continued examination of this area of study.

Informal independent analyses of speech samples

The findings related to the selected independent analyses of small speech samples, phonetic inventory and word shape analysis, were partially consistent with Morris (2009) [12] and Wittler and DeVeney (2016) [15]. Morris (2009) [12], who studied ten 18- to 22-month-olds with typical development, noted high reliability for two analyses not studied in the current project, syllable structure level and the index of phonetic complexity; however, she noted only moderate reliability for word-final phonetic inventory and word shape analysis, and word-initial phonetic inventory was not found to be a reliable measure. Additionally, she noted that the word shape analysis was reliable across data collection sessions. Wittler and DeVeney (2016) [15], who studied three 29- to 33-month olds with typical development, noted inconsistent word-initial and consonant cluster phonetic inventories for two of three participants, inconsistent word-final inventories for one participant, and consistency across analyzed word shapes for all participants.

Present study findings from three 24- to 31-month-olds identified as late talkers indicated inconsistent word-initial and word-final phonetic inventory for one of the three participants (P3) and two of the three participants (P2 and P3) demonstrated inconsistency on word shape results. These inconsistencies indicate the potential for unreliable use of these independent phonological measures obtained from play-based communication samples with a clinically-relevant population

of young children, late talkers.

One reason for these findings may be due to inherent variability in spontaneous communication samples, particularly across sample size. As Van Severen et al. (2012) noted [4], the size of the communication sample is strongly associated with the consonant inventory obtained from it. Namely, a larger sample is correlated with a larger consonant inventory [4]. Van Severen et al. (2012) [4] further determined that frequently occurring consonants are evident even in small sample sizes, but low incidence consonants need larger communication samples in order to be demonstrated and documented. Since communication samples are unstructured and do not require a standard set of stimulus items, unlike standardized measures of speech sound productions, there is also more room for production variance due to discrepancy in stimulus materials used as each can elicit different topics and, consequently, words and phones [2]. When sample size is not controlled, as was the case for Morris (2009) [12], Wittler and DeVeney (2016) [15], and the present study, reported inventories depend heavily on the size of the communication sample and may not be sufficiently reliable [4]. However, communication samples consisting of variable sizes, elicited through differing methods, are consistent with issues reflected in common clinical practices [3] and, as such, indicate an ecologically valid avenue of investigation. Ample sample size differences between Wittler and DeVeney (2016) [15] and the present study were indicated with 197-573 total words documented across participant samples and 20-86, respectively. Provided this variability in sample size, differences in documented consonant usage were not unexpected. However, given that instability in independent measures was noted across both large and small sample sizes for this age group, caution should be taken when using these measures for clinical use.

Impediments in early speech sound production associated with early language delay may also account for the present study findings of informal measure unreliability. Several researchers have noted the smaller, less complex array of phones used by young children with language delay compared with typical peers [21,22]. As with any small sample size, the communication samples of late talkers are vulnerable to individual differences. For example across context variances, engagement with one toy-stimulus item (i.e., toy car and garage) may elicit a narrow band of words and speech sound productions that may vary considerably from those produced when engaging with a different toy-stimulus item

(i.e., grocery set). Consequently, this clinically-relevant population may be more susceptible to changes in contextual settings than their peers with typical development who have a variety of words and speech sound productions that could be used to discuss either toy-stimulus item.

Informal relational analyses of speech samples

The present study findings indicated inconsistent findings across at least two of the three participants on both relational measures studied. For PMV, recorded responses for both P1 and P3 resulted in substantive differences across data collection sessions for substitution errors in place, manner, and voicing and place and voicing, respectively. Differences in PCC-R severity ratings were noted for both P2 and P3.

Because this study was the first of its kind in terms of an exploratory investigation into the reliability of informal relational analyses with young children identified as late talkers, it is difficult to identify a context for comparison from previous literature. Claessen et al. (2016) [10] theorized that PCC may be an unreliable predictive indicator of speech sound disorder for toddlers due to their many age-appropriate speech sound errors. And, as with informal independent measures, variabilities across sample size and contexts would likely be associated with differences in results. Additionally, delays and differences in early speech sound productions associated with late talkers compared to peers with typical development across syllable structure use in words [17], frequency of vocalizations [18, 19], and limited speech sound repertoires [20, 21] would undoubtedly affect a child's ability to accurately reproduce adult-standard word productions.

Inconsistencies in how these limitations are manifested may vary across sample size and sampling contexts. Although the targeted sample size for the present study was even greater than Shriberg and Kwiatkowski (1982) [31] and Shriberg et al. (1997) [30] recommendation in order to reliably measure the percentage of accurate consonant productions: 5- to 10-minutes of continuous speech. The number of words included in the 20-minute samples ranged from 20 to 86, as noted above, which is on the low end of the 50 to 100 utterances recommended for clinical practice [32]. In fact, only two of the six data collection sessions across the three participants included the recommended number of utterances and both were from the same the participant, P1. Given the constraints in obtaining a sample of appropriate size with late talkers who may only have 50 words in their entire expressive vocabulary, stable outcomes of the PCC-R measure with this population

present limited clinical utility. Shriberg et al. (1997) [30] presented transcription and metric reliability information on 33 children and adults for PCC and PCC-R calculations, the youngest of whom was 3;9 (45 months). The oldest participant from the present study, at 31 months, was still over a year younger than the youngest participant from Shriberg et al. (1997) [30]. Although the measure may be more appropriate for older preschool populations, use with toddlers should be interpreted with caution due to the potential for unreliable outcomes.

Limitations and Future Directions

A variety of limitations influence the findings of the present study. The small sample size of the clinically-relevant population cannot be generalized to the entire late-talking toddler population. Consequently, further research with a larger sample size would be needed to effectively generalize findings. However, given the intent of the present study to determine if further investigation is warranted, findings do indicate the need for continued study of informal analyses of speech samples and their use with late talkers in clinical settings. Similarly, future researchers may also consider utilizing additional clinically-relevant populations (e.g., young children with severe phonological disorders or those with apraxia of speech) to determine the presence of analogous and contrasting test/re-test reliability challenges across populations. Although present study procedures and analyses replicated those utilized by Morris (2009) [12] and Wittler and DeVeney (2016) [15], differences still exist in methodology, participant characteristics, and the other study conditions. Further investigations including replication and convergence are necessary to investigate test/re-test reliability of informal analyses and the various procedures that could be used to conduct such analyses beyond the four targeted in the present study.

Clinical Implications

This study has provided valuable information on test/re-test reliability of informal measures with late-talking toddlers and the potential for inconsistencies in analysis results. Investigating this issue with late talkers represents a clinical population whose results are more pertinent to the typical assessment and treatment decision-making within the speech-language pathology field than those of participants with typical development. Due to the reliance on informal measures for professionals documenting language and speech output during early childhood evaluations, these results indicate a potential

for inaccuracies during the data collection and interpretation process. Reliable assessment tools are essential for determining the need and goal for intervention. Therefore, caution is suggested when interpreting informal analyses of small speech samples obtained from the naturalistic communication of young children with early language delays. The present study also lends support to the notion of evidence-based practice in which informal analyses should be only one component of a comprehensive assessment and data from multiple sources should be compared for effective clinical decision-making.

CONCLUSIONS

The present study included three participants identified as late talkers and found inconsistencies in a variety of informal independent and relational analyses conducted using two conversational samples acquired one week apart under similar circumstances. Although further research is needed, the preliminary results of the current study suggest caution should be taken when utilizing results of these analyses to form goals and inform intervention with children who have a language delay. These analyses may be useful for diagnostic decision-making, but should be compared with other forms of data collection to determine a holistic view of a young child's communication abilities.

REFERENCES

1. Crais ER. Testing and beyond: Strategies and tools for evaluating and assessing infants and toddlers. *Language, Speech, and Hearing Services in Schools*. 2011 Jul 1;42(3):341-64.
2. Heilmann J, DeBrock L, Riley-Tillman TC. Stability of measures from children's interviews: The effects of time, sample length, and topic. *American Journal of Speech-Language Pathology*. 2013 Aug 1;22(3):463-75.
3. Pavelko SL, Owens RE, Ireland M, Hahs-Vaughn DL. Use of language sample analysis by school-based SLPs: Results of a nationwide survey. *Language, speech, and hearing services in schools*. 2016 Jul 1;47(3):246-58.
4. Van Severen L, Van Den Berg R, Molemans I, Gillis S. Consonant inventories in the spontaneous speech of young children: A bootstrapping procedure. *Clinical linguistics & phonetics*. 2012 Feb 1;26(2):164-87.
5. Goldman R, and Fristoe M. *Goldman-Fristoe Test of Articulation 3rd ed.* Circle Pines, MN: Pearson Education, Inc; 2015.
6. Tattersall P, Dawson J. *Structured Photographic Articulation Test - Third Edition featuring Dudsberry.* Dekalb, IL: Janelle Publications; 2016.
7. McIntosh B, Dodd BJ. Two-year-olds' phonological acquisition: Normative data. *International Journal of Speech-Language Pathology*. 2008 Jan 1;10(6):460-9.
8. Bernthal J, Bankson N, & Flipsen P. *Articulation and phonological disorders: Speech sound disorders in children.* Boston: Pearson; 2013.
9. Stoel-Gammon C. Phonological skills of 2-year-olds. *Language, Speech, and Hearing Services in Schools*. 1987 Oct 1;18(18):323-29.
10. Claessen M, Beattie T, Roberts R, Leitao S, Whitworth A, Dodd B. Is two too early? Assessing toddlers' phonology. *Speech, Language and Hearing*. 2017 Apr 3;20(2):91-101.
11. Dodd B. Assessment and intervention for 2 year olds at risk for phonological disorder. In C. Bowen (Ed.), *Children's speech sound disorders 2nd ed.* (pp. 88-94). Oxford: Wiley-Blackwell; 2015.
12. Morris SR. Test-retest reliability of independent measures of phonology in the assessment of toddlers' speech. *Language, Speech, and Hearing Services in Schools*. 2009 Jan 1;40(1):46-52.
13. Bauman J. *Articulatory and Phonological Impairments: A Clinical Focus.* 4th ed. Upper Saddle River, NJ: Pearson Education; 2012.
14. Limbrick N, McCormack J, McLeod S. Designs and decisions: The creation of informal measures for assessing speech production in children. *International Journal of Speech-Language Pathology*. 2013 Jun 1;15(3):296-311.
15. Wittler K, DeVeney SL. Test-retest reliability of independent phonological measures of 2-year-old speech: A pilot study. *Journal of Special Education and Rehabilitation*. 2016;17(3-4):71.
16. Rescorla L, Dale P. *Late talkers: language development interventions, and outcomes.* Baltimore: Brookes; 2013.
17. Williams AL, Elbert M. A prospective longitudinal study of phonological development in late talkers. *Language, Speech, and Hearing Services in Schools*. 2003 Apr 1;34(2):138-53.
18. Carson CP, Klee T, Carson DK, Hime LK. Phonological profiles of 2-year-olds with delayed language development: Predicting clinical outcomes at age 3. *American Journal of Speech-Language Pathology*. 2003 Feb 1;12(1):28-39.
19. Pharr AB, Ratner NB, Rescorla L. Syllable structure development of toddlers with expressive specific language impairment. *Applied Psycholinguistics*. 2000 Dec;21(4):429-49.
20. Paul R, Jennings P. Phonological behavior in toddlers with slow expressive language development. *Journal of Speech, Language, and Hearing Research*. 1992 Feb 1;35(1):99-107.
21. Rescorla L, Ratner NB. Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech, Language, and Hearing Research*. 1996 Feb 1;39(1):153-65.
22. Stoel-Gammon C. Prespeech and early speech development of two late talkers. *First language*. 1989 Jan;9(6):207-23.
23. Storkel HL, Morrisette ML. The Lexicon and Phonology Interactions in Language Acquisition. *Language, Speech, and Hearing Services in Schools*. 2002 Jan 1;33(1):24-37.
24. Girolametto L, Pearce PS, Weitzman E. Effects of lexical intervention on the phonology of late talkers. *Journal of Speech, Language, and Hearing Research*. 1997 Apr 1;40(2):338-48.
25. Fenson L, Dale P, Reznick J, Thal D, Bates E, Hartung J, et al. *MacArthur communicative development inventories.* San Diego, CA: Singular; 1993.

26. Zimmerman I, Steiner V, Pond R. *Preschool Language Scale – Fifth Edition*. San Antonio, TX: Psychcorp; 2011.
27. Robins DL, Fein D, Barton ML, Green JA. The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of autism and developmental disorders*. 2001 Apr 1;31(2):131-44.
28. Bricker D, Squires, J. *Ages & Stages Questionnaires*. Baltimore: Brookes; 2003.
29. Williams A. *Speech disorders resource guide for preschool children*. Clifton Park, NY: Singular Publishing Group; 2003.
30. Shriberg LD, Austin D, Lewis BA, McSweeney JL, Wilson DL. The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research*. 1997 Aug 1;40(4):708-22.
31. Shriberg LD, Kwiatkowski J. Assessing the severity of involvement. *Journal of speech and hearing disorders*. 1982 Aug;47:242-56.
32. Bliele K. *The manual of speech sound disorders: A book for students and clinicians*. 3rd ed. Stanford, CT: Cengage Learning; 2015.
33. Shriberg L. Diagnostic assessment of developmental phonological disorders. In M. Crary (Ed.). *Phonological intervention, concepts, and procedures*. San Diego, CA: College-Hill Inc; 1982.
34. Flipsen P. Severity and SSD: A continuing puzzle. In C. Bowen, *Children's speech sound disorders 2nd ed*. Oxford: Wiley-Blackwell; 2015.
35. Potter NL, Hall S, Karlsson HB, Fourakis M, Lohmeier HL, McSweeney JL, Tilkens CM, Wilson DL, Shriberg LD. Reference data for the Madison Speech Assessment Protocol (MSAP): a database of 150 participants 3-to-18 years of age with typical speech. Technical Report; 2012 May.
36. Kazdin A. *Single-case research designs: Methods for clinical and applied settings - Second Edition*. New York, NY: Oxford University Press; 2011.