Test-retest reliability of independent phonological measures of 2-year-old speech: A pilot study

Katherine Wittler
University of Nebraska at Omaha, kwittler@unomaha.edu

Shari L. DeVeney
University of Nebraska at Omaha, sdeveney@unomaha.edu

Follow this and additional works at: https://digitalcommons.unomaha.edu/spedfacpub

Part of the Special Education and Teaching Commons

Recommended Citation
https://digitalcommons.unomaha.edu/spedfacpub/23

This Article is brought to you for free and open access by the Department of Special Education and Communication Disorders at DigitalCommons@UNO. It has been accepted for inclusion in Special Education and Communication Disorders Faculty Publications by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.
TEST-RETEST RELIABILITY OF INDEPENDENT PHONOLOGICAL MEASURES OF 2-YEAR-OLD SPEECH: A PILOT STUDY

Katherine Marie WITTLER
Shari Leigh DEVENEY

Department of Special Education and Communication Disorders, University of Nebraska at Omaha, USA

Abstract

Introduction: Within the field of speech-language pathology, many assume commonly used informal speech sound measures are reliable. However, lack of scientific evidence to support this assumption is problematic. Speech-language pathologists often use informal speech sound analyses for establishing baseline behaviors from which therapeutic progress can be measured. Few researchers have examined the test-retest reliability of informal phonological measures when evaluating the speech productions of young children. Clinically, data regarding these measures are critical for facilitating evidence-based decision making for speech-language assessment and treatment.

Objectives: The aim of the present study was to identify the evidence-base regarding temporal reliability of two such informal speech sound measures, phonetic inventory and word shape analysis, with two-year-old children.

Rezime

Вовед: Во областа на патологијата на говорниот јазик, многумина претпоставуваат дека најчесто употребуваните неформални мерки на звукот на говорот се сигурни. Сепак, недостатокот на научни докази за поддршка на оваа претпоставка е проблематичен. Логопедите често користат звучни анализи на неформалниот говор за утврдување на основни-те однесувања според кои може да се мери терапевтскиот напредок. Неколкумина истражувачи ја испитале веродостојноста од тест-ретестирањето на неформалните фонолошки мерки при оценување на говорот кај мали деца. Од клинички аспект, податоците поврзани со овие мерки се од клучно значење за олеснување на донесувањето одлуки врз основа на докази а во однос на проценката и тетманот на говорниот јазик.

Цели: Целта на оваа студија беше да се идентификува базата на докази во врска со првата времената веродостојност на две такви неформални мерки на говорниот звук, фонетскиот инвентар и анализата на зборовната форма кај двегодишни деца.

Адреса за кореспонденција:
Шари ДЕВЕНЕИ
Department of Special Education and Communication Disorders, University of Nebraska at Omaha, 6005 Dodge Street, 512E Roskens Hall, Omaha, NE 68182; САД
Тел: (402) 554-2993, (402)968-6083
E-пошта: sdeveney@unomaha.edu

Corresponding address:
Shari DEVENEY
Department of Special Education and Communication Disorders, University of Nebraska at Omaha, 6005 Dodge Street, 512E Roskens Hall, Omaha, NE 68182; USA
Phone: (402) 554-2993, (402)968-6083
E-mail: sdeveney@unomaha.edu
Methods: The researchers examined analyses conducted from conversational speech samples taken exactly one week apart for three children 29- to 33-months of age. The videotaped 20-minute play-based conversational samples were completed while the children interacted with their mothers. The samples were then transcribed using the International Phonetic Alphabet (IPA) and analyzed using the two informal measures noted above.

Results: Based on visual inspection of the data, the test-retest reliability of initial consonant and consonant cluster productions was unstable between the two conversational samples. However, phonetic inventories for final consonants and word shape analyses were relatively stable over time.

Conclusion: Although more data is needed, the results of this study indicate that academic faculty, clinical educators, and practicing speech-language pathologists should be cautious when interpreting informal speech sound analyses based on play-based communication samples of young children.

Keywords: communication disorders, phonetic inventory, phonology, speech, speech-language pathology, word shape analysis

Introduction

Speech-language pathologists (SLPs) rely on formal (i.e., standardized, criterion-based testing) and informal (i.e., observations, clinical judgment, non-standardized) assessment measures to comprehensively evaluate a child’s speech and language development. These measures are used for determining eligibility for intervention services provided in educational or clinical settings. They also offer descriptive information critical to establishing a child’s baseline performance for therapeutic speech-language progress monitoring (1). Informal assessment tools are not intended to be compared to a larger group (i.e., not standardized or norm-referenced). Types of informal assessment measures include clinical observation, parent report, and analysis of a sample of the child’s speech obtained from a

JOURNAL OF SPECIAL EDUCATION AND REHABILITATION 2016; 17(3–4):71–88
DOI: 10.19057/jser.2016.11
Few studies have been conducted in order to examine the test-retest reliability of informal phonological measures like phonetic inventory and word shape analysis when evaluating the

...
The speech productions of young children. One researcher, Morris (3) evaluated the speech samples of ten typically developing 18-to-22 month old children using phonetic inventory and word shape analysis. Children were determined to be typically developing through use of a combination of language and vocabulary screening measures. Once eligibility was determined, each mother-child dyad participated in two 20-minute play sessions occurring exactly one week apart. The children’s speech was then analyzed from the videotaped samples using independent phonological analysis measures. Results of this study showed that the test-retest reliability of analyses conducted on two different speech samples collected from the same child one week apart were unstable and did not necessarily represent the same number or range of speech sounds produced. Particularly, Morris found that for children in this age range the number of initial consonant productions were the least stable measure over time. Final consonant productions and word shape analyses were determined to be moderately stable over the one-week time span, but neither reached a required level of significance. Morris noted that if the informal measures were not consistent over time, the perceived improvements during speech-language therapy may not represent true progress. Instead, it may be considered as “an artifact of an unstable measure” (3).

Preston, Ramsdell, Oller, Edwards, and Tobin suggested using a weighted measure for speech sound accuracy and referenced the Morris (3) study as a reason for doing so. The weighted measures used in their study, the Weighted Speech Sound Accuracy Measure (WSSA), was an independent phonological measure. (4) It functions as the Percent Consonants Correct (PCC) analysis, in which different types of errors are weighted differently. For example, unusual errors and

DOI: 10.19057/jser.2016.11
искористена за да се анализираат примероците на говор кај деца на различна возраст, и тоа и кај деца со нормален развој и кај деца со нарушена продукција на говорни звуци. Резултатите покажаа дека и веродостојноста и валидноста е висока кога WSSA се користи како мерка за анализи. Оваа истражувачка група овозможи дополнително оправдување за употреба на независни фонолошки мерки на анализи, но употребата на метриката на WSSA не била применета на пошироки контекsti на говор, вклучувајќи примероци од поврзан говор.

Друго истражување ја поддржа употребата на неформални релацијски мерки за фонолошка аналiza (на пример, споредување на исказот на детето на простор на зборовата форма на возрастните), бидејќи веродостојноста на тест-ретестирањето беше сила кога примероци од комуникацијата беа собрани од мали деца на училишна возраст, преку примена на интерви (6). Истражувачите собраа примероци на говор на 20 деца од градинка, коишто беа изложени на ризик од посепор развој на говор и/или јазик. Примероци беа земени со распокое од една неделa. Потоа, истражувачите ја измерија точноста на говорниот звук преку користење на неформални релацијски мерки. Наодите од студијата покажаа силна веродостојност на тест-ретестирањето на неформални релацијските анализи пресметани од структурирана примероци на комуникациjа за оваа возрастна група.

Van Severen, Van Den Berg, Molemans, и Gillis (7) ги проучувале ефектите од големината на разговорниот говор кај примероци од 30 холандски ученици на возраст од 6 до 24 месеци. Тие пресметале фонетски инвентари користејќи метода на повторување (на пример, тие постојано влечеле случајни примероци од разговорниот говор од еден поголем, подолг примерок). Истражувачите заклучиле дека биле пронајдени недоследности за оваа млада возраст и веродостојноста за понатамошни неформални аналиzi базирани врз примероци од разговорниот говор, во голема мера зависи од должината на вкупниот добиен примерок. Со оглед на претходните наоди на недоследности во врска со употребата на примероци од разговорниот говор на мали деца и веродостојноста на последователните неформални мерки, целта на овој наод беше основана. Оваа истражувачка студиjа имаше за цел да phoneme omissions are weighted more heavily than errors involving common substitutions. The WSSA was used to analyze speech samples from a variety of age ranges for both typically developing and disordered speech sound productions. Results indicated that both reliability and validity is high when the WSSA is used as an analysis measure. This research group provided further justification for the use of independent phonological analysis measures, but the use of the WSSA metric has not been applied to an extended variety of speech contexts including connected speech sampling.

Other research has supported the use of informal relational measures for phonological analysis (i.e., comparing the child’s utterance to the adult form of the word) because the test-retest reliability was strong when communication samples were collected for young school-aged children through a structured interview (6). Researchers collected speech samples from 20 kindergartners who were at risk for speech and/or language delays. The samples were taken one week apart. Then, the researchers measured speech sound accuracy using informal relational measures. The study findings indicated strong test-retest reliability for informal relational analyses calculated from structured communication samples for this age group.

Van Severen, Van Den Berg, Molemans, and Gillis (7) studied the effects of conversational speech sample size with 30 Dutch-speaking participants whose ages range from 6- to 24-months. They calculated phonetic inventories using a bootstrapping procedure (i.e., they repeatedly drew random conversational speech samples from a larger, lengthier sample). Researchers concluded that inconsistencies were found for this young age group and the reliability for further informal analyses based from the conversational speech sample depended heavily on the length of the overall sample obtained.

Given previous findings of inconsistencies related to the use of conversational speech samples of young children and reliability of subsequent informal measures, the aim of the present was established. This exploratory
go zgolemi deloto na Morris (3) преку одредување на тест-ретестирање на веродостојноста на независни фонолошки анализи кај малку постари деца, од 29 до 33 месеци, деца кои се 7 до 15 месеци постари од оние кои беа целна група во студијата на Morris (3). Со оглед на недостатокот на податоци во врска со веродостојноста на тест-ретестиранито фонолошки развој, во оваа област беше предложена пилот-работа од различни причини, вклучувајки и обезбедување на оправдување за понатамошна истрага преку собирање на предламарни податоци, проценка на одржливоста на техниките за анализа на податоци и проценка на потенцијалните различни исходи за идните студии.

Со цел постојано донесување одлуки врз основа на докази во текот на процесот на проценка и третман, SLP треба да иманат информации базирани на докази во врска со веродостојноста на неформалните независни фонолошки анализи собрани од комуникациските примероци на мали деца во различни фази на развојот на јазикот. Специфичната возрасна група беше цел на оваа студија, бидејќи во оваа фаза за време на типичниот јазичен развој, поголемиот дел од децата доживуваат брз раст на речникот што резултира со експресивна „експлозија“ на вокабуларот и почеток на искази составени од повече зборови. Освен тоа, многу мали деца кончтод одат на терапии за говор исто така почнуваат со транзиција од искази од еден кон искази со повече зборови и веродостојните мерки за документирање на фонолошките добивки во текот на овој процес се од суштинско значење за донесување одлуки, следење на напредокот, како и планирање на третманот.

Се осврнавме на следните истражувачки прашања:

1. Кај двегодишни деца, каква е краткочрочната (во рок од една недела) тест-ретестираната веродостојност на фонетскиот инвентар кога се пресметува со користење на 20-минутниот примерок од комуникација?

2. Кај двегодишни деца, каква е краткочрочната (во рок од една недела) тест-ретестираната веродостојност на анализата на зборовната форма кога се пресметува со користење на 20-минутниот примерок од комуникација?

study was aimed at extending Morris’ work by determining test-retest reliability of independent phonological analyses over time for a slightly older child population, 29- to 33-month olds, and children seven- to 15-months older than the sample used in the Morris study (3). In view of the scarcity of data regarding the test-retest reliability of phonological development, pilot work in this area was proposed for a variety of reasons including providing justification for further investigation through the collection of preliminary data, assessing the viability of data analysis techniques, and estimating potential outcome variability for future full-scale studies.

In order to consistently use evidence-based decision making throughout the assessment and treatment process, SLPs need to have evidence-based information regarding the reliability of informal independent phonological analyses collected from communication samples of young children across differing language development stages. The specific age range was targeted in the present study because at this stage during typical language development the majority of children experience rapid vocabulary growth resulting in an expressive vocabulary ‘explosion’ and the onset of multi-word utterances. Moreover, many young children receiving speech-language services are also beginning to transition from one- to multi-word utterances and reliable measures for documenting phonological gains during this process are essential for evaluation decisions, progress monitoring, and treatment planning. The following research questions were addressed:

1. For two-year-old old children, what is the short-term (within one week) test-retest reliability of phonetic inventory when calculated using a 20-minute communication sample?

2. For two-year-old old children, what is the short-term (within one week) test-retest reliability of word shape analysis when calculated using a 20-minute communication sample?
2. Materials and 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 #035-14-EP).

2.1 Participants

Participants included three children between the ages of 29- to 33-months of age. Children were identified as having no known delay in language development through completion and scoring of two screening measures. First, the Preschool Language Scale - Fifth Edition (PLS-5) was administered during the first of two experimental sessions (8). The PLS - 5 is a standardized, norm-referenced assessment instrument commonly used by SLPs in educational settings to evaluate the receptive and expressive language skills of preschool-aged children. In order to be included in the study as a child without a language delay, the participants needed to receive a standard score (M = 100, SD = 15) of 85 or above for total language development. The three participants scored above average on the expressive communication subtest (standard score range: 119-126) and one participant scored above average on the auditory comprehension subtest (standard score: 120). Other scores indicated language function within typical age-level expectations. In addition, the MacArthur Bates Communicative Development Inventory - Words and Sentences (CDI) was used as a screening measure (9). The CDI is a 680-word parent checklist that is standardized and norm-referenced to measure a young child’s expressive vocabulary. Again, this is a measure routinely used by practicing SLPs when evaluating the vocabulary skills of young children. Scores at or above the 25th percentile qualified a child to participate in the present study. All participants were monolingual native English speakers whose parents reported neither concerns for speech or language development nor hearing or vision abilities. See Tables 1 and 2 for descriptive participant information.
Табела 1. Описни информации за учесниците: стандардизирани мерки / 
Table 1. Participant Descriptive Data: Standardized Measures

<table>
<thead>
<tr>
<th>Опис / Descriptor</th>
<th>Учесник/Participant 1</th>
<th>Учесник/Participant 2</th>
<th>Учесник/Participant 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Возраст/Age</td>
<td>33 месяци / months</td>
<td>30 месяци / months</td>
<td>29 месяци / months</td>
</tr>
<tr>
<td>Пол/Gender</td>
<td>мачки / male</td>
<td>женски / female</td>
<td>женски / female</td>
</tr>
<tr>
<td>PLS-5 експ.² / PLS-5 Exp.²</td>
<td>126</td>
<td>119</td>
<td>119</td>
</tr>
<tr>
<td>Станд. резултат / Standard Score</td>
<td>95</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Процентил / Percentile</td>
<td>95</td>
<td>79</td>
<td>58</td>
</tr>
<tr>
<td>PLS-5 ауд.³ / PLS-5 Aud.³</td>
<td>120</td>
<td>112</td>
<td>103</td>
</tr>
<tr>
<td>Станд. резултат / Standard Score</td>
<td>91</td>
<td>79</td>
<td>58</td>
</tr>
<tr>
<td>CDI/CDI III⁴ / CDI/CDI III⁴</td>
<td>75 (CDI)/90 (CDI III)</td>
<td>90 (CDI)/90 (CDI III)</td>
<td>25</td>
</tr>
</tbody>
</table>

³ Предучилишна јазична скала – 5. издание, поттест за аудиоразбирање
⁴ MacArthur Bates Развој на инвентар за комуникација – зборови и реченици (CDI)/MacArthur Bates Развој на инвентар за комуникација – продолжение (CDI III) спроведено кај деца на возраст од над 30 месеци. Два учесника беа на возраст од 30 месеци или постари (P1 и P2), поради тоа и учесниците покажаа процентен опсег и за CDI и за CDI III.

Табела 2. Описни информации за учесниците: мерки базирани на примерок / 
Table 2. Participant Descriptive Data: Sample-based Measures

<table>
<thead>
<tr>
<th>Опис / Descriptor</th>
<th>Учесник/Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>ПДИ³ / MLU³</td>
<td>Учесник/Participant 1</td>
</tr>
<tr>
<td>Сесија 1 / Session 1</td>
<td>3.32</td>
</tr>
<tr>
<td>Сесија 2 / Session 2</td>
<td>3.70</td>
</tr>
<tr>
<td>Вкупно зборови / Total Words</td>
<td>402</td>
</tr>
<tr>
<td>Сесија 2 / Session 2</td>
<td>319</td>
</tr>
<tr>
<td>PI⁴</td>
<td>Серија 1 / Session 1</td>
</tr>
<tr>
<td>Серија 2 / Session 2</td>
<td>14</td>
</tr>
<tr>
<td>Последни согласни / Final Consonants</td>
<td>14</td>
</tr>
<tr>
<td>Серија 2 / Session 2</td>
<td>13</td>
</tr>
<tr>
<td>Групи согласни / Consonant Clusters</td>
<td>16</td>
</tr>
<tr>
<td>Серија 2 / Session 2</td>
<td>12</td>
</tr>
<tr>
<td>Форма на збор⁵ / Word Shape⁵</td>
<td>Серија 1 / Session 1</td>
</tr>
<tr>
<td>Серија 2 / Session 2</td>
<td>8</td>
</tr>
</tbody>
</table>

³ Просечна должина на изговор (ПДИ) 
⁴ Фонетски инвентар (ФИ) за продуктивни согласни употребени на определена позиција во најмалку два различни зборови во текот на примерокот. 
⁵ Форма на збор. Анализа на осум различни целини форми на зборови: V, CV, CVCV, VC, CVC, CCVC, CVCC и CVCVC

Табела 1. Описни информации за учесниците: стандардизирани мерки / 
Table 1. Participant Descriptive Data: Standardized Measures

<table>
<thead>
<tr>
<th>Опис / Descriptor</th>
<th>Учесник/Participant 1</th>
<th>Учесник/Participant 2</th>
<th>Учесник/Participant 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Возраст/Age</td>
<td>33 месяци / months</td>
<td>30 месяци / months</td>
<td>29 месяци / months</td>
</tr>
<tr>
<td>Пол/Gender</td>
<td>мачки / male</td>
<td>женски / female</td>
<td>женски / female</td>
</tr>
<tr>
<td>PLS-5 експ.² / PLS-5 Exp.²</td>
<td>126</td>
<td>119</td>
<td>119</td>
</tr>
<tr>
<td>Станд. резултат / Standard Score</td>
<td>95</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Процентил / Percentile</td>
<td>95</td>
<td>79</td>
<td>58</td>
</tr>
<tr>
<td>PLS-5 ауд.³ / PLS-5 Aud.³</td>
<td>120</td>
<td>112</td>
<td>103</td>
</tr>
<tr>
<td>Станд. резултат / Standard Score</td>
<td>91</td>
<td>79</td>
<td>58</td>
</tr>
<tr>
<td>CDI/CDI III⁴ / CDI/CDI III⁴</td>
<td>75 (CDI)/90 (CDI III)</td>
<td>90 (CDI)/90 (CDI III)</td>
<td>25</td>
</tr>
</tbody>
</table>

³ Предучилишна јазична скала – 5. издание, поттест за експресивна комуникација
⁴ MacArthur Bates Развој на инвентар за комуникација – зборови и реченици (CDI)/MacArthur Bates Развој на инвентар за комуникација – продолжение (CDI III) спроведено кај деца на возраст од над 30 месеци. Два учесника беа на возраст од 30 месеци или постари (P1 и P2), поради тоа и учесниците покажаа процентен опсег и за CDI и за CDI III.

Табела 2. Описни информации за учесниците: мерки базирани на примерок / 
Table 2. Participant Descriptive Data: Sample-based Measures

<table>
<thead>
<tr>
<th>Опис / Descriptor</th>
<th>Учесник/Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>ПДИ³ / MLU³</td>
<td>Учесник/Participant 1</td>
</tr>
<tr>
<td>Серија 1 / Session 1</td>
<td>3.32</td>
</tr>
<tr>
<td>Серија 2 / Session 2</td>
<td>3.70</td>
</tr>
<tr>
<td>Вкупно зборови / Total Words</td>
<td>402</td>
</tr>
<tr>
<td>Серија 2 / Session 2</td>
<td>319</td>
</tr>
<tr>
<td>PI⁴</td>
<td>Серија 1 / Session 1</td>
</tr>
<tr>
<td>Серија 2 / Session 2</td>
<td>14</td>
</tr>
<tr>
<td>Последни согласни / Final Consonants</td>
<td>14</td>
</tr>
<tr>
<td>Серија 2 / Session 2</td>
<td>13</td>
</tr>
<tr>
<td>Групи согласни / Consonant Clusters</td>
<td>16</td>
</tr>
<tr>
<td>Серија 2 / Session 2</td>
<td>12</td>
</tr>
<tr>
<td>Форма на збор⁵ / Word Shape⁵</td>
<td>Серија 1 / Session 1</td>
</tr>
<tr>
<td>Серија 2 / Session 2</td>
<td>8</td>
</tr>
</tbody>
</table>

³ Просечна должина на изговор (ПДИ) 
⁴ Фонетски инвентар (ФИ) за продуктивни согласни употребени на определена позиција во најмалку два различни зборови во текот на примерокот. 
⁵ Форма на збор. Анализа на осум различни целини форми на зборови: V, CV, CVCV, VC, CVC, CCVC, CVCC и CVCVC
2.2 Setting and Procedures

All data collection was conducted and recorded at a university campus speech-language pathology clinic. At this site, graduate students in speech-language pathology typically provided assessment and therapeutic services to a community-based clinical population. In the clinic, standard small individual therapy rooms included an adjustable table, three to four chairs, and a small cabinet. In addition, each therapy room was fitted with a remote recording system, the Interactive Session Recorder (ISR) system, set up to record audio and visual documentation of clinical interactions through video cameras secured in the ceiling.

After the first author administered the screening measures, two 20-minute conversational speech samples were collected from each child while he/she was interacting with his/her parent with toys available for play. Two different sets of toys (e.g., farm, grocery, kitchen, cars with garage) were randomly assigned to each participant for each of the two sessions conducted exactly one week apart. For each participant, the same two sets of randomly assigned toys were available each week. The child and caregiver were instructed to utilize the toys either simultaneously or one at a time based on their child’s interests. The sessions were recorded using the ISR system present in all clinic rooms which includes adjustable wall/ceiling mounted video and audio recorders. These recordings were later reviewed by research personnel for transcription and analysis. Each participant’s speech productions were transcribed using the International Phonetic Alphabet (IPA) and analyzed using two separate informal speech sound analyses; phonetic inventory and word shape. Reliability procedures were aligned with those used by Morris (3) in that all vocalizations from each session were reviewed and transcribed. Transcriptions were completed by the first author and a second transcriber. Second transcribers were students majoring in speech-language pathology trained in IPA transcription. In contrast to Morris, the researchers did attempt to achieve phoneme-by-phoneme agreement between the two
transcribers. Although, the researchers noted concern about lower transcription reliability using this method, the motivation for achieving a precise and accurate measure of each phonemic production was paramount. When instances of disagreement occurred between the first author and a second transcriber related to a particular phoneme, together the two reviewed the child’s production of the disagreed upon utterance and reached an agreement. If an agreement could not be reached after three review attempts, the utterance was not used in the final analysis. Initial inter-rater reliability was just above 62% (38% disagreement); however, after reviewing utterances with disagreement, transcribers resolved 100% of disagreements by discussing the videos together. Low inter-rater transcription reliability is a documented concern in the field of speech-language pathology, particularly when phonetically transcribing the connected speech of young children whose limited phonological development results in speech samples that are quite different from an adult’s sample of mature and well-formed speech production (10). Even with the more precise agreement requirements, the initial transcription reliability of the present study was 11% higher than that of Morris who noted 27% initial disagreement (3). The primary reason for the increase in disagreement is likely two-fold: (1) the present study researchers’ attempt to achieve phoneme-by-phoneme agreement and (2) data recording instrumentation. Morris used a combination of freestanding and remote microphones, while the present study used wall/ceiling mounted recording devices (3). Although used in a quiet, small room with little background noise interference, the distance of the recording device from the sound-source may have resulted in lower-quality sound capture for later transcription, which could have resulted in increased transcription variability.

The two informal speech sound analyses, phonetic inventory and word shape analysis, were conducted for each 20-minute speech sample for each participant. Results were compared between each child’s first and second sample through visual analysis.
3. Results

3.1 Phonetic Inventory

When calculating participant phonetic inventories, consonants were considered ‘productive’ for a particular word position if the child demonstrated use of the consonant in two different words during the sample. Consonants were considered ‘emerging’ if the child produced them in only one word or position during the sample. The visual analysis of the results for productive, emerging, and total consonants used (see Figure 1) indicated test-retest inconsistencies for the calculated phonetic inventories. An inconsistency was indicated if a difference of three or more consonant productions in a target word position were present. This cutoff point was selected because it represented a difference of just over one standard deviation for both initial (2.75 for session 1 and 2.91 session 2) and final consonants (1.48 for session 1 and 2.05 for session 2) in the study findings from Morris (3). Specifically, as evident in Table 2 and Figure 1, the second and third participants (P2, P3 respectively) produced an inconsistent number of productive initial consonants between the two sessions (15 in the first session; 18 in the second sessions and 4;10, respectively), while the first participant (P1) produced a consistent number of productive initial consonants (15;14). Visual analysis of productive final consonant productions by all three participants remained relatively consistent between the two sessions (14;13, 12;10, 7;7, respectively). However, inconsistencies were noted in the number of productive consonant clusters produced between the two sessions in two of the three participants. P1 and P2 used an inconsistent number of productive consonant clusters (16;12, 5;13), while P3 demonstrated the use of a consistent number of consonant clusters (4;4) across the two sessions.
Фонетски инвентар: Почетни согласки / Phonetic inventory: Initial consonants

Фонетски инвентар: Последни согласки / Phonetic inventory: Final consonants

Фонетски инвентар - Групи на согласки / Phonetic inventory - Consonant clusters

Слика 1. Фонетски инвентар на почетни согласки, последни согласки и групи согласки по сесија.

Figure 1. Phonetic inventories of initial consonants, final consonants, and consonant clusters by session.
3.2 Word Shape Analysis

When calculating the presence of target word shapes in the communication samples, participants were credited for a word shape if the sequence of sounds were produced in at least two different words across the sample. The researchers specifically analyzed eight different target word shapes including V, CV, CVCC, VC, CVC, CCCV, CCVC and CVCC. As shown in Table 2, no substantive differences were found across the two sessions for any participant. All participants produced at least two different words in each of the eight target word shape categories. Consequently, word shape analysis findings were consistent for all participants across the two sessions.

Discussion

The present study represents an early attempt to begin to determine the test-retest reliability of two different informal independent phonological analyses, phonetic inventory and word shape analysis for 2-year-old children with no known language delay. Clinically, data concerning these measures are quite valuable for facilitating evidence-based decision making for speech-language diagnosis and intervention. In the present study, results for the first research question regarding the test-retest reliability of phonetic inventory of 29- to 33-month old children calculated from a 20-minute communication sample indicated mixed findings. The results indicated partial support for previous conclusions regarding inconsistent measures of phonetic inventory (3, 7). Two of the three participants obtained inconsistent phonetic inventories for word-initial sound productions and consonant cluster productions (i.e., production of two adjacent consonant sounds such as “sn” in “snake” or “pl” in “plate”), while two of the three indicated consistent profiles for word-final sound productions over the two data collection sessions. These findings indicated that two of the three participants produced approximately the same number of productive final consonants in Session 1 as they did in Session 2, but did not produce approximately the same number of productive initial consonants or consonant clusters across sessions. Morris found that for 18- to 24-month old children, final sound productions were more stable than initial...
The second research question addressed in the present study regarding the test-retest reliability of a word shape analysis calculated using the 20-minute communication sample. In this instance, findings were consistent with Morris (3) in that all participants demonstrated consistent word shapes over the two sessions. The present study findings indicated that measure used for word shape analysis was more temporally reliable than phonetic inventory over a one-week period with three 29- to 33-month old children.

The use of a 20-minute communication sample, which was consistent with Morris’ procedures, may be at cause for these differences in measurement consistencies (3). As noted, a standard communication sample measurement time has not been established and applied consistently in the speech-language pathology professional community (7). The present findings support further investigation into the optimal communication sample time used for analysis since the findings were inconsistent across studied measurement analyses. For instance, the 20-minute play-based sample time allotted adequate time for each participant to demonstrate multiple uses of a number of different word shapes, but may not have allowed adequate time for each to demonstrate the use of all of the speech sounds in their expressive repertoire. More research into these important variables is needed, particularly as they relate to the analysis of speech for young children.
4.1 Educational and clinical significance

Since no measurement method – formal or informal - will have perfect agreement across administrations (e.g., the CDI has a correlation across one-month test-retest administrations that ranges from 0.61-0.95) (9), the finding that a few instances of difference across sampling times does not directly provide evidence that the measure is inappropriate for clinical use. Rather, when informal phonological measures are used in clinical settings for evaluations, goal development, and/or progress monitoring over time, clinicians should use caution in interpretation. Multiple sources of data should be considered during an evaluation and when monitoring therapeutic progress so that decisions and reports are based on more than one converging measurement tool. Additionally, the researchers urge caution among speech-pathology academic and clinical educators when advising graduate students to use informal independent phonological measures calculated from a young child’s communication sample. Because the temporal reliable for phonetic inventories has not be established for children under three-years of age, we should use caution when advocating for their use with this young population. In addition, practicing SLPs working with a young clientele should exert similar restraint when using phonetic inventories for baseline or descriptive information. While these measures provide descriptive information helpful for baseline and therapeutic progress documenting, inconclusive test-retest reliability indicators imply that use of other alternative assessment tools may be necessary or multiple baseline measures undertaken to provide a more representative sample of a child’s speech production abilities. Not all informal independent measures are created equally and test-retest reliability does seem to vary across analyses with word shape analyses appearing to be more stable over time than phonetic inventories. Finally, SLPs need to be aware and mindful of the factors that may influence the reliability of these informal measures including age of child, length of sample collected, phonetic materials, corpus size, as well as data collection and transcription procedures.
4.2 Limitations and future directions

A number of factors limit the potential for generalization of the present study results including the small sample size and data analysis techniques. Consequently, the present study does not yet provide clear clinical guidelines for use of informal phonological measures with young children nor does it sufficiently expand the evidence base for clinical decision making in speech-language pathology. Rather, the present study provides justification for further investigation in the area and an increased awareness of data analysis technique to be considered in future studies.

Although these preliminary findings indicated partial support for Morris’ outcomes extending to slightly older children with more advanced language skills, the small sample size severely restricted the generalization of the findings to the general population. For instance, due to the small sample size, the authors could only visually analyze the data rather than test the findings using statistical methods. In order to conduct analyses beyond visual display comparisons and more accurately account for support of Morris’ findings, a larger sample size is needed. Future studies along this vein that include additional age ranges - both above and below those included in the present study - would provide information on informal phonological analyses for a wide range of young children at differing developmental language stages. The inclusion of children with language delays would facilitate generalizable findings to clinical populations who may perform differently than the children represented in the present study.

The data analysis techniques utilized in the present study included a measure of word shape use. All three participants performed at ceiling on this measure. Ceiling-level performances cause practical problems when calculating reliability because they prevent a variable from being measured or estimated above a certain level, or as in this case, beyond the distribution of eight different target word shapes. The particular word shape analysis utilized in the present study was selected in an effort to adhere closely with measured used by Morris. However, given the age and language development status of the Morris participants compared to those of the present participants, a
ците на Morris во споредба со сегашните учесници, може да се користи покомплексна техника за мерење на перформансите. Идините студии кои вклучуваат деца над 24-месечна возраст може да вклучуваат посуптливи мерки за анализ на зборовната форма со цел по-добро да се долги различното изразување. Дополнителна забелешка на сегашните наоди вклучени во постапките кои се користат за собирање и транскрибирање на примероците од комуникацијата е, дека, и покрај тоа што беа реплицирани повеќето процедури кои се користат од страна на Morris во оваа студија, постои широк спектар на процедурали различни примероци кои се користат во областа за каков било број на примероци од комуникацијата и неформални фоностолски мерки. Во оваа студија, почетната веродостојност беше 62%, прилично ниска кореспонденција со оглед на тоа што 80% веродостојност е често цел во кодирањето на веродостојноста. Оваа се покаја како проблем во областа на логопедијата билејки фоностолската транскрипција на говорите примероци често не користи за да се утврдат соодветните цели за интервенција, како за следење на напредокот со текот на времето (10). Друг придонес за варијабилноста при собирање и транскрибирање на примерокот од комуникацијата е изборот на различни сетови на играчки, укажувајќи на потенцијалот за различни фоностолски материјали кои би можеле да влијаат врз бројот на зборовите кои децата ги произведуваат и врз нивните фоностолски инверти. Многу SLP различни го користат фоностолскиот материјал кој се користи во сесии, во зависност од одговорите на детето на понудениот материјал. Според тоа, иако изборот на играчки може да биде избор на варијабилност во оваа студија, тој служи како еколошки валидна застапеност на промената која постои во клиничката пракса. Конечно, врз фоностолскиот инветар може да влијае големината на корпусот, па така алукаата на истражувачите да се користи целото транскрибиран примерок, а не само првите 100 збора, исто така би можело да придонесе за варијацији, и покрај тоа што оваа процедура била применета подеднакво кај сите учесници. Идините студии може да се осврнат на некои од овие потенцијални извори на варијабилност и да одговорат на потребата за повеќе истражувања за должината на примерокот од говорот и процедурите за фоностолска транскрипција кои се користат за клиничкото и нормативното население, со цел да се прочистат процедурите за докосување оценка врз основа на докази во областа на логопедијата.

An additional caveat to the present findings involved the procedures used to collect and transcribe the communication sample. Although in the present study, most procedures used by Morris were replicated, there exists a wide variety of procedural differences utilized in the field for any number of communication sample collections and informal phonological measures. For instance, low inter-judge reliability when transcribing speech samples of young children occurs because the speech to be transcribed is usually quite different from adult-like forms. In the present study, initial inter-rater reliability was 62%, a fairly low correspondence considering 80% reliability is often the goal in reliability coding. This has proven to be a concern in the field of speech-language pathology as phonetic transcriptions of speech samples are often used to determine appropriate goals for intervention as well as monitor progress over time (10). Other contributions to variability in collection and transcribing the communication sample include choice of differing toy sets, indicating the potential for different phonetic materials which could affect the number of words children produce and the phonetic inventories elicited. Many practicing SLPs vary the phonetic material used across elicitation sessions depending on child responses to provided material. Therefore, although toy selection may be a source of variability in the present study, it serves as an ecologically-valid representation of the variability that exists in clinical practice. Finally, phonetic inventory may be affected by corpus size, so the researchers’ decision to use the entire transcribed sample rather than only the first 100 words may also have contributed to variance even though this procedure was applied consistently across participants. Future studies may address some of these potential sources of variability and address the need for more research into the length of speech sample collected and phonetic transcription procedures used for clinical and normative populations in order to refine evidenced-based assessment procedures for the speech-language pathology profession.
In conclusion, for a small sample of two-year-old children without known language delays, the test-retest reliability of initial consonant and consonant cluster productions was unstable between two conversational samples obtained one week apart under near-identical conditions. However, phonetic inventories for final consonants and word shape analyses appeared relatively stable over the one-week time period. Although more data is needed, the preliminary results of this pilot study indicated that academic faculty, clinical educators, and practicing SLPs should be mindful that while a finding of a few instances of difference across informal phonological measurement samplings does not mean the measure itself is inappropriate, it does indicate that outcomes should be cautiously interpreted for young children.

Conflict of interests

Authors declare no conflict of interests.

References