**Photocopy and Use Authorization**

# Phonetic Transcription in Development: When is Reliability Achieved? Part 2

by

Lani Roemer

A thesis to be

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in the Department of Communication Sciences and Disorders

Idaho State University

Summer 2018

**Committee Approval**

To the Graduate Faculty:

The members of the committee appointed to examine the thesis of Lani Roemer find it satisfactory and recommend that it be accepted.

Heather Ramsdell-Hudock,
Thesis Advisor

Mary Von Donsel,
Committee Member

Nancy Devine,
Graduate Faculty Representative

## Human Subjects Committee Approval

September 13, 2017

Lani Roemer
Comm Sci Disorders/Deaf Educ
MS 8116

RE: regarding study number IRB-FY2018-51: Phonetic Transcription in Development: When is Reliability Achieved? Part 2

Dear Ms. Roemer:

I have reviewed your request for expedited approval of the new study listed above. This is to confirm that I have approved your application.

Notify the HSC of any adverse events. Serious, unexpected adverse events must be reported in writing within 10 business days.

You may conduct your study as described in your application effective immediately. The study is subject to renewal on or before Sep 13, 2018, unless closed before that date.

Please note that any changes to the study as approved must be promptly reported and approved. Some changes may be approved by expedited review; others require full board review. Contact Tom Bailey (208-282-2179; email humsubj@isu.edu) if you have any questions or require further information.

Sincerely,

Ralph Baergen, PhD, MPH, CIP
Human Subjects Chair

**Table of Contents**

**List of Figures**

**Phonetic Transcription in Development: When is Reliability Achieved? Part 2**

**Thesis Abstract--Idaho State University (2018)**

Phonetic transcription, a popular method for analyzing speech, is not reliable for examining vocalizations of young infants. This study is to determine what age phonetic transcription reliability is achieved. Results were combined with a prior phonetic transcription reliability study involving seven children, with each child's vocalizations recorded from 7 to 18 months of age. For this present study, vocalization samples were obtained from 12 children, two at each month of age between 19 and 24 months, then 20 utterances from each child were transcribed by two researchers and compared using a weighted approach to measure inter-transcriber reliability. The result of this study is phonetic transcription reliability increases and reaches a reliability value of 0.8 or greater by the time children are 21 months old. Conclusions from this study will facilitate speech-language pathologists selecting the most appropriate method for documenting speech sounds at various ages in clinical practice or research.

Keywords: phonetic transcription reliability, infant vocal development, speech language pathology

Phonetic Transcription in Development: When is Reliability Achieved? Part 2

Early language development is critical to infant and child development as a whole, as it can form foundations for later success throughout all aspects of life, including academic, employment, social, and psychological wellbeing (Sosa, 2015). Early intervention has been demonstrated to positively impact developmental outcomes for young children who have, or are at risk, for delay or disorder (Paul & Roth, 2011). Accordingly, it is vital to identify infants and toddlers in need of services as early as possible; this ensures that intervention is provided during pivotal developmental periods, when the brain is at its highest degree of plasticity and is most capable of change (Hebbeler, 2009).

Since identifying infants and young children in need of additional support is essential, having a reliable method for speech-language pathologists (SLPs) to document speech sounds is crucial. The preferred method by many SLPs for this task is phonetic transcription, however clinical and research expertise indicate that phonetic transcription is not well developed for documenting prelinguistic speech sounds (Cucchiarini, 1996; Ramsdell, Oller, & Ethington, 2007; Stockman, Woods, & Tishman, 1981). Being able to examine infant vocalizations is critical in tracking development and identifying children who are in need of early intervention programs.

Fortunately, besides transcription, there are numerous other options available to study infant vocal development (Lieberman & Lohmander, 2014; Serkhane et al., 2007; Xu et al., 2014). However, it is necessary to determine the age at which phonetic transcription becomes reliable, so that clinicians and researchers know when phonetic transcription is a viable option for documentation of child speech sound production. The alternative would be to utilize other methods, such as caregiver report, to identify children for early intervention programs (Ramsdell, Oller, Buder, Ethington, & Chorna, 2012).

**Stages of Infant Vocal Development**

In children who are typically developing, speech abilities change dramatically and are rapidly evolving throughout the first few years of life (Goldstein & Schwade, 2008). According to Oller (2000), there are five stages in infant vocal development, which include the phonation, primitive articulation, expansion, canonical, and integrative stages.

**Phonation stage.** Phonation is the initial stage of development and occurs from birth until the child is 2 months old (Oller, 2000). According to Oller, infants produce vegetative sounds such as cries, sighs, and burps, in addition to quasivowels during the phonation stage. Quasivowels are vowel-like utterances that are not fully formed. He states that these vowel sounds occur when the vocal tract is at rest, and do not incorporate the precise tongue and lip positions of mature vowel sounds. Quasivowels are typically nasal consonants or nasalized mid or high vowels (e.g., “mmmm” or “ooooo”) and are produced with the newborn’s mouth in a closed or nearly closed position (Nakatani-Murai, 2008).

**Primitive articulation stage.** The primitive articulation stage takes place when infants are between 2 and 4 months of age (Oller, 2000). According to Oller, this is when infants begin cooing, or displaying periods of smooth voicing. Infants combine consonant sounds, frequently the “k” or “g” sound, with quasivowel productions (e.g., “kooo” or “goo”) to create syllable-like sounds (Nakatani-Murai, 2008). Some limited articulation occurs during this stage, as infants begin to experiment with tongue and lip movements while vocalizing (Oller, 2000). However, quasivowels remain prominent, as vocalizations are still often involuntary at this stage (Oller, 2000).

**Expansion stage.** Many new sounds appear during the expansion stage, the third stage in infant vocal development (Oller, 2000). Oller states that this stage occurs between 4 and 7 months of age and is when infants develop full vowel sounds. The infant is now able to fully articulate his/her speech production mechanism by opening the vocal tract completely and positioning the lips and tongue to produce various sounds (Oller, 2000). Other sounds that emerge during this stage due to the infant's greater use of their vocal mechanism are raspberries, squeals, whispers, and growls (Nakatani-Murai, 2008). Later in this stage, infants will begin to demonstrate marginal babbling, which is when full vowel sounds and closed sounds, such as bilabial consonant-like sounds or raspberries, are connected together (Oller, 2000). Marginal babbling is marked by slow transitions between the vowel and closed sounds, particularly when compared with more mature vocalizations. Accordingly, the babbling is not fully formed.

**Canonical stage.** The canonical stage occurs from 7 to 12 months of age and is when infants master babbling (Oller, 2000). A canonical syllable is a prototypical consonant-vowel syllable containing well-formed consonants and vowels with timely transitions between the two. This stage begins with infants producing reduplicated babbling of canonical syllables, which is when the same sequence of sounds is said in repetition (e.g., "mamama" or "papa"). As they begin to progress, Oller tells that infants will transition to more variegated babbling, which is when they begin to connect different sequences of sounds in a variety of orders (e.g., "degati" or "madobee"). The infants' vocal productions in this stage sound incredibly speech-like and often follow mature intonation patterns (Oller, 2000).

**Integrative stage.** The integrative stage occurs between 12 and 18 months of age (Oller, 2000). Infants typically produce their first word by the time they are 12 months old (Fenson et al., 1994). Therefore, their first word has emerged by this stage. In addition to the development of a few true words, infants continue to babble and also develop protowords (Oller, 2000). According to Oller, protowords are words that are used by the child with a consistent phonetic form and semantic meaning, but they do not follow the traditional pronunciation of the adult target (e.g. "bibi" for pacifier). As children develop more true words, their use of babbling begins to decrease (Oller, 2000).

**The Importance of Early Intervention**

When infants' development does not align with that of their peers who are typically developing, it can be beneficial for them to begin therapy in an early intervention program (Paul & Roth, 2011). Enrolling these children in early intervention programs as early as possible increases the likelihood that they will develop effective communication skills (Paul & Roth, 2011), due to the fact that the brain is more flexible and adaptable to change in the early years (Hebbeler, 2009). Accordingly, the sooner infants or toddlers are enrolled in services, the more beneficial therapy will be. Early intervention programs in speech-language pathology strive to be family centered, culturally appropriate, applicable to the child's natural environment, and interdisciplinary when appropriate in order to set the child up for success (Paul & Roth, 2011). Interdisciplinary approaches involve professionals from a variety of disciplines, often including physical therapists, occupational therapists, physicians, nurses, and psychologists. Paul and Roth (2011) also indicate that it is imperative that early intervention programs utilize evidence-based practice, so therapy will be as effective as possible. In regards to identifying young children who are at risk for developing speech sound disorders, the clinician must know what methods are most effective for analyzing speech sounds and developmental behaviors in order to accurately identify children for these services.

**Phonetic Transcription Reliability for Infants**

Phonetic transcription is widely used to document speech production by professionals in the field of speech-language pathology, and is a familiar and comfortable method for many therapists. Since having a technique for recording and analyzing speech sounds in infants is crucial, it would seem that phonetic transcription is the logical choice. However, as mentioned earlier, phonetic transcription is not reliable for young infants (Cucchiarini, 1996; Ramsdell et al., 2007; Stockman et. al., 1981).

The International Phonetic Alphabet, which is the system of symbols developed to represent speech sounds produced around the world, was developed to document mature speech sounds, and therefore does not include sounds that are unique to immature infant productions (Oller & Ramsdell, 2006). In the phonation and primitive articulation stages of development, quasivowels in particular are notoriously difficult to transcribe (Oller, 2000). Therefore, phonetic transcription will likely not accurately represent infant vocalizations until the speech sound productions become more adult-like. There is evidence that transcription of canonical syllables, with well-formed consonant and vowel segments and timely transitions between the two, results in greater reliability than transcription of sounds from earlier in development (Ramsdell et al., 2007). Until vocalizations begin to include primarily canonical syllables and development is more mature, phonetic transcription will not be accurate[1].

Additional research demonstrates that transcription reliability increases with infant age and maturing vocal development. One study found that reliability in phonetic transcription

[1] Throughout this paper, the terms "transcription agreement" and "transcription reliability" will be used interchangeably, though it should be noted that agreement is technically a subtype of reliability (Cucchiarini,1996). Reliability or agreement refer to the degree of similarity or difference between two transcribers, or one transcriber at different points in time. This differs from "transcription accuracy", which refers to how close a transcription matches its target. Transcription accuracy can be described as how correct a transcription is as it relates to its target.

increased around 20 to 21 months of age (Stockman et al., 1981), however no specific age in which phonetic transcription is fully reliable has been determined. Determining the age at which an infant’s speech is mature enough for phonetic transcription to be a reliable method will help both clinicians and researchers examine the speech sounds of these young children.

In addition, prior research conducted in the Infant Vocal Development Laboratory at Idaho State University (ISU) as part of a master’s thesis has shown that mean inter-transcriber reliability values for infant vocalizations increased from 0.52 at 7 months of age to 0.66 at 18 months of age (Schroeder, 2016). An important question is how high of a reliability value is acceptable to establish an age with which phonetic transcriptions can be used? Naturally, a higher reliability is better than a lower reliability. Lance, Butts, and Michel (2006) discuss that a reliability value of 0.7 is commonly thought to be acceptable. However, the reference from Nunnally (as cited in Lance et al., 2006) elaborates and says to only use a reliability value of 0.7 when conducting research in the early stages. Nunnally (as cited in Lance et al., 2006) also argues that for applied research, a reliability of 0.8 is the lowest acceptable value, with reliability values of 0.9 or 0.95 more desirable. The highest reliability value in the Schroeder (2016) study was 0.657, therefore the commonly accepted reliability value of 0.7 or the lowest acceptable value for applied research of 0.8 was not reached. Though the study was able to show increasing reliability for phonetic transcription between 7 through 18 months of age, the results obtained did not indicate that phonetic transcription is an adequate tool for documenting vocal development through 18 months of age.

**Other Methods for Measuring Infant Vocalizations**

Phonetic transcription is not the only option when it comes to analyzing speech sounds; there has been a surge of research on various techniques to assess prelinguistic vocalizations.

There are many alternatives to phonetic transcription that have proven to be effective for analyzing infant speech. These methods are particularly effective at ages when phonetic transcription is not reliable. Once a precise age is determined for acceptable phonetic transcription reliability, clinicians and researchers will be able to use the research to determine when to utilize their phonetic transcription skills and when to rely on one of the proven alternative methods listed below in documenting speech sound productions.

**Articulatory model.** The articulatory model is a possible alternative to phonetic transcription for analyzing speech sounds in infants. This model provides a visual representation of articulatory movements and is a method to analyze the role of the articulators in speech sound production. Though exploring their sensory-motor system and learning the limits of their vocal mechanism is a vital aspect of development for infants, strategies such as transcription and direct measurement do not assess these skills.

A study was conducted that examined the effectiveness of an articulatory model to measure infant vocal development (Serkhane, Schwartz, Boe, Davis, & Matyear, 2007). The study examined 24 infants who were 4 months of age and 3 infants who were 7 months of age. The researchers compared formant values (F1 and F2) to place of articulation using an articulatory acoustic model. A computerized model was used to measure the infant's vocal tract as it grew. Results indicated that for 4-month-olds, the jaw plays a minor role in articulatory exploration, while for 7-month-olds, the jaw plays a more predominant role. The utility of an articulacy model in depicting such change demonstrates that it is appropriate for analysis of infant vocalizations. However, it is not necessarily a clinically efficient methodology.

**Observation.** Observation is another technique that is used by clinicians to analyze infant speech. In these instances, professionals who are trained in vocal development clinically observe the child and listen for specific speech sounds and watch for certain developmental behaviors without formally transcribing the child's vocalizations. Research looking specifically at the development of babbling and consonant sounds in infants using an observational method has shown that observation is a reliable technique (Lieberman & Lohmander, 2014). This study provided a comparison between the phonetic transcription and observational method. Participants consisted of 29 children who were 12 months of age and 38 children who were 18 months of age. All of the children engaged in child/caregiver interactions, during which (and immediately following) an SLP completed various observation forms. Two blinded coders transcribed the utterances, and then the transcriptions were compared to measure inter-rater reliability. When examining babbling structure, observation and transcription had high agreement values. Observation was also shown to be a valid measurement for several types of consonant sounds, including stops, bilabials, and dentals, but not glottal consonants. Accordingly, perhaps clinicians could forgo the tedious process of phonetic transcription when assessing a child's babbling abilities, and simply rely on skilled observation.

**Automated computational measure.** Another method for analyzing infant vocalizations that has been proven to be effective is using an automated computer system to analyze recordings, rather than using phonetic transcription (Xu, Richards, & Gilkerson, 2014). In this method, researchers would identify the different sound segments in a recording by using a segmentation algorithm and then phonemic units were identified using speech recognition software. Researchers were able to draw various conclusions from using this method with 106 children who were typically developing, 71 children with autism spectrum disorder, and 48 children with a language delay that was unrelated to autism spectrum disorder. The children were between 8 to 48 months of age. For example, using this method it was determined that children who are typically developing used more consonant and vowel sounds than both the children with a language delay and the children with autism spectrum disorder. It was also determined that children with a language delay produced more consonant sounds, but not more vowels, than children with autism spectrum disorder.

The conclusions gathered from this method indicate that an automated computational measure would be a valuable tool in a research setting and is a viable alternative to phonetic transcription for studying infant vocal development. However, this method is quite technical and time consuming, so SLPs may not find it to be practical in the clinical setting.

**Measuring Phonetic Transcription Reliability**

Measuring phonetic transcription reliability can be a complex task. Inter-rater reliability is a common approach to determining the reliability of phonetic transcription. In this approach, two individuals transcribe a production, and then the transcriptions are compared. The more similar the transcriptions, the higher the reliability. There have been multiple studies done to determine how to classify errors in transcriptions in order to most accurately measure the reliability of a transcription. This information is beneficial for any phonetic transcription task, but particularly for when determining the reliability of infant transcriptions.

Traditionally, researchers have used an unweighted approach for measuring transcription reliability, in which all errors in the transcription are marked as equal. However, Oller and Ramsdell (2006) conducted a study in which they compared the traditional unweighted approach to that of a weighted approach, in which three different types of phonetic transcription agreement were analyzed. The three types were global structural agreement, featural agreement, and overall transcription agreement. This allowed the researchers to compare more features of phonemic segments than a traditional unweighted approach would allow. An infant, a toddler, an adult who spoke American-English, an adult who spoke Korean, and an adult who spoke Ukrainian participated in the study. The weighted approach was shown to be a better measure of phonetic transcription reliability than the unweighted approach for all participants.

Cucchiarini (1996) also demonstrated a way to improve measurement of phonetic transcription reliability by using different features, along with diacritic markers, in assessing how similar two sounds are. The features examined for consonant sounds were nasality, voice, place, lateral, glide, stop, fricative, trill, distribution, and height. For vowels, the features examined were lip rounding, tongue height, and tongue advancement (horizontal position toward the front or the back of the mouth). These features, along with diacritic markers, allowed for much more detailed transcriptions of the speech sounds and provided researchers with more information to analyze. By examining these different features when comparing phonetic transcriptions, researchers are able to increase the effectiveness of measuring inter-rater reliability.

**Goals and Rationale**

Accordingly, the *long-term goal* of this research is to determine what methods of tracking speech sounds are most appropriate at the various stages of development. The *objective* of the present study is to determine the age at which phonetic transcription becomes reliable for documenting speech sound production. Lance and colleagues (2006) indicated that a reliability of 0.8 is the lowest acceptable value. Accordingly, the *central hypothesis* is that phonetic transcription reliability will increase to a reliability value of 0.8 as children age from 7 to 24 months. This hypothesis has been formulated on the basis that phonetic transcription reliability increases as children age and speech sound productions become more adult-like (Ramsdell et al., 2007; Stockman et all., 1981). The *rationale* for the proposed research is that, once the age at which phonetic transcription is reliable is established, alternate methods can be used prior to that age. This knowledge will help clinicians be able to identify children for early intervention programs more effectively and efficiently, and help researchers to conduct more accurate research in the area of infant vocal development.

Accordingly, in children who are typically developing from 7 to 24 months of age, our aim is to identify inter-transcriber reliability patterns using a weighted reliability measure. The *working hypothesis* for this aim is that inter-transcriber reliability will increase to an acceptable value of 0.8 by 24 months of age.

## Methods

This study is a continuation of research presented in a thesis by Schroeder (2016) in order to determine at what age phonetic transcription reliability is achieved. Schroeder examined children from 7 to 18 months of age, while the present study continued with children from 19 to 24 months of age. Procedures and methods of analysis remained as close as possible to those utilized by Schroeder in order for researchers to merge the results together.

### Participants

To complete this project, an additional sample of vocalizations was obtained from 12 children, two at each month of age between 19 and 24 months. Five in-person participants were recruited from the community to come into the ISU Infant Vocal Development Lab to participate in the study. In addition to being between 19 and 24 months of age, inclusion criteria for the in-person participants required: normal hearing, speaking primarily English in the home, no pre/perinatal difficulties, no significant ENT problems, no siblings and/or parents with a speech and/or language disorder, being from a home of middle socio-economic status, and typical speech/language development. Normal hearing was determined through a hearing screening in the ISU Audiology Clinic, while the remaining requirements were determined through parent self-report. Caregivers gave informed consent (previously approved by Human Subjects at ISU) to participate in the study. One of the in-person participants was female, while the other four were male.

For the additional seven participants, recording files from an online database were used to supplement the in-person participants. The PhonBank database, which is a publicly available child phonology database, was used to locate these audio files (Rose & MacWhinney 2014). The PhonBank database is part of the broader system TalkBank, which is an international database intended for research consisting of video recordings, audio recordings, and transcripts of recordings of vocal interactions (MacWhinney, n.d.). The Davis Corpus from PhonBank was selected for use in the present study (Davis, MacNeilage, & Matyear, 2002). Files were chosen only from one corpus in order to limit the number of extraneous variables for this study. The audio files that were selected were recordings from children who were between 19 and 24 months of age and who were born in the 2000s. Participants from the Davis Corpus received a hearing screening using sound field techniques, had normal development established through a parent case history repot, completed the *Battelle Developmental Screening Inventory*, and were recorded during natural interactions with their families in their home environment (Davis et al., 2002). According to the *TalkBank Code of Ethics* (MacWhinney, n.d.), participants included in the TalkBank databases signed an Informed Consent that was approved by the researcher's local Institutional Review Board (IRB) (unless the data was collected before the IRB was established, from non-funded work, or from speakers of indigenous and endangered languages). Participants included in the Davis Corpus were recorded after the IRB was established and spoke English in the recordings, though it was not specified in TalkBank if this collection was funded or non-funded. From the selected audio files for this study, four of the participants were female and three were male.

Data from both the in-person participants and the PhonBank participants were combined in order to form a complete data set with two children at each month of age between 19 and 24 months. In total, five of the participants were female and seven of the participants were male.

**Procedure**

To collect the vocalizations from the in-person participants, the children were video and audio recorded for 20 minutes while participating in a period of free play with one or more of their parents/guardians in the ISU Infant Vocal Development Laboratory. The same selection of toys was available for all in-person participants. The room was equipped with eight Sony EVI-D70/W cameras, which were mounted on the wall of the lab. Additionally, audio from the session was recorded for infants using a wireless microphone housed in a vest, and for caregivers using a wireless lapel microphone. During each session, all recordings were transmitted to an adjacent control room where laboratory staff chose two cameras angles to capture. Staff chose the best view of the child's face to aid in future location of utterances. In an attempt to acquire a natural interaction, the lab was set up as similarly as possible to a typical child's room/ nursery and the parents were told to interact with their child as they would at home.

For both the in-person and online database recordings, a member of the laboratory staff selected 20 utterances from each recording based on a breath group criterion (Oller & Lynch, 1992). Selected utterances were then transcribed by two transcribers who were trained in the International Phonetic Alphabet. The transcribers worked independently, did not view acoustic displays, listened to each utterance no more than six times, and included exotic sounds in their transcriptions that may not be part of the phonemic repertoire of General American English (e.g., /ʙ, ʀ, ɸ, β/, etc.).

Once the transcriptions were completed, they were systematically aligned according to principles established in previous research using 4 alignment principles: the strict order principle, matched segment principle, minimum discrepancy principle, and nucleus alignment first principle (Oller & Ramsdell, 2006; Ramsdell et al., 2007). The strict order principle required that segments in a transcription were not reordered and remained in their original sequence. Vowel-like and consonant-like segments, that were in the same order, were matched together in aligned transcriptions as called for in the matched segment principle. The minimum discrepancy principle described the approach to take when there were different numbers of vowel-like or consonant-like segments, or those segments were not ordered in the same way. This principle required that these segments be aligned in such way as to create the most phonetically similar segment matches, without reordering any segments as this would violate the strict order principle. Since vowels are perceived as the center of the syllable, vowel-like segments were aligned first in accordance with the nucleus alignment first principle.

After the transcriptions were aligned, a program written in LIPP™ (Logical International Phonetics Program) analysis language (Oller & Delgado, 1999) calculated the weighted reliability. This program is designed to rate each segment of a transcription on a scale of 0-1, with scores closer to one indicating similarities between segments and scores near zero indicating substantial differences. Comparing the segments from two aligned transcriptions in this way determines the degree of match between transcribed segments and was how the weighting was calculated. There were three types of agreement used in this approach: global structural agreement, featural agreement, and overall transcription agreement (Oller & Ramsdell, 2006). In global structure agreement, a score of one indicates perfect reliability for matched segments and zero indicates there is no segment, or an orphaned segment. For featural agreement, higher reliability is awarded for more similar segments (e.g., /p/ versus /b/) and lower reliability for less similar segments (e.g., /p/ versus /n/). The global structural agreement and the featural agreement are multiplied together to calculate the overall weighted transcription agreement.

Consider the following example discussed by Schroeder (2016):

| | | | | |
|---|---|---|---|---|
| Coder A | [p | ĩ | n | ] |
| Coder B | [b | i | d | i] |

In the provided example, the first three segments would receive a global structural agreement of 1 and the last segment would receive a global structural agreement of 0, as Coder A did not transcribe a segment while Coder B did. The global structural agreement for the entire vocalization is determined by calculating the mean of all global structural agreement values for each segment. The global structural agreement for the entire vocalization in the above example would be 0.75. For featural agreement, the similarities or differences between the segments of the two transcriptions would determine the scores. For example, a voicing difference in the first segments gives the two segments a featural agreement of 0.67. In the second segments, while both have the same symbol, the top transcription includes the nasalization diacritic, which gives the segments a featural agreement of 0.9. Both transcriptions have the same place of articulation, but a different manner of articulation in the third segments, which results in a featural agreement of 0.67. Featural agreement is only calculated in slots that both coders transcribed segments. Therefore the mean featural agreement, as determined by calculating the mean of the values from the first, second, and third segments in this example, is 0.74. The mean featural agreement and the mean global structural agreement are multiplied together to get the overall transcription agreement, which is 0.56.

**Design**

To evaluate the variables of interest, a repeated-measures analysis of variance (ANOVA) was used. Variables of interest are presented in Figure 1. The dependent variable of interest was phonetic transcription reliability, determined through the weighted reliability measure in LIPP™. The child’s age, monthly from 7 through 24 months of age, was the independent variable of interest.

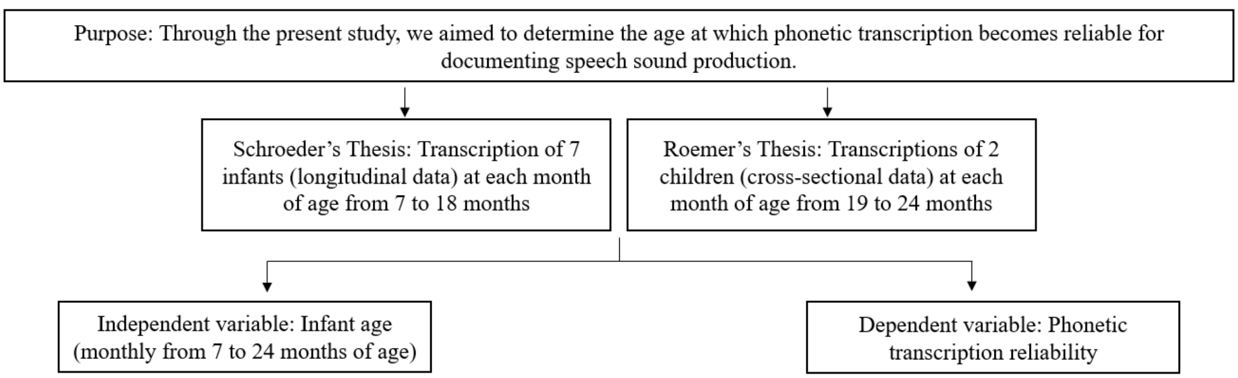


*Figure 1.* Purpose, participants, and variables of interest.

## Results

The percentage of segments that were a perfect match in the transcriptions increased as the child aged from 7 to 24 months of age, with the mean unweighted reliability at 7 months of age being 0.218 and the mean unwieghted reliability at 24 months of age being 0.644. Mean proportion perfect for all ages studied are displayed in Figure 2. While the conclusions of this study will be drawn from weighted reliability versus unweighted reliability, this increase in proportion perfect does provide additional evidence that phonetic transcription reliability does increase with the child’s age.

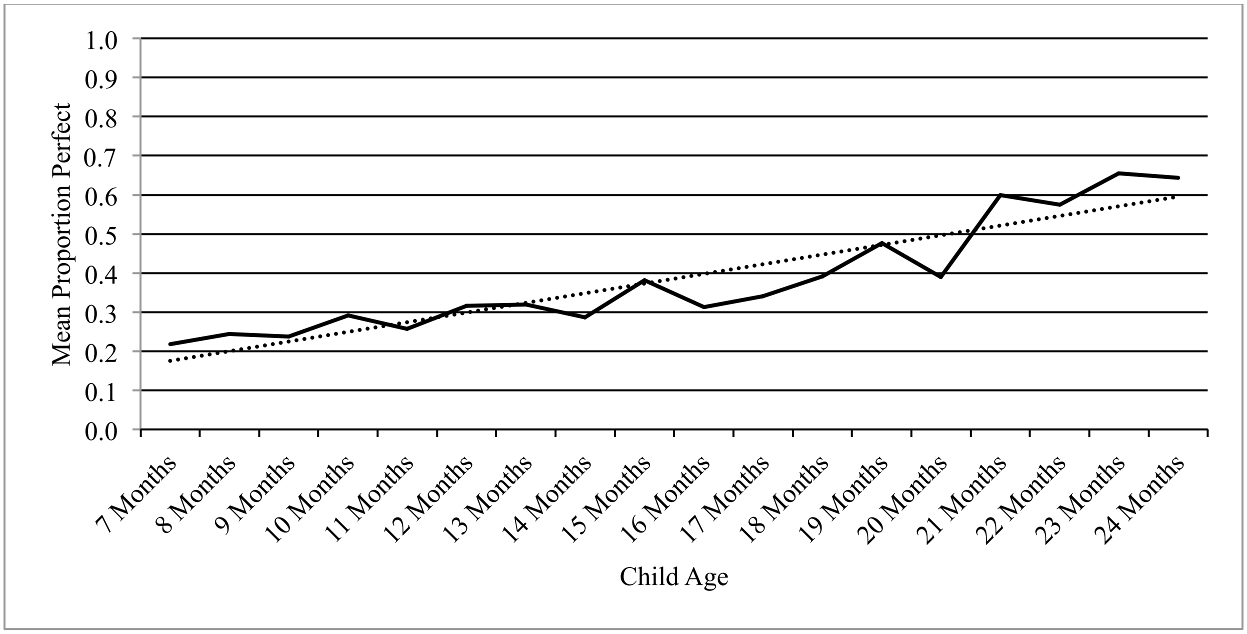


*Figure 2.* Mean proportion perfect from 7 through 24 months of child age.

Given that a weighted approach has been shown to be a more reliable method to test phonetic transcription reliability than an unweighted approach (Oller & Ramsdell, 2006), global structural agreement and featural agreement were calculated to determine the weighted reliability. Global structural agreement increased from a mean of 0.648 at 7 months of age to a mean of 0.917 at 24 months of age. Mean global structural agreement across ages is displayed in Figure 3. Featural agreement increased from a mean of 0.797 at 7 months of age to a mean of 0.912 at 24 months of age. Mean featural agreement for all the ages between 7 to 24 months is displayed in Figure 4. The values for global structural agreement and featural agreement indicate that both the degree of match between the segments and the similarity of the segments increased as the children aged.

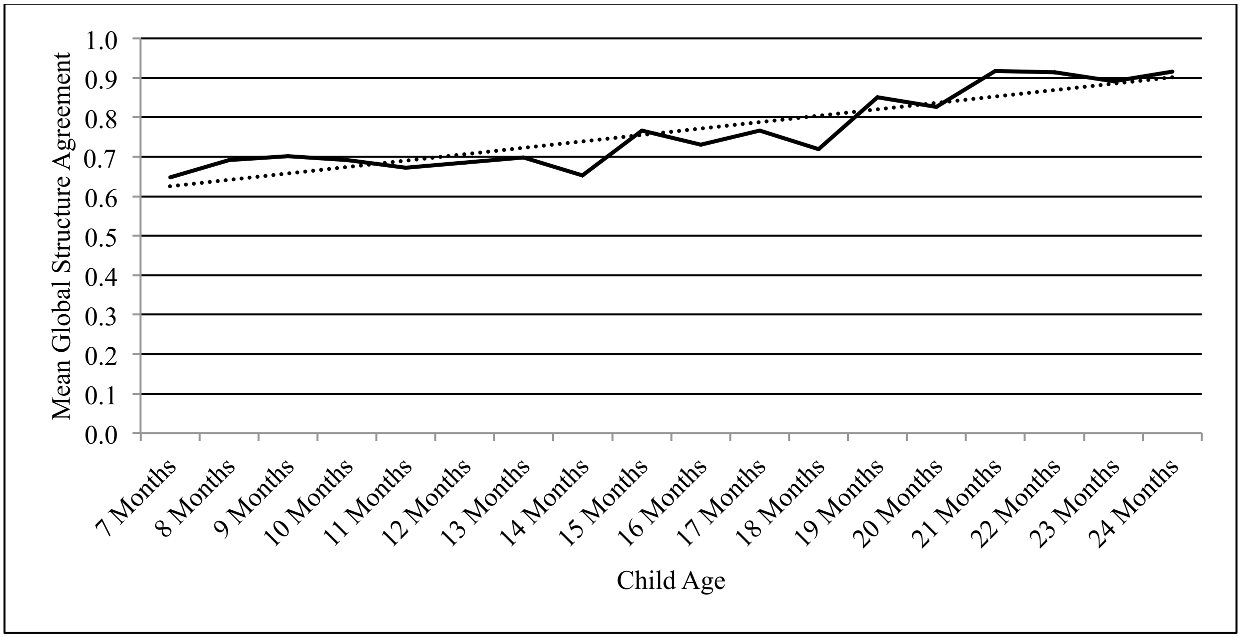


*Figure 3.* Mean global structural agreement from 7 through 24 months of child age.

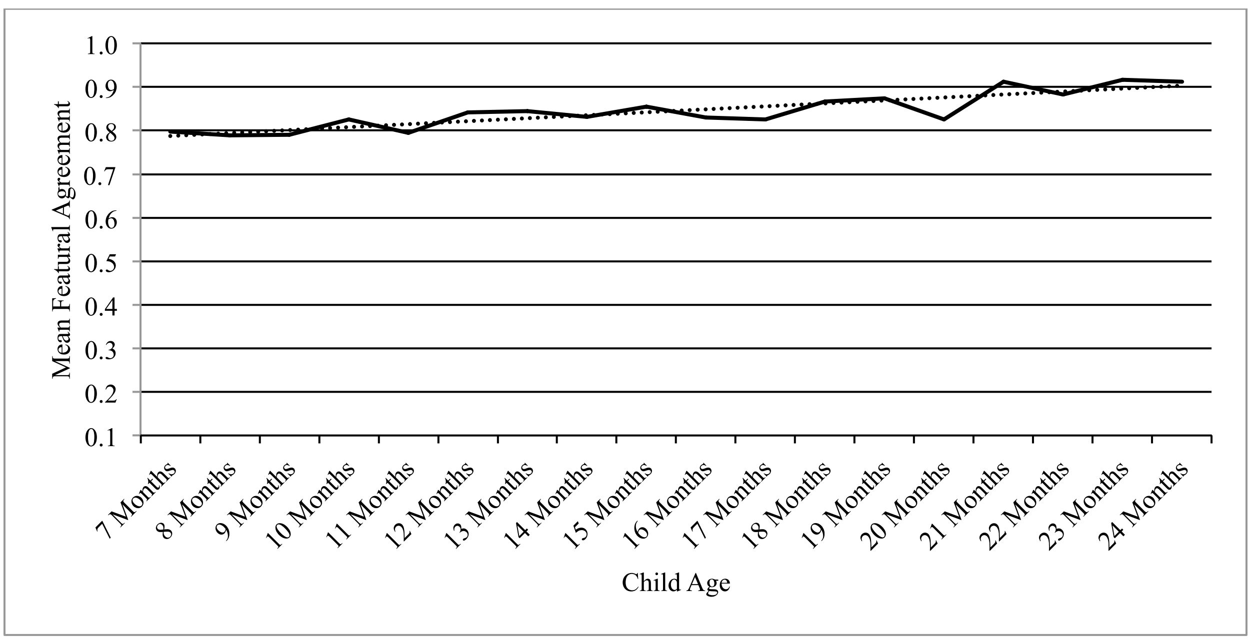


*Figure 4.* Mean featural agreement from 7 through 24 months of child age.

The global structural agreement and the featural agreement were multiplied together to calculate the overall transcription agreement, or the weighted reliability. A repeated measures ANOVA with sphericity assumed showed that weighted transcription reliability differed statistically significantly across child ages [$F$ (17, 17) = 31.595, $p < 0.001$), as displayed in

Figure 5. Post hoc tests using Tukey's LSD revealed that transcription reliability was statistically significantly lower at 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, and 18 months than at 21 and 24 months. Further, transcription reliability was statistically significantly lower at 15 and 20 months than at 21 months. Thus, it can be concluded that child age, and production of more canonical and linguistic vocalizations, leads to a statistically significant increase in weighted transcription reliability, though significant increases will only be observed after 21 months of child age.

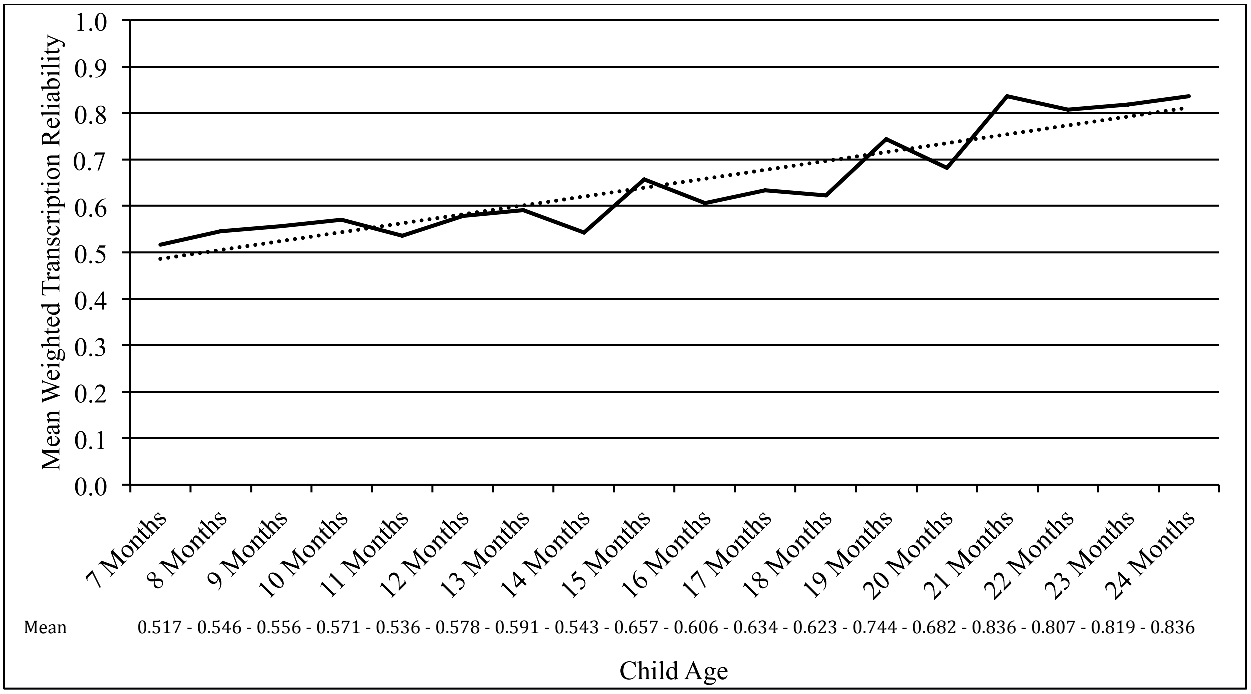


*Figure 5.* Mean weighted transcription reliability from 7 through 24 months of child age.

For transcriptions of children between 7 through 18 months of age, an independent-samples *t*-test was used to compare the weighted transcription reliability between Coder A and B and between Coder A and C. There were three coders in this portion of the study and each child was transcribed twice. As described by Schroeder (2016), Coder A's transcriptions were compared with Coder B's and C's; Coder A transcribed the utterances from all of the children, Coder B transcribed four of the seven children, and Coder C transcribed the three remaining children. The results indicated that there was not a statistically significant difference in the weighted transcription reliability between Coder A and B (*M* = 0.598, *SD* = 0.002) and between Coder A and C (*M* = 0.569, *SD* = 0.002), *t* (10) = 2.092, *p* = 0.063 (Schroeder, 2016). This suggests that for the children between 7 through 18 months of age, there were no substantive differences between the transcribers.

For transcriptions of children between 19 through 24 months of age in the present study, a third coder transcribed 20 utterances randomly selected from one child. An independent-samples *t*-test was conducted to compare the weighted transcription reliability between Coder A and C and between Coder B and C for these transcribed utterances. There was a statistically significant difference in the transcriptions between Coder A and C (*M* = 0.782, *SD* = 0.171) and between Coder B and C (*M* = 0.757, *SD* = 0.171), *t* (3) = 6.054, *p* = 0.009. These results indicate that there may have been a substantive difference between the coders who transcribed the vocalizations from the children between 19 through 24 months of age for this study.

## Discussion

The objective of this study was to determine the age at which phonetic transcription becomes reliable for documenting speech sound production. The hypothesis was that transcription reliability would increase to a reliability value of 0.8 as infants age from 7 to 24 months. Through analyzing inter-rater reliability using a weighted approach for children from 7 to 24 months of age, the hypothesis was shown to be correct. The mean reliability value at 21 months of age was 0.84, and the mean reliability values at 22, 23, and 24 months were all above 0.8 as well. When the reliability value is 0.8, this indicates that phonetic transcription is an acceptable method to be used by clinicians and researchers (Lance et al., 2006). This data demonstrates that phonetic transcription becomes a reliable method for documenting speech sound production at 21 months of age.

The increase in reliability as a child ages may be due to the fact that as infants age, their speech becomes more adult-like and begins to contain more canonical syllables. Evidence suggests that transcribing canonical syllables leads to a higher reliability than transcribing earlier developing vocalizations (Ramsdell et al., 2007). Additionally, prior research by Stockman and colleagues (1981) has indicated that the reliability for phonetic transcription increased around 20 to 21 months of age. Though the methods for calculating reliability and the reliability values obtained differed between the present study and that by Stockman and colleagues (1981), both studies align in that they found a significant change in reliability around 21 months of age. Having an age for which phonetic transcription is reliable will allow clinical and research professionals to have more evidence as to what method will be most appropriate to select for documentation of speech sounds at various ages. Specifically, prior to 21 months of age clinicians and researchers should utilize alternative methods to phonetic transcription (such as caregiver report) for analysis of infant vocalizations.

**Clinical Implications**

Identifying when phonetic transcription is reliable for documentation of infant/early child productions is extremely important for clinical purposes in speech-language pathology. When conducting an evaluation on a new client, SLPs typically have limited time completing the assessment. Knowing which method is most appropriate for recording and analyzing speech sounds at which age is essential for efficiency during the assessment process. Being able to select the most appropriate method, whether it is phonetic transcription, an articulatory model, observation, or an automated computational measure, will make the evaluation more thorough and accurate. This will help SLPs identify more children who are candidates for early intervention programs.

Since assessment sets the foundation for therapy, a stronger assessment will help clinicians to develop effective treatment plans that will provide the most benefit for the client. Being able to analyze the speech sounds of infants will provide clinicians with detailed information on which speech sounds the child is able to produce, and which ones have not developed yet. This will allow SLPs to get a jump-start on introducing new sounds to young children in early intervention programs. Use of phonetic transcription prior to 21 months of age is not appropriate for such analysis; alternative methods should be implemented. As previously mentioned, the sooner an early intervention program starts, the greater communication gains the child will make (Paul & Roth, 2011).

**Research Implications**

From observation to articulatory models to automated computational measures, there are many different ways researchers have to study infant vocal development (Lieberman & Lohmander, 2014; Serkhane et al., 2007; Xu et al., 2014). It will be helpful for researchers to know at what age phonetic transcription is an additional method that can be used to analyze speech sounds in early development.

Having a plethora of research in the area of infant vocal development is needed in order for SLPs to make educated, evidenced-based decisions in their clinical practice and provide effective therapy for their clients. There is still so much to learn in the area of infant vocal development. Knowing that phonetic transcription is reliable across transcribers from 21 months and up can help pave the way for future research in the area. Next, we must determine the effectiveness and clinical efficiency of other methods for analysis of vocal development prior to 21 months of age. Having a larger pool of research to draw from will make speech-language therapy more effective in early intervention programs.

**Limitations**

While results supported the study hypothesis, several limitations can be discussed. One limitation to this study is the small sample size. For the children from 7 to 18 months of age, Schroeder (2016) studied vocalization from seven children at each month of age. For the 19 to 24 month old children, a sample of vocalizations was obtained from 12 children, two at each month of age. Therefore, there were 7 participants in each group for the 7 to 18 month olds and 2 participants in each group for the 19 to 24 month olds. The low number of children at each month of age could have affected the results of this study. Another limitation to this study is, due to time constraints, a relatively low number of utterances were transcribed for each participant. Twenty utterances were transcribed for each participant at each month of age. Increasing the number of utterances transcribed could potentially improve this study. Another limitation to this study is that only inter-rater reliability was analyzed; intra-rater reliability was not considered. Examining intra-rater reliability would strengthen the results of this study.

Though efforts were taken to ensure that methods were as consistent as possible throughout this study, there are several differences between the methods for certain groups that create extraneous variables in this experiment. For instance, though one of the transcribers transcribed all of the utterances from 7 to 24 months of age, the remaining transcribers for the participants between 7 to 18 months of age differed from the transcribers for the participants between 19 to 24 months of age. All transcribers were trained in the International Phonetic Alphabet and in phonetic transcription. They also followed the same guidelines and procedures to limit the effect of having different transcribers for different age groups. However, all transcribers did differ in terms of their years of experience with phonetic transcription and their exposure to sounds that are non-native to English that were included in the transcriptions, so variability between transcribers could have affected the results as well.

Another limitation to this study is that data demonstrated there was a significant difference between the transcribers for the children from 19 to 24 months of age, which could have impacted the inter-transcriber reliability. However, this difference could be due to the small sample size. The third transcriber from which this this value was calculated only transcribed 20 of the utterances and only transcribed utterances from one participant. This small sample makes it challenging to draw conclusions about the difference between the transcribers, so more data would be needed in order to make this determination.

Additionally, in the 19 to 24 month old group, data was gathered in-person while other data was gathered from the online research database PhonBank. The in-person participants were required to meet specific inclusion criteria (see methods), while online participants only met some of the criteria to our knowledge. All of the online participants received a hearing screening, had normal development established through a parent repot, and were administered the *Battelle Developmental Screening Inventory*. These measures helped to ensure that the online participants met the normal hearing requirement and the typical speech and/or language requirement. However, it was not feasible to ensure that online participants met the remaining requirements that the in-person participants were required to meet. Additionally, all of the in-person participants were recorded in the ISU Infant Vocal Development Lab during a period of free play with one or more of their parents/guardians. Therefore, all of these participants were interacting in the same setting, with the same selection of toys and activities, and with similar types of communication partners. However, the online participants were recorded during natural interactions with their families in their home. Therefore, recordings took place either indoors or outdoors, during many different activities, and with various different communication partners. This variability between participants could have impacted the results of the study.

One of the guidelines that was established for the transcribers stated that they were only permitted to listen to an utterance six times before transcribing. This guideline was in place in order limit the variability between the transcribers' procedures, however this requirement proved to be a challenge when transcribing utterances from older children. As children aged, their utterances began to increase in length, and transcribers agreed that it was occasionally difficult to transcribe these longer utterances when only listening to the utterance six times. Therefore, this may have negatively impacted the inter-transcriber reliability for this study at the older ages. Increasing the number of times that a transcriber could listen to the utterances to ten times may have increased reliability further yet.

**Future Research**

Based on the results and limitations of this study, it is recommended that this study be replicated with a larger sample size in order to confirm this study's finding that phonetic transcription reaches a reliability value of greater than 0.8 at 21 months of age. Having a larger sample of participants, particularly at the ages between 19 to 24 months of age, will provide further evidence as to when phonetic transcription is an appropriate method for both clinical and research use. It is also recommended that in replicating this experiment, transcribers be allowed to listen to utterances up to ten times. This will aid in phonetic transcription reliability when the child is in the older months and is producing more lengthy utterances. Additionally, having consistent methods across age groups to limit extraneous variables is recommended.

**Conclusion**

This study concludes that based on a weighted approach using inter-rater reliability, phonetic transcription becomes a reliable method to use when children are 21 months of age. Accordingly, other methods should be used for analyzing vocalizations prior to 21 months of age. Clinicians and researchers can use this research to support their decisions in utilizing phonetic transcription when children are 21 months or older, and in using alternate methods of examining speech sounds in the younger ages.

References

Cucchiarini, C. (1996). Assessing transcription agreement: Methodological aspects. *Clinical Linguistics & Phonetics, 10*(2), 131-155.

Davis, B. L., MacNeilage, P. F., & Matyear C. L. (2002). Acquisition of serial complexity in speech production: A Comparison of Phonetic and Phonological Approaches to First Word Production. *Phonetica*, 59, 75-107.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 1–173.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science, 19*(5), 515-523.

Hebbeler, K. (2009). *First five years find briefing. Presentation given at a Congressional briefing on June 11, 2009, to discuss education that works: The impact of early childhood intervention on reducing the need for special education services.* Retrieved from http://www.sri.com/neils/pdfs/FFYF_Briefing_Hebbeler_June2009_test.pdf

Knight, R. (2010). Transcribing nonsense words: The effect of numbers of repetitions and voices. *Clinical Linguistics & Phonetics, 24*(6), 473-484. doi:10.3109/02699200903491267

Lance, C. E., Butts, M. M., Michels, L.C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, 9(2), 202–220. doi: 10.1177/1094428105284919

Lieberman, M., & Lohmander, A. (2014). Observation is a valid way of assessing common variables in typical babbling and identifies infants who need further support. *Acta Paediatrica: Nurturing the Child, 103*, 1251-1257. doi:10.1111/apa.12776

MacWhinney, B. (n.d.). *TalkBank Code of Ethics*. Retrieved from https://talkbank.org/share/ethics.html

Nakatani-Murai, T. (2008). On the properties of infant vocalization in the Japanese pre-verbal stage. *Nanzan Linguistics, 3*(2), 123-137.

Oller, D. K. (2000). The emergence of the speech capacity. Mahwah, N.J: Lawrence Erlbaum Associates.

Oller, D. K., & Delgado, R. E. (1999). *Logical International Phonetics Program* (Version Windows). Miami: Intelligent Hearing Systems Corp.

Oller, D.K., & Lynch, M.P. (1992). Infant vocalizations and innovations in infraphonology: Toward a broader theory of development and disorders. In C. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 509-536). Parkton, M.D.: York Press.

Oller, D. K., & Ramsdell, H. L. (2006). A weighted reliability measure for phonetic transcription. *Journal of Speech, Language, and Hearing Research, 49*(6), 1391-1411.

Paul, D. & Roth, F. P. (2011). Guiding principles and clinical applications for speech-language pathology practice in early intervention. *Language, Speech, and Hearing Services in Schools, 22,* 320-330. doi:10.1044/0161-1461(2010/09-0079)

Ramsdell, H. L., Oller, D. K., Buder, E. H., Ethington, C. A., & Chorna, L. (2012). Identification of prelinguistic phonological categories. *Journal of Speech, Language, and Hearing Research*, *55*(6), 1626-1639. doi:10.1044/1092-4388(2012/11-0250)

Ramsdell, H. L., Oller, D. K., & Ethington, C. A. (2007). Predicting phonetic transcription agreement: Insights from research in infant vocalizations. *Clinical Linguistics & Phonetics, 21*(10), 793-831.

Rose, Y., & MacWhinney, B. (2014). The PhonBank Project: Data and software-assistend methods for the study of phonology and phonological development. In J. Durand, U. Gut & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 380-401). Oxford, UK: Oxford University Press.

Schroeder, K. (2016). *Phonetic transcription in vocal development: When is reliability achieved?* (Master's Thesis).

Serkhane, J. E., Schwartz, J. L., Boe, L. J., Davis, B. L., & Matyear, C. L. (2007). Infants' vocalizations analyzed with an articulatory model: A preliminary report. *Journal of Phonetics, 35*(3), 321-340.

Sosa, A.V. (2015). Association of the type of toy used during play with the quantity and quality of parent-infant communication. *The Journal of the American Medical Association (JAMA) Pediatrics, 170,* 132-137.

Stockman, I. J., Woods, D. R., & Tishman, A. (1981). Listener agreement on phonetic segments in early infant vocalizations. *Journal of Psycholinguistic Research, 10*(6), 593-617.

Xu, D., Richards, J. A., & Gilkerson, J. (2014). Automated analysis of child phonetic production using naturalistic recordings. *Journal of Speech, Language, and Hearing Research, 57*(5), 1638-1650. doi:10.1044/2014_JSLHR-S-13-0037