

INVESTIGATION OF SPEECH SAMPLES FROM
TYPICALLY DEVELOPING PRESCHOOL AGE CHILDREN:
A COMPARISON OF SINGLE WORDS AND IMITATED SENTENCES
ELICITED WITH THE PABA-E

by

Matthew William Olsen

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Speech and Hearing Sciences

Thesis Committee:
Christina Gildersleeve-Neumann, Chair
Mary Gordon-Brannan
Amy Donaldson

Portland State University
2010

ABSTRACT

Assessment of speech sound production in young children provides the basis for diagnosis and treatment of speech sound disorders. Standardized single-word articulation tests are typically used for identification of speech sound errors because they can provide an efficient means of obtaining a speech sample for analysis and comparison to same-age peers. A major criticism of single-word articulation tests is that they may not accurately reflect speech sound production abilities in conversation. Comparison of performance in single-word and conversational contexts has produced conflicting results in the available research.

The purpose of the present study was to compare speech samples obtained using an extensive single-word naming task with samples of continuous speech elicited by sentence imitation. It was hypothesized that there would be differences in overall speech sound production accuracy as well as differences in types and frequency of errors across the two sampling conditions. The present study is a pilot investigation as part of the development of the *Phonological and Bilingual Articulation Assessment*, English Version (PABA-E; Gildersleeve-Neumann, unpublished).

Twelve preschool children ages 3;11 to 4;7 (years;months) from the Portland Metropolitan area participated in this study. Participants were monolingual native English speakers and exhibited typical speech sound development as measured by

the GFTA-2 (Goldman-Fristoe, 2000). Hearing acuity for participants was within acceptable limits, and participants' families reported no significant illnesses or developmental concerns that would impact speech sound production abilities.

Mean *t*-scores for percentage of consonants correct (PCC) in the single-word samples were significantly higher at the .05 level than those for the sentence imitation samples. There was no significant difference between the percentage of vowels produced correctly (PVC) in the two sampling conditions. Similar types of error patterns were found in both the single-word and continuous speech samples, however error frequency was relatively low for the participant population. Only the phonological process of stopping was found to be significantly different across sampling conditions. The mean frequency of occurrence for stopping was found to be significantly higher in continuous speech as compared with the production of single-words.

TABLE OF CONTENTS

Abstract.....	i
List of Tables.....	iv
List of Figures.....	v
Chapter 1: Introduction.....	1
Chapter 2: Literature Review.....	2
Chapter 3: Method.....	20
Chapter 4: Results and Discussion.....	32
Chapter 5: Conclusion and Implications.....	61
References.....	67
Appendix A: Parent Survey of Language Development.....	72
Appendix B: PABA-E Single-Word and Imitated Sentences Lists.....	74

LIST OF TABLES

Table 1: Demographic Data and GFT-2 Scores.....	21
Table 2: IPC Scoring Grid.....	28
Table 3: Total Transcription Reliability.....	30
Table 4: Transcription Reliability of Individual Samples.....	31
Table 5: PCC and Within-Subject Tests of Difference in Proportion.....	32
Table 6: Group Comparison of Percentage Accuracy for Consonants and Vowels.....	32
Table 7: Percentage of Consonants Produced Correctly by Word Position.....	34
Table 8: Group Comparison of Percentage Accuracy for Consonants across Word Positions.....	34
Table 9: Group Comparison of Frequency Occurrence for Phonological Error Patterns.....	37
Table 10: Occurrence of Word Structure Errors.....	38
Table 11: Occurrence of Glide and Liquid Errors.....	40
Table 12: Occurrence of Substitution Errors.....	43
Table 13: Group Comparison of Index of Phonetic Complexity (IPC).....	44
Table 14: Administration and Transcription of Speech Samples.....	46

LIST OF FIGURES

Figure 1
Mean Percentage of Consonants Correct (by Word Position)..... 33

CHAPTER 1: INTRODUCTION

The goal of a speech sound assessment is to identify, describe, and classify a child's speech patterns (Elbert & Gierut, 1986; Grunwell, 1997; Williams, 2002). Typically, children's speech production abilities are compared to the adult standard in the target language (relational analysis), but the children's abilities may be identified and described independently of any standard competency (independent analysis) depending upon their developmental level and the extent of their speech production repertoire (Bernthal & Bankson, 2004; Stoel-Gammon & Dunn, 1985; Williams, 2003). Based on analysis of the child's speech patterns, the clinician must determine whether intervention is warranted and then select and organize targets for remediation (Bernthal & Bankson, 2004; Bleile, 2004). This entire process is dependent upon accurate analysis of a speech sample that is representative of the child's abilities.

Practices and procedures for assessing children suspected of having speech sound disorders have been a major topic of discussion over the past several decades. Issues regarding single-word productions and connected speech have been examined extensively in an attempt to determine the optimal method for collecting a speech sample. Methodological issues related to the influence of elicitation mode on the responses obtained are complex, requiring consideration of a number of important and related factors.

CHAPTER 2: LITERATURE REVIEW

Not only is speech sampling a critical aspect of assessment, but the need for efficient procedures has been supported by a number of recent investigations. The American Speech-Language-Hearing Association (2008a, 2008b) reported that speech-language pathologists working in a variety of school-based settings serve an average of 50 clients per week, including a substantial percentage of children with speech sound disorders. Respondents reported devoting an average of 4 hours per week to assessment and diagnostic activities, while the majority of time was spent providing direct intervention. Skahan, Watson and Lof (2007) reported comparable data from their national survey of speech-language pathologists serving large caseloads of preschool and school-age populations. Information provided by participants indicated that activities such as data analysis and paperwork frequently consume the majority of assessment time. Furthermore, the investigators suggested that direct assessment practices of their survey respondents were influenced to a great extent by state and federal guidelines for qualifying children to receive speech intervention through special education. Time spent interacting with the child and administering formal and informal tests appeared to be tailored to meet the criteria for determining eligibility and establishing goals and objectives for treatment. These reports underscore the clinical value in continuing to explore efficient methods for assessing speech in order to address the demands of heavy caseloads and inevitable time constraints. At the same time, any methodology must be sufficiently

comprehensive to ensure an appropriate diagnosis resulting in intervention that is both warranted and effective.

COMPARISON OF TWO SAMPLING CONDITIONS

The issue of efficiency has been the central argument in favor of obtaining a speech sample through elicitation of single-word productions, and perhaps the greatest advantage of this method is the ease of administering and scoring a single-word naming test. The speech sample can be elicited quickly, and analyses of responses are generally straightforward and have been adapted to a variety of theoretical paradigms (Williams, 2003). Because the stimuli are controlled by the examiner, single-word naming provides an opportunity to test all the sounds of the target language across relevant word positions in a relatively short amount of time. Control of content also has the advantage of a known referent which can be essential when attempting to transcribe the productions of a child who is highly unintelligible (Paden & Moss, 1985). Additionally, single-word naming tests allow for standardization of response criteria and comparisons of performance across children. Obtaining a standard score may be a necessary requirement when the priority is determining eligibility for services (Khan, 2002). The score is also a useful measure of baseline performance. Multiple scores obtained with the same probe can be used to document progress over the course of an intervention.

Single-word articulation tests typically require the examinee to spontaneously name a series of objects or pictures. Single-word tests traditionally assess consonant

singletons in initial, medial, and final word positions, with opportunities to evaluate a limited number of consonant clusters. Procedures for analyzing speech production errors vary, but are typically based on a single attempt for each target to document the level of mastery of the speech sound. Standard procedures usually score the child's production as either correct or incorrect for each stimulus item with the option of rating the severity of any errors noted during examination. The clinician may choose to augment the pass/fail analysis with a transcription of the child's production of the target, thus providing supplemental information regarding specific error types that can be helpful in planning treatment.

Single-word articulation tests have been criticized for their limited size and scope (Klein, 1984; Miccio, 2002). Following standard procedures, mastery of a target sound may be assumed from a single attempt, potentially ignoring any influence of phonetic context on speech production accuracy and making it difficult to determine the consistency of any given response. Furthermore, articulation tests typically assess only a few salient consonant clusters, and assessment of vowels and diphthongs are typically limited with regard to available standardized protocols. According to Klein (1984) and others, single-word tests may result in overestimation of speech production abilities (DuBois & Bernthal, 1978; Faircloth & Faircloth, 1970; Smith & Ainsworth, 1967; Stoel-Gammon & Dunn, 1985).

A major disadvantage of single-word samples is that they may not be an accurate reflection of articulatory performance in everyday speech (Dubois &

Berenthal, 1978; Faircloth & Faircloth, 1970; Johnson, Winney & Pederson, 1980; Smith & Ainsworth, 1967; Stoel-Gammon & Dunn, 1985). Conversely, a sample of conversational speech may be a more representative test of a child's performance in real-life communication. Morrison and Shriberg (1992) argued that the interaction of speech and language observed in social discourse "when one is talking to be understood" (p. 259) provides the most ecologically valid context for assessing speech production. While the influence of phonetic context on articulation has been well documented (e.g., Gallagher & Shriner, 1975a, 1975b; Hoffman, Schuckers & Ratusnik, 1977; Kent, 1982; Prather & Kenney, 1986; Zehel, Shelton, Arndt, Wright & Elbert, 1972), research has also shown that pragmatics, semantics, linguistic stress, syntax and morphology all contribute to articulatory performance (e.g., Campbell & Shriberg, 1982; Klein & Spector, 1985), adding support to the view that continuous running speech may reveal more typical patterns of speech production. Additionally, suprasegmental factors that affect conversational intelligibility such as prosody and rate can be evaluated only in connected speech (Bernhardt & Holdgrafer, 2001a, 2001b; Masterson, Bernhardt, & Hofheinz, 2005; Shriberg, 1993).

Overall speech intelligibility has been reported to be highly variable in young children, both typically developing and those exhibiting extensive speech production errors (Gordon-Brannan & Hodson, 2000). According to Kwiatkowski and Shriberg (1992), speech sound errors account for only a small percentage of the variance observed in intelligibility (see also Shriberg, Kwiatkowski, Best, Hengst, & Terselic-

Weber, 1986). Many other contextual and linguistic variables contribute to the perception of speech intelligibility, including prosody, syntax, morphology and semantics (Gordon-Brannan & Hodson, 2000; Kent, Miolo, & Bloedel, 1994; Kwiatkowski & Shriberg, 1992; Weston & Shriberg, 1992). A sample of continuous speech would appear to provide the most comprehensive information for judging intelligibility given the complexity of factors that must be accounted for in an evaluation. Support for this notion was reported by Gordon-Brannan and Hodson (2000) who found that samples of continuous speech were strongly correlated with four independent measures of intelligibility, while tasks involving single-words resulted in considerably more variability across measures.

Unlike single-word responses, the information obtained from a connected speech sample allows for an integrated analysis of speech, language, and prosodic factors. At the same time, intelligibility in conversation is often an important consideration in clinical judgments of speech sound disorder severity and is critical for decisions regarding intervention (Bernthal & Bankson, 2004). In these respects, assessing connected speech may be more efficacious when planning treatment (Andrews & Fey, 1986; Gierut, 1998; Kwiatkowski & Shriberg, 1998; Morrison & Shriberg, 1992; Schwartz, 1992; Shriberg, 1993; Shriberg & Kwiatkowski, 1985). However, under certain conditions, this method may not be ideal. Conversational samples obtained from highly unintelligible children may be difficult or impossible to transcribe reliably without knowledge of the intended utterance or a referent for the

exact target. Some children may be uncooperative or hesitant to engage in a spontaneous interaction with the clinician, and limited speech output will likely yield a sample insufficient for comprehensive analysis (Wolk & Meisler, 1998). Even when a sample is representative of a child's everyday speech, the resulting data may not be representative of the target language (Masterson et al., 2005). Missing exemplars of sound segments and syllable shapes could be a major source of bias in the diagnostic evaluation.

A number of researchers have documented fewer opportunities to analyze certain sounds, sound clusters, and complex word shapes in samples of natural, spontaneous speech when compared with elicited single-word samples. Morrison and Shriberg (1992) found a significant difference in the distribution of intended word shapes between the two sampling modes. The conversational samples they analyzed contained a larger proportion of simple word shapes and significantly fewer examples of two-syllable words as compared to the single-word samples. Differences were also observed in the distribution of intended consonant targets across sampling modes. Fewer opportunities were available to analyze certain speech sounds in the conversational context. Similar differences in the distribution of sounds and word shapes across the two sampling modes have been reported by other researchers (Masterson et al., 2005; Wolk & Meisler, 1998). Wolk and Meisler (1998) suggested that observed differences may be a consequence of a child's tendency to avoid certain sounds and word shapes that are too difficult to produce. However, based

solely on samples of conversational speech, it is difficult to determine whether the differences represent avoidance of certain targets or a discrepancy in the number of opportunities to produce them.

Lack of control over content may be a trade off for the advantage of ecological validity in a sample of connected speech, and two important clinical implications emerge with respect to speech sound assessment. First, the sample of connected speech may not provide sufficient information about a child's speech production capabilities. Investigators have attempted to develop a variety of protocols for collecting connected speech samples which allow the clinician some degree of control over the interaction in order to ensure the sample will contain sufficient opportunities to test all salient sounds across relevant word positions (Shriberg & Kwiatkowski, 1985). For example, the child may be asked to spontaneously describe a picture or tell a story from a visual stimulus constructed to elicit specific targets. The clinician may attempt to conduct a structured interview of the child, with specific questions directed to elicit a desired response. Objects or pictures may be combined with verbal prompts from the clinician during a play routine or other interaction to exert some level of control over the child's responses. Alternately, the clinician may administer a delayed imitation task such as the Sounds in Sentences subtest from the *Goldman-Fristoe Test of Articulation, Second Edition* (GFTA-2; Goldman & Fristoe, 2000). In this procedure, the clinician tells a story based on a series of illustrations, and the child is asked to retell the story from memory with

support from the visual aids. The more control exhibited by the clinician over a connected speech sample, the less the representative that sample may be of a child's everyday speech (Shriberg & Kwiatkowski, 1985). However, the likelihood of obtaining the desired targets for analysis is increased.

The second issue related to the content of a connected speech sample is standardization. A standardized assessment provides a valid and reliable means for determining whether any errors observed in speech sound production differ from what is expected in typical development. Comparison of a child's performance to normed response criteria is useful for identifying the presence or absence of a disorder. Individual variability typical in samples of connected speech, however, illustrates the difficulty of establishing standardized norms with this procedure. The difficulty of such a task is highlighted in a study conducted by Masterson et al., (2005) who analyzed single-word and connected speech samples of 14 children and found that some participants in their study did not attempt to produce certain segments in conversation. These results were consistent with Morrison and Shriberg (1992) and Wolk and Meisler (1998) who also found individual differences in the sounds and sound clusters attempted by participants in conversational speech. With limited clinician control over content, there appears to be less consistency in the corpus of items across samples of conversational speech, and intersubject comparisons of performance are potentially less reliable.

SPEECH SAMPLING FOR CLINICAL ASSESSMENT

Obtaining a standardized test score may be an important outcome of a speech sound assessment. State and federal guidelines established by the Individuals with Disabilities Education Act (IDEA; 2006), as well as third party payers, require evidence of the presence of a disorder to determine whether a child is eligible to receive services. To meet this requirement, obtaining a standardized score from a single-word articulation test is often the focus of an initial assessment (Khan, 2002). Because of the lack of standard procedures for conducting a connected speech assessment and the difficulty of collecting normative data for reference, the use of this procedure is likely to require a significantly greater investment of time to determine whether a child qualifies.

The amount of time required to elicit, transcribe, and analyze a sample of connected speech presents the greatest challenge to its clinical utility. DuBois and Bernthal (1978) reported a significant difference in the amount of time required to obtain a speech sample of 20 stimulus items when they compared elicitation of continuous speech to a single-word picture naming task. Paden and Moss (1985) examined three speech sound assessment procedures and found that collecting and analyzing samples of connected speech required approximately two to four times longer than comparable single-word procedures that were collected, transcribed, and analyzed in just under an hour. Conducting a comprehensive assessment of spontaneous conversational speech may not be practical in clinical settings that serve

large numbers of children, particularly when reporting a standard score is a priority and the amount of time available is limited (Khan, 2002; Klein, 1984).

Maximizing the utility of multiple sampling modes may increase the efficiency and thoroughness of a functional speech sound assessment. A number of researchers and clinicians have recommended collecting both single-word and connected speech samples for evaluation (Bernhardt & Holdgrafer, 2001a; Bleile, 2002; Dyson & Robinson; 1987; Hodson, Scherz, & Strattman, 2002; Khan, 2002; Klein, 1984; Miccio, 2002; Tyler & Tolbert, 2002; Wolk & Meisler, 1998), citing the compelling advantages of each elicitation mode and the optimal use of complimentary procedures. Whole-word transcription of responses from a single-word test, as suggested by Klein (1984), maximizes the amount of data from the sample available for analysis. Additional opportunities to analyze production accuracy of sounds and sound sequences across word positions can provide the clinician with important information regarding consistency of errors. There appears to be a general consensus around the use of a conversational speech sample to confirm the presence of any errors observed in single-word productions. In the interest of efficiency, however, it has been suggested that detailed transcription and analysis of connected speech may not be part of the initial assessment (Klein, 1984; Skahan et al., 2007). Instead, an elicited sample of connected speech can provide information about prosodic, linguistic, and contextual influences on speech production without the need for detailed, time-consuming analysis. In addition, the connected speech sample can be

used to estimate intelligibility and, when combined with data from single-word responses, as the basis for clinical judgments regarding severity.

ELICITATION METHODS: A REVIEW OF PREVIOUS FINDINGS

A number of investigations have attempted to answer the question of whether speech sound errors observed in samples of single-word responses accurately represent speech production ability in conversation. The majority of research has focused on comparisons between elicited single-word responses and samples of conversational speech to investigate the theoretical and clinical consequences associated with the choice of sampling method. Outcome measures have been primarily concerned with examining consistency of errors across sampling modes and any differences in the selection of potential remediation targets.

Regardless of sample type, most researchers have attempted to elicit spontaneous speech productions whenever possible rather than relying on imitation tasks. Preference for spontaneously produced samples is based on the assumption that responses are more likely to resemble utterances produced in non-test contexts (Stoel-Gammon & Dunn, 1985). However, the use of imitation as an elicitation technique is a potentially efficient method for gathering data, allowing the clinician more control over content of the speech sample and selection of specific targets. Moreover, issues related to limited lexical knowledge or limited intelligibility may potentially be avoided when the clinician constructs both the referent and the desired response.

Only a few studies have systematically explored the outcome of speech tasks involving imitation. Bankson and Bernthal (1982) observed only small, nonsignificant differences in the errors identified in imitated words and imitated sentences. Direct comparison of imitated versus spontaneous productions have produced conflicting results. DuBois and Bernthal (1978), Faircloth and Faircloth (1970), and Smith and Ainsworth (1967) reported a higher frequency of speech production errors in both single-word and continuous speech samples elicited spontaneously as compared to imitation tasks. On the other hand, Paynter and Bumpas (1977) and Siegel, Winitz, and Conkey (1963) found no significant differences between imitated and spontaneous responses. Test-retest performance scores were found to be stable in a study conducted by Haynes and Stead (1987) when they examined speech production accuracy on consonants embedded in a sentence imitation task. Furthermore, they concluded that performance on the imitation task was strongly correlated with performance in conversational speech.

Researchers have consistently found differences in speech production accuracy depending on the type of sample analyzed. Morrison and Shriberg (1992) reviewed over 50 studies representing a variety of methodologies related to sample size, elicitation mode and linguistic content. Despite the differences reported in speech production profiles obtained from analysis of single-word versus continuous speech contexts, some common trends emerged. Most studies found that continuous speech appears to be associated with a greater number and variety of errors

compared to production of single words (e.g., Andrew & Fey, 1986; DuBois & Bernthal, 1978; Faircloth & Faircloth, 1970; Healy & Madison, 1987; Klein, 1984; Watson, 1989). Reduction and deletion of unstressed syllables, deletion of initial and final consonants, reduction of clusters, stopping, assimilation and vowel neutralization have all been reported with greater frequency in samples of connected speech (DuBois & Bernthal, 1978; Faircloth & Faircloth, 1970; Klein, 1984; Watson, 1989), although in general, similar error patterns were found in single-word contexts.

Some studies, however, have reported that errors related to place and manner of production were observed with greater frequency in single-words (Paden & Moss, 1985; Watson, 1989). Healy and Madison (1987) investigated types of errors produced by 20 children with speech sound disorders and found that nearly 35% of all errors observed in samples of connected speech were realized as different error types in single-word samples. In this study, sounds omitted in continuous speech were more likely to be realized as substitutions or distortions in single word contexts. Differences in the type of error produced within a specific context may be related to decreased rate of speech and more deliberate articulation in single word tasks (DuBois & Bernthal, 1978). The implications of these findings suggest that samples of single-word productions may fail to identify clinically significant errors.

Significant differences in the speech profiles obtained from the two sampling conditions were confirmed by Morrison and Shriberg (1992). Analyses of transcripts from 61 moderate to moderate-severe children with speech delays showed

significantly greater production accuracy for vowels/diphthongs and consonant singletons in continuous speech than in responses from single-word articulation testing. The significance of these findings was supported for individual participants wherein 47 out of 61 participants (77%) performed better in conversation as measured by Percent Consonants Correct (PCC; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997). The continuous speech samples contained a greater proportion of early-developing sounds in simple word shapes, perhaps contributing to higher performance scores in conversation, whereas stimuli in the single-word task were characterized by more complex word shapes and contained more instances of later-developing sounds. Morrison and Shriberg found that the types of errors characterizing the two sampling conditions were especially similar. However, production errors involving clusters, unstressed syllables and assimilations occurred with significantly greater frequency in single-word responses, whereas rates of stopping and final consonant deletion were significantly higher in conversation. Selection of treatment targets based on the results of this study would have resulted in similar clinical decisions for only two-thirds of the errors examined. Morrison and Shriberg concluded that the choice of sampling condition may fail to identify clinically significant errors.

More recent investigations comparing speech elicitation methods for speech sound assessment have attempted to tailor the single-word tasks to ensure a sufficiently diverse and representative assessment of articulatory performance. Wolk

and Meisler (1998) developed an extensive 162-item picture-naming task for eliciting single-word responses. Their test was carefully constructed to elicit multiple exemplars of all phonemes across word positions as well as a large number of consonant clusters within a variety of one-, two- and three-syllable words. Responses from 13 participants with speech sound disorders were compared to samples of continuous speech. Wolk and Meisler found a significant positive correlation between the speech profiles obtained from the single-word and continuous speech tasks, but the single-word samples yielded a significantly greater number of production errors. They concluded that the single-word task was a more comprehensive test of participants' speech production abilities, yielding a more extensive evaluation of sounds and word shapes.

In a study of 20 participants with speech sound impairments, Masterson et al. (2005) elicited single-word samples of 46 core stimulus items supplemented with additional target words chosen in response to participants' performance in the initial sample. The entire single-word sample was then compared to conversational samples elicited from each participant. Speech production requirements across the single-word and continuous speech tasks were found to be comparable. That is, individual speech sounds and word shapes occurred with similar frequency in each sampling condition. In general, PCC was found to be higher in conversation indicating better production accuracy when compared with responses from the tailored single-word task. Analyses of word shapes indicated that most structures were produced

more accurately in the single-word context. Masterson and colleagues argued that consonant deletions, voicing errors, and glottal replacement are often acceptable features of word-final phonemes in conversation, making it difficult to accurately transcribe continuous speech. Their revised analysis demonstrated these characteristics of conversational speech may have significantly affected comparisons of performance data. When these word-final features were no longer counted as errors in their single-word samples, differences between single-word and connected speech were no longer significant. Consistent with findings reported by Morrison and Shriberg (1992), certain sound and syllable errors, particularly those occurring at segment and word boundaries, may be considered acceptable consequences of continuous conversational speech. Masterson and colleagues concluded that regardless of sampling mode, potential targets selected for treatment would have been the same for the most participants.

THE PURPOSE OF THE PRESENT STUDY

Given the various methodologies used to compare continuous speech with single-word responses, it is difficult to determine which elicitation mode more accurately reflects the speech production performance of the participants assessed. There appears to be no clear characterization of findings presented in the research literature to date. Rather, potential sources of variance are likely related to a number of factors such as test construction, method of presentation, size and content of the sample, structural differences in the stimulus items, linguistic effects, and the

response definitions that guide transcription. There does appear to be some support that single-word tests, when carefully constructed, can provide a comprehensive evaluation of a child's speech production abilities (Masterson, et al., 2005; Wolk & Meisler, 1998). However, these tests provide limited information about prosodic, linguistic, and contextual influences on speech production. The development of a second, complementary elicitation procedure to address these additional variables may prove to be clinically beneficial. Because of the critical need for efficient assessment procedures, the use of an imitated sentences task to elicit a sample of connected speech may be a clinically useful means of evaluating speech production performance within a limited time frame.

The purpose of the present study was to compare speech samples obtained from the *Phonological and Bilingual Articulation Assessment, English Version* (PABA-E; Gildersleeve-Neumann, unpublished) using two elicitation procedures: an extensive single-word naming task (SWT) and an imitated sentences task (IST). The project is a pilot investigation as part of the development of the English sentences task. Because the investigation was intended to be exploratory, speech samples were collected from typically developing children in order to test some general characteristics of the experimental stimuli. This study compared broad measures of overall speech sound production accuracy in speech samples elicited by each procedural method. In addition, comparisons were made of the number and types of errors elicited using the two sampling conditions.

It was hypothesized that there would be measurable differences observed between the speech samples collected with the two elicitation procedures. The question posed by this study was whether speech sound accuracy would be better in either the single-word or the connected speech context. Previous research comparing spontaneously generated conversational speech to single-word naming procedures found fewer errors in the connected speech tasks (Masterson et al., 2005; Morrison & Shriberg, 1992; Wolk & Meisler, 1998). However, differences in the phonetic complexity of the experimental tasks used in these studies may have contributed to speech performance since the spontaneously generated conversational speech was reported to contain fewer complex segments and structures than the extensive naming tasks constructed by the researchers. In the present study, sentence imitation, rather than spontaneously generated conversation, was used to elicit connected speech samples. Since sentence imitation allowed a great deal of control over the samples elicited, it was suspected that the phonetic complexity of the SWT and the IST would be closely matched. In the present study, both the SWT and the IST were carefully constructed to contain multiple opportunities for testing similar segments and structures. It was suspected that poorer speech production accuracy in the IST would be the result of linguistic and contextual influences on speech production abilities.

CHAPTER 3: METHOD

PARTICIPANTS

Thirteen typically-developing preschool children were initially recruited as participants for this project. Ages ranged from 3;3 to 4;7 (years;months). Participants were recruited through contacts with the author's previous graduate school supervisors. Flyers were distributed at Helen Gordon Child Development Center, The Emerson School, and Buckman Arts Focus Elementary School seeking preschool aged children to take part in this pilot investigation. Twelve families responded during the recruitment phase resulting in 13 participants due to the inclusion of one set of fraternal twins. The participants self-selected to take part in this study by responding directly to the author.

Data collected from one of the participants were not used in the analyses for this project. The child was the youngest participant (3;3) recruited, 8 months younger than the next youngest participant. This participant demonstrated some difficulty completing the experimental tasks and did not complete the initial screening procedures used for this project. In addition, the author determined that the participant's speech samples were too different from the samples collected from the older participants, likely a result of the difference in ages and development. For these reasons, the participant's samples were excluded from analyses.

Nine female and 3 male children participated in this investigation (see Table 1). Ages ranged from 47 to 55 months (mean age of 50.92 months). All participants

were monolingual English speakers from monolingual native English-speaking homes. One participant was adopted into the United States at the age of 7 months 6 days. Parents were asked to complete a short developmental history questionnaire for each participant (Appendix A). The questionnaire was used to screen for children reported to be typically developing by their parents and to rule out any parental concerns regarding speech development. No major concerns were documented, and all participants were reported to be typically developing. No major illnesses were reported to have a negative impact on speech sound development.

Table 1
Demographic Data and GFTA-2 Scores

Part.	Sex	Age (years;months)	GFTA-2	
			Stnd Score	%ile
P01	M	3;11	114	82
P02	F	4;6	116	87
P03	F	4;2	109	63
P04	F	4;3	116	82
P05	M	4;5	112	83
P06	F	4;5	112	70
P07	F	4;2	113	74
P08	F	4;7	114	74
P09	F	3;11	117	87
P10	F	4;1	112	72
P11	M	4;3	119	98
P12	F	4;3	120	93

Mean Age (in months) = 50.92; Standard Deviation = 2.54
Mean Score (GFTA-2) = 114.5; Standard Deviation = 3.21; S.E. = 0.91
Mean %ile (GFTA-2) = 80.42; Standard Deviation = 10.14; S.E. = 2.93

To determine whether participants demonstrated typically developing speech sound production skills, each participant was administered the *Goldman-Fristoe Test*

of Articulation, Second Edition Sounds-in-Words Subtest (GFTA-2; Goldman & Fristoe, 2000). This standardized assessment was used to ensure that all participants demonstrated speech sound production skills considered typical in comparison to other children of similar ages. Standard scores ranged from 109 to 120 (mean = 114.5, standard deviation = 3.21) suggesting that the participants all performed in the high-average range, well within acceptable limits measured in typically developing peers. Eleven participants passed a bilateral hearing screening at 20 dB for 1000, 2000, and 4000 Hz. Participant 12 passed the hearing screening for the right ear at 20 dB and responded at 25 to 30 dB for the left ear. The GFTA-2 and the hearing screenings were administered by the author of this study.

Based on the information obtained from parents as well as the GFTA-2 and the hearing screenings, the examiner determined that the participants met the eligibility criteria for participation in this study.

MATERIALS

The GFTA-2 was administered as a standardized metric of speech production accuracy. The GFTA-2 was chosen as the standardized measure because of its similarity to the experimental procedures, straightforward administration/scoring, moderately comprehensive design, and general popularity with practicing speech-language pathologists (Skahan, et al., 2007).

The *Phonological and Bilingual Articulation Assessment, English Version* (PABA-E; Gildersleeve-Neumann, unpublished) was used as the experimental

procedure in this study (Appendix B). The PABA-E consisted of two tasks designed to elicit speech samples from each participant in two different contexts for this pilot investigation. The English single-word list (SWT) consisted of 135 items representing objects, actions and attributes familiar to young children. The list was constructed to test English consonants, vowels, and consonant clusters in a variety of word positions and contexts.

The PABA-E sentence list (IST) was designed as a complementary procedure to the SWT. The IST consisted of 18 sentences to evaluate speech sound production in continuous speech. Many of the same target words contained in the SWT were used in construction of the IST.

PROCEDURE

Each participant was administered the GFTA-2 according to the standard protocol. The PABA-E SWT and IST were administered to each participant as the experimental tests. The procedures were conducted over two sessions, and each individual session lasted approximately 30 minutes. A minimum of one day and a maximum of seven days elapsed between sessions. In one of the sessions, the participant received a hearing screening and was assessed with the GFTA-2. In the other session, the participant was administered the single-word and imitated sentences tasks of the PABA-E. Order of administration of the GFTA-2 and the experimental tasks was counterbalanced across participants in order to control for any potential learning effects that might influence performance and the resulting

samples. There were 4 possibilities for ordering the sampling tasks: (1) GFTA-2, SWT, IST; (2) GFTA-2, IST, SWT; (3) SWT, IST, GFTA-2; and (4) IST, SWT, GFTA-2.

Elicitation of the speech samples occurred in a variety of environments. Participant recruitment sites, participants' homes, and the PSU Speech and Hearing Clinic were all used as data collection sites. Some of the participants were accompanied by a parent or familiar adult during data collection. Four of the participants were assessed in pairs, each participant taking turns. In these cases, the examiner scheduled the tasks to be different for each participant during the session. All speech samples were collected by the author of this study in an available area that was quiet and free of distraction.

The SWT was administered according to the following procedure which has been used in previous research with this experimental protocol. A participant was shown a series of pictures, each representing a stimulus item. Spontaneous responses were elicited by asking "What's this?" or "This is a ___?" For no response to an initial bid, an attempt was made to elicit a response using delayed imitation (e.g., "This is a ____. We use it for x. What did I say it is?"). With some of the first participants, attempts were made to prompt the desired response by offering a choice between the actual target and an absurd alternative (e.g., "Is this a fork or a car?"). The examiner determined that this elicitation procedure often resulted in intentionally incorrect responses from the participants. Elicitation procedures were eventually simplified to "This is a ____. What did I say this is?"

The IST was administered via direct imitation. Participants were instructed on how to “play a game” with the examiner by listening to what the examiner says and then repeating the same words. Each of the 18 stimulus sentences was illustrated by a corresponding picture that depicted the major features of the target sentence. The participant was shown a stimulus picture while the examiner read the corresponding sentence. The participant was then asked to repeat the examiner’s words. Three practice sentences were used to teach the task before implementing the experimental stimuli for a total of 21 imitated sentences. If a participant did not respond to an initial request or if the response did not contain all the target words, additional attempts were made to elicit the target sentence. Most participants responded correctly after the initial bid. Only rarely was a third attempt required to elicit the desired response.

Speech samples were recorded on a Sony Handycam DCR SX40, and the resulting video files were downloaded and saved onto a computer for later transcription. Phonetic transcription was completed from the video files during playback using VLC Media Player on a laptop computer with Sennheiser HD 280 Pro headphones.

Speech samples were transcribed using the International Phonetic Alphabet (IPA) and entered into Logical International Phonetic Program (LIPP) software for analysis (Oller & Delgado, 2000). Relational analyses were performed by comparing participants’ responses with the adult targets represented in the PABA-E. For both

the SWT and the IST, participants were not required to produce the stimulus items verbatim. The comprehensive construction of the experimental procedures allowed for some minor deviations from the desired targets. When the gloss of a participant's response resulted in an intended utterance that differed from the target stimuli, response criteria and the adult target used for comparison were altered to reflect the participant's intended utterance.

DATA ANALYSIS

Speech samples obtained from the SWT and the IST were compared for overall production accuracy. PCC was calculated for each sample based on the number of consonants produced correctly out of the total number of consonants attempted. The Percentage of Vowels Correct (PVC; Shriberg et al., 1997) was also calculated for each sample as a measure of the number of vowels produced correctly out of the total number of vowels attempted. When appropriate, vowel targets were changed to match dialectical variations within an intended utterance and were counted as correct. The total number of consonants and vowels attempted varied by participant and elicitation method. In addition to overall accuracy of speech sounds produced, samples from the SWT and the IST were compared with respect to percentage of consonants produced accurately in initial, medial, and final word positions.

Speech samples obtained with the SWT and the IST were compared for the frequency of occurrence of phonological error patterns. Several error patterns were

investigated in this study. Eight phonological error patterns were described by Shriberg and Kwiatkowski (1980) including cluster reduction, liquid simplification, stopping, velar fronting, final consonant deletion, palatal fronting, unstressed syllable deletion, and assimilation. A similar method for categorizing error patterns was attempted in the analyses of speech samples collected for this investigation. LIPP analyzed each of the 24 samples for specific phonological error patterns. For each sample, the number of opportunities for a specific error pattern to occur was calculated along with a count of the actual number of occurrences for that pattern. The number of possible occurrences for each phonological error pattern in a given sample varied across participants.

Phonetic complexity of the samples obtained in each elicitation condition was compared using Jakielski's index of phonetic complexity (IPC; Howell, Au-Yeung, Yaruss, & Eldridge, 2006). Eight separate variables were used to score each word in a given sample (see Table 2). The IPC scoring rubric accounts for phonetic difficulty with respect to consonant features (place and manner), vowel class (monophthongs and diphthongs), word shape and length, presence of contiguous consonant strings, and the motoric complexity of a given string of consonants. One point is attributed for each phonetically difficult feature present in a word. A summary of these points results in a composite score for the entire word. Composite scores were then averaged to provide a measure of the phonetic complexity of each sample analyzed.

A higher IPC average is associated with greater degree of phonetic difficulty in terms of production demands.

Finally, the amount of time required to elicit and transcribe each sample was recorded. A comparison was made between the time required to collect and prepare the single-word speech samples for analyses versus the time spent on the same activities for the samples of connected speech.

Table 2
IPC Scoring Grid

Factor	No Score	One Point Each
1. Consonant by Place	Labials, coronals, glottal	Dorsals
2. Consonant by Manner	Stops, nasals, glides	Fricatives, affricates, liquids
3. Singleton C by Place	Reduplicated	Variegated
4. Vowel by Class	Monophthongs, diphthongs	Rhotics
5. Word Shape	Ends with a vowel	Ends with a consonant
6. Word Length (syllables)	Mono-, disyllables	≥ 3 syllables
7. Contiguous Consonants	No clusters	Consonant clusters
8. Consonant by Place	Homorganic	Heterorganic

Statistics

A series of two-tailed *t*-tests were conducted on paired samples to compare a number of variables across the different elicitation tasks. Raw data used for the *t*-tests were derived by calculating the occurrence of a specific variable out of the total number of possible opportunities. Original calculations resulted in percentages or frequencies of occurrences. Percentages were then converted into proportions for statistical analyses. In some instances data were presented as percentages to clarify reporting of results. Additionally, Pearson Product Moment Correlations were

computed for PCC and PVC to measure the strength of the relationships between these variables across sampling conditions.

Paired sample *t*-tests were conducted in consultation with Bret Fuller, Ph.D. Large effect sizes were desirable for reporting meaningful differences from the small number of participants. Because multiple comparisons were made from the data, there was a high chance of inferring significant differences in error. To control for this error rate, a Bonferroni Correction was used to calculate the significance level (.05/15=.0036) for the paired samples tests. In essence, variables that were considered significant were those that were found to be significant at .000 ($p < .0004$). Acknowledging that this investigation is exploratory, significant differences were also reported using less stringent criteria ($\alpha = .05$). In such cases, this distinction is clearly stated.

Within-subject comparisons were made to test differences in proportion of PCC for the SWT and the IST. Within-subject statistical analyses were conducted in consultation with Doug Neeley, Ph.D., PSU Statistical Consulting Lab. The standard normal distribution was used to test the hypotheses for proportions, and the level of significance was set at $\alpha = .05$ (C.V. = +/- 1.96). The following formula was used to calculate *z*-scores for each participant where *P* equals the proportion of consonants produced correctly and *n* equals the total number of opportunities:

$$z = \frac{P(SWT) - P(IST)}{\sqrt{\frac{P(SWT)[1 - P(SWT)]}{n(SWT)} + \frac{P(IST)[1 - P(IST)]}{n(IST)}}$$

Reliability

The 24 speech samples used in this investigation were transcribed by the author of the study and entered into LIPP. Before the data were analyzed, the author's supervisor reviewed all 24 transcriptions using the participants' video files and the information in LIPP. The author's supervisor made the final decision regarding phonetic data used for analyses. Based on the total number of changes made to the 24 speech samples during the second transcription, the inter-rater reliability of the total transcriptions was 99.35% (Table 3). The number of differences found for each sample are presented in Table 4.

Table 3

Total Transcription Reliability

	Changes	Phonemes Transcribed	Percentage
Single-Words	76	7028	98.92%
Sentences	48	5129	99.06%
Total	124	12157	98.98%

Table 4*Transcription Reliability of Individual Samples**

Subject	Single-Word Task (SWT)			Imitated Sentences Task (IST)		
	Changes	Phonemes Transcribed	Percentage Reliability	Changes	Phonemes Transcribed	Percentage Reliability
P01	14	581	97.59%	9	418	97.85%
P02	2	581	99.66%	0	426	100.00%
P03	8	595	98.66%	5	431	98.84%
P04	4	578	99.31%	1	430	99.77%
P05	5	599	99.17%	4	429	99.07%
P06	4	583	99.31%	6	433	98.61%
P07	21	587	96.42%	5	430	98.84%
P08	5	580	99.14%	2	432	99.54%
P09	6	586	98.98%	8	423	98.11%
P10	3	589	99.49%	2	424	99.53%
P11	2	587	99.66%	3	430	99.30%
P12	2	582	99.66%	3	423	99.29%

* (Phonemes Transcribed – Changes)/Phonemes Transcribed = Percentage Reliability

CHAPTER 4: RESULTS AND DISCUSSION

RESULTS

Statistical Comparisons: PCC and PVC

Data on the percentage of consonants produced correctly (PCC) for individual participants are presented in Table 5. For all participants, PCC was higher when producing single words as opposed to connected speech. Within–subject comparisons of the two sampling conditions resulted in statistically significant differences in the proportion of correctly produced consonants for 7 out of 12 participants. P03, P04, P06, P09, P10, P11, and P12 produced significantly more consonants correctly in the SWT ($\alpha = .05$, C.V. = ± 1.96 , two-tailed). For these participants, speech production was better when citing single words.

Table 5
PCC and Within-Subject Tests of Difference in Proportion

Part.	SWT		IST		z-scores	$\alpha = .05$
	PCC SWT	Raw Data	PCC IST	Raw Data		
P01	82.10%	307/374	78.70%	203/258	1.05	
P02	92.80%	348/375	92.00%	241/262	0.37	
P03	92.30%	348/375	87.10%	230/264	2.10	*
P04	94.70%	350/379	89.80%	238/265	2.24	*
P05	91.90%	354/374	91.70%	242/264	0.09	
P06	92.80%	354/385	80.50%	214/216	4.44	*
P07	89.60%	349/376	86.40%	214/266	1.22	
P08	95.50%	335/374	92.50%	229/265	1.55	
P09	90.40%	340/376	81.30%	246/266	3.20	*
P10	91.60%	347/379	86.20%	224/260	2.10	*
P11	94.20%	355/377	89.80%	238/265	1.99	*
P12	95.70%	360/376	90.00%	235/261	2.67	*

* Significant at $\alpha = .05$, C.V. = ± 1.96 (two-tailed)

Group comparisons of PCC and percentage of vowels produced correctly (PVC) are presented in Table 6. Participants produced more consonants correctly in the SWT (M = 91.97%) than in the IST (M = 87.17%). Mean PCC was higher for single words and was statistically significant ($t = 5.010$, $df = 11$, $p < .0004$). Participants exhibited greater production accuracy for consonants in the single-word context. No significant difference was found in the proportion of vowels produced correctly for the two sampling conditions. Overall, fewer vowel errors were observed in the single-word samples. The mean PVC was only slightly higher in the IST (M = 94.18%) than in the SWT (M = 92.83%). A Pearson Product Correlation showed a strong association in performance between the two sampling conditions for PVC ($r = 0.979$, $p < .0004$). The relationship between observed performance on the SWT and the IST for the group demonstrated a moderate positive correlation for PCC ($r = 0.715$, $p < .0004$).

Table 6
Group Comparison of Percentage Accuracy for Consonants and Vowels

Variable	t-test	SWT		IST		t	df	Sig. (2-tail)
		M	SD	M	SD			
PCC	SWT-IST	0.919667	0.036423	0.871700	0.036423	5.010	11	*0.000
PVC	SWT-IST	0.928333	0.076605	0.948142	0.058174	-3.000	11	0.120

* Significant with Bonferroni Correction ($p < .0004$)

Word Position Analysis

A detailed summary of consonant accuracy by word position is presented in Table 7.

Group comparison shows that fewer consonant errors were made in the SWT across word positions, though differences varied and were greatest for word-initial and word-final speech sounds (Figure 1). The largest discrepancy between single-word and connected speech samples was observed in word-initial consonants. Paired samples *t*-tests (Table 8) showed a significant difference between the SWT (M= 95.50%) and the IST (M = 87.99%) for word-initial consonants ($t = 7.644$, $df = 11$, $p < .0004$). Significantly fewer initial consonant errors were found in the context of citing single words. The average difference for final consonants in the SWT (M = 89.82%) and the IST (M = 83.72%) was not significant with the Bonferroni adjustment ($p < .0004$). Using less stringent criteria, the mean difference for final consonants did reach significance ($t = 2.572$, $df = 11$, $p < .05$), showing that a significantly greater number of final consonant errors were observed in connected speech as opposed to single-words.

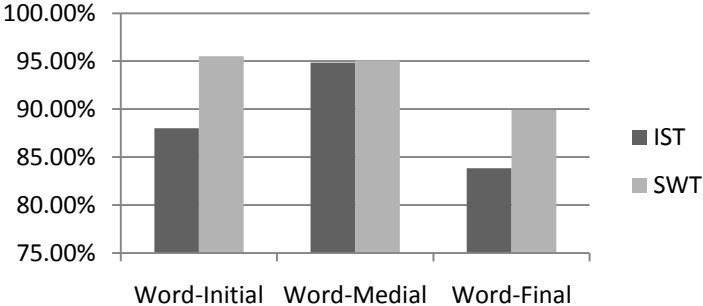


Figure 1 Mean Percentage of Consonants Correct (by Word Position)

Table 7*Percentage of Consonants Produced Correctly by Word Position (SWT)*

Part.	Initial	Raw Data	Medial	Raw Data	Final	Raw Data
P01	91.84%	90/98	90.91%	50/55	77.11%	64/83
P02	94.00%	94/100	90.00%	45/50	91.76%	78/85
P03	97.03%	98/101	96.61%	57/59	95.12%	78/82
P04	96.00%	96/100	93.75%	45/48	93.10%	81/87
P05	96.04%	97/101	98.08%	51/52	93.02%	80/86
P06	95.05%	96/101	96.00%	48/50	95.40%	83/87
P07	97.00%	97/100	94.55%	52/55	82.72%	67/81
P08	98.00%	98/100	98.00%	49/50	91.86%	79/86
P09	94.95%	94/99	92.73%	51/55	87.06%	74/85
P10	94.95%	94/99	94.34%	50/53	86.75%	72/83
P11	97.03%	98/101	98.08%	51/52	89.53%	77/86
P12	94.06%	95/101	98.04%	50/51	94.38%	84/89

Percentage of Consonants Produced Correctly by Word Position (IST)

Part.	Initial	Raw Data	Medial	Raw Data	Final	Raw Data
P01	83.33%	60/72	89.74%	35/39	74.07%	40/54
P02	92.00%	69/75	94.87%	37/39	88.14%	52/59
P03	90.54%	67/74	97.44%	38/39	89.83%	53/59
P04	92.00%	69/75	87.18%	34/39	81.67%	49/60
P05	93.24%	69/74	97.44%	38/39	94.92%	56/59
P06	84.00%	63/75	97.44%	38/39	68.97%	40/58
P07	89.04%	65/73	94.87%	37/39	88.52%	54/61
P08	90.54%	67/74	94.87%	37/39	91.53%	54/59
P09	83.78%	62/74	92.31%	36/39	77.59%	45/58
P10	85.33%	64/75	92.31%	36/39	83.93%	47/56
P11	84.21%	64/76	100.00%	38/38	81.36%	48/59
P12	87.84%	65/74	100.00%	39/39	84.21%	48/57

Table 8*Group Comparison of Percentage Accuracy for Consonants across Word Positions*

Variable	t-test	SWT		IST		T	df	Sig. (2-tail)
		M	SD	M	SD			
Initial	SWT-IST	0.954958	0.017123	0.879875	0.036998	7.644	11	*†0.000
Medial	SWT-IST	0.950908	0.028570	0.948725	0.039429	0.261	11	0.799
Final	SWT-IST	0.898175	0.055434	0.837283	0.075357	2.572	11	†0.026

* Significant difference with Bonferroni Correction ($p < .0004$)† Significant without Bonferroni Correction ($p < .05$)

Phonological Error Pattern Analysis

Ten phonological error patterns were selected for analyses of individual and group comparisons for frequency of occurrence across sampling conditions. The criterion used to select these specific error patterns was based on evaluation of the speech samples from each elicitation procedure. A series of *t*-tests for repeated measures was conducted on error patterns occurring more than 2% of the time in at least one sampling condition. The following error patterns met the criteria for inclusion in this report: cluster reduction, initial consonant deletion, final consonant deletion, total consonant deletion, vocalization, gliding, derhotization, stopping, lisping, and glottal replacement of stops. Cluster reduction, gliding, vocalization, and stopping are considered typical errors in speech development for the ages of the participants in this study (Stoel-Gammon & Dunn, 1985).

Visual inspection of phonological error pattern data for individual participants showed that some error types occurred with greater frequency in the SWT while others were observed more often in the IST. Cluster reduction, gliding, derhotization, and glottal replacement of oral stops occurred more frequently in single-word samples. Participants demonstrated initial consonant deletion, final consonant deletion, vocalization, stopping, and lisping with greater frequency in connected speech. Overall, more total consonants were deleted in the IST as compared to the SWT. Initial consonant deletion was observed only once in the single-word context:

P07 exhibited initial consonant deletion in the SWT on 1 out of 100 opportunities (1%).

Group comparisons are presented in Table 9. The frequency of occurrence for phonological error patterns in general was relatively low for the group. Some of the participants did not exhibit a given process in one or both sampling conditions. In addition, the number and types of error patterns observed were inconsistent for participants and contexts.

Table 9
Group Comparison of Frequency Occurrence for Phonological Error Patterns

Variable	SWT		IST		t	df	Sig. (2-tail)
	M	SD	M	SD			
CR	0.021600	0.037691	0.010608	0.021110	1.948	11	0.077
ICD	0.000833	0.002887	0.021325	0.029846	-2.416	11	†0.034
FCD	0.025983	0.030967	0.050042	0.030535	-1.798	11	0.100
TOTCD	0.013700	0.015137	0.028175	0.014921	-2.382	11	†0.036
VOC	0.847893	0.092793	0.872549	0.111737	-0.590	11	0.561
GLIDING	0.142650	0.166544	0.126542	0.165900	0.912	11	0.381
DERHOT	0.348400	0.456584	0.320900	0.457778	2.226	11	†0.048
STOPPING	0.016300	0.016209	0.076617	0.036254	-6.657	11	*†0.000
LISPING	0.046667	0.094370	0.067033	0.148308	-0.969	11	0.354
GL4STP	0.025850	0.018328	0.022683	0.014573	-0.530	11	0.607

Cluster reduction (CR); initial consonant deletion (ICD); final consonant deletion (FCD); total consonant deletion (TOTCD); vocalization (VOC); gliding (GLIDING); derhotization (DERHOT); stopping (STOPPING); lispings (LISPING); glottal for oral stop (GL4STP)

* Significant difference with Bonferroni Correction ($p < .0004$)

† Significant without Bonferroni Correction ($p < .05$)

Individual data on word structure errors is presented in Table 10. In general, the frequency of occurrence of error patterns affecting word structure (cluster reduction and consonant deletions) was higher in the IST than in the SWT, though

differences observed in the mean occurrence of structural error patterns was not statistically significant with the Bonferroni Correction ($p < .0004$).

Table 10
Occurrence of Word Structure Errors

Part.	CR		ICD		FCD		TOTCD	
	SWT	IST	SWT	IST	SWT	IST	SWT	IST
P01	13.04%	6.25%	0.00%	1.39%	4.82%	3.70%	1.69%	1.82%
P02	1.43%	0.00%	0.00%	0.00%	1.18%	5.08%	0.85%	2.31%
P03	0.00%	0.00%	0.00%	1.35%	2.44%	6.78%	1.65%	3.49%
P04	1.45%	0.00%	0.00%	0.00%	3.45%	6.67%	2.55%	2.87%
P05	0.00%	0.00%	0.00%	0.00%	1.16%	1.69%	0.84%	0.58%
P06	1.45%	0.00%	0.00%	2.67%	0.00%	12.07%	0.42%	5.23%
P07	1.47%	2.13%	1.00%	4.11%	11.11%	3.28%	5.51%	2.89%
P08	0.00%	0.00%	0.00%	1.35%	0.00%	1.69%	0.00%	1.74%
P09	1.45%	0.00%	0.00%	2.70%	0.00%	1.72%	0.00%	1.75%
P10	5.63%	4.35%	0.00%	10.67%	2.41%	3.57%	0.85%	5.88%
P11	0.00%	0.00%	0.00%	0.00%	3.49%	6.78%	1.67%	2.31%
P12	0.00%	0.00%	0.00%	1.35%	1.12%	7.02%	0.41%	2.94%

Cluster reduction (CR), initial consonant deletion (ICD), final consonant deletion (FCD), total consonant deletion (TOTCD)

Only cluster reduction was found to be higher in the SWT ($M = 2.16\%$) than in the IST ($M = 1.06\%$). Consonant clusters involving /s/ and /l/ emerged as the most consistently difficult in both tasks, but clusters that occurred across word boundaries posed difficulties unique to the continuous speech context. Initial and final consonant deletion in the IST often resulted in the reduction of consonant strings spanning word boundaries. Some instances of cross-boundary consonant omissions in the IST were related to the re-articulation of a consonant occurring at the end of one word-boundary and the beginning of the next. For example, in the sentence, *Three sheep*

played in the flowers, the phoneme /p/ occurs twice in succession across the word boundary between *sheep* and *played*. Not all participants re-articulated matched phonemes across word boundaries which may be considered typical for adult speakers as well. When participants did not clearly mark 2 phonemes in succession with either a pause or a re-articulated feature, the word-final segment was counted as a consonant deletion because of the inconsistency observed across participants.

Omission of /ð/ in function words such as *the* and *these* contributed to the observed frequency of initial consonant deletion in the continuous speech context. Certain types of consonant omissions contributed to simplified clusters in connected speech. Reduced clusters by omission of /s/ characterized some of the word-onset errors observed in the IST. Final stops were deleted at the end of consonant strings between adjoining words, usually preserving the continuant elements of the clusters (e.g., / tuθpest/ → [tuθpes]). Final /s, z/ were occasionally dropped from word-final position, though /z/ was also likely to be devoiced which did not result in altered word structures.

Three phonological error patterns that affected production of liquids and glides occurred with greater than 10% frequency for the group and are presented in Table 11. This increased frequency of occurrence may be considered more clinically significant than the 2% criteria used initially for the selection of error patterns. Error

patterns involving liquids and glides resulted in the highest percentages of occurrence in both single-word and continuous speech samples, though mean differences between sampling contexts did not reach significance with the Bonferroni adjustment ($p < .0004$). Mean frequency of occurrence of vocalization was higher in the IST ($M = 87.26\%$, $SD = 11.18\%$) than in the SWT ($M = 84.79\%$, $SD = 9.28\%$). In the SWT, gliding ($M = 14.26\%$, $SD = 16.65\%$) and derhotization ($M = 34.84\%$, $SD = 45.65\%$) occurred with greater frequency as compared to the IST ($M = 12.65\%$, $SD = 16.60\%$ and $M = 32.09\%$, $SD = 45.78\%$ respectively).

Table 11
Occurrence of Glide and Liquid Errors

Part.	VOC		GLIDING		DERHOT	
	SWT	IST	SWT	IST	SWT	IST
P01	3.45%	0.00%	2.70%	3.33%	2.94%	0.00%
P02	0.00%	0.00%	18.92%	3.45%	0.00%	0.00%
P03	85.19%	50.00%	48.65%	37.93%	100.00%	95.00%
P04	0.00%	0.00%	2.63%	0.00%	5.88%	0.00%
P05	64.52%	70.00%	39.47%	41.38%	97.14%	100.00%
P06	62.07%	71.43%	24.32%	31.03%	91.43%	91.30%
P07	65.52%	58.82%	23.68%	27.59%	97.14%	89.47%
P08	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
P09	11.11%	0.00%	0.00%	0.00%	9.09%	0.00%
P10	0.00%	0.00%	5.41%	7.14%	0.00%	4.55%
P11	6.90%	0.00%	2.70%	0.00%	11.43%	4.76%
P12	0.00%	0.00%	2.70%	0.00%	3.03%	0.00%

Vocalization (VOC), gliding (GLIDING), derhotization (DERHOT)

Seven participants demonstrated errors involving the substitution of a vowel for a liquid phoneme in word-final position (vocalization). The frequency of occurrence of vocalization was higher in the SWT for P01, P03, P07, P09, and P11 as

compared to the IST. For P01, P09, and P11, vocalization did not occur in the IST, and the frequency of occurrence of vocalization in the SWT was relatively low for these same participants (approximately 3-11%). The frequency of occurrence of vocalization for P03, P05, P06, and P07 was greater than 50% for both tasks, but the participants were split as to which context resulted in the highest observed frequency for this error pattern. For P03, the frequency of occurrence of vocalization was higher in the SWT (85.19%) than in the IST (50.00%). The performance of P07 was similar (65.52% in the SWT as compared with 58.82% in the IST). For P05 and P06, the frequency of occurrence of vocalization was higher in the IST (P05 70.00% and P06 71.43%) as compared to the SWT (P05 64.52% and P06 62.07%).

No clear trend emerged with regard to the context in which gliding was exhibited more frequently. The frequency of occurrence of gliding was higher in the SWT for P02 (18.92%), P03 (48.65%), P04 (2.63%), P11 (2.70%), and P12 (2.70%) as compared to the frequency observed in the IST for these same participants (P02 = 3.45%, P03 = 37.93%, P04 = 0.00%, P11 = 0.00%, P12 = 0.00%). For P05, P06, P07, and P10, the frequency of occurrence of gliding was higher in the IST (P05 41.38%, P06 31.03%, P07 27.59%, and P10 7.14%) than in the SWT (P05 39.47%, P06 24.34%, P07 23.68%, and P10 5.41%). Participant P03, P05, P06, and P07 exhibited the highest incidence of gliding. For P03 and P04, the frequency of occurrence of gliding was 10-15% higher in the SWT than in the IST. Though gliding occurred more frequently in

the IST for P05, P06, and P07 as compared to the SWT, the observed difference between the tasks was smaller (approximately 2-7%).

More participants exhibited errors involving distortion of consonantal /ɹ/ or loss of vocalic /ɹ/ coloring (derhotization) in the SWT. Using less conservative criteria for group comparisons, derhotization demonstrated a significant difference between samples obtained from single-word productions as opposed to connected speech. The mean frequency of occurrence of derhotization was higher in the SWT (M = 34.84%, SD = 45.66) as compared to the IST (M = 32.09%, SD = 45.78%) and was statistically significant without the Bonferroni Correction ($t = 2.226$, $df = 11$, $p < .05$). No clear trend characterized the differences observed across sampling contexts. In some cases, labial distortion of consonantal /ɹ/ characterized the single-word productions whereas gliding or omission of /ɹ/ appeared more prevalent in the IST. However, deletion of /ɹ/ in consonant blends and loss of /ɹ/ coloring in rhotic vowels was also observed in the SWT, even for participants who otherwise demonstrated mastery of /ɹ/.

Individual data on errors involving consonant substitutions is presented in Table 12. Of all the errors characterized by consonant substitutions, stopping was the only error pattern that reached statistical significance with the Bonferroni Correction. The frequency of occurrence of stopping was higher in the IST (M = 7.66%, SD =

3.63%) than in the SWT (M = 1.63%, SD = 1.62%), and this difference was statistically significant ($t = -6.657$, $df = 11$, $p < .0004$). All participants were more likely to replace target phonemes with stops in connected speech. Stopping of lingua-dental fricatives ($\theta \rightarrow d/\text{d}$) in function words such as *the* or *these* represented the majority of stopped consonants in continuous speech. Stopping of consonants / θ , v , n / was also observed in some instances. Simplification of some cross-boundary consonant strings resulted in glottal stops in continuous speech.

Table 11
Occurrence of Substitution Errors

Part.	STOPPING		LISPING		GL4STP	
	SWT	IST	SWT	IST	SWT	IST
P01	5.26%	13.64%	25.00%	23.53%	2.11%	3.30%
P02	0.85%	5.62%	0.00%	0.00%	6.34%	3.26%
P03	0.00%	4.49%	0.00%	0.00%	0.00%	0.00%
P04	0.86%	7.87%	7.27%	5.56%	4.20%	2.15%
P05	0.00%	4.49%	0.00%	0.00%	2.78%	3.23%
P06	0.84%	8.99%	0.00%	0.00%	3.57%	3.19%
P07	0.00%	4.55%	0.00%	0.00%	3.57%	2.11%
P08	2.56%	6.74%	0.00%	0.00%	3.47%	1.06%
P09	3.33%	10.87%	23.73%	48.65%	0.72%	4.60%
P10	2.50%	3.49%	0.00%	2.70%	0.71%	0.00%
P11	0.84%	14.29%	0.00%	0.00%	2.84%	1.09%
P12	2.52%	6.90%	0.00%	0.00%	0.71%	3.23%

Stopping (STOPPING), lispings (LISPING), glottal replacement of oral stops (GL4STP)

Four participants produced a dentalized / ʃ / (lispings) in their speech samples.

The frequency of occurrence of lispings was higher in the SWT for P01 (25%) and P04 (7.27%) than in the IST (P01 23.53%, P04 5.56%). In the IST, lispings was observed

more frequently for P09 (48.65%) and P10 (2.7%) compared to the SWT (P09 23.73%). Participant 10 did not exhibit lisping in single-words.

Glottal replacement of oral stops was observed more frequently in the SWT (M = 2.59%, SD = 1.83%) than in the IST (M = 2.27, SD = 1.46%). The observed difference was not significant. In many instances, glottal stops were not considered errors when the intended production matched conventions for use of glottal stops in the target adult speech.

The participants in this study exhibited a variety of phonological error patterns across the sampling conditions. In general, the error patterns observed were similar for both tasks. Where they differed, no clear trend emerged in which one context presented a better understanding of a participant's speech production abilities.

Comparison of Phonetic Complexity

Table 13 shows a comparison of the single-word and continuous speech samples in terms of phonetic complexity. Composite word scores were derived using the index of phonetic complexity (IPC) developed by Jakielski (Howell et al, 2006). These scores were then averaged across the entire sample to obtain an overall IPC score. The results show a significant difference for IPC in the SWT (M = 2.661250, SD = 0.151230) as compared to the IST (M = 2.030917, SD = 0.111523). The single-word samples were phonetically more complex than the continuous speech samples, and this difference was statistically significant ($t = 29.813$, $df = 11$, $p < .000$).

Table 13*Group Comparison of Index of Phonetic Complexity (IPC)*

Variable	<i>t</i> -test	SWT		IST		<i>t</i>	df	Sig. (2-tail)
		M	SD	M	SD			
IPC	SWT-IST	2.661250	0.151230	2.030917	0.111523	29.813	11	*0.000

* Significant with Bonferroni Correction ($p < .0004$)*Administration and Transcription Record*

The amount of time required to collect and transcribe the speech samples varied for each participant (Table 14). The time required to obtain single-word samples ranged from 9:41 (minutes:seconds) to 19:15 (M = 14:35, SD = 2:55). For the IST, values ranged from 3:13 to 7:15 (M = 04:35, SD = 01:27). Administering the SWT took significantly longer than obtaining the samples of connected speech ($t = 12.344$, $df = 11$, $p < .0004$). There was less discrepancy in the amount of time spent transcribing speech samples obtained from the two elicitation methods. The difference between how much time was spent transcribing samples from the SWT (M = 24:22, SD = 5:52) as opposed to the IST (M = 20:16, SD = 6:37) was not found to be statistically significant. While the samples of connected speech could be obtained relatively quicker than those from the single-word naming task, the amount of time spent transcribing samples from the two conditions was similar.

Transcription and analyses of speech samples from both the IST and SWT was aided by the use of Logical International Phonetic Program (LIPP) software (Oller & Delgado, 2000). Transcription of the samples was completed on templates that

contained the phonetic targets from the test stimuli. This required the transcriber to mark deviations from the intended targets on the templates rather than complete a segment by segment transcription of the samples.

Table 14

Administration and Transcription of Speech Samples (minutes:seconds)

Part.	Administration Time for SWT	Transcription of SWT*	Administration Time for IST	Transcription of IST*
P01	19:15	39:25	06:20	34:17
P02	19:11	25:05	03:31	15:46
P03	13:33	23:57	06:20	26:31
P04	14:05	27:10	03:15	18:09
P05	10:32	18:41	03:26	18:05
P06	09:41	24:34	03:13	25:14
P07	14:06	28:53	03:35	24:12
P08	14:26	19:27	03:19	12:12
P09	14:35	21:17	07:15	22:41
P10	17:23	24:31	04:50	19:32
P11	14:41	22:14	05:40	14:58
P12	13:30	17:06	04:15	11:30

* Initial transcription of samples

DISCUSSION

Production Accuracy

The purpose of this study was to compare samples of continuous speech elicited with a sentence imitation task to samples obtained with a single-word naming task. The IST is currently under development and testing as part of the *Phonological and Bilingual Articulation Assessment, English Version (PABA-E)*, and this project represents the pilot investigation of the IST under experimental

conditions. The SWT was designed to assess a corpus of single words elicited with spontaneous citation of pictured stimuli. Similar to most standardized articulation assessments, the SWT uses isolated words to assess speech sound production abilities. Single-word naming provides a method for representing a variety of desired phonetic targets in a controlled sample.

The SWT has been previously tested on many children (300+), but this is the first comparison to the complementary sentence task. The IST was developed as a potential measure of speech production abilities in continuous speech. The IST consists of a series of sentences, each describing a specific picture. During administration, the examiner reads each sentence aloud and the examinee repeats the examiner's words. This method allows for controlled content of the resulting sample.

As part of developing the IST protocol for the PABA-E, the decision was made to pilot the investigation of both the English single-word and sentences tasks with preschool aged children who demonstrated speech production skills within the average range for their age. Standard scores from the GFTA-2 showed that participants all exhibited speech production abilities in the high-average range. Five of the 12 participants scored more than 1 standard deviation better than the mean ($M = 100$, $SD = +/- 15$). The lowest score on the GFTA-2 was 109. Speech samples were obtained from 12 preschool children who participated in this project. Each participant was administered the SWT and the IST from the PABA-E for the purpose

of comparing broad measures of phonetic accuracy and the frequency of selected error patterns across experimental speaking contexts. It was hypothesized that differences would be observed in the samples obtained with the two elicitation procedures.

Results indicated that the participants made fewer consonant errors in the production of single-words than in the imitated sentences. Using a broad measure of phonetic accuracy, the present study found that the percentage of consonants produced correctly (PCC) was significantly better for 7 out of 12 participants in the single-word samples, although a moderately high correlation was found overall between the two tasks. Group comparison of PCC confirmed that continuous speech samples yielded a significantly greater number of consonant errors.

The present findings are not consistent with the most recent reports for children with speech sound disorders. Masterson, et al. (2005), Morrison and Shriberg (1992), and Wolk and Meisler (1998) reported that participants exhibiting speech sound disorders demonstrated significantly fewer consonant errors in conversation as compared to single-word naming. These researchers found that PCC was higher in samples of spontaneous conversation as compared to production of single words. The elicitation mode used for collection of connected speech may account for some of the difference in findings between the previous research and the present study. The phonetic complexity of the sentence imitation used to elicit samples of connected speech in the present study may have presented more

opportunities for errors than the participants would have exhibited spontaneously. For example, Morrison and Shriberg (1992) reported a greater number of early developing sounds and simple word shapes in the spontaneous conversation of their subjects as compared to a single-word naming task. Similarly, Wolk and Meisler (1998) found more opportunities for consonant and word shape errors in their single-word picture naming task as compared to spontaneously elicited conversation.

The present findings appear to support some early research suggesting that samples of connected speech were associated with a greater number and variety of errors (Andrew & Fey, 1986; DuBois & Bernthal, 1978; Faircloth & Faircloth, 1970; Healy & Madison, 1987; Klein, 1984; Watson, 1989) though some errors related to place and manner occurred more frequently in single-word samples (Paden & Moss, 1985; Watson, 1989). The results of the present study are generally consistent with these earlier findings. However, this comparison should be viewed with caution given the use of sentence imitation to elicit connected speech samples for the present study. Early research findings were based on comparing spontaneous conversation to spontaneous single-word naming.

When phonetic accuracy was analyzed by word position, the present data continued to support the general findings that PCC was higher in single words for the participants in this study. The largest discrepancy between the SWT and the IST occurred for word-initial and word-final consonants. The observed mean difference in speech production accuracy for word-medial consonants in the two sampling

conditions was relatively small and appears to support word boundary characteristics affecting performance data. However, post hoc comparison of word-medial consonant accuracy between the SWT to the IST showed only a moderate relationship using a Pearson Product Correlation ($r = 0.6788$).

Further analysis of the data included a comparison of the percentage of vowels produced correctly (PVC) in the SWT and the IST. In this study, PVC was found to be only slightly higher in the continuous speech task as compared to the SWT. The difference was not found to be significant for this group. The two sampling conditions were highly correlated in terms of vowel accuracy. None of the research reviewed for this project reported values for PVC. The results of this study indicated that PVC was not influenced by sampling condition study and that fewer errors were observed for vowels than consonants overall.

Given the present findings that differences observed between samples were significant for word-initial consonants and nearly significant for word-final consonants, it is worth considering how the discrepancies tended to be most meaningful at word boundaries. Although Masterson et al. (2005) found a higher PCC in continuous speech, their reported differences in consonant accuracy were no longer significant using a revised scoring rubric that ignored word-final consonant deletions, voicing errors, and glottal replacement in single-word samples. Final consonant deletions, initial consonant deletions resulting from reduced clusters across word boundaries, and word-final voicing errors occurred in the IST samples

from the present study and contributed to the greater number of consonant errors observed.

Morrison and Shriberg (1992) argued that certain sound errors occurring at word boundaries may be acceptable consequences of conversational speech. Stopping of fricatives, consonant omissions, simplification of consonant clusters, or reduction of certain consonant features were argued to be more frequent as the attention of the speaker shifts from a lexical reference in single-word naming to the flow of ideas and a more conceptual frame in conversation. Phonetic accuracy may be less important in continuous speech, whereas single-word tasks may encourage exaggerated features and substitutions.

The work of Masterson et al. (2005) and Wolk and Meisler (1998) focused on construction of sufficiently extensive single-word elicitation procedures that adequately capture speech production abilities in light of the many advantages discussed with respect to single-word tests. Masterson et al. (2005) and Wolk and Meisler (1998) found fewer consonants were produced correctly in single-word samples when the elicitation stimuli were extensively comprehensive or tailored to an individual's abilities. Wolk and Meisler (1989) showed that development of an adequately extensive speech task would tax speech abilities in ways similar to spontaneous speech. The demands of these tests may push the boundaries of speech production abilities in ways that are different than participants themselves would tax their own phonological systems in a spontaneously generated sample. Masterson et

al. reported that their single-word task sampled more fricatives, affricates, and liquids, while presenting increased demands for some complex word shapes (e.g., CVCVCV). The spontaneous conversation samples obtained by Morrison and Shriberg (1992) contained proportionately greater exemplars of early developing sounds and simple word shapes. Spontaneously elicited speech samples may be less phonetically demanding and, therefore, result in a greater percentage of consonants produced correctly. Findings from studies that found fewer consonant errors in continuous speech samples as compared to single-word productions were not consistent with the present findings that connected speech elicited via sentence imitation was associated with a higher frequency of errors as compared to single-word naming. While no attempt was made in the present study to compare samples of spontaneous conversation to samples of connected speech elicited via sentence imitation, it can be argued that the present findings do add some support to assertions that contextual and linguistic factors influence phonetic accuracy in connected speech in ways that single-word tasks fail to capture (e.g., Campbell & Shriberg, 1982; Gallagher & Shriener, 1975a, 1975b; Hoffman et al., 1977; Kent, 1982; Klein & Spector, 1985; Morrison & Shriberg, 1992; Prather & Kenney, 1986; Zehel et al., 1972). The influence of contextual and linguistic factors on speech production has been a major criticism against the use of single word naming tasks for speech sound assessment since, by design, production of single words is not influenced by such factors to the same degree.

One challenge of developing a comprehensive sampling protocol that is representative of the target language and simultaneously captures the various contextual and linguistic influences on speech production may be an increased phonetic complexity in the elicitation procedures. Constructing a representative sample of the target language may necessitate a measure of complexity in the choice of target words for the elicitation procedures that goes beyond the typical speech used by the intended participants. In the present study, an extensive SWT was used to elicit single-word samples, and many of the content words in the IST were matched to stimuli in the SWT. These content words became the basis for constructing the IST. Findings of the present study showed significantly greater phonetic complexity in the SWT than the IST. The index of phonetic complexity (IPC) used in this study measured phonetic difficulty as a function of motoric constraints on speech production (Howell et al., 2006). The IPC scoring rubric accounted for place and manner of production, variegated and heterorganic motor planning in consonant singletons and clusters, word shape and length (in syllables), and the presence of consonant strings that challenge oral-motor abilities.

Even though the phonetic difficulty of the SWT was greater, performance on the IST resulted in a greater number of consonant errors. One possible conclusion is that the increased linguistic and contextual demands in the IST affected participants' performance. On the other hand, participants may have relaxed performance expectations in the IST associated with increased focus on the linguistic aspects of

the sentences and its similarity to naturalistic conversational speech (Morrison & Shriberg, 1992).

Even though the participants in this study all exhibited typical speech development, certain phonological error patterns may be considered acceptable for the ages of the participants (Stoel-Gammon & Dunn, 1985). Even so, examination of the data showed that relatively few error patterns were observed with sufficient frequency to draw any meaningful conclusions regarding their occurrence across conditions. Vocalization, gliding, and derhotization were the most frequently occurring processes, yet their occurrence in the naming task was not significantly different than that observed in continuous speech. Of the 10 phonological error patterns examined in this study, no clear trend emerged with respect to frequency of occurrence in relation to elicitation mode. Rather than one elicitation mode providing a clearer picture of the participants' speech production abilities, comparison of the tasks reinforced the influence of motoric, linguistic, and contextual factors on speech production.

It was expected that stopping would occur more frequently in the IST (Morrison & Shriberg, 1992), and this was the only process to differ significantly between tasks. In addition, stopping of continuants was the only error pattern that was consistently more frequent in the continuous speech context. Similar to findings of several studies, stopping was relatively rare in the SWT, and participants in this

investigation stopped more continuants in the sentence task (Andrews & Fey, 1986; Faircloth & Faircloth, 1970; Healy & Madison, 1987; Morrison & Shriberg, 1992).

Other studies found similar rates of stopping in both single-word and continuous speech samples (DuBois & Bernthal, 1978; Kenny, Prather, Mooney, & Jeruzel, 1984; Klein, 1984; Paden & Moss, 1985). One major difference in this investigation was that the continuous speech task was constructed rather than spontaneous. The significant variance observed between rates of stopping in single words and continuous speech may imply that this error pattern was partly influenced by contextual and/or linguistic factors embedded in the experimental task. The majority of stopped consonants in the IST were lingua-dental fricatives ($\theta \rightarrow d/d$). Function words such as *the* and *there* were not features of the SWT but were represented in the continuous speech task.

One way to characterize the findings of the present study with regard to phonological error pattern trends is that certain sampling contexts may be better suited to capturing certain error patterns. In general, error patterns affecting word structure (consonant deletions and cluster reduction) were more frequent in continuous speech. More consonants were deleted in the IST, and this difference was arguably significant using less stringent statistical criteria. It is of interest to note that consonant cluster errors occurring across word boundaries in the IST were captured in the phonological error patterns of initial and final consonant deletion. Strictly speaking, cluster reduction was observed with greater frequency in the SWT,

however scoring was confined to consonant clusters within word boundaries.

Motoric demands of the within-word consonant strings in the SWT may have been greater as evidenced by higher IPC scores in the single-word samples.

Differences in phonetic complexity found between the SWT and the IST may offer some support for context-sensitivity to certain error patterns. At least in some instances, error patterns involving liquids and glides were related to cluster reduction or simplification. In both the SWT and IST, participants tended to have difficulty with clusters such as /fɹ/ → [f], /fɹ/ → [fw], and /θɹ/ → [θw]. With the IPC model, consonant strings consisting of a labial- or lingua-dental fricative plus a liquid (rhotic) would receive a high IPC score. Gliding and derhotization were observed more frequently in the single-word context likely as a result of greater complexity in the sampling demands. Wolk and Meisler (1998) also reported that error patterns involving liquids, glides, and reduced clusters were more frequent in naming.

It was expected that one of the sampling conditions would better capture the speech production abilities of at least some of the individuals who participated in this study. Seven of the 12 participants demonstrated significantly fewer consonant errors overall in the SWT, yet the phonological error patterns that characterized the types of errors observed in individual samples were not so clearly distributed. In general, word structure and substitution errors were more frequent in continuous speech, while errors affecting glides and liquids were captured in single-word

naming. Only cluster reduction, which affected word structure accuracy, was more frequent in single words overall.

Subject 10 showed the most consistent profile: only cluster reduction occurred more frequently in the SWT. Subject 09 exhibited more cluster deletions and consonant substitutions in continuous speech; cluster reduction, vocalization and derhotization were more frequent in single words. On the other hand, vocalization and gliding were more frequent in the IST for Subject 06.

Only 2 participants demonstrated lispings, categorized as a substitution error pattern, with moderate frequency (> 20%). Lispings were slightly higher in the SWT for Subject 01, though the other substitution errors (stopping and glottal replacement of oral stops), as well as gliding, were more frequent in the IST. Cluster reduction, final consonant deletion, vocalization, and derhotization were more frequent in the SWT for P01.

Subject 09 exhibited a speech profile fairly consistent with the general findings. That is, substitution errors (stopping, lispings, and glottal replacement) and consonant deletions (word structure errors) were more frequent in continuous speech. Vocalization and derhotization occurred more often in single-word naming. As with general trends reported here, cluster reduction was more frequent in SWT (13.04%) as compared to the IST (6.25%).

Sentence Imitation as an Elicitation Method

In this study, imitation was used to elicit samples of continuous running speech (IST). After participants listened to stimulus sentences read aloud by the examiner, they were asked to repeat those sentences verbatim. Visual stimuli were used to illustrate the target sentences. The IST was composed of 18 target sentences plus an additional 3 sentences that were used to train the participants on the task. None of the participants demonstrated difficulty completing the IST task. When necessary, repetition of the stimulus sentences by the examiner was allowed in order to provide the participants with sufficient opportunities to produce the intended targets. When participants' intended words did not match the exact target, targets were changed to reflect the intended utterance, and these changes were not counted as errors.

Continuous speech samples elicited with the IST were compared to samples obtained in response to a single-word spontaneous naming task (SWT). The results of this study showed significantly fewer consonants errors in the IST than in the SWT. These findings were not consistent with several previous studies that reported a higher frequency of speech production errors in spontaneous speech as compared with imitation tasks (DuBois & Berthal, 1978; Faircloth & Faircloth, 1970; Smith & Ainsworth, 1967). However, other studies have reported no significant differences in spontaneous versus imitated speech (Paynter & Bumpas, 1977; Siegel et al., 1963). In the present study, imitated and spontaneous speech samples were shown to have a

moderately strong correlation. This finding is consistent with one of the few studies to systematically explore performance in imitated versus spontaneous reported similar results (Haynes & Stead, 1987).

A major objection to the use of imitation as an elicitation technique is the assumption that spontaneous samples are more likely to resemble speech in nontest contexts (Stoel-Gammon & Dunn, 1985). This distinction may have particular merit when speech sample must be collected for the purpose of assessing speech production abilities for diagnosis and treatment of disorders. The present investigation did not address this question directly, but the findings do suggest that the number and types of errors made by the participants in this study were similar.

Eliciting speech samples through imitation provided control over the corpus of the continuous speech samples obtained in this study. While relying on imitation to elicit these samples, review of the data showed that overall consonant accuracy was similar to expected assumptions with regard to single-word versus continuous speech (Morison & Shriberg, 1992). One of the major criticisms of single-word tests has been that they may overestimate speech production abilities. Several studies comparing spontaneous conversation to single-word naming found PCC to be higher in the conversation tasks (Masterson et al., 2005; Morrison & Shriberg, 1992; Wolk & Meisler, 1998). This is not consistent with the findings from this investigation. Wolk and Meisler (1998) suggested the possibility that the single-word tasks used in their study taxed participants' speech production abilities to a greater extent than the

spontaneously generated samples provided by their research participants. The present study also employed an extensive single-word task, yet the data appeared to capture at least some of the contextual and linguistic factors that may influence speech production in conversational contexts. Use of matched target words in the construction of the SWT and the IST provides some support for this possibility.

Collection and Transcription

Another benefit of eliciting speech samples through imitation is the potential efficiency with which continuous speech samples may be obtained. In this study, results indicated that significantly less time was required to obtain the continuous speech samples compared to the single-word responses. This difference may be meaningful for clinicians who operate with limited contact time. On the other hand, the amount of time spent transcribing the single-word and sentence tasks in this study did not differ significantly. Masterson et al. (2005) reported that certain characteristics of conversational speech make transcription more difficult. Specifically, consonant deletions, voicing errors, and glottal replacement all occurred in word-final boundaries with increased frequency in continuous running speech. In this investigation, the greatest discrepancies in consonant accuracy between the two conditions appeared at word boundaries. The difficulty of transcribing word boundaries may have influenced the time requirements for transcription in this study.

CHAPTER 5: CONCLUSION AND IMPLICATIONS

This study was a pilot investigation of the sentence imitation portion of the PABA-E. Samples of spontaneously produced single words were compared to samples of continuous speech elicited via imitated sentences. It was expected that differences in error frequency would be observed across the two sampling conditions. The participants in this study produced fewer phonetic errors in single-word responses. Overall, consonant errors occurred with greater frequency when participants produced connected speech. While differences in the frequency of occurrence for some errors were found to be statistically significant, the clinical significance of the observed differences is questionable. Differences in the percentage of occurrence between sampling conditions were relatively small in most cases.

The majority of errors occurred at word boundaries, and comparison of single-word and continuous speech samples showed a greater discrepancy in consonant accuracy at word boundaries. It was expected that certain error types would be more apparent in continuous speech as opposed to single-word naming. Since participants demonstrated typical speech development, relatively few errors were seen overall. Overall, the types of errors observed were similar across samples, albeit infrequently.

In general, error patterns involving liquids and glides were more frequently observed in single-word contexts. Consonant substitutions and errors affecting word structure were more prominent in continuous speech samples. Cluster reduction

proved to be the exception, occurring more frequently in single words. However, consonant strings across word boundaries were not counted as cluster errors. Since continuous speech resulted in a greater number of consonant deletions and substitutions at word boundaries, the frequency of errors affecting cross-boundary consonant strings may have actually been similar to, or even exceeded, rates observed in single words. This would be more consistent with previously reported findings.

The phonetic complexity of the SWT was found to be significantly more difficult than the IST. This finding appears to be counterintuitive to indications that consonant accuracy was higher in the SWT. Since certain consonant clusters, specifically those containing liquids and glides, score high on the IPC rubric used to assess phonetic complexity in this study, it is not surprising that these errors were observed with greater frequency in the SWT.

Imitation provided a relatively quick means for obtaining samples of continuous speech. However, the amount of time spent analyzing samples from the two conditions did not vary significantly though transcription was made easier by the use of target templates created with the LIPP software (Oller & Delgado, 2000). One possible explanation for this is that increased errors are characteristic of conversational speech, thus requiring additional analytical resources. Another possibility is that increased errors are observed in continuous speech as a consequence of linguistic and contextual influences that single-word tasks fail to

capture. No specific conclusions can be drawn from the results of this study, and the question of whether the increased errors at word boundaries are an acceptable consequence of conversational speech or represent clinically significant deviance remains open.

LIMITATIONS OF THE PRESENT STUDY

One limitation of the current study was that the participants all exhibited typical speech development. In addition, many participants demonstrated above-average performance on a standardized assessment. None of the participants scored in the low-average range. Children exhibiting typical speech development were recruited in order to present a general picture of the experimental task under construction, yet the resulting data may not be representative of other typically developing children of similar ages. There is also no way to predict whether the current findings would generalize to populations of children with speech sound disorders.

The participants in this study were all preschool age children. Performance on the tasks, specifically the imitation task, may be different for older or younger children. There is some evidence to support this. Thirteen participants were initially recruited to participate in this project. Data from the youngest participant were not used in the analyses. The youngest participant was not able to complete all the tasks, and the data were considered too different from that provided by the older participants. Hence, only 12 participants participated in this investigation, and the

resulting conclusions were drawn from a small pool of sample data. While caution was used in the statistical analyses of the data presented, it is possible that reported differences in speech production between the two tasks were characteristic of the specific population used in the investigation.

Finally, the experimental protocols used in this investigation are still under development and testing. This was the first investigation comparing the imitated sentences to single-word naming using the PABA-E. The results should be viewed with caution until the present findings are replicated through further testing of these procedures. The results of this investigation support continued exploration of the sentence imitation task with broader populations, including children with speech sound disorders. A more in-depth analysis of the IST as compared to samples of spontaneous speech would provide additional support to the current findings.

IMPLICATIONS

While spontaneous conversation has been referred to as the most naturalistic context for assessing connected speech, lack of control over the content of spontaneous conversation samples creates some challenges. Avoidance of certain sounds and segments by the speaker, as well as unknown referents within unintelligible utterances, make transcription and analysis of spontaneous conversation difficult. Still, the influence of linguistic and contextual factors on speech production can only be observed in connected speech, and their influence may

be an important consideration in the identification, description, and classification of a child's speech patterns.

The use of sentence imitation to elicit samples of connected speech may hold some promise as part of a comprehensive speech sound assessment. Findings from the present study suggest that the IST captured many of the word-boundary and simplification errors characteristic of connected speech. At the same time, careful design of the sentence stimuli ensured that the types of sounds and segments were similar to those in the SWT. Even though the phonetic complexity of the SWT was found to be statistically higher than the IST in this investigation, whether this difference would be clinically significant awaits further study.

In general, similar types of errors were observed in the single-word and sentence imitation tasks used in this study. Consonant accuracy was found to be better in the SWT, and this finding would be expected given the increased demands of the connected speech task. The phonetic targets were similar in both the SWT and the IST, but the contextual and linguistic factors exclusive to the IST would be expected to influence the speech production abilities of the speaker in ways consistent with the findings of the current study. Caution should be used when interpreting the observed differences in error frequency between the single-word and imitated sentences tasks reported in this study. Even those differences found to be statistically significant may not have reached a level of significance that would be considered clinically relevant.

The question of whether the use of both methods may suggest a more comprehensive picture of speech production abilities awaits further study. The principal findings of this investigation suggest the potential utility of the SWT and the IST as complementary procedures for speech sound assessment in young children. No clear trend emerged with respect to which sampling condition provided a more consistent picture of speech production abilities. On the contrary, performance varied for the individuals who participated in this study even though all the children demonstrated better consonant accuracy in the SWT as compared to the IST. For some individuals the frequency of occurrence of a certain phonological error pattern was higher in the SWT than in connected speech. For other participants, the connected speech context appeared to better capture the occurrence of a given error pattern. These observations support the need for multiple procedures in speech sound assessment since comparison of performance across sampling contexts provided the most comprehensive picture of ability. It may be that one context is better suited to the needs of a given individual, depending on the developmental level and extent of the child's speech production repertoire. Information obtained through the trial of both complementary procedures may present the most representative description of an individual's speech patterns.

REFERENCES

- American Speech-Language-Hearing Association. (2008a). *2008 Schools survey report: Caseload characteristics*. Rockville, MD: Author.
- American Speech-Language-Hearing Association. (2008b). *2008 Schools survey report: Caseload characteristics trends, 1995-2008*. Rockville, MD: Author.
- Andrews, N., & Fey, M. (1986). Analysis of the speech of phonologically impaired children in two sampling conditions. *Language, Speech, and Hearing Services in Schools, 17*, 187-198.
- Bankson, N., & Bernthal, J. (1982). A comparison of phonological processes identified through word and sentence imitation tasks of the PPA *Language, Speech, and Hearing Services in Schools, 13*, 96-99.
- Bernhardt, B., & Holdgrafer, G. (2001a). Beyond the basics I: The need for strategic sampling for in-depth phonological analysis. *Language, Speech, and Hearing Services in Schools, 32*, 18-27.
- Bernhardt, B., & Holdgrafer, G. (2001b). Beyond the basics II: Supplemental sampling for in-depth phonological analysis. *Language, Speech, and Hearing Services in Schools, 32*, 28-37.
- Bernthal, J., & Bankson, N. (2004). *Articulation and phonological disorders* (5th ed.). Boston, MA: Allyn and Bacon, Inc.
- Bleile, K. (2004). *Manual of articulation and phonological disorders: Infancy through adulthood* (2nd ed.). Clifton Park, NY: Thompson/Delmar Learning.
- Bleile, K. (2002). Evaluating articulation and phonological disorders when the clock is running. *American Journal of Speech-Language Pathology, 11*, 243- 249.
- Campbell, T., & Shriberg, L. (1982). Associations among pragmatic functions, linguistic stress, and natural phonological processes in speech-delayed children. *Journal of Speech and Hearing Research, 25*, 547-553.
- DuBois, E., & Bernthal, J. (1978). A comparison of three methods for obtaining articulatory responses. *Journal of Speech and Hearing Disorders, 43*, 295-305
- Dyson, A., & Robinson, T. (1987). The effect of phonological analysis procedure on the selection of potential remediation targets. *Language, Speech, and Hearing Services in Schools, 18*, 364-377.
- Elbert, M, Dinnsen, D., Swartzlander, P, & Chin, S. (1990). Generalization to conversational speech. *Journal of Speech and Hearing Disorders, 55*, 694-699.
- Elbert, M., & Gierut, J. (1986). *Handbook of clinical phonology approaches to assessment and treatment*. San Diego, CA: College-Hill Press.
- Faircloth, M., & Faircloth, S. (1970). An analysis of the articulator behavior of a speech-defective child in connected speech and in isolated-word responses. *Journal of Speech and Hearing Disorders, 35*, 51-61.

- Fey, M. (1986). Articulation and phonology: Inextricable constructs in speech pathology. *Language, Speech, and Hearing Services in Schools*, 23, 25-232.
- Gallagher, R., & Shriner, T. (1975a). Articulatory inconsistencies in the speech of normal children. *Journal of Speech and Hearing Research*, 18, 168-175.
- Gallagher, R., & Shriner, T. (1975b). Contextual variables related to inconsistent /s/ and /z/ production in the spontaneous speech of children. *Journal of Speech and Hearing Research*, 18, 623-633.
- Gierut, J. (1998). Treatment efficacy: Functional phonological disorders in children. *Journal of Speech, Language, and Hearing Research*, 14, 85-101.
- Gildersleeve-Neumann, C. *Phonological and Bilingual Articulation Assessment, English Version (PABA-E)*. Unpublished assessment, Portland State University.
- Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation, Second Edition (GFTA-2)*. Minneapolis, MN: Pearson Assessments.
- Gordon-Brannan, M., & Hodson, B. (2000). Intelligibility/severity measurements of prekindergarten children's speech. *American Journal of Speech-Language Pathology*, 9, 141-150.
- Grunwell, P. (1997). Developmental phonological disability: Order in disorder. In B. Hodson & M. Edwards (Eds.), *Perspectives in applied phonology*, (pp. 61-103). Gaithersburg, MD: Aspen Publications.
- Haynes, W., & Steed, S. (1987). Multiphonemic scoring of articulation in imitative sentences: Some preliminary data. *Language, Speech, and Hearing Services in Schools*, 18, 4-14.
- Healy, T., & Madison, C. (1987). Articulation error migration: A comparison of single words and connected speech samples. *Journal of Communication Disorders*, 20, 129-136.
- Hodson, B., Scherz, J., & Strattman, K. (2002). Evaluating communicative abilities of a highly unintelligible pre-schooler. *American Journal of Speech-Language Pathology*, 11, 236-242.
- Hoffman, P., & Norris, J. (2002). Phonological assessment as an integral part of language assessment. *American Journal of Speech-Language Pathology*, 11, 230-235.
- Hoffman, P., Schuckers, G., & Ratusnik, D. (1977). Contextual-coarticulatory inconsistencies of /r/ misarticulation. *Journal of Speech and Hearing Research*, 20, 631-643.
- Howell, P., Au-Yeung, J., Yaruss, S., & Eldridge, K. (2006). Phonetic difficulty and stuttering in English. *Clinical Linguistics & Phonetics*, 20(9), 703-716.
- Individuals with Disabilities Education Act (IDEA), 34 CFR § 300.301 (2006).
- Johnson, J., Winney, B., & Pederson, O. (1980). Single words versus connected speech articulation testing. *Language, Speech & Hearing Services in Schools*, 11, 175-179.

- Kenny, K., Prather, E., Mooney, M., & Jeruzel, N. (1984). Comparisons among three articulation sampling procedures with preschool children. *Journal of Speech and Hearing Research, 27*, 226-231.
- Kent, R. (1982). Contextual facilitation of correct sound production. *Language, Speech, and Hearing Services in Schools, 13*, 66-76.
- Kent, R., Miolo, G., & Bloedel, S. (1994). The intelligibility of children's speech: A review of evaluation procedures. *American Journal of Speech-Language Pathology, 3*(2), 81-95.
- Khan, L. (2002). The sixth view: Assessing preschool children's articulation and phonology from the trenches. *American Journal of Speech-Language Pathology, 11*, 250-254.
- Klein, H. (1984). Procedure for maximizing phonological information from single-word responses. *Language, Speech, and Hearing Services in Schools, 15*, 267-274.
- Klein, H., & Spector, C. (1985). Effect of syllable stress and serial position on error variability in polysyllabic productions of speech-delayed children. *Journal of Speech and Hearing Research, 50*, 391-402.
- Kwiatkowski, J., & Shriberg, L. (1992). Intelligibility assessment in developmental phonological disorders: Accuracy of caregiver gloss. *Journal of Speech and Hearing Research, 35*, 1095-1104.
- Kwiatkowski, J., & Shriberg, L. (1998). The capability-focus treatment framework for child speech disorders. *Journal of Speech-Language Pathology, 7*(3), 27-38.
- Masterson, J., Bernhardt, B., & Hofheinz, M. (2005). A comparison of single words and conversational speech in phonological evaluation. *American Journal of Speech-Language Pathology, 14*, 229-241.
- Miccio, A. (2002). Clinical problem solving: Assessment of phonological disorders. *American Journal of Speech-Language Pathology, 11*, 221-229.
- Morrison, J. & Shriberg, L. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research, 35*, 259-273.
- Oller & Delgado. (2000). *Logical International Phonetic Program (LIPP) (Version 2.02)*. Miami, FL: Intelligent Hearing Systems.
- Paden, E. & Moss, S. (1985). Comparison of three phonological analysis procedures. *Language, Speech, and Hearing Services in Schools, 16*, 103-109.
- Paynter, E., & Bumpas, T. (1977). Imitative and spontaneous articulatory assessment of three-year-old children. *Journal of speech and Hearing disorders, 17*, 119-125.
- Pena-Brooks, A., & Hegde, M. (2000). *Assessment and treatment of articulation and phonological disorders in children*. Austin, TX: Pro-Ed, Inc.
- Prather, E., & Kenney, K. (1986). Coarticulation testing of kindergarten children. *Language, Speech, and Hearing Services in Schools, 17*, 285-291.

- Schwartz, R. (1992). Clinical applications of recent advances in phonological theory. *Language, Speech, and Hearing Services in Schools, 23*, 269-276.
- Shriberg, L. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research, 36*, 105-140.
- Shriberg, L, Austin, D., Lewis, B, McSweeny, J, & Wilson, D. (1997). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*, 708-722.
- Shriberg, L., & Kwiatkowski, J. (1980). *Natural processes analysis: A procedure for phonological analysis of continuous speech samples*. New York: Macmillan.
- Shriberg, L., & Kwiatkowski, J. (1985). Continuous speech sampling for phonologic analysis of speech-delayed children. *Journal of speech and hearing disorders, 50*, 323-334.
- Shriberg, L., Kwiatkowski, J., Best, S., Hengst, J., & Terselic-Weber, B. (1986). Characteristics of children with phonologic disorders of unknown origin. *Journal of Speech and Hearing Disorders, 51*, 140-186.
- Siegel, R., Winitz, H., & Conkey, H. (1963). The influence of testing instruments on articulatory responses in children. *Journal of Speech and Hearing Disorders, 28*, 67-76.
- Skahan, S., Watson, M., & Lof, G. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology, 16*, 246-259.
- Smith, M., & Ainsworth, S. (1967). The effects of three types of stimulation on articulatory responses of speech defective children. *Journal of Speech and Hearing Research, 10*, 333-338.
- Stoel-Gammon, C., & Dunn, C. (1985). *Normal and disordered phonology in children*. Austin, TX: Pro-Ed Press.
- Tyler, A., & Tolbert, L. (2002). Speech-language assessment in the clinical setting. *American Journal of Speech-Language Pathology, 11*, 215-220.
- Watson, M. (1989). Comparison of three methods for eliciting phonological processes. *Perceptual and Motor Skills, 69*, 771-778.
- Weston, A., & Shriberg, L. (1992). Contextual and linguistic correlates of intelligibility in children with developmental phonological disorders. *Journal of Speech and Hearing Research, 35*, 1316-1332.
- Williams, A.L. (2002). Epilogue: Perspectives in the assessment of children's speech. *American Journal of Speech-Language Pathology, 11*, 259-263.
- Williams, A.L. (2003). *Speech disorders resource guide for preschool children*. Clifton Park, NY: Singular.
- Wolk, L., & Meisler, A. (1998). Phonological assessment: A systematic comparison of conversation naming and picture naming. *Journal of Communication Disorders, 31*, 291-313.

Zehel, Z., Shelton, R., Arndt, W., Wright, V., & Elbert, M. (1972). Item context and /s/ phone articulation test results. *Journal of Speech and Hearing Research*, 15, 852-860.

APPENDIX A: PARENT SURVEY OF LANGUAGE AND DEVELOPMENT

Participant # _____

Child's Name: _____

Your Name: _____ Your Relationship to Child: _____

Date of Birth: _____ Date of Testing: _____ Age: _____

Person Completing Survey: _____

Section 1. DEVELOPMENT HISTORY. These questions help us understand your child's development. If you have questions or concerns about a question, please feel free to not answer or to ask for clarification.

Family History:

	Name	Age	Occupation	Education Level
Mother				
Father				
Sisters/ Brothers				

1. Where was your child born?
2. Where were the child's parents born?
3. What language(s) do the child's parents speak?
4. How old was your child when he or she first babbled (eg., say *bababa* or *dadada*)?
5. How old was your child when he or she first spoke three different words? What were they?
6. How old was your child when he or she started saying 2 and 3 word sentences on a regular basis?
7. How old was your child when she or he first spoke in sentences, even though some of the words in the sentence may have been missing?
8. Has your child ever had his or her hearing checked? What were the results?
9. What schools has your child attended?
10. Has your child been treated for ear infections? If yes, how many times? When were they?
11. Has your child ever had a serious illness or been hospitalized? If yes, please explain.
12. Does your child have any diagnosed medical conditions? If so, please explain.
13. Does your child have any difficulty eating or drinking?
14. Who are the main people your child interacts with?
15. Do you have any concerns about your child's general development? If so, what are they?

Section 2. SPEECH DEVELOPMENT QUESTIONS. These questions help us understand how clearly your child speaks, as well as whether you have concerns about your child's speech development. **Please circle one for each question.**

1. Is your child's pronunciation difficult to understand? *(Circle one)*

1-Never 2-Rarely 3-Sometimes 4-Frequently 5-All the time

2. In comparison to other children his/her age, do you think your child is difficult to understand? *(Circle one)*

1-Never 2-Rarely 3-Sometimes 4-Frequently 5-All the time

3. Do other people think your child is difficult to understand? *(Circle one)*

1-Never 2-Rarely 3-Sometimes 4-Frequently 5-All the time

4. Does your child have difficulty pronouncing words? *(Circle one)*

1-Never 2-Rarely 3-Sometimes 4-Frequently 5-All the time

5. Does your child have problems producing certain sounds? *(Circle one)*

1-Never 2-Rarely 3-Sometimes 4-Frequently 5-All the time

6. Does your child leave out sounds when he/she speaks? For example, saying "ca" for "cat", or "pato" for "plato?" *(Circle one)*

1-Never 2-Rarely 3-Sometimes 4-Frequently 5-All the time

7. Does your child change sounds when he/she speaks? For example, saying ""too" for "shoe" or "wun" for "run?" *(Circle one)*

1-Never 2-Rarely 3-Sometimes 4-Frequently 5-All the time

8. Is your child frustrated when he/she speaks? *(Circle one)*

1-Never 2-Rarely 3-Sometimes 4-Frequently 5-All the time

9. In comparison to other children his/her age, do you think your child has speech problems? *(Circle one)*

1-No 2-Probably not 3-Maybe 4-Probably 5-Yes

10. Do other people think your child has speech problems? *(Circle one)*

1-No 2-Probably not 3-Maybe 4-Probably 5-Yes

APPENDIX B: PABA-E SINGLE-WORD AND IMITATED SENTENCES LISTS

apples	chair	fork	lamp	red	string
baby	chicken	four	leaf	ring	swinging
balls	chocolate	french fries	lips	school	teacher
banana	clouds	frog	little	scissors	teeth
bath	cookies	gate	milk	seven	ten
bathtub	cup	giraffe	mirror	sheep	this/that
big	dinosaur	girl	mountains	ship	three
bike	dog	glasses	mouse	sick	throwing
bird	door	goose	mouth	six	tiger
black	drinking	grapes	nest	skates	toe
boat	drum	guitar	nine	slide	tongue
books	eggs	hamburger	one	smile	toothpaste
boot	eight	hammer	orange	smoke	toys
boy	elephant	hand	pancakes	snake	truck
brother	face	helicopter	pants	soap	two
brush	feather	horse	pencil	sock	vacuum
butterfly	feet	house	pet	soup	watch
cake	finger	ice cream	pie	spider	waterfall
candle	fire truck	jacket	pig	spoon	yellow
carrots	fish	jet	planes	squirrel	zipper
cars	five	juice	pool	stars	
cat	flower	jumping	puppy	stop	
caterpillar	foot	kitchen	rainbow	strawberry	

- a. He is wet.
- b. One cat sat on a ball.
- c. He found a puppy in a cup.
1. The yellow duck is swimming fast.
2. A big frog jumped over the bathtub.
3. I love milk and cookies.
4. Our younger brother has a broken foot.
5. Spotted elephants like green bananas.
6. There are five candles on my birthday cake.
7. I read a small book about a dinosaur.
8. Three sheep played in the flowers.
9. She's trying to cut paper with good scissors.
10. The toy firetruck drove over the mountains.
11. The chicken laid an orange egg outside.
12. We eat yogurt through a straw.
13. She brushed her teeth with chocolate toothpaste.
14. The young girl likes blueberry jam.
15. Her school teacher is wearing noisy shoes.
16. The boy's jacket looks warm.
17. Six rabbits hopped across the street.
18. Seven kids are making peanut butter and jelly sandwiches.