Dynamic Assessment in Phonological Disorders: The Scaffolding Scale of Stimulability

Amy M. Glaspey

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

Doctor of Philosophy

University of Washington

2006

Program Authorized to Offer Degree:
Department of Speech and Hearing Sciences

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Amy M. Glaspey

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

Carol Stoel-Gammon

Reading Committee:

Carol Stoel-Gammon

Lesley B. Olswang

Owen White

Judith Stone-Goldman

Date: 5/18/06

University of Washington

Abstract

Dynamic Assessment in Phonological Disorders: Scaffolding Scale of Stimulability

Amy M. Glaspey

Chair of the Supervisory Committee:
Professor Carol Stoel-Gammon
Department of Speech and Hearing Sciences

The Scaffolding Scale of Stimulability (SSS) is a dynamic assessment of phonological change. Using the SSS, clinicians evaluate phonological skills by documenting a child's response to cues and manipulations of the linguistic environment. The SSS differs from standard phonological assessments, which only assess productions elicited in a static manner without assistance. When administering the SSS, clinicians rate each target production using a 21-point ordinal scale that represents the number of cues and assistance or "scaffolds" that could be used to assist the child in production. Over time, children need less scaffolding and this change is reflected by a lower score on the measure. In the current study, sensitivity and patterns of change on the SSS were compared to one dynamic assessment, the Carter-Buck Articulation Assessment (Carter & Buck, 1958), and three static assessments: a probe of single-words individualized to the children's errors, the *Hodson Assessment of Phonological Patterns* (HAPP-3)(Hodson, 2004), and Percent Consonants in Error (adapted from Shriberg & Kwiatkowski, 1982). The measures were administered to eight preschool age children

with moderate-to-profound phonological disorder across six months of Cycles treatment. The measures were administered prior to treatment, after three months of treatment, and after six months of treatment. Results indicated that the SSS showed the greatest percentage of change across participants more frequently than the other measures. Patterns of change on the SSS were more even across two cycles of treatment or showed a greater percentage of change during Cycle 1. The Carter-Buck, another dynamic assessment, showed patterns of change consistent with the static assessments, in particular, the HAPP-3. The static assessments more frequently showed a greater change during Cycle 1 or Cycle 2, rather than even change. Percentage of change on the probe measure varied across the children with no consistent pattern. The PCE was most unusual because seven of the eight participants showed regression on this measure. Overall, the SSS was considered the most sensitive measure for documenting change across time in children with phonological disorders.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Number | Page

LIST OF TABLES

ACKNOWLEDGEMENTS

I wish to thank Dr. Carol Stoel-Gammon for her guidance and support in the completion of my doctoral program. It has been an honor and a privilege to be mentored by this extraordinary researcher. Her encouragement has been endless and has given me confidence to pursue a research career. I also wish to acknowledge my committee members: Dr. Lesley Olswang who guided my thinking in search of the ultimate research question, Dr. Owen White who provided alternative perspectives in assessment design and instruction, and Dr. Judith Stone-Goldman who encouraged clinical research and also precision in writing.

Further support from the Department of Speech and Hearing Sciences came from Dr. Christopher Moore in research presentation and the pursuit of external funding. Many students contributed their time, labor, and friendship, and I thank: Amber Franklin, Minjung Kim, Andrea MacLeod, Anna Sosa, Beate Peter, Amy Donaldson, Curtis Billings, Chris Sanford, Cathy Off, Sasha Tishler, Alex Pear, Kacie Fornia.

I was privileged to work with a dedicated group of families who made accommodations around work, new babies, and even household moves, yet still persisted to bring their children for treatment twice a week. I wish to thank the University of Washington Speech and Hearing Clinic and the Issaquah School District for their cooperation; both agencies provided space for treatment and assisted my participants.

I appreciate my family who has always supported my educational pursuits especially my grandparents Fred and Mabel Glaspey, my parents Bill and Kathy, and my brother Bo. I wish to acknowledge my husband, Brett Leischner, who observed the life of an academic and still agreed to marry me one year ago. He has given me all the support I could need to complete this endeavor.

Finally, this research was funded by individual and institutional research training grants through the National Institutes of Health/National Institute on Deafness and other Communication Disorders, the Department of Speech and Hearing Sciences, and the Graduate School at the University of Washington.

DEDICATION

To my Mom,

who modeled the value of higher education by completing a nursing degree while raising teenagers and working nights.

To my Dad,

who had faith in me and told me that I would grow up to be a doctor or lawyer

. . . and buy him a condo on Maui and support him in his old age.

CHAPTER ONE: LITERATURE REVIEW

". . . if it were possible to devise a scale for measuring degrees of misarticulation, this scale would enable experimenters to compare the effectiveness of therapeutic procedures, and would enable clinicians to describe results of corrective procedures in more exact terms." Curry, Kennedy, Wagner, and Wilke (1943)

The above observation is interesting because it was written in an article over 60 years ago noting a measurement dilemma for documentation of treatment efficacy in articulation disorders, a problem that is still current today. The authors suggest a potential solution for documenting treatment efficacy, the development of a scale that could describe speech samples and show changes in samples across time. Still after more than 60 years, researchers and clinicians continue to struggle with accurately measuring phonological change across time.

In the 1940's Curry and colleagues did develop a scale for evaluating change, "a phonographic scale for measuring defective articulation." Their scale included an audio record of ten different speech samples of children with articulation disorders that were put in order by severity as measured by intelligibility. Clinicians could then be trained to match their own client's connected speech samples to the original recordings and rate the severity on a scale. Unfortunately, wide-spread duplications of the recordings were unavailable to clinicians "due to war-time efforts." Other factors that have slowed the process of measurement development, other than World War II, are inherently the result of the nature of phonological disorders and are further considered.

Proof of treatment efficacy in phonological disorders is problematic due to the slow nature of phonological change and due to the insensitivity of measures for capturing phonological change. Current measures may take *months* before minimal changes are documented and *years* before maximal changes have occurred. By the time change occurs, determination of the source of change, whether from the treatment or outside factors, is difficult to ascertain. A sensitive measure would allow for the documentation of phonological change in small increments, for the study of specific treatment programs, and for the manipulation treatment approaches for the development efficacious treatments. In response to these problems, a new phonological measure, the Scaffolding Scale of Stimulability (SSS) (Glaspey & Stoel-Gammon, 2001, 2002, 2005), was developed with the goal of achieving greater sensitivity to incremental change than previously used measures. The SSS provides opportunities for the study of the nature of phonological change by measuring changes in *stimulability*. Comparisons across measures must first be made to support the reliability and validity of the SSS. Once the benefits of the SSS are established, then future studies can use the SSS to document the efficiency of treatment methods and determine which treatments produce optimal change. Ultimately, with improved treatment methods, faster and broader phonological change may be induced in children with phonological disorders and allow them to acquire age-appropriate speech more rapidly.

Traditional Assessment versus Dynamic Assessment

Traditional forms of assessment typically measure a child's abilities in an unassisted context; in other words, they are *static* assessments. In the area of articulation and phonological disorders, individual phonemes and speech patterns are assessed in a static manner by showing the child a series of pictures or objects and instructing the child to name the pictures, thus producing a target sound in a single-word production. During the assessment process, the child does not receive any feedback regarding the accuracy of productions. In fact, clinicians are trained to avoid any types of reinforcement indicators that might skew the results of the assessment. The clinician may comment to the child about good behavior attending to the task or participation, but not anything specific about the phonemes themselves. Furthermore, the assessment is administered in a rigid fashion and the protocol is not modified in any way based on the child's responses; for example, if a child produces a [t] for /k/ in a target word, the clinician simply transcribes the production and continues the assessment. Several examples of commonly used static assessments include the *Goldman-Fristoe Test of Articulation* (Goldman & Fristoe, 2000), the *Hodson Assessment of Phonological Patterns* (Hodson, 2004), *The Photo Articulation Test* (Lippke, Dickey, Selmar, & Soder, 1997), or the *Assessment Link between Articulation and Phonology* (Lowe, 1986).

Static assessments have been traditionally used by speech-language pathologists and provide many benefits in the treatment process. Static assessments typically include simple materials (e.g., manipulatives or pictures books) and an established protocol for elicitation, scoring, and analysis of target speech productions (Velleman, 1998; Williams, 2000). Elicitation may include spontaneous or imitated naming tasks using single words or sentences; the data collected includes opportunities for productions across a wide range of skills. Results from static assessments may inform the clinician regarding severity levels, differential diagnosis, or treatment procedures. Static assessments offer many other advantages: (1) Procedural reliability, in general, is high. (2) Administration for most phonological assessments is completed in less than 20-30 minutes. (3) Simple repetition and naming tasks place little stress on the child. (4) A score or rating is typically generated which can be used to indicate change in performance across time, to compare skills across children, and to compare normative data. (5) Static assessments identify specific error patterns and/or occurrences of specific error types.

Another type of static assessment is an untreated single-word probe. Probes are a form of measurement frequently used by researchers and clinicians to document treatment efficacy by assessing a child's ability to transfer skills learned on specific words to words that have not been treated. The probes are based on elicitation of words that contain targeted phonological patterns or phonemes; however, the selected words are not used as part of treatment. Probes are designed to determine how well sound patterns are learned and then applied in new contexts.

In previous research, the use of probes has varied. The timing of administration of probes may differ from several weeks to months. Differences are also evident in scoring procedures: some probes are scored by correct patterns, correct classes, or even just percentage of consonants correct. For example, Gierut, Elbert, and Dinnsen (1987) used a comprehensive probe with elicitation of 310 words, which sampled all ambient

English consonants in each relevant word position in a minimum of five different exemplars. In a later study, Gierut (1996) used a similar protocol with 198 words, which was administered pre-treatment; a subset of these words, based on the child's individual errors, was elicited prior to treatment phase shifts, and two weeks and two months post-treatment. The use of probes as part of an assessment offers many benefits including individualization to the participant and increased opportunities for assessment of each error pattern. Additionally, once children begin to generalize skills from treated targets to untreated targets, the occurrences of correct productions can be linked to aspects of treatment. In this way, treatment efficacy is documented through changes on the untreated single-word probes in relation to treatment.

The question then arises, if static assessments have been used for so long and provide information to the clinician, why should other assessment procedures be considered? While static assessments provide useful information, they do not provide a complete picture of a child's phonological abilities. Several disadvantages of using static assessments become apparent. First, the ecological validity may be questioned because productions for many assessments are only at the single word level; some children with phonological disorders may easily produce single-words that they cannot produce in connected speech. For example, Morrison and Shriberg (1992) compared standard assessments and conversational speech samples and found an increased frequency of errors and types of errors on conversational speech samples that were not evident on the citation assessments. Second, perhaps the biggest disadvantage of a static assessment is the length of time required before significant improvements can be observed; these assessments do not readily allow for measurement of small incremental steps that may occur prior to transfer of skills from treatment to the assessment. Finally, static assessments may not readily discriminate subtle differences across children that could indicate a differential diagnosis or phonological profile that would guide decision making in the treatment process.

Clinicians seeking to gather more detailed information about the child's skills and the potential for the child to make progress in response to treatment may use a *dynamic assessment.* Dynamic assessment is a socio-cultural perspective that encompasses Vygotsky's model of cognitive development (Bain & Olswang, 1995; Vygotsky, 1978). Vygotsky's work early in the 1920's and 1930's included the study of higher psychological processes that could document and differentiate humans from animals. With this goal in mind, Vygotsky studied the differences in development, tool use, memory, attention, and social interactions. He observed that the most commonly used assessments included descriptions of external features through documentation of stimulus and response, a unidirectionally reactive approach. While this type of assessment did inform psychologists regarding an individual's mental capacity, the assessment did not clearly differentiate individuals or give indicators of future progress; furthermore, the stimulus-response framework could not capture higher specifically human forms of behavior.

Vygotsky suggested that there were two developmental levels: the *actual* developmental level and the *potential* developmental level. The actual developmental level is the performance of an individual when no assistance is given, as described above under static assessments. In contrast, the potential developmental level is the level of performance that an individual achieves when given assistance. The assessment includes analysis of the process, not the object and begins to explain the differences across individuals. Vygotsky claimed that individuals who performed in the same manner on basic stimulus-response tests, could in fact, be differentiated by the amount of support that they needed through the assessment of their potential. Vygotsky called the region between the actual development level and the potential development level the "zone of proximal development" (ZPD). In this way he advocated for the study of both the product and the process of learning. Furthermore, dynamic assessment for the ZPD depends upon the relationship between the adult and the child and their interactions together because children are motivated through social needs. The assessment procedures described by Vygotsky were analogous to solving a math problem and showing the work involved in achieving the solution. Instructors can differentiate children based on their process of problem-solving.

Dynamic assessment may also be used to help determine whether a child is ready to learn. Vygotsky claimed that once the ZPD was documented, the assessment could be used to guide the learning process. For optimal learning to occur, instruction should be above the child's actual level but within the ZPD and that learning creates the ZPD. In phonology, a child who cannot produce a phoneme on a static assessment, and who cannot produce a phoneme with assistance, may not be ready to learn the phoneme. Another child who cannot produce a phoneme on a static assessment, but can produce a phoneme when given some support may show readiness in the learning process. More specifically, two children could both score 0/10 on a static picture naming task for the target /k/. Yet, when the two children are given some help in the production of /k/ they may suddenly appear very different. One child may produce /k/ when given only a verbal model as support, but the other may not produce the target at all even when given support in the form of instructions, a verbal model, or tactile cues. The child who could produce the /k/ with some minimal support would likely acquire the sound quickly in treatment. The static assessment for these two children may have misled the clinician in treatment planning.

If a child's phonological skills are assessed in an environment that is unsupported, it may seem as though the child is not capable of producing a phoneme or pattern; in fact, the child could be very close to producing a target phoneme if only given a little help. A child acquiring new information initially needs maximal adult assistance, however, as the child internalizes the information less adult assistance is needed until the child can perform independently. The goal is for the adult to systematically move through the zpd to induce the greatest change.

Stimulability

Although terminology and theories of dynamic assessment are most often associated with language or cognitive development (Bain & Olswang, 1995; Lidz, 1991; Peña, Iglesias, & Lidz, 2001) similar concepts have been applied in phonological intervention, but through assessment of *stimulability.* Even in Vygotsky's works, references were made that the purpose of dynamic assessment was "to evaluate the response of a stimulating situation in a controlled way." In the assessment of phonology, a "stimulating situation" is created when testing stimulability, as the name implies.

Methods of Stimulability

Stimulability has been used for over 75 years as part of the assessment and treatment process (Powell, 1996; Travis, 1931). Stimulability refers to testing the participant's ability to produce a misarticulated sound in an appropriate manner when "stimulated" by the clinician to do so (Bauman-Waengler, 2000). A broad and general definition has been selected here because of the variability in which the assessment of stimulability has been conducted in the past. Typically, stimulability assessment occurs following the administration of a static assessment where targets have been assessed without assistance. One example of a published assessment that includes stimulability testing is the *Goldman-Fristoe Test of Articulation* (Goldman & Fristoe, 2000). Using this measure, clinicians assess phonemes with a spontaneous elicitation procedure and then a stimulability subtest is administered on any targets produced in error. The stimulability subtest includes a verbal model of the target phoneme in syllables, words, and sentences. If the child can produce the target phoneme with a verbal model, then the child is said to be stimulable.

Most often, however, clinicians assess stimulability with a variety of informal procedures; variability of methods across clinicians includes differences in the *environments* assessed and the types of *cues* that are given. Clinicians may choose different linguistic environments such as isolation, syllables, words, sentences, or connected speech. In addition, different types of cues may be given in each of these environments. The cues might include placement instructions (Rvachew, Rafaat, & Martin, 1999), a verbal and/or visual model (Carter & Buck, 1958), or tactile cues (Bain, 1994). The most commonly used cues consist of a verbal model and visual model where children are instructed to watch the clinician's face when the verbal model is presented (Carter & Buck, 1958; Goldman & Fristoe, 2000; Lof, 1996; Milisen, 1954; Scott & Milisen, 1954).

The assessment of stimulability may further vary in number of opportunities that are presented to the child. Lof (1996) assessed stimulability by giving participants 10 opportunities to produce a phoneme in a CVC environment (initial error consonant + /ʌ/ + a stop or nasal consonant). The child was cued to listen, watch, and repeat the clinician's model. The child was considered stimulable if the target was produced in at least two out of ten opportunities. Lof further reported factors that most influenced stimulability: visibility of the sound, age, socioeconomic status, and overall imitation skills. The method of administration may further vary based on the number of times that the target is modeled prior to the child's response.

One final variation in the assessment of stimulabilty involves the use of a scale. Rvachew, Rafaat, and Martin (1999) used a 3-point stimulability scale to examine the relationship between pretreatment stimulability and speech perception. Syllable-initial targets were assessed in three environments: isolation, word, and sentence. Productions correctly produced in isolation were rated with a 1, in words with a 2, and in sentences with a 3. Participants were given a verbal model of the target in each of the three environments.

Purpose of Stimulability Testing

Historically, the assessment of stimulability has been used as a prognostic measure, as a diagnostic tool, and as a treatment guide. As a prognostic tool, the assessment of stimulability has a long history of use (Carter & Buck, 1958; Diedrich, 1983; Farquhar, 1961; Irwin, West, & Trombetta, 1966; Kisatsky, 1967; Powell & Miccio, 1996; Snow & Milisen, 1954; Sommers et al., 1967). Caseload size for speech-language pathology has been a constant issue; in the area of articulation and phonology, testing of stimulability was initially suggested as a method for determining which students should be prioritized for speech therapy services (Carter & Buck, 1958).

Stimulability and Prognosis

Carter and Buck (1958) analyzed the concept of stimulation to determine which first grade children should receive priority for speech treatment based on differences in performance at the spontaneous word, imitated word, and imitated non-sense syllable levels. Their objective was to distinguish children who would independently self-correct without intervention from the children who would not progress without speech therapy. Children with "defective articulation" were divided into a treatment group and a non-treatment group over the course of the school year. Three subtests were administered that included spontaneous single-word productions, imitated single-word productions, and imitated syllable productions across 13 selected phonemes at the beginning and end of the school year. Errors were counted on each subtest, converted to percentages, and comparisons were made based on the percentage of correction for each subtest. Results showed that in the no-treatment group, the higher the correction between syllable productions and spontaneous productions, the more likely the children improved. In contrast, children who scored under 25% correction made little to no improvement without speech therapy. Children in the treatment group were much more likely overall to make a larger correction. Results regarding prognosis suggested that clinicians should use the imitated-syllable subtest in comparison with the spontaneous subtest and that those children who achieved 75% correction or more on the imitated-syllable test would likely mature independently without speech therapy.

Stimulability and Diagnosis

As a diagnostic tool, assessment of stimulability has been used by clinicians to support standardized testing in the diagnosis of a phonological disorder. Clinicians have made statements about a child's skill levels by identifying phonemes missing from a child's phonetic inventory and by determining whether these sounds were stimulable. Children with a large number of unstimulable sounds were often diagnosed with a phonological disorder and selected for speech services because of a lack of stimulability. By comparison, children who were stimulable did not qualify for services because it was believed that their speech sound repertoire would continue to develop without intervention.

The *Goldman-Fristoe Test of Articulation* (GFTA) (Goldman & Fristoe, 2000) provides a stimulability subtest that may assist clinicians in making diagnostic decisions. The GFTA provides a simple hierarchy with syllable, word, and sentence level tasks with explicit directions for administration; children receive on verbal model and visual cues at each level. Although the GFTA is highly regarded for its reliability and validity, the stimulability subsection of the newest version does not result in a score and does not have reliability and validity data. Clinicians can only use the subsection to make some general statements regarding diagnosis and prognosis.

The decision to provide services for children with phonological disorders based on the results of stimulability testing is complicated by studies showing that a child who is stimulable has good prognosis in therapy and a child who is not stimulable may be more resistant to treatment (Bain, 1994). The literature remains mixed regarding the results of stimulability and the impact on outcomes for children (Diedrich, 1983; Powell & Miccio, 1996; Rvachew et al., 1999). Studies have documented that without treatment, children who were stimulable made more progress than those who were not stimulable (Carter & Buck, 1958; Farquhar, 1961; Kisatsky, 1967; Snow & Milisen, 1954), and that with treatment, children who were stimulable made less progress than those who were not stimulable (Carter & Buck, 1958; Sommers et al., 1967). Other studies, in contrast, have documented that when treatment was administered, children who were stimulable made the greatest gains (Irwin et al., 1967) and that all participant groups who received treatment, whether stimulable or not, made improvement compared to those who did not (Sommers, et al., 1967). Many of these discrepancies could be related to measurement differences and that children with more severe disorders may also have the most room for growth.

Stimulability and Treatment Planning

Once clinicians have completed the diagnostic phase and prognostic phase and enrolled a child in speech therapy, some clinicians use the results from the assessment of stimulability to select treatment targets and develop treatment programs. What is interesting is that some clinicians choose treatment targets based on those that are the highly stimulable versus other clinicians who choose targets based on those that are the least stimulable. Milisen (1954) advocated using highly stimulable targets when using Integral Stimulation Treatment and Van Riper recommended selecting targets that were highly stimulable in conjunction with other characteristics such as the frequency of use in English and earlier developing phonemes in what has been called the "traditional" treatment approach (Secord, 1989). Furthermore, in Metaphon treatment (Howell & Dean, 1994), stimulable targets are treated in the second phase of treatment following the initial phase of concept instruction.

Other treatment approaches advocate the treatment of targets that are least stimulable, or in other terms are of "least phonological knowledge" (Gierut, 1998; Gierut, Morrisette, Hughes, & Rowland, 1996). The unstimulable phoneme that best represents the missing articulatory features within the child's phonological system is selected for treatment; it is believed that generalization is greater when these sounds are treated (Gierut, 1998; see also Powell, Elbert, and Dinnsen, 1991). Many other researchers also suggest treating unstimulable phonemes because these are least likely to improve without the help of the treatment (e.g., Miccio, Elbert, & Forrest, 1999; Powell, Elbert, & Dinnsen, 1991; Powell & Miccio, 1996).

One treatment that incorporates both stimulable and potentially non-stimulable targets is Hodson's Cycles Approach. This treatment is based on "cycles" in which error patterns are targeted for a short period of time and then, if still in error at the end of a cycle, they are targeted again in the next cycle (Hodson & Paden, 1991). The notion of "cycles" in phonological treatment is based on the hypothesis that treatment of one exemplar of each error pattern will induce generalization to the entire pattern. In the early phases of treatment, Hodson targets the most stimulable targets, yet also advocates treatment of consonant clusters and liquids, which for some children may not necessarily be stimulable.

Overall, many questions still remain regarding the treatment efficacy of these different treatment approaches and whether any one treatment is more efficient than another. Given the variability of participants, treatment approaches, and measures used to document efficacy, it is difficult to make comparisons. A recent study (Rvachew, Rafaat, & Martin, 1999) "loosely" used a Cycles Approach while manipulating stimulability; however, the treatment paradigm was so extensively modified that much of the essence of "cycles" was lost. Findings indicated that targets that were stimulable at the start of treatment showed greater improvement than those that were less stimulable.

Merging Stimulability and Dynamic Assessment

While the uses of stimulability have varied, the theoretical construct of dynamic assessment was linked with stimulability by Bain (1994) who proposed a framework for applying the strategies of dynamic assessment to phonological disorders using stimulability. She applied the constructs of dynamic assessment in a previous study of language skills and suggested that the same could be done in phonology (Bain & Olswang, 1995). Bain proposed a tool that could help clinicians make better decisions during treatment by focusing on the clinical questions: "Who to treat?" "When to treat?" "How to treat?" and "What will be the prognosis for treatment?"

Bain's framework included suggestions for the following elements: (1) manipulation of antecedents, (2) manipulation of responses, and (3) manipulation of consequences. First, the antecedents for manipulation included the different ways that the target could be presented to the child. During dynamic assessment, the clinician could present the target in a minimal pair, give an auditory model, give a visual model (or a combination of both), manipulate the frequency of stimulus presentation, alter the prosodic emphasis, describe visual imagery, suggest placement cues, or manipulate the articulators. Second, the responses could be manipulated in terms of linguistic complexity or through interactions with language components. Linguistic complexity varied from isolation, nonsense syllables, monosyllable words, multi-syllable words, and sentences or phrases. This hierarchy stems from concepts within the "traditional approach" of speech treatment (Secord, 1989). Further interaction could occur with manipulation of the language components of syntax, semantics, pragmatics, and phonology. Finally, the consequences of the child's response could be altered and manipulated through variations in schedules and types of reinforcement.

Overall, Bain suggests that the clinician should begin with the least supportive cues within the hierarchy and proceed to the most supportive cues until a correct production is reached. A weighted scoring system was suggested to document the differences in support needed across children. With a valid and reliable system such as this in place, clinicians could begin to answer many clinical questions.

Bain's framework of dynamic assessment was applied by Perrine (1999) with a study that assessed the construct validity of a cueing hierarchy for phonology. Perrine developed and analyzed a cueing hierarchy for nine target phonemes: /k, g, f, v, s, l, ʃ, tʃ, dʒ/. The targets were selected based on normative data for the age group and were balanced to represent targets across sound classes. Fifteen children between the ages of 4;7 and 6;5 with phonological delays participated in the study. Dynamic assessment was completed with the children on two or three phonemes that were missing from their phonemic inventories; the phonemes assessed differed across the children. The cueing hierarchy progressed from least-supported to most-supported and included seven antecedent cues: (1) indirect auditory model, (2) indirect auditory and visual model, (3) direct auditory model, (4) direct auditory and visual model, (5) placement cues with direct auditory model, (6) placement cues with direct auditory and visual model, and (7) tactile cues plus placement cues with auditory and visual model. Three linguistic environments were assessed: CVC words, CV nonsense syllables with /a/ vowel, and isolation.

During assessment, the child was given each cue only one time while attention was maintained. Accuracy feedback was given following each response along with intermittent feedback regarding attention. Overall, 21 forms of scaffolding were administered in sequence for each phoneme beginning with the least supportive forms. Construct validity was suggested because accuracy was higher at the most supported levels of the hierarchy and the children's scores followed the expected hierarchy. The scores for each child were added together to create a composite score of stimulability. A degree of stimulability was achieved through a rank ordering of the participants' scores. Construct validity was further established because the scores fell along a range that suggested that the children's abilities could be differentiated.

Dynamic Assessment of Stimulability as an Outcome Measure

The work of Bain (1994) and Perrine (1999) supports the theoretical constructs of stimulability as a dynamic assessment for developing treatment programs and for making prognostic statements about a child's productions, and consequently, supports previous studies; however, a novel and more advanced use of the assessment of stimulability is as an outcome measure for treatment, rather than just as a prognostic or diagnostic indicator (Glaspey & Stoel-Gammon, 2001, 2002, 2005). Thus, the Scaffolding Scale of Stimulability was created (see Figure 1.0).

CUES / ENVIRONMENTS	*Level 0*	*Level 1* Instructions	*Level 2* Instructions, verbal model	*Level 3* Instructions, verbal model, segmented, simultaneous production, tactile cues
Not Stimulable				21
Isolation				20
Word	16	17	18	19
Carrier Phrase	12	13	14	15
Novel Phrase	9	10	11	
One-target sentence	6	7	8	
Two-target sentence	3	4	5	
Connected Speech	1	2		

Figure 1.0. The Scaffolding Scale of Stimulability: Cues and Environments.

The Scaffolding Scale of Stimulability is a dynamic assessment of phonological change. The measure is based on a 21-point scale that is used to assess children's speech productions. The measure could potentially be used for diagnostic and prognostic purposes, but was developed with treatment efficacy in mind. Targets are assessed through the manipulation of linguistic environments and manipulation of cues. Conceptually, the SSS is highly related to Perrine's measure; however, the environments that were selected and the cues provided to the child are slightly different. First, the SSS comprises seven different environments: isolation, word, carrier phrase, novel phrase, one-target sentence, two-target sentence, and connected speech. Second, the SSS includes a greater range of antecedent cues that are grouped together in four levels: Level 0—spontaneous or delayed model; Level 1—verbal instruction about articulatory placement; Level 2—verbal instruction plus verbal model; Level 3—verbal instruction plus verbal model plus prolongation, segmentation, simultaneous production, or tactile cues. A third difference is that the range of targets included on the SSS is greater; the SSS assesses of all phonemes in initial and final position of words based on the phonotactics of English (see Figure 1.0). A fourth difference on the SSS is that unlike Perrine's study, not all 21 levels of the assessment are administered for every phoneme. The assessment actually begins in the middle of the scale and progresses up or down depending on the responses of the child. Assessment rules were developed to include stopping points within the measure (see Methods chapter for complete administration procedures).

With the SSS, the traditional definition of stimulability is expanded to fit the framework of dynamic assessment. Often clinicians describe stimulability in a binary fashion, where a child is stimulable or not stimulable on a particular sound; however, with the SSS, the clinician is allowed to describe a child's productive abilities along a gradient of 21-points, or in Vygotsky's terms, a 21-point zone of proximal development. The 21-points represent the number of scaffolds available to the child in the production of a phoneme; consequently, the goal is to remove the scaffolds, thereby decreasing the amount needed. The clinician must find the "threshold" of correct speech production where the linguistic environment is the most complex and the cues are the least supportive. As shown in Figure 1.0, improvement on the SSS is represented by a decreasing score: a score of 1 indicates that the child produces the target sound independently in connected speech (i.e., with no instructions). Scores on the SSS can be evaluated for individual phonemes or added together for a composite score across phonemes.

Using the SSS, clinicians may document small changes throughout the entire course of treatment; such changes could potentially show shifts in acquisition as a function of the amount of time spent working on a target. The scores could be generated across sounds prior to the implementation of treatment. With treatment, children need less and less scaffolding until they produce independent productions. Re-administration of the SSS following treatment, or at specified intervals during treatment, may show changes across time.

In order to illustrate the potential of the SSS, a feasibility study was conducted. Glaspey & Stoel-Gammon (2002) documented changes in stimulability during treatment with two preschool boys with severe phonological disorder and compared the results to three phonological measures. Child A was six years, six months of age after completion of seven cycles of treatment (cycles were approximately three months). Treatment targets included velars, stridents, consonant clusters, and liquids. Child S was four years, eight months of age following five cycles of treatment with similar target patterns: velars, stridents, consonant clusters, and liquids. Both children performed within normal limits in all developmental areas with the exception of phonology; in addition, they both passed hearing screenings, although Child S had a history of ear infections. A Cycles treatment program was implemented with four patterns of treatment and two exemplars of each pattern per treatment cycle; in other words, targets changed weekly over a three month period.

The experimental assessments included a dynamic assessment, the Scaffolding Scale of Stimulability, and three static assessments: a single-word probe, *the Assessment* *of Phonological Processes-Revised* (APP-R), and an intelligibility rating scale. The single-word probe included 60 words that were individualized to the children's error patterns; the words were elicited via a picture naming task with flash cards and error patterns were analyzed. The APP-R included 50 words that were elicited using objects; phonological processes were analyzed in accordance with the standard directions. The intelligibility rating scale was a seven-point descriptive judgment made by the investigator of the child's connected speech that ranged descriptively from no noticeable differences from normal to unintelligible. The assessments were administered before and after every three months of treatment.

Results indicated that scores on the SSS changed immediately after initiation of treatment and showed continuous change throughout the entire course of treatment. The pattern of change was, in general, a steady slope across time. Larger changes on the SSS occurred prior to large changes on the other measures, suggesting that the SSS was a more sensitive instrument of phonological change than the *static* assessments. The static assessments remained unchanged for longer periods of time: the probe was the first to change following the SSS, the APP-R changed following the probe, and the rating scale showed virtually no change until the very end of the treatment period for these two children.

In addition to the above assessments, the probe was administered weekly to evaluate changes in relation to treatment targets and additional analyses were conducted. Results indicated that accuracy increased or decreased in relationship to two treatment components: (a) the exemplar being treated and (b) the degree of stimulability of the child's productions. High stimulability scores co-occurred with higher accuracy of single-word productions. A multiple baseline approach showed that scores on the probe remained flat until an exemplar of the pattern was treated. During the week of treatment, or one week after treatment, accuracy of single-word production often increased for the treated pattern. Typically, after treatment for the pattern was removed and another pattern introduced, single-word production accuracy decreased for the previous pattern, and increased for the new pattern until it stabilized. After several cycles of treatment, however, with increased scores on stimulability, the single-word productions reached a high level accuracy and accelerated to a level of complete acquisition. Targets that were not stimulable showed little change. These findings suggest that as stimulability levels increase during treatment, the accuracy of single-word productions also increases.

The feasibility study provided promise regarding the documentation of small increments of change through the administration of the SSS; however, only two participants were included and the reliability and validity of the measures were not rigorously assessed making it difficult to draw firm conclusions. Further research must be conducted to determine whether the SSS can truly inform investigators regarding the nature of phonological change.

Goals and Research Questions

Many questions remained unanswered following the completion of the feasibility study. For example, is the SSS the best measure for documenting phonological change across treatment in children with phonological disorders? Is the SSS a valid and reliable assessment choice when compared to other phonological measures across an increased number of children? The long-term objective for the current study is to trace the course of phonological change during treatment of phonological disorders through the use of the new dynamic assessment, the Scaffolding Scale of Stimulability, in comparison to other measures in order to begin to answer the question: **What is the nature of phonological change?**

One concern regarding the SSS is whether other measures already provide the same information. This issue is addressed by comparing SSS to another dynamic assessment, the Carter-Buck Articulation Assessment to evaluate concurrent validity. The next objective to consider is whether or not the SSS shows change in a more sensitive manner than static assessments and whether predictions can be made regarding a relationship between the measures. The following specific research questions are addressed in the current study:

(1) How does performance change on five different measures of phonology during six months of treatment when each measure is considered independently?

a) The Scaffolding Scale of Stimulability (SSS)

b) The Carter-Buck Articulation Assessment

c) The single-word probe

d) The Hodson Assessment of Phonological Patterns

e) The Percent Consonants in Error in connected speech

(2) How does accuracy of production on the Scaffolding Scale of Stimulability change in relation to the Carter-Buck Articulation Assessment?

(3) How does accuracy of production on the Scaffolding Scale of Stimulability change in relation to the Hodson Assessment of Phonological Patterns, the probe, or the Percent Consonants in Error?

CHAPTER TWO: METHODS

Design

"It is possible to create designs that literally have no name, as would be the case with a particular combination of elements never before used. There is no special vulnerability in using a 'new design' in such a situation as long as the design elements are used appropriately." (Hayes, Barlow, Nelson-Gray, 1999, p.225). The design of the current study is unique and is described as follows.

A time-series analysis of performance and change in performance across time and participants under constant therapeutic intervention was conducted. The time-series study compared five outcome measures of phonological change across eight participants. The results of the five measures were compared within each assessment time period and across assessment time periods. The comparisons across time periods evaluated the change in performance across the eight participants.

Participants

Eight participants between the ages of three years one month and four years six months were selected based on inclusionary and exclusionary criteria. All eight participants met the following inclusionary criteria: (1) were monolingual English speakers; (2) were aged three to five years old; (3) scored in the moderate to profound range for phonological disorder based on the *Hodson Assessment of Phonological Patterns* (Hodson, 2004); (4) exhibited a Percentage of Consonants Correct (Shriberg & Kwiatkowski, 1982) of 75% or less on a connected speech sample, (5) scored above the

20th percentile on receptive language and cognitive abilities based on measures from the *Preschool Language Scale-4* (PLS-4) (Zimmerman, Steiner, & Pond, 2002) and the *Peabody Picture Vocabulary Test* (PPVT) (Dunn & Dunn, 1997); (6) performed adequately on a structural-functional examination adapted from the *Oral and Speech Motor Control Protocol* (Robbins & Klee, 1987) and *Motor Speech Examination* (Strand & McCauley, 1999); (7) exhibited voice and fluency skills within normal limits, and (8) passed a hearing screening. Participants met the following exclusionary criteria: (1) did not exhibit the prime characteristics of developmental apraxia of speech and/or dysarthria; and (2) did not receive additional treatment services from other sources during the study period. Table 2.0 provides a summary of inclusionary and exclusionary criteria.

Table 2.0

Inclusionary assessments and criteria

Category	Criteria
Native language	Monolingual English
Age	3-5 years
Phonological disorder on HAPP-3	Moderate to profound
PCC	<75%
Receptive language on PLS-4	>20 percentile rank
Receptive language on PPVT	>20 percentile rank
Oral-motor exam	Within normal limits
Voice and fluency	Within normal limits
Hearing screening	Pass
Apraxia and dysarthria	Ruled-out based on oral-motor exam
Treatment	One setting only

The inclusionary and exclusionary criteria were determined in order to control variables that may affect outcomes on the experimental measures and were based on the following rationale: (1) Participants from monolingual English speaking homes were selected because bilingual speakers or monolingual speakers of other languages may demonstrate different developmental patterns in their phonological systems. (2) Participants between three and five years of age were selected because the children would more likely fall into similar developmental stages and because large age differences could result in highly variable outcome scores; furthermore, children over five and adults do not exhibit the specific linguistic characteristics of interest for this study. (3) Participants with moderate to profound phonological disorder were selected because this group of children was most likely to show slow progress that is difficult to document across time. Phonological disorder was measured with the severity rating from the *Hodson Assessment of Phonological Patterns* (Hodson, 2004). In addition, Percent Consonants Correct (PCC) (Shriberg & Kwiatkowski, 1982) confirmed the severity of phonological disorder. Participants with PCC scores below 75% were selected; furthermore, the PCC reflects good ecological validity because a connected-speech sample is used. (4) Participants with receptive language and cognitive skills within normal limits were selected because language and cognitive deficits may negatively influence the children's understanding of directions on the experimental measures and result in lower overall scores. The *Preschool Language Scale-4* (Zimmerman, et al., 2002) was administered to assess language abilities and the Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 1997) was administered because scores on the PPVT are highly correlated with scores of intelligence. (5) Participants with adequate oral-motor skills were selected because children with severe motor-deficits such as dysarthria or suspected childhood apraxia of speech would likely demonstrate different developmental patterns and show no change across assessments. (6) Participants with adequate voice and fluency abilities were selected because these disorders may negatively impact articulation testing and skew overall intelligibility scores. The connected speech sample from the PCC was used to make these judgments. (7) Participants with hearing skills within the normal range were selected because a hearing loss may alter speech productions and intelligibility across the assessments. Pure tone hearing screenings were used to determine these criteria.

The exclusionary criteria were based on the following rationale: (1) Participants with suspected childhood apraxia of speech or dysarthria were not selected because children with these characteristics were not likely to respond to treatment in the same manner as other children. (2) Participants who received treatment outside of the study were not selected for participation because the amount and method of treatment was controlled across participants between administrations of the experimental measures.

The results of the initial assessments for the participants are summarized in Table 2.1. Four boys and four girls between 3;1 and 4;6 qualified as participants. Severity ratings on the HAPP-3 include: mild, moderate, severe, or profound; scores on the PCC could range between 0 and 100 percent correct; scores on the PLS-4 include percentile ranks that could range between 0 to 100 with a score of 50 being average; scores on the PPVT also include percentile ranks that could range between 0 to 100 with a score of 50 being average; ratings on the hearing screening could be pass or fail; and scores on the oral-motor assessment could range up to a maximum of 136.

Table 2.1

Results of inclusionary assessments

Participant	Gender	Age	HAPP-3 Rating	PCC	PLS-4 (PR)	PPVT (PR)	Hearing Screen	Oral Motor Score
3	F	3;7	Moderate	65	75	68	Pass	123
4	M	3;5	Severe	59	90	75	Pass	130
5	F	4;6	Profound	37	73	68	Pass	118
7	F	4;1	Profound	45	39	47	Pass	125
8	M	3;1	Severe	49	82	32	Pass	115
9	M	3;6	Moderate	75	55	30	Pass	123
10	M	4;0	Moderate	55	53	21	Pass	129
12	F	4;2	Severe	66	66	25	Pass	120

Recruitment

Participants were recruited throughout the greater Seattle Metropolitan area. Flyers were circulated at the UW Speech and Hearing Clinic, in the local public schools, and at private practices. The investigator first solicited permission from the local school districts and clinics to distribute flyers to speech-language pathologists informing them of the study (see Appendix A). The speech-language pathologists could then determine which families might benefit from the study and forward the information (see Appendix B). The participants were not directly solicited by the investigator; rather, potential participants contacted the investigator after reading the flyer. At the University of Washington Speech and Hearing Clinic (UWSHC), clinical supervisors were given flyers to distribute to families, and a letter to parents describing the study was posted on the UWSHC website. The flyer was also posted on the Washington Speech and Hearing Association listserv and was distributed to all members who subscribe.

The Issaquah School District further cooperated by providing a setting within their district for treatment to occur. Families who volunteered to participate in the study received treatment within their local district in lieu of the treatment provided by the local school district.

Demographics

The gender distribution was expected to include six boys and two girls, based on a study to determine risk factors for speech-delay of unknown origin in 3-year-olds (Campbell, et al., 2003). Campbell and colleagues reported that 70% of the children in their study with speech delay were male and 30% were female with a boy:girl ratio of 2.3:1. They reported that this ratio was consistent with other studies that included ranges from 2:1 to 3:1 in samples of preschool children with speech delay; furthermore, being male was reported as a risk factor for speech delay.

In the current study, gender was not used as part of exclusionary criteria; 50% of the participants were male and 50% were female. Of the families who contacted the investigator, 60% included children who were male and 40% were female.

Participants were sought with linguistically similar speech characteristics, but diverse racial and ethnic characteristics from the greater Seattle metropolitan area. Based on US Census from 2000 (see Table 2.2), the investigator expected to include children by ethnic category: 1 Hispanic or Latino and 7 non-Hispanic or Latino; and by racial categories: 6 children who were white, 1 child who was Asian, and 1 child who was black or African American.

Table 2.2

US Census 2000. Demographics for Seattle Metropolitan Area

	American Indian or Alaskan Native	Asian	Pacific Islander	Black, not of Hispanic origin	Hispanic	White, not of Hispanic origin	Some other race	Two or more races
Male (49.8%)	0.5%	4.5%	0.2%	2.3%	2.8%	39.1%	1.3%	1.9%
Female (50.2%)	0.5%	4.9%	0.2%	2.1%	2.4%	39.5%	1.0%	1.9%

Efforts were made to encourage racially and ethnically diverse individuals into the study. The school setting in the Issaquah School District was one of the most diverse in the area. It was thought that this would encourage participation by improving access for lower income families. However, the demographics did not match the described criteria (see Table 2.3). No children from diverse families chose to participate in the study. All participants were white and appeared to be from middle to upper income homes. One family of Vietnamese descent inquired about the study, but chose not to participate.

Table 2.3

Study demographics for gender, race, and ethnicity.

	American Indian or Alaskan Native	Asian	Pacific Islander	Black, not of Hispanic origin	Hispanic	White, not of Hispanic origin	Some other race	Two or more races
Male (50%)	0%	0%	0%	0%	0%	50%	0%	0%
Female (50%)	0%	0%	0%	0%	0%	50%	0%	0%

Setting

Participants received treatment in one of two settings: the University of Washington Speech and Hearing Clinic, or the Issaquah School District. At the University of Washington Speech and Hearing Clinic, participants were seen in a small therapy room with a one-way mirror so that parents could watch the sessions from the other side of the glass. In the Issaquah School District, participants were seen at Apollo Elementary School during the regular school session and at Tiger Mountain High School during the summer months. A small classroom was made available for the treatment sessions. The setting in the Issaquah School District was temporarily changed to allow for the presence of an on-site administrator during the school's vacation period. Parents were given the option of sitting at the back of the room to observe the sessions. When sessions were missed, the parents were given the option of rescheduling the session at the university setting. In all settings the children worked individually with the investigator and the treatment routine was the same.

Procedure

Initial Contact

Parents contacted the investigator via telephone or email to inquire about their child's participation in the study. The investigator described the study to the parents, summarized criteria for the target participants, and allowed the parents to determine whether they would like to pursue assessment procedures for their child's qualification in the study. When parents indicated interest, an appointment was made for initial testing.

Consent and Assent

At the first appointment, the consent and assent forms were presented and discussed (see Appendixes C, D, E, F, & G). Informed consent procedures, including HIPAA, were followed in accordance with the approved guidelines of the Human Subjects Division at the University of Washington (#01-9930-E/C 04 Stimulability as a measure of phonological change). Approved consent forms were read to the parents and assent forms were read to the participants by the investigator prior to any research activities. The consent and assent forms emphasized that participation in the study was entirely voluntary. The investigator explained the purpose and the procedures of the study; no deception procedures were conducted. The parents and participants indicated their approval by signing the consent and assent forms. In addition to the consent and assent forms for study participation, parents and participants could give their consent or assent for the use of their videos in settings outside of research purposes: in the classroom, at conferences, on the web, or in a text book.

Recording

All data collection sessions were video-recorded for data analysis and completion of reliability measures. The sessions were video-recorded using a Sony Super Steady Shot Digital Handycam-DCR-VX2000 NTSC video recorder. All transcriptions, scores, and judgments were made using the video recordings.

Initial Assessments

Following consent and assent for participation, the initial assessment procedures were implemented. Parents were asked if they agreed to provide the results of any previous testing to reduce participant fatigue. Scores from tests administered by other speech-language pathologists within the previous six months were used to determine inclusion within the study. Scores from experimental assessments were not used and the assessments were re-administered. Throughout the first session, fatigue was monitored and testing was discontinued when the child appeared to have difficulty attending. Two to three sessions were used to complete the initial assessments. Some children could complete the testing within a few sessions; others needed several sessions to complete all of the initial assessments.

A hearing screening was completed to rule out hearing deficits that may impact speech production. A pure-tone screening was administered in both the right and left ears at 20dB at 500, 1000, 2000, 4000hz. The participants at the University of Washington were assessed by an audiologist in a sound-proof booth at the University of Washington Speech and Hearing Clinic. The participants at the Issaquah School District were assessed by speech-language pathologists within the district in an open room.

A structural-functional examination was also completed as part of the initial assessment procedures. The examination procedures were selected from *The Oral and Speech Motor Control Protocol* (Robbins & Klee, 1987) and subsections of the *Motor Speech Examination* (Strand & McCauley, 1999) (see Appendix H). The procedures were used to rule out the presence of significant motor deficits related to suspected childhood apraxia of speech and dysarthria. The assessment generates two scores: a structural score and a functional score. The structural score has 24 points and summarizes the adequacy of the structures used for articulation, whereas the functional score has 112 points and summarizes the movement of the articulators in speech and non-speech tasks. The two scores are added together for one composite score with a total possible of 136. The test includes speech and non-speech tasks, diadokokinesis, measures of duration, whole word productions, and increased length of utterances. Children were required to perform adequately on the structural-functional assessment and demonstrate that they did not exhibit deficits related to suspected childhood apraxia of speech or dysarthria. Scores of the participants are presented in Table 2.1.

The next test to be administered was the *Peabody Picture Vocabulary Test-III* (PPVT) (Dunn & Dunn, 1997). The PPVT was used to assess vocabulary skills that correlate with cognitive abilities. On the PPVT, each participant is asked to point to a picture that best represents the word the child hears. The child is shown a page with four black and white line drawings and the investigator says, "Show me the ____." Once the child responds, then the page is turned and another word is prompted. The administration continues until the child reaches a ceiling of 8 errors in a block of 10 target words. The raw score is converted to a standard score, percentile rank, and age-equivalent. A standard score of 85-115 and a percentile rank between 15 and 85 reflects the average range. The children were required to score above the 20^{th} percentile for participation in the study. The PPVT was administered first because it requires a minimal response by the child and gave the child a chance to become familiar with the administrator and the testing procedures before continuing with the next assessments (see Table 2.1).

After the PPVT, the Preschool Language Scale-4 (PLS-4) (Zimmerman et al., 2002) was administered to determine the participant's receptive and expressive language skills. The PLS-4 uses pictures and objects as prompts to assess the understanding and use of language. Each child is asked to follow directions, point to pictures, and manipulate objects. A standard score of 85-115 and a percentile rank between 15 and 85 reflects the average range. Although the assessment is divided into a receptive language assessment and an expressive language assessment, only percentile rank scores on the receptive portion of the test were used as part of the inclusionary criteria. The participants were required to score above the 20^{th} percentile in order to participate (see Table 2.1).

The *Hodson Assessment of Phonological Patterns* (HAPP-3) (Hodson, 2004) was administered to assess the occurrence of phonological error patterns. During administration of the HAPP-3, children name 50 objects and produce single-word utterances. From the utterances, percentages of occurrence of phonological deviations are calculated. The patterns that are assessed include: syllables, consonant clusters, singletons, liquids, nasals, glides, stridents, velars, anterior non-stridents. The raw scores for each pattern are added together into a summative score that is converted to a standard score, percentile rank, and age-equivalent. In addition, a rating is generated that ranges from mild, moderate, severe, or profound. The inclusionary criteria required that all children be rated as moderate, severe, or profound (see Table 2.1). The scores from this administration of the HAPP-3 were used in the experimental assessment portion of the study.

Percent Consonants Correct (PCC) (Shriberg & Kwiatkowski, 1982) was calculated from a 100-word sample of connected speech. During administration, the investigator gives the participant a wordless book by Mercer Mayer and instructs the child to tell a story. The investigator explains to the child that there are no words in the book and that the words needed to be made up. The instructions are followed by a model of story telling for the first page and the child is encouraged to tell about the subsequent pictures. In order to score the assessment, the child's language productions are audio- and video-recorded for later play back.

Once the sample is collected, the investigator watches the video and glosses 100 intelligible words in accordance with Shriberg and Kwiatkowski (1982). Any words that are completely unintelligible and uncertain are not included in the assessment. In the second step, the investigator phonetically transcribes the intelligible words as an adult would pronounce them; for example, the child's production, [ɪʔə ta] would be glossed as "It's a car," and the adult transcription would be /ɪtsə kar/. Words produced with grammatical errors are also transcribed with the grammatical errors as an adult would pronounce them; for example, if the child says, [i toud əweɪ], the adult gloss would be considered, "he goed away" and the adult transcription would be / hi goud əweɪ /. After the investigator transcribes the child's actual productions, the Percent Consonants Correct is calculated by comparing the consonants in the child's productions to the consonants in the adult form. The number of consonants that the child produces correctly is divided by the number of correct and incorrect consonants and multiplied by 100. Judgments and scores are based on procedures developed by Shriberg and Kwiatkowski (1982). The inclusionary criteria required that all participants score less than 75% consonants correct (see Table 2.1). The scores from this assessment of PCC were used as part of the experimental measures and were converted to Percent Consonants in Error (PCE). In addition to phonological assessment, the language sample generated for PCC was also used to make judgments regarding fluency and voice skills.

Conclusion of Initial Assessments

The scores from the initial assessments were used to determine whether each child met the criteria for participation in the study were reviewed with the participant's parents. All eight children who completed the assessment procedures met the inclusionary criteria and their parents allowed them to participate in the study. Two additional families contacted the investigator regarding the study and attended one assessment session. The two families declined to continue with study; one family could not follow through with two sessions per week because of work conflicts, and the second family did not return for a second scheduled session or respond to the investigators phone calls. The reason for discontinuation was unknown. Following the completion of the initial assessments and enrollment, the experimental measures were administered (see Table 2.4).

Experimental Measures

Five measures of phonological change were used during the study: (1) the Scaffolding Scale of Stimulability (SSS), (2) the Carter-Buck Articulation Assessment (Carter & Buck, 1958); (3) a probe, (4) the *Hodson Assessment of Phonological Patterns* (Hodson, 2004), and (5) Percentage of Consonants in Error (adapted from Shriberg & Kwiatkowski, 1982). Each measure is described below.

Table 2.4

Initial assessments and experimental measures administered.

Initial Assessments	**Experimental Measures**
Hearing Screening	Scaffolding Scale of Stimulability
Structural-functional Examination	Carter-Buck Articulation Assessment
Peabody Picture Vocabulary Test	Probe
Preschool Language Scale-4	Percent Consonants in Error*
Percent Consonants Correct*	Hodson Assessment of Phonological Patterns**
Hodson Assessment of Phonological Patterns**	

Note: *The scores from Percent Consonants Correct collected during the initial assessments were converted to Percent Consonants in Error for administration 1 of the experimental measures. **The scores from the Hodson Assessment of Phonological Patterns collected during the initial assessments were used for administration 1 of the experimental measures.

(1) The Scaffolding Scale of Stimulability (SSS)

Targets Assessed

The Scaffolding Scale of Stimulability (SSS), (Glaspey & Stoel-Gammon, 2001, 2002, 2005) is a dynamic assessment of phonological change. For this study, the SSS was used as an outcome measure to document change across treatment. The measure comprises 48 target items: 5 consonant clusters and 43 singleton consonants (see

Appendix I for list of target words). The consonant clusters are initial /tr/, /pl/, /sp/, /str/ and final /ts/. The consonant singletons are all phonemes of English in initial and final position of words in accordance with the phonotactics of English. The 22 consonants in initial position are: /m, n, p, b, t, d, k, g, f, v, θ, ð, ʃ, tʃ, dʒ, s, z, h, w, j, l, r/. The 21 consonants in final position are: /m, n, ŋ, p, b, t, d, k, g, f, v, θ, ð, ʃ, ʒ, tʃ, dʒ, s, z, l, r/.

Scale

The SSS uses dynamic assessment to rate the production of each target sound on a scale from 1 to 21 (see Figure 2.0) with 21 possible scaffolds that support the child's production of a target. A score of 1 (being best) represents the least amount of support needed by the child and indicates high stimulability. A score of 21 means the child needs the highest level of support and indicates no stimulability. A gradation of stimulability is scored between 1 and 21.

CUES ENVIRONMENTS	***Level 0***	***Level 1*** **Instructions**	***Level 2*** **Instructions, verbal model**	***Level 3*** **Instructions, verbal model, segmented, simultaneous production, tactile cues**
Not Stimulable				21
Isolation				20
Word	16	17	18	19
Carrier Phrase	12	13	14	15
Novel Phrase	9	10	11	
One-target sentence	6	7	8	
Two-target sentence	3	4	5	
Connected Speech	1	2		

Figure 2.0. The Scaffolding Scale of Stimulability: Cues and Environments.

The scale is presented on a grid with manipulation of two factors: the *environment* and the *cues* (see Figure 2.0). The *environment* refers to the linguistic context in which a phoneme is produced, i.e., isolation, words, sentences. The investigator changes the environment in response to each of the child's productions. If the child's production is correct, then the environment is made more complex and support is removed; conversely, if the child's production is in error, then the environment is made less complex and support is increased.

Environments

The SSS has seven environments for eliciting target sounds or patterns; they are described in sequence from most supportive to least supportive: isolation, words, sentence/carrier phrase, sentence/novel phrase, sentence/one-target, sentence/two-target, and connected speech.

(1) *Isolation*. In the isolation environment the target is elicited alone or in a syllable with /ʌ/ for a CV or VC production. The articulatory features determine whether or not the target is combined with a vowel. Consonants that can be produced in a continuous manner, such as /s/ or /z/, do not need to be produced with the vowel; however, stop sounds such as /p/ or /b/ require the addition of the /ʌ / and are elicited as /pʌ/or /bʌ/. In the isolation environment, the investigator gives a verbal model of the sound with or without placement instructions. The instructions are individualized to the child's errors.

(2) *Word.* In the word environment, the target is elicited in a CVC word with the exception of targets for syllables and clusters, which are more complex. The word

environment is more difficult because of the additional non-target phoneme in the word which may cause interference in production. On the SSS, dynamic assessment of each target begins at the word level. The investigator shows the child a picture and asks the child, "What's this?" For example with the target initial /m/, the child is shown a picture of a *mop* and asked to name *mop*.

The next four environments include variations of sentence productions: carrier phrase, novel phrase, one-target sentence, or a two-target sentence. The sentences progress from most to least supported. After the child correctly produces the target at the word level, the child is instructed, "Tell me about this" or "Tell me about these."

(3) *Sentence: Carrier Phrase*. The first sentence production is a carrier phrase, a repetitive phrase that starts with, "I can . . ." or "It's a . . ." The context is determined by whether the target word is a noun or verb; for example, the target noun "mop" is produced as "It's a *mop*," but the target verb "hug" is produced as, "I can *hug*."

(4) *Sentence: Novel Phrase.* The novel phrase is a three-word sentence with the target consonant in a word at the end of the sentence. A novel phrase is different from a carrier phrase because the words within novel phrase change from one target to the next; for example, the novel phrase for the target word-initial /m/ is, "Use the *mop*," and for the target word-final /n/ is, "Eat the *bean*." Changing the composition of the phrases makes this environment slightly more difficult than the repetitive carrier phrase.

(5) *Sentence: One-target*. The fifth environment is a one-target sentence with the target in a word in the middle of the sentence. The sentence is made up of four words with at least one word on either side of the target. The word with the target is either the

second or third word in the sentence, but does not occur at the beginning or end of the sentence; for example, the one-target sentence for word-initial /m/ might be, "The *mop* is dirty." Embedding the target word within the middle of the sentence makes the production slightly more difficult.

(6) *Sentence: Two-target.* The sixth environment is a two-target sentence with at least one of the targets occurring in the middle of the sentence. The sentence is made up of four to five words and the words with target phonemes can occur anywhere in the sentence. The second target word is selected with the same considerations as the first described above. The investigator shows the child two pictures and says, "Make up something about these." For example, a two-target sentence for initial /m/ might be, "The *mop* is by the *map*."

(7) *Connected Speech.* The last and most complex environment involves the spontaneous production of the target consonant in at least two different words while describing a picture scene in connected speech. The child's utterances must include at least two word combinations; single-word productions such as labeling do not count and are not scored as correct. Each picture scene has several opportunities for productions, although some scenes have more opportunities than others. The child is shown a scene and instructed, "Tell me a story about this picture." The investigator may also point to the target words to indicate some of the most important items to discuss. For example with initial /k/, the child might say, "What a funny picture! The *kitty* is furry. A *king* won't leave the *cowboy*." In this example, the child receives credit for *kitty, king,* and *cowboy.* Correct production of only one target in connected speech does not count as

successful production of this environment; for example, "What a funny picture! The *kitty* is furry. A *ting* won't leave the *towboy*." In addition, if the child did not produce any opportunities for the target, the child is not given credit; however, this production is noted with a (+) because it may have been possible for the child to continue.

Target Word and Sentence Selection

On the SSS, specific words and sentences were selected to assess the above targets. The investigator chose words, in part, that could be easily pictured and identified by preschool-age children. For consonants, several factors within the phonetic environment were considered to increase the likelihood that a child would produce each target correctly (Bankson & Bernthal, 2004, p. 253). First, the syllable shape was controlled and consonants were elicited in CVC words rather than in words with multiple syllables, clusters, or blends. In addition, vowel and consonant combinations were selected for each target that would likely be most facilitative; for example, to increase the correct production of the velars /k/ and /g/, words with back vowels were selected. These vowels would most likely influence a backward movement of the tongue for the velar production. Non-target consonants within each word were selected that would not likely cause interference with the target sound; for example, with velar targets /k/ and /g/, the phonemes /t/ and /d/ were avoided as non-target consonants because they are the most common error substitutions for velars. Based on assimilation processes, the child may have more difficulty producing /g/ correctly in the word "goat" than in "gob." Finally, consonants from the same sound class were also avoided in later developing targets, and early developing sounds were most often selected as the non-target consonants. Overall, the phonetic environment was carefully considered for optimal stimulability.

As with target words, the sentence environments were carefully developed for each of the consonants and patterns with stimulability in mind. When possible the investigator chose words for the elicitation sentences following many of the above principles considered in the selection of individual words. For example, when the target word was a strident, then stridents were avoided in the other words to reduce any interference in the productions from coarticulation.

Cue Levels

The *cues* include instructions and modeling techniques that are used to assist the child with the correct production of consonants and patterns across the environments (see Figure 2.0). The cue levels are the second factor manipulated on the SSS grid. The investigator begins each elicitation with spontaneous production of the targets without cues. When the child produces error responses at an environment, the investigator modifies the elicitation by adding cues. The investigator begins by giving as little support as possible and gradually adds support as needed. The SSS includes four contexts of cueing which are labeled from zero to three. In the following paragraphs, the cues are described in sequence from least amount of support to most amount of support. The cues are cumulative and additive; once a cue is used, it may be used again as needed in the next level. Cues cannot be removed once they are added.

Cue Level 0: Spontaneous Picture. The first context for cueing is labeled with "0" because no cues are given other than a picture elicitation prompt. The clinician shows the child a picture that represents a word and asks the child to name it; the child is not given any other assistance. A picture is always present, and as environments change, a second picture is added or the stimulus is changed to a picture scene. In some cases, the presentation of the target picture may facilitate correct productions, especially when a child has been in speech therapy for some time and is familiar with the act of practice and concentration when a picture is presented. If the child produces the target correctly, no cues are added; however, if the child produces the target in error, the child is given cueing at Cue Level 1.

Cue Level 1: Instruction. When a child produces a target in error at the spontaneous picture level (Cue Level 0), support is given to the child with instructions for correct production. The addition of cueing that a child receives first occurs in the form of an instruction, the least supportive form of cueing. The investigator gives an instruction that might help the child produce the target accurately. The nature of the instruction is dependent on the child's previous error response. The instruction may include feedback about the error response and then a suggestion about how to alter the articulators to improve the production. For example, "Not quite, pull your tongue back. Try again." If the child responds again in error, more cueing is added from Cue Level 2.

Cue Level 2: Verbal Model. In the next level of cueing (Cue Level 2), the investigator gives a verbal model of the target production with the previous instructions and gradually adds support. The investigator models the target in its environment for the child to repeat immediately. The investigator also cues the child to watch; for example, the investigator says, "Look at me. Say, mop." The child says, "Mop." If the child responds again in error, such as, "op," then cueing is added from Cue Level 3.

Cue Level 3: Segmentation, Prolongation, Simultaneous Production, Tactile. Cue Level 3, the most supportive cue level, includes multiple choices for additional cues depending on the child's error and the characteristics of the target. The cues within this level are considered hierarchical in presentation, although they are scored the same. One type of cue that may be selected is segmentation with which the investigator separates the target consonant from the rest of the word; for example, in the word "mop," the target "m" is produced in isolation, followed by the rest of the word, "op." The child is asked to repeat after the investigator: "m" "op." The second type of cue in Cue Level 3 is prolongation. Using the prolongation cue, the investigator emphasizes the target sound and stretches the length of the sound within the word: "m-m-mop." The child is asked to repeat after the investigator. At Cue Level 3, the investigator begins with either segmentation or prolongation to emphasize the target consonant; most often stops are segmented and fricatives are prolonged. If this cue is unsuccessful, the investigator tries another cue from Cue Level 3: simultaneous production or tactile cue.

The third type of cue in Cue Level 3 is simultaneous production, which is effective with continuant targets such as /f/ in the following example. With simultaneous production, the investigator begins by producing the target sound in a prolonged manner and then instructs the child to accompany: "Join in with me, f-f-f-f." Once the investigator and the child are producing the sound at the same time, the investigator says the rest of the word while the child is still saying the initial consonant, and encourages the child to finish the word: the investigator says, "in," while the child says, "f-f-f-f," and finally, the child says, "in." If the production of "fin" is still in error, the final cue type, tactile cue, from Cue Level 3 is added.

The fourth type of cue in Cue Level 3 is a tactile cue. Tactile cues are any physical manipulation of the articulators, such as touching the lips on the sides to encourage retraction, or a tactile representation of the sound, such as a tap on the hand for a stop or a slide down the arm for a fricative. All cues used within Cue Level 3 are scored the same no matter which cues are used by the investigator.

Administration Procedures: Phonemic Inventory

The first step of the SSS is to document the child's phonemic inventory from a connected speech sample. For this study, the speech sample that was collected for PCC was also used to evaluate the phonemic inventory. Phonemes are evaluated by initial and final position, and a phoneme is considered part of the child's inventory when it occurs at least two times per position in the connected speech sample and only if it is used correctly (adapted from Stoel-Gammon, 1985). Phonemes that occur in the inventory are not evaluated on the SSS. The purpose of documenting the phonemic inventory is to reduce the number of targets administered to the child and ultimately reduce assessment fatigue by eliminating targets the child produces correctly. Each item on the SSS takes several minutes to assess and the entire assessment could take up to an hour to administer. The phonemic inventory procedure reduces the administration time for targets that do not need to be assessed. All phonemes in the inventory as described above are given a score of "1" on the SSS. Phonemes that do not have an opportunity to occur in the connected speech sample are assessed in full on the SSS; furthermore, any consonants that are in question, perhaps because of a distortion or microphone noise, are also assessed on the SSS.

Based on data from the phonemic inventory, the administration of the SSS begins with the earlier developing sounds that are missing from the inventory or had no opportunity to be produced, and progresses through the later developing sounds. Once the child's inventory is established, subsequent administrations of the SSS include only phonemes and patterns in error. The phonemic inventory is analyzed only during the first assessment.

Administration Procedures: Sequencing Rules

The next step in the administration of the SSS differs slightly from a traditional dynamic assessment approach. In dynamic assessment, the assessment process typically begins in the least supported environment and then support is gradually and systematically added. The SSS is slightly different because the goal of the first step is to find the least supportive environment. The SSS includes several spontaneous environments in which phonemes are elicited from the participant: word, sentences, two-target sentence, and connected speech. The SSS begins with the most supportive environment, word level, but no cues are provided. If the child is successful, a more complex environment is tested, still without providing cues. Once the child produces an error, the cues are systematically added. The dynamic assessment of the SSS is on two separate but interconnected levels. The environments are changed from most-supported to least-supported while at the same time cues are added to change the assessment from least-supported to most-supported.

CUES / ENVIRONMENTS	*Level 0*	*Level 1* Instructions	*Level 2* Instructions, verbal model	*Level 3* Instructions, verbal model, segmented, simultaneous production, tactile cues
Not Stimulable				21
Isolation				20
Word	16	~~17~~	~~18~~	19
Carrier Phrase	12	13	14	15
Novel Phrase	9	10	11	
One-target sentence	6	7	8	
Two-target sentence	3	4	5	
Connected Speech	1	2		

Figure: 2.1. SSS: Error Response. The investigator increases support to the child by moving to the *right* and *up* on the SSS grid.

The dynamic assessment for each target always begins at the word level without any cues (Cue Level 0) to assess spontaneous productions. The investigator shows the child a picture and says, "What's this?" The child names the target word. Once the child responds, the investigator makes an immediate judgment about the target production. If the child makes an *error response*, the investigator moves to the *right* on the hierarchy grid (to Cue Level 1) and adds instruction cues from Cue Level 1 (see Figure 2.1). The investigator gives an instruction regarding the participant's production, which may include feedback such as, "Pull your tongue back, try again." If the child continues to make errors, the investigator again moves to the *right* along the hierarchy and adds verbal models, Cue Level 2 and Cue Level 3, and with continued errors, moves *up* to the Isolation environment. Once a cue is given, the investigator may not move back to the left on the hierarchy.

CUES / ENVIRONMENTS	*Level 0*	*Level 1* Instructions	*Level 2* Instructions, verbal model	*Level 3* Instructions, verbal model, segmented, simultaneous production, tactile cues
Not Stimulable				21
Isolation				20
Word	(16)	17	18	19
Carrier Phrase	12	13	14	15
Novel Phrase	9	10	11	
One-target sentence	6	7	8	
Two-target sentence	3	4	5	
Connected Speech	1	2		

Figure: 2.2. SSS: Correct Response. The investigator decreases support to the child by moving *down* on the SSS grid.

In contrast, if the child makes a *correct response*, the investigator changes the environment, removes support, and moves *down* the hierarchy to a more complex environment (see Figure 2.2). The investigator prompts the child, "Tell me about that," and elicits a sentence production. If the child produces a sentence correctly, the investigator again moves *down* to the two-target sentence environment. At Cue Level 0, the child is not required to produce every sentence environment. With continued success, the investigator prompts the child to tell a story about a picture scene. If the child produces two targets correctly in connected speech, testing for the target is complete.

The investigator ends the assessment when a threshold is reached. Ultimately, the child is scored for each target based on the least supportive environment and cue level needed for a correct production.

Sample Scoring

The following section provides a sample administration of cueing and environmental manipulation for eliciting a target. The target selected for illustration is word-initial /f/: first with a correct response, and second with an error response. In Figure 2.3, the investigator's antecedent cues are described and sample responses from the child are given for a correct response on the first stimulus presentation. Following Figure 2.3, Figure 2.4 shows the corresponding steps along the SSS hierarchy. Figures 2.5 and 2.6 illustrate the same target elicitation, but for an error response.

Step	Investigator's antecedent	Child's production	Plan for next step
1	Show picture of *fin* and say: "What's this?"	[fɪn] Correct.	Reduce support by changing the environmental complexity.
2	"Great sounds!" Show a picture of *fin* and say: "Tell me about that."	[hi haɪdə fɪn hɪr] Correct.	Reduce support by changing the environmental complexity.
3	"You said all the sounds!" Show a picture of a *fin* and a *phone* and say: "Tell me about these."	[hi poun wɪʔ ə fɪn] Error.	Increase support by adding instruction cues from Level 1 at the same environmental complexity
4	Say, "Nice try, but remember to make the sound with bunny teeth on this word too. Say that again" (point to phone).	[hi poun wɪʔ ə fɪn] Error.	Increase support by adding cues at the same environmental complexity.
5	Say, "You're trying very hard. Look at me. Say, a *phone* won't need a *fin*.	[ə foun wount nid ə fɪn] Correct.	End of manipulation. The child Scores a **5.**

Figure 2.3. The SSS. Sample administration sequence for word-initial /f/ with a correct response in the first presentation.

CUES / ENVIRONMENTS	*Level 0*	*Level 1* Instructions	*Level 2* Instructions, verbal model	*Level 3* Instructions, verbal model, segmented, simultaneous production, tactile cues
Not Stimulable				21
Isolation				20
Word	(16) 1	17	18	19
Carrier Phrase	12	13	14	15
Novel Phrase	9	10	11	
One-target sentence	(6) 2	7	8	
Two-target sentence	(3) 3 →	(4) 4 →	(5) 5	
Connected Speech	1	2	Stop	

Figure 2.4. The SSS. The above sequence for word-initial /f/ with a correct response in the first presentation as shown in Figure 2.3.

Step	Investigator's antecedent	Child's production	Plan for next step
1	Show picture of *fin* and say: "What's this?"	[pɪn] Error.	Increase support by adding cues from Cue Level 1 at the same environmental complexity.
2	Say, "Nice try, Make the sound with bunny teeth."	[pɪn] Error.	Increase support by adding cues from Cue Level 2 at the same environmental complexity.
3	Say, "Not quite. Watch me. Say, *fin*."	[pɪn] Error.	Increase support by adding cues from Cue Level 3 at the same environmental complexity
4	Say, "Try again. Watch me. Say, ffffffin.	[ffffɪn] Correct.	Remove support by changes the environmental complexity.
5	"You said the sound! Now watch me and say, It's a ffffffin."	[ɪtsə pɪn] Error.	End of manipulation. The child Scores a **19.**

Figure 2.5. The SSS. Sample administration sequence for word-initial /f/ with a error response in the first presentation.

CUES / ENVIRONMENTS	*Level 0*	*Level 1* Instructions	*Level 2* Instructions, verbal model	*Level 3* Instructions, verbal model, segmented, simultaneous production, tactile cues
Not Stimulable				21
Isolation				20
Word	16 [1] →	17 [2] →	18 [3] →	19 [4]
Carrier Phrase	12	13	14	15 ← ● [5]
Novel Phrase	9	10	11	Stop
One-target sentence	6	7	8	
Two-target sentence	3	4	5	
Connected Speech	1	2		

Figure 2.6. The SSS. The above sequence for word-initial /f/ with an error response in the first presentation as shown in Figure 2.5.

Reinforcement

During the dynamic assessment children are given reinforcement in several ways. One type of reinforcement occurs when the investigator gives verbal praise and verbal feedback regarding speech productions. Children are given praise for correct productions and feedback about error productions. The verbal reinforcement is administered with each step of the dynamic assessment. In addition to the verbal responses, the investigator gives the children token reinforcement for attending to the task and responding to the elicitations. The amount of token reinforcement is variable throughout the sessions depending on each child's interest level and ability to stay on task. Over the course of the session, the token reinforcement often is increased as the children become fatigued.

Simple activities are used for token reinforcement. The children receive: a chip to put on a picture full of circles, a frog to put on a lily pad, a chip to feed a frog in its mouth, a bone to feed a dog, a piece of cheese to put on a magnetic board with a mouse, a wand to blow bubbles, a dauber to put a spot of paint on a flower cutout, a sticker to put on a chart, a turn at choosing a duck to find a matching color, or a turn to manipulate an interactive toy. Children are given approximately five to eight targets for each kind of reinforcement activity before it is changed to another activity; rapidly switching the activities facilitates attention for up to an hour.

The length of administration of the SSS varies depending on the stimulability levels, age, and attention abiltities of each child. Typically, the assessment takes up to an hour to administer. Children with low stimulability complete the entire measure in less than an hour and children with higher stimulability take longer because of the connected speech task with the picture scenes. In general, younger children take longer for administration because more targets need to be assessed and their attention abilities require more reinforcement activities and breaks.

Alterations

During assessment with the SSS, sometimes the child does not respond in the expected manner and the investigator is required to alter the prescribed procedures. Ideally, the child responds with the correct target production in the correct word. Occasionally, problems arise during the spontaneous portion of the test. Sometimes, the child does not know the target word and does not correctly name it. In this situation, the investigator first attempts to elicit the word by using cloze procedures; however, if this approach is not successful, the investigator gives a delayed model. The investigator names the picture, says a filler sentence, and prompts the child to say the target again, as in, "Fin. Fish use it to swim. You say it." A delayed model is not considered to be a change of environment along the cueing hierarchy and it is only used at the word level.

A second issue arises when the child names the target correctly at the word level, but does not spontaneously create a sentence when prompted. Often younger children may not generate their own sentences without a verbal model. If the child does not respond to the prompt, "Tell me about that," at the sentence level, the investigator "jumps over" the instructions from Cue Level 1 and moves directly to the verbal model from Cue Level 2. Once the child fails to respond spontaneously across three consecutive targets, on the fourth target the investigator discontinues spontaneous sentence level elicitations and continues testing at the verbal model from Cue Level 2.

The child's attention levels during dynamic assessment also impact the administration of the test. When a child appears inattentive or unresponsive, the investigator is allowed to repeat a cue up to three times at Cue Level 1 and Cue Level 2, and up to four times at Cue Level 3. The investigator judges when the child appeares inattentive and repeats the cues; most often, the cues are only repeated two times.

Scoring the composite

On the SSS, the targets are given individual scores that are added together to create a composite value. With 48 consonant targets (22 initial and 21 final), the participants could score between 48 (stimulable for all phonemes) and 1008 (not stimulable for any phonemes). The scores reflect the number of cues and environmental supports that the child needs to produce all of the sounds.

(2) The Carter-Buck Articulation Assessment

The Carter-Buck Articulation Assessment (Carter & Buck, 1958) is the second experimental measure administered over the course of the study. The Carter-Buck is considered a dynamic assessment because targets are assessed in two environments—words and syllables, and across two cue levels—spontaneous and imitated. Thirteen sounds are assessed in initial, medial, and final position across 112 target words and 117 syllables. The sounds included: /s, z, l, f, v, k, g, r, ʃ, tʃ, dʒ, θ, ð/.

The Carter-Buck comprises three subtests. On Subtest 1, targets are elicited in spontaneous single-word productions of 112 words. The investigator shows each participant pictures and elicits the single-word productions by saying, "I'll show you some pictures and I want you to name them for me. What is this?" The child responds by naming the picture. The investigator is not allowed to model the target words for the child. When the child does not know the target word, the investigator may give clues in the form of cloze procedures or descriptions that might help elicit the target; however, if the child still cannot name the target words, the elicitation ends and the target word is skipped. Once all of the target words are elicited, the investigator ends the session for the day.

Subtests 2 and 3 are administered during the subsequent session. On Subtest 2 the same 112 target words are elicited as on Subtest 1; however, during this elicitation task the investigator names the target word (i.e., provides a verbal model) and the child imitates the word. Immediately following Subtest 2, Subtest 3 is administered. Subtest 3 involves elicitations of 117 syllables for the same 13 target sounds in initial, medial, and final position of words. The investigator models syllables and the child imitates. The child produces each target sound in each position with the three different vowel contexts: /i/, /æ/, /a/; for example, the target /f/ included the syllables /fi/, /fæ/, /fa/, /ifi/, /æfæ/, /afa/, /if/, /æf/, /af/. In the present study, when both subtests are completed, the session is discontinued for the day.

During the assessment, the investigator gives reinforcement to the child as needed. The child is periodically praised for good memory, naming, or attention, but is not given reinforcement for correct productions or feedback regarding accuracy. In addition to verbal reinforcement, the investigator gives the child token reinforcers after completion of a block of target words. The token reinforcers include bubbles, pieces to art projects, stickers, or other items of interest similar to those listed under the SSS administration.

Scoring

Several scores are generated from the Carter-Buck including percentages of errors for comparison within the current study and original calculation procedures from Carter and Buck (1958). First, each subtest was analyzed separately. On each subtest, the number of error consonants were added, divided by the total number of consonants, and multiplied by 100. A percentage of error productions was calculated for each subtest.

Second, the number of error consonants from all three subtests were added together, divided by the total number of combined consonants, and multiplied by 100 to generate a total percentage of error across all three subtests. Third, a percentage of correction was calculated in accordance with Carter and Buck (1958) between different subtests and time interval combinations. The lower raw error score was divided by the higher raw score, multiplied by 100, and then subtracted from 100. The percentage of correction was calculated between Subtest 1 and 2, Subtests 1 and 3 across three time intervals. A percentage of correction was also calculated between Subtest 1: Time 1, Time 2, and Time 3; Subtest 2: Time 1, Time 2, and Time 3; and Subtest 3: Time 1, Time 2, Time 3.

(3) Probe

The third experimental assessment is a probe of target sounds and patterns in single-word productions elicited with pictures. The probe was designed by the investigator and individualized to each child. The full version of the probe comprises 72 single-words with three opportunities for /k, g, f, v, s, z, ʃ, tʃ, dʒ, l, r/ to be produced in word-initial and word-final position; three phonemes, /θ/, /ð/, and /ʒ/ have fewer than three opportunities. Cluster targets include: s-clusters in initial and final position, three-consonant s-clusters in initial position, l-clusters in initial position, r-clusters in initial position, and glide clusters in initial position. Overall, more than one target occurs in each word and the words are CVC with the exception of the cluster targets. The words on the probe are not used in treatment and do not occur on any of the other assessments.

Pictures are displayed one at a time on a laptop computer screen using a Powerpoint presentation. The presentation was created with clipart from The Print Shop Deluxe Suite version 2.0. The investigator begins the assessment by telling the child, "I'm going to show you some pictures and I want you to name them for me. What's this?" The investigator or the child clicks the cursor to change the picture after each one is named; the presentation is not timed and is completed between three to ten minutes. When the child is uncertain of the name of a picture, the investigator names the picture for the child, says a filler sentence, and prompts the child to say the word again.

Scoring

In order to score the probe, target phonemes are transcribed and transferred to a data sheet which is divided into target consonants. The targets are scored as correct or incorrect with distortions counted in error. The error consonants are summed to generate a composite score that was used for comparison with the other measures.

(4) The Hodson Assessment of Phonological Patterns

The fourth experimental assessment is the *Hodson Assessment of Phonological Patterns* (HAPP-3) (Hodson, 2004). The HAPP-3 was used during the initial assessment phase and the score from the first administration is carried forward to the experimental portion of the study; it is not administered again prior to treatment. The HAPP-3 is scored as described in the initial assessment section of this document.

(5) Percent Consonants in Error

The fifth experimental assessment was the Percent Consonants in Error (PCE). The procedures for this assessment are the same as the calculation of Percent Consonants Correct described earlier except that the inverse, percent consonants in error, is calculated. The PCE is calculated instead of the PCC so that the data across assessments may be compared. In this way, all assessments showed progress visually with a downward slope. The speech sample from PCC during the initial assessment phase was used, converted to a PCE score, and carried forward to the experimental portion of the study for the first administration.

Summary of Experimental Measures and Schedule of Administration

Five experimental measures were selected to document phonological change: The Scaffolding Scale of Stimulability, The Carter-Buck Articulation Assessment, the probe, *The Hodson Assessment of Phonological Patterns,* and Percent Consonants in Error. The five assessments were administered at three time intervals: (1) prior to treatment, (2) after three months of treatment and completion of cycle one, (3) after six months of treatment and completion of cycle two. The treatment was controlled in the sense that all children received the same type of treatment; however, the treatment was not experimentally manipulated.

Treatment

During the study period, Cycles treatment was administered to all of the participants in the study. Two cycles were completed across six months with four patterns targeted during each cycle. The children began treatment at different times; however, they received the same number of sessions. The patterns were different for each participant and were individualized to the child's errors. The percentages of occurrence scores from the HAPP-3 were used to select the treatment patterns; priority was given to those occurring greater than 40% of the time. Each pattern was targeted for four sessions, and within that time, two exemplars of each pattern were treated; for example, the error pattern of velar fronting resulted in treatment of /k/ in word-final position for two sessions and treatment of /g/ in word-initial position for two sessions. Overall, the cycle lasted for 16 sessions with a different target every two sessions (see Table 2.5 for a sample cycle).

Table 2.5

Sample of one treatment cycle with four patterns and two exemplars of each pattern across 16 sessions.

	1-2*	3-4	5-6	7-8	9-10	11-12	13-14	15-16
Pattern 1: Velar	-k	g-						
Pattern 2: Stridents			-f	v-				
Pattern 3: Liquids					l-	r-		
Pattern 4: Clusters							sp-	-ts

*session number

The investigator administered the treatment to all of the participants and used practices common to Cycles method of phonological treatment (Hodson & Paden, 1991). Each of the sessions was 50-minutes in length with approximately two sessions scheduled per week. Each session began and ended with two minutes of auditory bombardment; the children listened under amplification as the investigator read a list of words with the session's target. Next, the participants made four to six picture cards that they used during the following activities: story time, game, and art. Naturalistic play activities, using objects and manipulatives as reinforcement, were combined with drill and practice game-activities, using picture cards to elicit correct productions. The children were given their picture cards and projects for practice at home. Parents were instructed to practice with the cards, but no data were collected to document how much practice was actually completed at home.

In general, the sessions were scheduled with two per week (see Figure 2.7); however, sometimes the children missed sessions due to illness or family issues. When children missed sessions, then the schedule of targets was paused and continued at the next session. All of the children received the exact same number of treatment sessions, except for the two children who discontinued early. The time it took to complete all of the sessions varied across the children.

		Cycle 1					**Cycle 2**				
Phase	A	B				A	B				A
Pattern	BL	1				WD	1				WD
	BL		2			WD		2			WD
	BL			3		WD			3		WD
	BL				4	WD				4	WD
Measures	Probe, SSS, C-B, HAPP-3 PCE					Probe, SSS, C-B, HAPP-3 PCE					Probe, SSS, C-B, HAPP- 3 PCE
Sessions	5	4	4	4	4	5	4	4	4	4	5

Figure 2.7. Sample cycle and assessment schedule.

Behavioral strategies were also implemented during the treatment sessions to increase participation. A visual schedule of the day's activities was used and was posted across the top of the treatment table where the children could follow along. In addition, children were given praise for participation and extrinsic motivators in the form of stickers, turns at games, and choices of some activities. They were given choices as often as possible regarding the activities so that they felt ownership in the sessions. At the end of the sessions, the target and strategies for production were reviewed with the parent and the child for reinforcement at home.

Reliability

Initial assessments and experimental measures were analyzed for both *procedural* reliability and *measurement* reliability. Procedural reliability assesses whether the procedures were administered consistently as planned (Billingsley, White, & Munson, 1980), whereas measurement reliability assesses whether the measures were consistently scored (Bain & Olswang, 1995). All assessments were video taped and viewed by the coders to complete the analysis of reliability. An individual unfamiliar with the study procedures selected data sets for review to eliminate potential bias from the investigator. The data for review was selected in a counterbalanced method so that all of the children who participated in the study and all three intervals of time were represented (see Table 2.6).

Table 2.6

Participants and assessments used for procedural reliability.

	Time 1	Time 2	Time 3
Initial assessments	5F (PLS-IV)	8M (PPVT)	3F (SFE)
SSS	9M	10M	7F
Carter-Buck	4M	7F	12F
Probe	8M	9M	4M
HAPP-3	10M	3F	5F
PCE	12F	5F	3F

Procedural Reliability

Initial Assessments

Procedural reliability was completed on the initial assessments by an expert coder who was doctoral student from the University of Washington and was currently finishing her clinical fellowship year. In this section, three assessments are discussed and referred to as the initial assessments: *The Peabody Picture Vocabulary Test (PPVT), The Preschool Language Scale-4 (PLS-4)*, and the structural-functional examination. The *Hodson Assessment of Phonological Patterns* and the Percent Consonants in Error, which were considered both initial assessments and experimental measures, are described below in the experimental measures section. One sample of each of the initial assessments was assessed for procedural reliability, which equaled 13% of the overall data. The coder was given a video, instructions for viewing a video, and an observation checklist (see Appendix J). The coder tallied the number of procedures that the investigator completed. From the checklist, a percentage was calculated across all three assessments for the overall accuracy of procedural reliability for the initial assessments.

Scaffolding Scale of Stimulability

Procedural reliability for the SSS was calculated by an expert coder who was highly familiar with the theories of dynamic assessment and was a certified speech-language pathologist with a Ph.D. The coder was trained in the administration of the SSS during a two-hour training session. The investigator explained the scoring procedures to the coder and both practiced matching scores as they watched video clips together (see Appendix K for data sheet and Appendix L for coding instructions). None of the video clips observed during the training sessions were the same as those used to calculate reliability. Once the coder was familiarized with the scoring system, she was given a video tape of 15 targets to score independently; 88% agreement was achieved. The coder and the investigator met for a second session to compare scores and to discuss discrepancies across the 15 targets. Following the session, the coder was given a second set of 10 targets for independent practice; 89% agreement was achieved. Discrepancies between the coder and the investigator most often included decisions regarding whether a verbal model had actually been prolonged.

Once the training was complete, the coder independently scored the procedural reliability for three SSS sessions including participants 9M: Time 1, 10M: Time 2, and 12F: Time 3. Overall, 14% of the total data was observed. The final score for procedural reliability was calculated by adding the number of correct steps that the investigator followed, divided by the total number of steps, and multiplied by 100 to generate a percentage of procedural accuracy. Only the scores documented by the coder, who was considered unbiased, were used to calculate the final percentage of procedural reliability.

Carter-Buck, Probe, HAPP-3, and PCE

Procedural reliability for the Carter-Buck, probe, HAPP-3, and PCE was calculated by two undergraduate students who were unfamiliar with the assessments. During a 30-minute session, the coders were trained to observe the steps of administration for each assessment. The assessments were divided between the two coders and each coder was given a checklist of procedures and a video tape of three complete sessions for their assigned measures (see Appendixes M, N, O, and P for checklists). The coders viewed the video tapes and tallied the number of procedures that were completed by the investigator. An overall percentage was calculated for three of the 22 administrations of each measure, which included 14% of the total data. Only the scores documented by the coders, who were considered unbiased, were used to calculate the final percentage of procedural reliability.

Measurement Reliability

Measurement reliability was conducted across all initial assessments and experimental measures to determine whether the tests had been scored appropriately. The following section describes the manner in which the tests were rescored for reliability. The data that were extracted for reliability were counterbalanced across children and time periods and were selected by an unbiased individual not familiar with the study (see Table 2.7).

Table 2.7

Data samples from each assessment that was used to assess measurement reliability.

	Time 1	**Time 2**	**Time 3**
Initial Assessments	5F (PLS-IV)	8M (PPVT)	3F (SFE)
SSS	8 tokens	8 tokens	6 tokens
Carter-Buck	3F	3F	3F
Probe	12F	12F	12F
HAPP-3	5F	5F	5F
PCE	7F	7F	7F

Initial Assessments

Measurement reliability for the initial assessments was completed by the same doctoral student who calculated the procedural reliability. The PPVT, PLS-4, and the structural-functional examination (SFE) were grouped together as the initial assessments in this analysis. One assessment out of the total eight was re-scored for measurement reliability, which included 13% of the total data. The coder was given protocols from each of the assessments that were filled out, but not scored. The coder was instructed to score the assessment. The coder's scores were compared to the investigator's scores for overall percentage of agreement.

Scaffolding Scale of Stimulability

Measurement reliability for the SSS was completed by an expert coder who was a doctoral student at the University of Washington with a Master's degree in Linguistics. One token was extracted from each of the 22 SSS assessments that were completed during the course of the study. The tokens were counterbalanced across targets. Only the final steps of the assessment that would result in the final score were recorded. During a one-hour session, the coder was trained to score the children's productions on the SSS. The coder and the investigator watched videos and scored tokens together. Once the coder was familiarized with the coding and performed accurately on greater than 90% of the trial tokens, she was given a video tape of the 22 tokens. The coder's scores were compared to the investigator's original scores and a percentage of agreement was calculated.

Carter-Buck

Measurement reliability for the Carter-Buck was completed by an expert coder who was a professor at the University of Washington with a Ph.D. in linguistics. The data from the Carter-Buck were selected from one participant across the three time intervals, including 14% of the total data. The coder was given a video of the assessments and blank protocols, and was instructed to phonetically transcribe the targets using broad transcription as per the instructions for the Carter-Buck. The coder's transcription and the investigator's transcription were matched for point-to-point agreement on the accuracy of the targets and also assessed for agreement on correct/incorrect productions.

Probe

Measurement reliability for the probe was completed by an expert coder who was a doctoral student at the University of Washington and a certified speech-language pathologist. The data from one participant across three time intervals were assessed including 14% of the total data. The coder was given a video, data sheets, and instructions to transcribe the target words using broad phonetic transcription. The target consonants were then compared for point-to-point percentage of agreement and for agreement on correct/incorrect productions.

Hodson Assessment of Phonological Patterns

Measurement reliability was completed by an expert coder who was a doctoral student at the University of Washington and was in the process of completing her clinical fellowship year. The data from one participant across three time intervals were assessed, which included 14% of the total data. A formal training was not completed; however, the coder was instructed to re-score the HAPP-3 assessments as per the instructions from the standardized test. The coder used the original transcriptions of three assessments for scoring. The coder's scores and the investigator's scores were then compared for percentage of agreement across target patterns.

Percent Consonants in Error

Measurement reliability for the PCE was completed by an expert coder who was a doctoral student at the University of Washington with a master's degree in linguistics. The data from one participant across three time intervals were assessed, which included 14% of the total data. The coder did not participate in formal training, but was given the gloss of 100 words from each PCE. With the gloss, the coder watched the videos and phonetically transcribed the 100 words. Point-to-point agreement was calculated between the coder's transcription and the investigator's transcription. In addition, agreement was calculated on the correct/incorrect consonant agreement between the coder's transcription and the investigator's transcription.

Post-treatment Procedures

Participants eligible for the proposed study were enrolled for 6 months of treatment. After two cycles of treatment and with parent approval, the participants were referred to the UWSHC waiting list. Children within the Issaquah School District were transferred back to their original speech-language pathologist within the school district. All families received a final progress report and summary of all assessments and treatment with their child.

Research Questions

Thus far in chapter two, the testing and treatment methods have been described; the following section describes the approach adopted to answer the research question: ***What is the nature of phonological change?*** This question has three parts, presented below.

(1) How does performance change on five different measures of phonology during six months of treatment when each measure is considered independently? (a) The Scaffolding Scale of Stimulability, (b) The Carter-Buck Articulation Assessment, (c) The single-word probe, (d) The Hodson Assessment of Phonological Patterns, (e) The Percent Consonants in Error in connected speech.

To answer this question, each assessment was analyzed separately and the composite scores for each of the eight participants were plotted per measure. In addition, the percentage of change during the two treatment periods for each participant was calculated and plotted per assessment. The patterns of change on each measure were evaluated and described via visual inspection.

(2) How does accuracy of production on the SSS change in relation to the Carter-Buck Articulation Assessment?

To answer this question, the original composite scores from the SSS were compared to scores from the Carter-Buck. Patterns of change for raw scores and percentage of change were evaluated across participants via visual inspection. Secondary analyses were conducted on the Carter-Buck to determine the relationship between subtests on the Carter-Buck and the SSS. In addition, because the Carter-Buck and the SSS include different target phonemes, the SSS was rescored using only the targets on the Carter-Buck to better compare the concurrent validity of the two measures. Kendall's *tau* b correlations were calculated to observe the relationship between the rankings on the two assessments at each time interval and across intervals.

(3) How does accuracy of production on the SSS change in relation to the Hodson Assessment of Phonological Patterns, the probe, and Percent Consonants in Error?

In order to answer this question, the composite scores of the SSS were compared to the scores from the HAPP-3, the probe, and PCE. Patterns of change for raw scores and percentage of change were evaluated across participants via visual inspection. Kendall's *tau* b correlations were calculated to observe the relationship between the rankings on the two assessments at each time interval and across intervals.

CHAPTER THREE: RESULTS

The results from the assessment of procedural and measurement reliability are presented; these results pertain to all of the subsequent assessments and experimental measures that are compared in this study. Following this report on reliability, the research questions are addressed individually and further results are presented in relation to each of the research questions.

Procedural Reliability

The results of the procedural reliability were calculated across all of the initial assessments and the experimental measures.

Initial Assessments. Procedural reliability on the initial assessments (PPVT, PLS-4, and the structural-functional examination) was 93%. The investigator completed 13/14 of the expected procedures.

SSS. Procedural reliability on the SSS was 89%. The investigator administered the correct sequence of steps on the SSS 435 out of 491 times.

Carter-Buck. Procedural reliability on the Carter-Buck was 100%. The investigator completed 27/27 of the expected procedures.

Probe. Procedural reliability on the probe was 100%. The investigator completed 18/18 of the expected procedures.

HAPP-3. Procedural reliability on the HAPP-3 was 100%. The investigator completed 18/18 of the expected procedures.

PCE. Procedural reliability on the Percent Consonants in Error was 100%. The investigator completed 18/18 of the expected procedures.

Measurement Reliability

The results of the measurement reliability were calculated across all of the initial assessments and the experimental measures.

Initial Assessments. Measurement reliability on the initial assessments (PPVT, PLS-4, and the structural-functional examination) was 83%. The coder and the investigator agreed on 15 out of 18 scores final scores.

SSS. Measurement reliability on the SSS was 82%. The investigator and the coder agreed on 18 out of 22 of the final scores. The four tokens in disagreement were all based on differences in judgment regarding the accuracy of phonemes in isolation and the scores varied by only one point between 20 or 21. Scores differed by one step or less on the scale 100% of the time.

Carter-Buck. Measurement reliability on the Carter-Buck was 72% for point-to-point agreement on consonants in word productions with broad transcription. The investigator and the coder agreed on 483 out of 672 of the target consonants. Reliability increased to 90% for agreement on whether target consonants were correct or incorrect across all three subtests. The investigator and coder agreed on 919 out of 1023 target consonants.

Probe. Measurement reliability on the probe was 76% for point-to-point agreement of consonants with broad transcription. The coder and the investigator achieved point-to-point agreement on 357 out of 471 of the target consonants. Measurement agreement improved to 89% when voicing differences were not considered and improved to 95% when agreement was based on correct or incorrect consonant production.

HAPP-3. Measurement reliability on the HAPP-3 was 95%. The coder and the investigator agreed on 829 out of 873 error codes.

PCE. Measurement reliability for the PCE was 87% agreement on point-to-point transcription of supraglottal consonants. In addition, the coder and the investigator achieved 94% agreement on judgments of correct or incorrect consonant scores.

Research Question 1

(1) How does performance change on five different measures of phonology during six months of treatment when each measure is considered independently?

The results pertaining to Research Question 1 are presented in the following section. The changes on each of the five experimental measures are shown individually: SSS, Carter-Buck, Probe, HAPP-3, and PCE. The data from the participants are presented in the same sequence across all of the graphs with the participants who completed both cycles of treatment first, 3F, 4M, 5F, 7F, 10M, 12F, followed by the two participants who only completed only one cycle of treatment, 8M, 9M. The measures were administered across three time intervals: Time 1 (prior to treatment), Time 2 (after three months of treatment, and Time 3 (after six months of treatment). Data are compared by per administration and also by the percent change between the administrations. The change between zero and three months of treatment is referred to as Cycle 1 and the change between three months and six months is referred to as Cycle 2 (see Figure 3.0).

Scaffolding Scale of Stimulability (SSS)

On the SSS, each target was scored from 1 to 21 indicating the number of cues and environmental supports needed to produce each target sound; to obtain a composite score for each administration of the SSS, the investigator added all the target scores from the error consonants together to create one score (see Appendix Q for individual target scores). The composite scores could range from 0 to 756, with a low score being best. The SSS was administered to six participants three times and to two participants two times; the change on the SSS for all eight participants is shown across the administrations in Figure 3.0. The data for each participant are grouped together and the participants are shown side-by-side rather than layered across time to better show similarities and differences in change over time.

During Cycle 1, all eight participants showed improvement on the SSS: one participant's score changed less than 50 points (8M), four participants' scores changed between 50-99 points (3F, 7F, 9M, 12F), two participants' scores changed between 100-150 points (4M, 5F), and one participant's score changed greater than 150 points (10M). During Cycle 2, all six remaining participants showed improvement: two participants' scores changed less than 50 points (7F, 12F), two participants' scores changed between 50-99 points (3F, 10M), and two participants' scores changed between 100-150 points

(4M, 5F). The average change for all eight participants was 104 points for Cycle 1, and 77 points for the six remaining participants during Cycle 2.

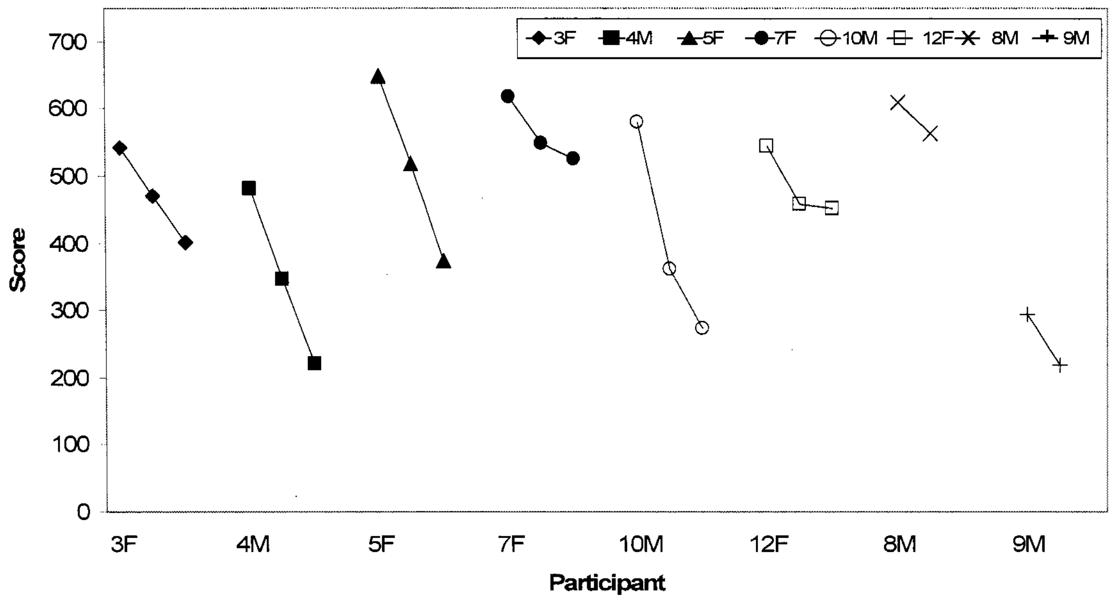

Figure 3.0. Scaffolding Scale of Stimulability: Scores for eight participants across three administrations at Time 1, Time 2, and Time 3 are presented, with low scores being best. Scores for 8M and 9M are provided for Time 1 and Time 2 only.

Percentage of change was also calculated across participants. In Figure 3.1, the percentage of change per treatment cycle is shown. The value is the participant's score relative to the total of possible errors for the targets that were assessed. Participants 4M and 10M made the highest percentage of change on the SSS overall during treatment; participants 12F and 8M showed the lowest percentage of change. The average percent of change across participants was 17% in Cycle 1 and 12% in Cycle 2.

On the SSS, all participants showed change during every cycle. The patterns of change in general show a steep downward slope with only two participants (7F and 12F) showing a very small change in the second cycle. Participant 12F showed such a small amount of change during Cycle 2 that this might not be considered clinically significant; in contrast, four out of the six children who received two cycles of treatment showed relatively similar amounts of change across both cycles. This finding indicates that the SSS is sensitive to change across time and for most children shows change continuously and incrementally. Ceiling and floor effects were not observed because at least 100 points above and 200 points below the participants' scores remained where changes in scores could still be observed given a longer treatment period. The two participants who were seen for only one cycle both showed change. Steady change may be indicative of a measure with a more sensitive range; children were not judged on one specific task that had to be mastered before a change registered.

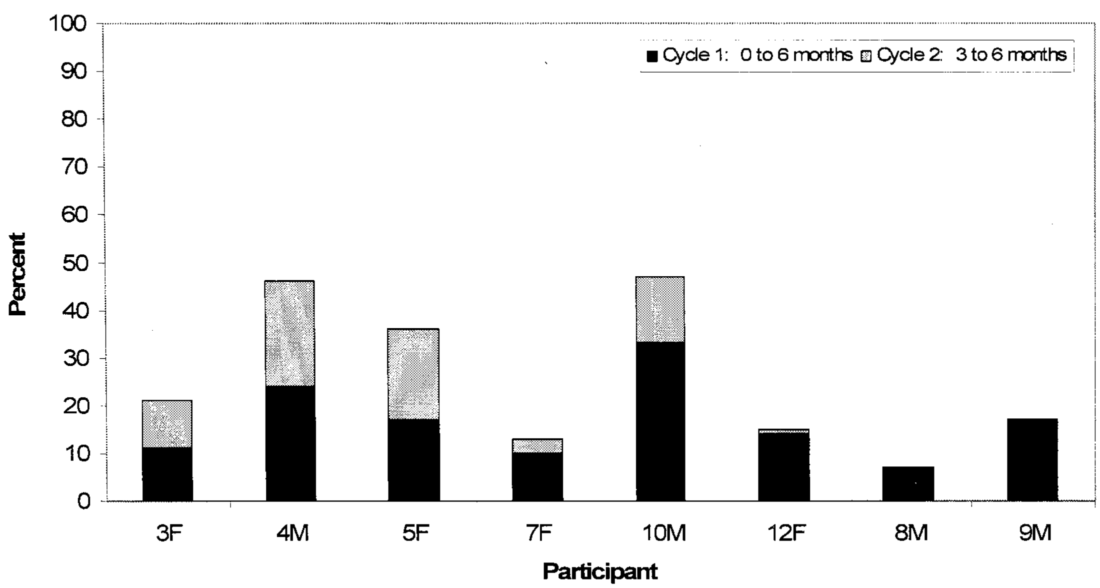

Figure 3.1. Scaffolding Scale of Stimulability: Percent change per treatment cycle relative to the total number of possible errors.

Carter-Buck Articulation Assessment

The Carter-Buck Articulation Assessment includes the assessment of 13 phonemes on three subtests: spontaneous word, imitated word, and imitated syllables. On each subtest, the number of errors were counted; the scores from each subtest were added together to create a composite score for comparison. Scores can range from 0 to 341 with a low score being best. The Carter-Buck was administered to six participants three times and to two participants two times; the change on the Carter-Buck for all eight participants is shown in Figure 3.2. The data for each participant are grouped together and the participants are shown side-by-side rather than layered across time to better show similarities and differences in change over time.

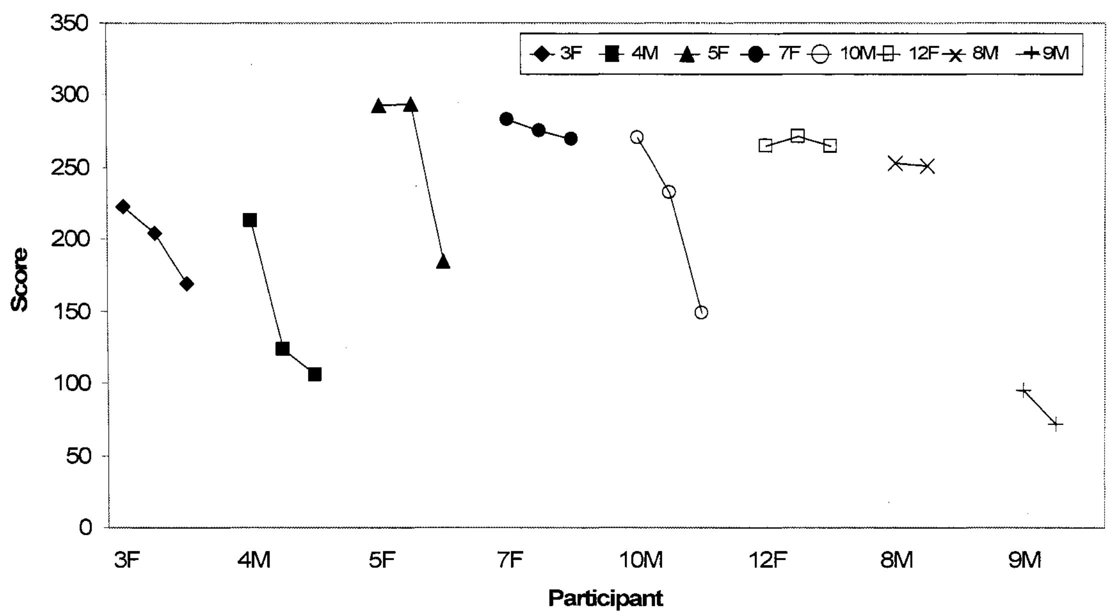

Figure 3.2. Carter-Buck: Scores for eight participants across three administrations at Time 1, Time 2, and Time 3 are presented with low scores being best. Scores for 8M and 9M are provided for Time 1 and Time 2 only.

During Cycle 1, six out of eight participants' scores showed improvement on the Carter-Buck: two participant's scores changed between 1-20 points (3F, 7F, 8M), three participant's scores changes between 21-40 points (9M, 10M), and one participant's score changed greater than 40 points (4M) (Percentage of change is presented later). Two participants' scores regressed between 0-10 points (5F, 12F). During Cycle 2, all of the remaining participants showed improvement on the Carter-Buck: three participants' scores changed between 1-20 points (7F, 12F, 4M), one participant's score changed between 21-40 points (3F), and two participants' scores changed greater than 40 points (5F, 10M). The average change for all eight participants was 22 points for Cycle 1 and 43 points for the six remaining participants during Cycle 2.

The percentage of change was also calculated for the Carter-Buck, as was done for the SSS. In Figure 3.3, the percentage of change per treatment cycle is shown. The value is the participant's score relative to the total number of possible errors. Participants 4M, 5F, and 10M showed the highest percentage of change during treatment, whereas participants 8M and 12F showed the lowest percentage of change. The average percentage of change across participants was 6% for Cycle 1 and 9% for Cycle 2.

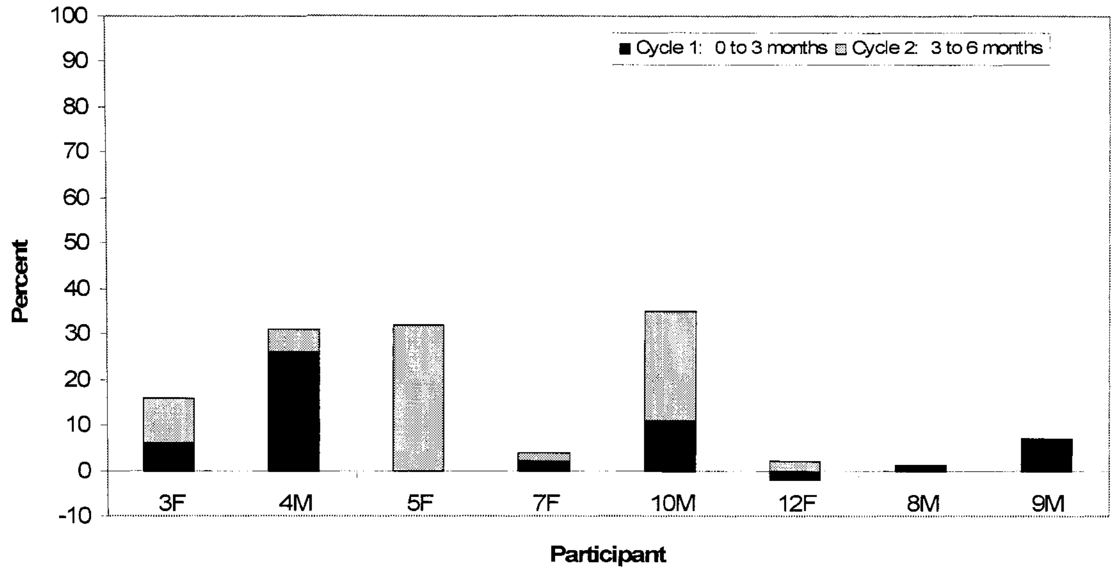

Figure 3.3. Carter-Buck: Percentage of change per treatment cycle relative to the total number of possible errors.

Figure 3.3 shows that the amount of change for 3F, 4M, 5F, and 10M varied substantially from one cycle to the next with 4M's change being greater for Cycle 1 while the other participants exhibited greater change in Cycle 2. Scores for 7F and 12F, in contrast, show little or no change across both cycles. A floor effect may be occurring for participants 5F and 10M who show smaller amounts of change during the first cycle and then a huge change in the second cycle. The Carter-Buck may not be sensitive enough to capture small changes early in the treatment process, which makes it appear as though the children have made a sudden improvement because they have crossed a threshold of performance. In addition, the small changes observed for 7F and 12F across both cycles may also be the result of a more severe floor effect. The documentation of a third cycle with large gains may have supported this concept if the children suddenly made improvements. Participant 8M was only seen for one cycle; a second cycle of treatment may have also supported the concept of a floor effect for some children. The floor effect could be caused by the selection of later developing targets for assessment that were selected by Carter and Buck (1958) for first grade children.

Two additional calculations are presented for the Carter-Buck: percentage of errors overall and analysis across subtest. The scores on the Carter-Buck were calculated in reference to the original study in 1958 (see Figure 3.4). The percentage of change of errors was calculated relative to the number of errors in the first administration rather than the number of total possible errors. The patterns of change for the participants are similar in Figure 3.3; however, the overall scores are slightly higher in Figure 3.4 because a lower denominator was used to calculate the percentages.

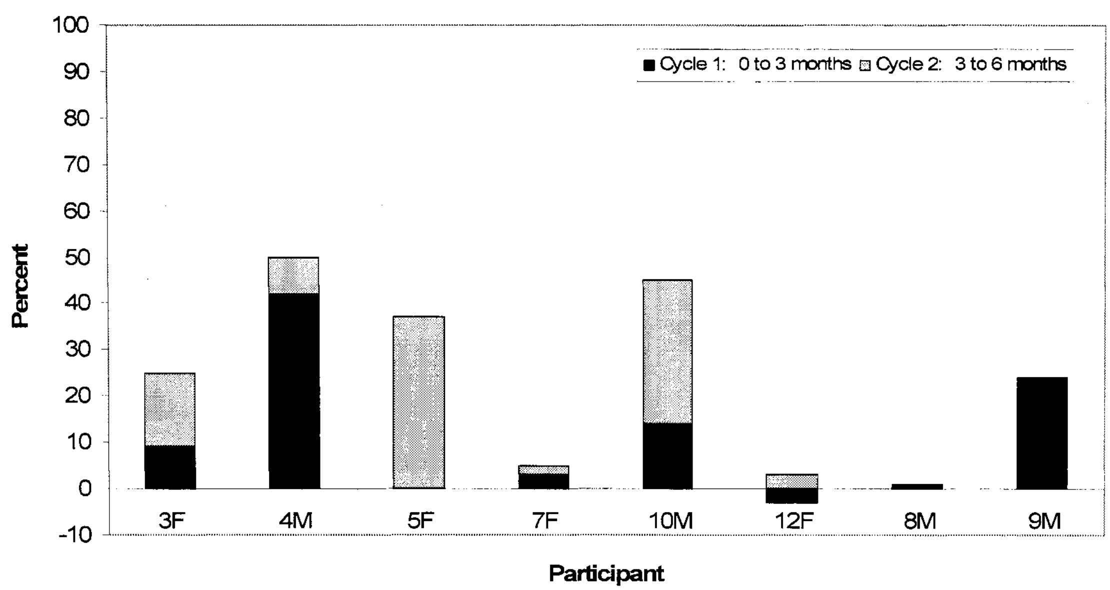

Figure 3.4. Carter-Buck: Percentage of change of errors relative to the original number of errors (in accordance with Carter and Buck, 1958).

The hierarchy on the Carter-Buck across the three subtests was evaluated. In general, the scores improve across the subtests and reflect the hierarchy; the spontaneous subtest is the least supported and shows the highest number of errors, followed by the imitation subtest, and finally, the syllable subtest. The errors were added together across all of the participants and administrations: the spontaneous word subtest had the highest number of errors (1600/2464), followed by the imitated words subtest (1584/2464), and finally the imitated syllables subtest (1471/2574).The patterns of change are relatively similar for the spontaneous word subtest and the imitated word subtest; however, the interesting result is that the syllable subtest reflects a more linear patterns of change and greater gains for six out of the eight children.

Probe

The probe provided information about how each child functions when asked to produce single-words without any support. On the probe, single-words were selected based on the children's errors patterns. All of the sound classes were represented with multiple opportunities to produce each phoneme; multiple consonants were assessed in each word. A child became more successful at producing sounds in spontaneous productions of single-words over time and the error scores on this assessment decreased; however, the child may still produce errors in other environmental contexts. The investigator tallied the number of error consonants that each child produced to generate an overall score. The score could range from 0 to 132 possible errors with a low score being best. The results from the probe are presented in Figure 3.5. The data for each participant are grouped together and the participants are shown side-by-side rather than layered across time to better show similarities and differences in change over time.

During Cycle 1, six out of eight participants showed improvement on the probe: three participant's scores changed between 1-20 points (8M, 10M, 12F), one participant's score changed between 21-40 points (3F), and one participant's score change between 41-60 points (4M). In contrast, one participant regressed (7F) and one participant showed no change (5F). During Cycle 2, all of the remaining six participants showed improvement on the probe: five participants' scores changed between 1-20 points (3F, 4M, 7F, 12F, 8M) and one participant's score changed between 20-40 points (5F). The average change for all eight participants was 17 points for Cycle 1 and 13 points for the six remaining participants during Cycle 2.

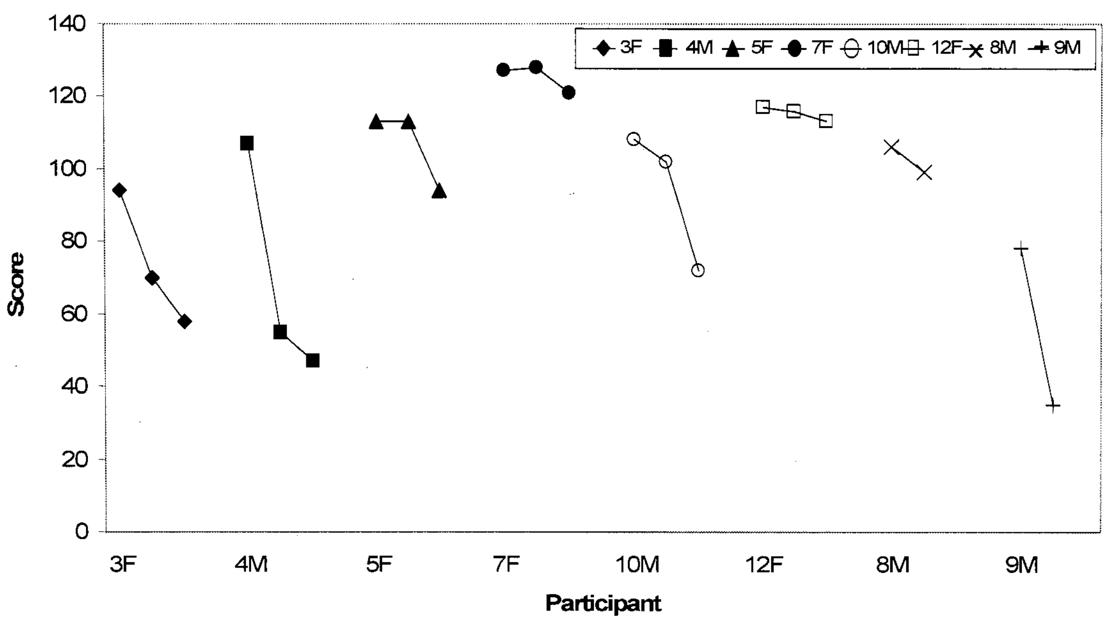

Figure 3.5. Probe: Scores for eight participants across three administrations at Time 1, Time 2, and Time 3 are presented with low scores being best. Scores for 8M and 9M are provided for Time 1 and Time 2 only.

The percent change was also calculated for the probe, as was done with the previous measures. In Figure 3.6, the percent change per treatment cycle is shown. The value is the participant's score relative to the total number of possible errors. Participants 4M and 9M showed the highest percentage of change, whereas participants 7F and 12F showed the lowest percentage of change during treatment. The average percent change across participants was 13% for Cycle 1 and 10% for Cycle 2.

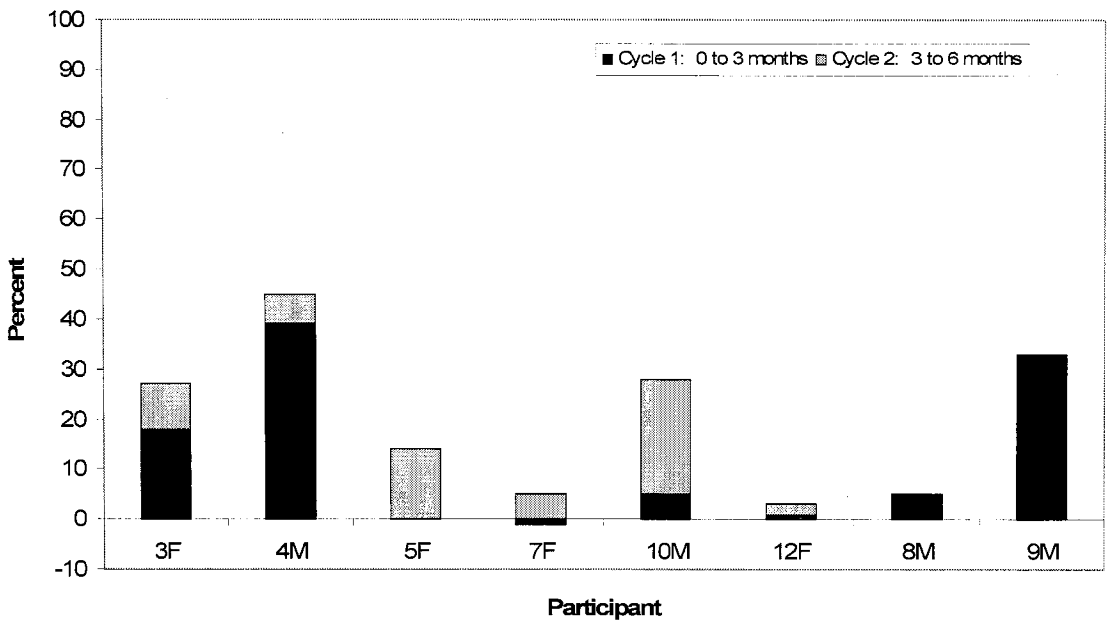

Figure 3.6. Probe: Percent change per treatment cycle relative to the total number of possible errors.

On the probe, the pattern of change for four out of six participants (3F, 4M, 5F, 10M) showed a contrast between the two cycles (see figures 3.5 and 3.6); in other words, one cycle greatly dominated over the other cycle in the amount of change. Floor and ceiling affects may have influenced this pattern. The extent of change exhibited by participants 5F and 10M may be the result of a floor effect because no change occurred for 5F and a small change occurred for 10M during Cycle 1, followed by large change in Cycle 2; furthermore, the regression observed for 7F and small change for 12F may also be the result of a floor effect, which could have been supported with one more cycle of data. In addition, 8M made very little change on this measure during the one cycle that was treated indicating a floor effect. The children may not have been ready for independent productions and, thus, scores showed little change on this measure. A different pattern of change occurred with participants 4M and 3F who showed a large change during the first cycle, but a small change in the second cycle. This type of change could suggest a ceiling effect; however, 9M made great gains on this measure in only one cycle of treatment. The probe gave little room for additional gains; participant 4M made great gains during the first cycle, but, did not change much during the second cycle. This could be due to a ceiling effect at the single word level for earlier developing sounds, or a floor effect for the later developing sounds that he could not yet produce independently.

Hodson Assessment of Phonological Patterns

The *Hodson Assessment of Phonological Patterns-Edition 3* (HAPP-3) documents changes in phonological patterns in spontaneous single-word productions. The scores that are presented in this study are the "Total of Major Phonological Deviations (TOMPD)." Different error patterns such as deletions of syllables or consonants and deviations relative to sound class are summed together to create the TOMPD score. The scores can range from 0 to 307 with a low score being best. The results of the HAPP-3 are presented in Figure 3.7. The scores for each participant are grouped together so that the pattern of change may be more easily observed rather than layered together across time.

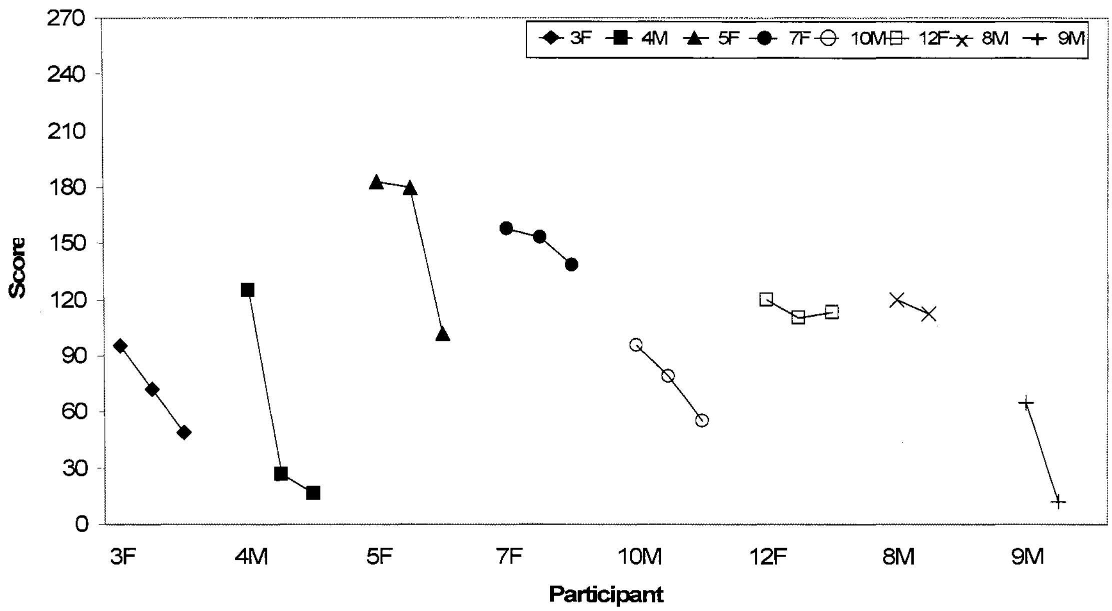

Figure 3.7. Hodson Assessment of Phonological Patterns: Scores for eight participants across three administrations at Time 1, Time 2, and Time 3 are presented, with low scores being best. Scores for 8M and 9M are provided for Time 1 and Time 2 only.

During Cycle 1, all eight participants showed improvement on the HAPP-3 (see Figure 3.7): four participants' scores changed between 1-10 points (5F, 7F, 8M, 12F), two participant's scores changed between 11-20 points (10M, 12F), one participant's score changed between 21-30 points (3F), and two participants' scores changed greater than 31 points (9M, 4M). During Cycle 2, five out of six of the remaining participants showed improvement on the HAPP-3: two participants' scores changed between 11-20 points (4M, 7F), two participants' scores changed between 21-30 points (3F, 10M), and one participant's score changed more than 31 points (5F). In contrast one participant's score regressed by three points (12F). The average change for all eight participants was 27 points for Cycle 1 and 25 points for the six remaining participants during Cycle 2.

The percentage of change was also calculated for the HAPP-3, as was done for the previous measures. In Figure 3.8, the percentage of change per treatment cycle is shown. The value is the participant's score relative to the total number of phonological deviations possible. Participants 4M and 5F showed the highest percentage of change during treatment, whereas participants 7F and 12F showed the lowest percentage of change (8M was also low, but was not assessed for a second cycle). The average percentage of change across participants was 10% for Cycle 1 and 9% for Cycle 2.

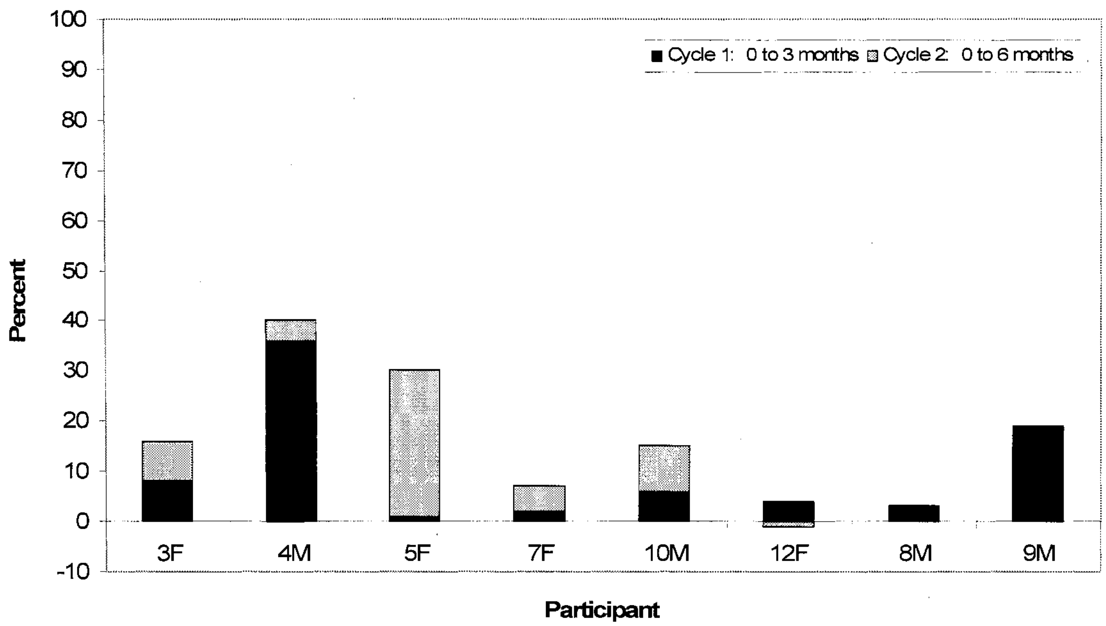

Figure 3.8. Hodson Assessment of Phonological Patterns: Percent change relative to the total number of possible errors.

On the HAPP-3, the patterns of change for the scores differed across the children. Three out of the six participants seen for two cycles showed a large contrast between the two cycles (4M, 5F, 7F); one cycle greatly dominated over the other cycle in the amount of change. The timing of change differed for these children: participants 5F and 7F showed little change during the first cycle, but a proportionately larger change during the second cycle (especially true for 5F). This suggests that the children may not have been ready to produce targets in spontaneous single-word productions and provides possible evidence of a floor effect; furthermore, two participants showed little or no change (12F, 8M) that could also be a more severe floor effect, but would require additional documentation across time to demonstrate.

Participant 4M showed a different pattern of change with a large gain during the first cycle followed by a proportionately smaller gain in the second cycle. The pattern for 4M may reflect a ceiling effect because there is little room left on the measure for him to improve. Participant 9M, who was only seen for one cycle, made large gains as well during the one cycle. Ceiling effects would likely be evident in an additional cycle of treatment for participant 4M and 9M who only had 10-15 points left to improve on this measure. One other pattern of change emerged for participants 3F and 10M; the steep and evenly spread pattern of change for participants 3F and 10M suggests that the HAPP-3 was a sensitive measure of change for the developmental stages of these two children.

Percent Consonants in Error

Percent Consonants in Error was calculated based on a connected speech sample gathered during a story-telling activity. The child's speech was glossed and transcribed as an adult would produce the phonemes and the child's actual productions were transcribed. The child's consonants were compared to the adult forms and the total number errors were divided by the total number of possible consonants to generate the PCE score. The results of the children's scores are presented in Figure 3.9. The data for each participant are grouped together and the participants are shown side-by-side rather than layered across time to better show the similarities and differences in change over time.

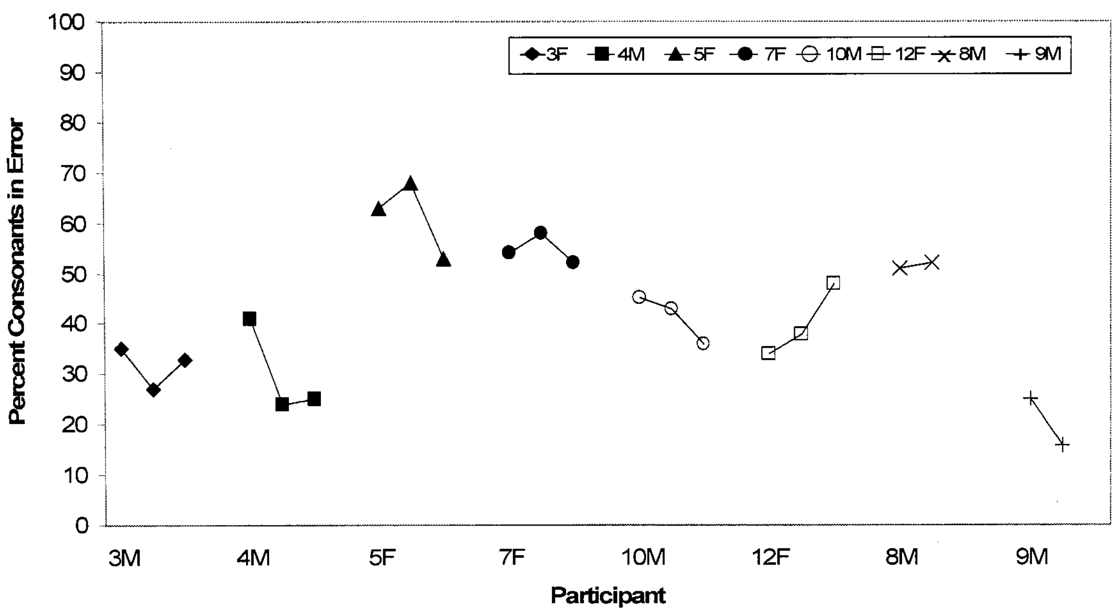

Figure 3.9. Phonological change on the PCE: Scores for eight participants across three administrations at Time 1, Time 2, and Time 3 are presented with low scores being best. Scores for 8M and 9M are provided for Time 1 and Time 2 only.

Scores on the Percent Consonants in Error (PCE) showed improvement for some children (fewer errors) and regression for others. The change on the PCE for all participants across both time periods is shown in Figures 3.9 and 3.10. During Cycle 1, four out of eight participants' scores showed improvement on the PCE: three participants' scores changed between 1-10 percentage points (3F, 9M, 10M) and one participant's score changed between 11-20 percentage points (4M). In contrast, four participants' scores increased (more errors) between 1-10 percentage points (5F, 7F, 8M, 12F). After Cycle 2, three out of the six remaining participants' scores showed improvement on the PCE: two participants' scores changed between 1-10 percentage points (7F, 10M) and one participant's score changed between 11-20 percentage points, while three participants regressed between 1-10 percentage points (3F, 4M, 12F).

The percent change was also calculated for the PCE, as was done with the previous measures. In Figure 3.10, the percent change per treatment cycle is shown. The values on this assessment are already plotted in percentage values, so the values in this chart are shown for the differences in percentages between 0 to 3 months and 3 to 6 months. Participants 4M and 5F show the highest overall improvement, whereas 12Fand 8M show the lowest improvement (8M was only treated for one cycle). The average percent change was 3% for Cycle 1 and 2% for Cycle 2.

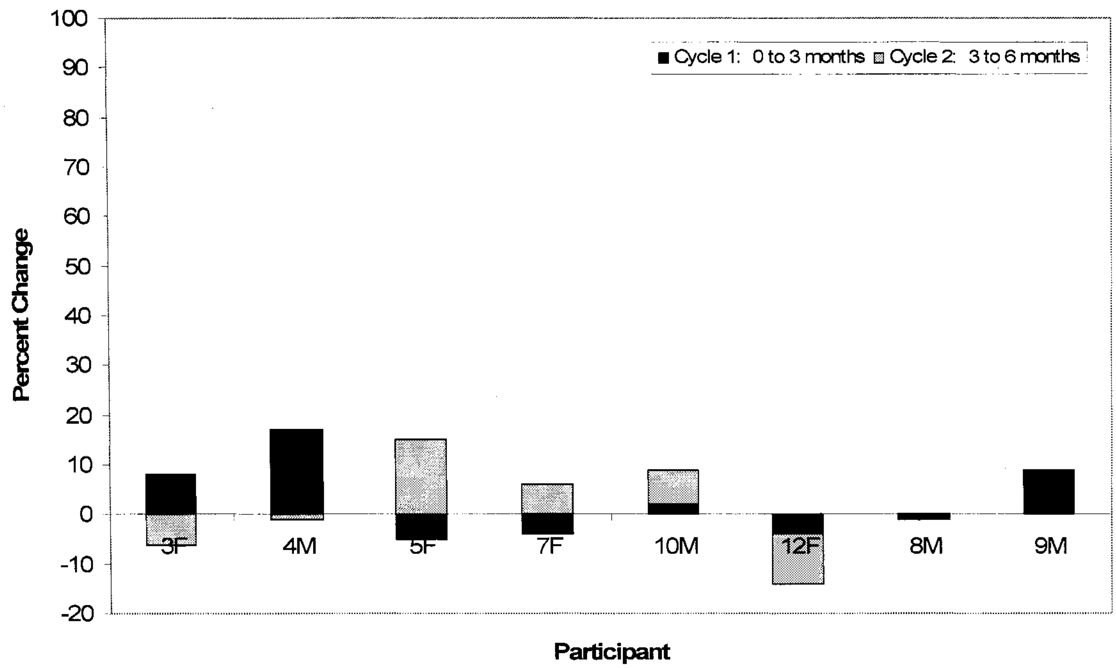

Figure 3.10. Percent Consonants in Error: Change in percentage per treatment cycle.

The scores from the PCE are unusual because only participant 10M showed progress across both cycles of treatment. Six out of eight participants showed regression on at least one cycle and participant 12F showed regression on both cycles. The periods of the regression make it difficult to determine whether ceiling or floor effects have influenced the scores; however, the impact of language development and use may be a secondary variable affecting the scores.

The results across all five experimental measures were presented individually to show the different ways that each measure captures and presents phonological change. The measures showed many similarities and differences. In the following section, comparisons will be made between the experimental measures. In particular, the differences with dynamic assessments and static assessments will be examined.

Research Question 2

How does accuracy of production on the Scaffolding Scale of Stimulability change in relation to the Carter-Buck Articulation Assessment?

With the goal of evaluating concurrent validity, the results were compared for two dynamic assessments: the SSS and the Carter-Buck Articulation Assessment. First, the composite scores were compared including percentages of change and correlations of raw scores. Second, the subtests of the Carter-Buck were compared to the SSS scores to evaluate similarities in patterns of change. Third, the test scores were revised to determine whether test results were more similar when the same targets were evaluated.

Composite scores were first compared for patterns of change that would indicate a relationship between the two measures. Recall that the composite score for the SSS was created by summing all of the individual phonemes scores (from 1 to 21) and that the SSS composite scores could fall between 0 and 756; the composite score for the Carter-Buck was created by summing the errors produced on each of the three subtests and that the scores could fall between 0 and 341.

It was hypothesized that the two measures would show some overlap in the patterns of change because they both assess more than one environment and include additional support in the form of cues; however, the patterns of change for the two measures are very different when percentage of change per cycle for composite scores were compared (see Figure 3.11). First, the combined percentage of change is greater on the SSS than on the Carter-Buck for all eight participants. If cycles are compared individually (a total of 14 cycles for the 8 children), the Carter-Buck changed more than on the SSS on only four cycles; overall, changes on the SSS were steadier across time.

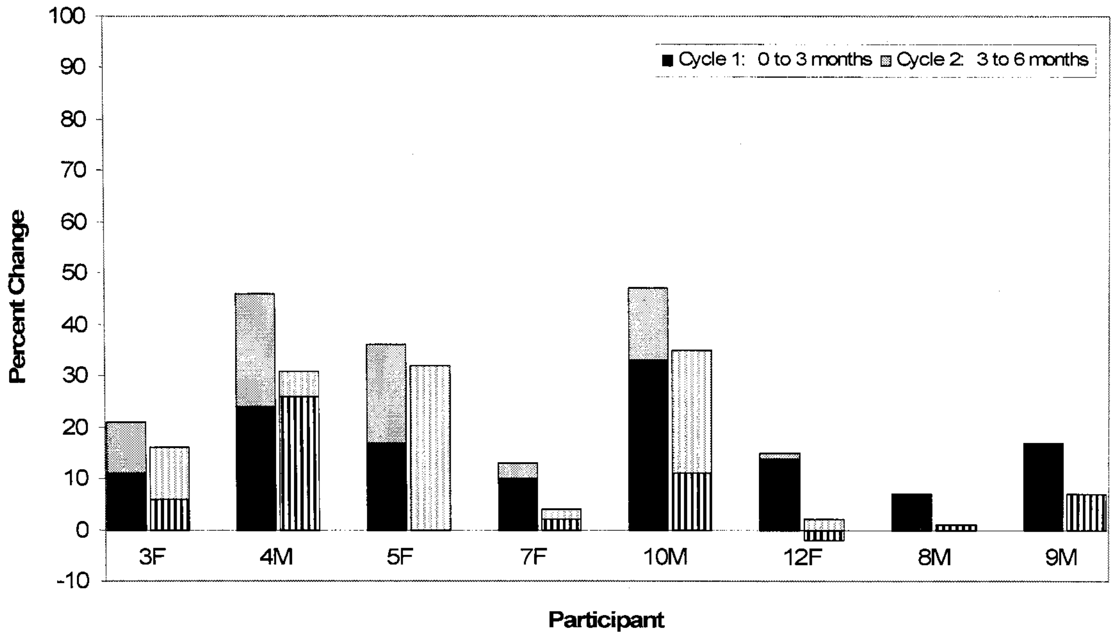

Figure 3.11. Percentage of change scores for the SSS and the Carter-Buck are presented by participant. The first and solid column of each pair represents the changes on the SSS and the second and striped column represents changes on the Carter-Buck. The first cycle is represented in black and the second cycle is represented in gray.

Second, the patterns of change on the SSS and the Carter-Buck differed across the children. The SSS showed a steady and relatively even change across both cycles for 4M and 5F, but a possible ceiling effect was evident on the Carter-Buck for 4M whose progress seemed to stop during the second cycle and a floor effect for 5F who showed 0% change during the first cycle but nearly caught up with the SSS in overall percentage of change during the second cycle. The scores for 10M are interesting because the scores on the SSS suggest a greater percentage of change during the first cycle, yet on the Carter-Buck the scores suggest a greater percentage of change on the second cycle. The scores for 12F are similar, but show regression on the Carter-Buck during the second cycle.

Correlations were calculated between the SSS and Carter-Buck using Kendall's *tau* b, a rank order correlation statistic. The results of the Kendall's *tau* b correlations are presented in Table 3.0. Although the two measures looked very different via visual inspection of the data, correlations were significant between the SSS and the Carter-Buck at each of the three administrations. The correlations may be related because both measures do show improvement across time and the children's skills are spread across severity so that both measures may be ranking the measures in the same way.

Table 3.0

Kendall's tau b correlations between Carter-Buck and SSS administrations.

	SSS: Time 1	SSS: Time 2	SSS: Time 3
Carter-Buck: Time 1	.857**	.429	.067
Carter-Buck: Time 2	.857**	.571*	.067
Carter-Buck: Time 3	.200	.067	.867*

Note: *. Correlation is significant at the .05 level (2-tailed). **. Correlation is significant at the .01 level (2-tailed). The numbers 1, 2, 3 represent the time periods that the measures were administered: prior to treatment, after 3 months, and after 6 months of treatment. Although the correlation at Time 3 is the highest number, there are only six children's scores compared at this time period which decreases the significance value.

Overall, the patterns of change on the SSS and the Carter-Buck were different via visual inspection. Two possible reasons for these differences were considered: subtest changes on the Carter-Buck were possibly lost when summed for a composite score, and the Carter-Buck assesses different and later developing targets.

Because subtest scores on the Carter-Buck were summed, it was hypothesized that potential changes on individual subtests could have been lost that certain subtests may show patterns of change would show a stronger relationship between the SSS and the Carter-Buck. Similarities and differences are summarized in Table 3.1 using data from Figures 3.0 and 3.5. First, SSS scores were compared to those from the *Syllable subtest* of the Carter-Buck and the general pattern of change was similar for four out of eight participants on the two tests; this was the most similarity across all three of the subtests. Participants 8M and 9M looked similar on both assessments, but were only assessed for one cycle which could be a little misleading. Second, patterns of change were compared between the SSS and the *Imitated words subtest* of the Carter-Buck; only participants 7F and 9M looked similar on both tests. Finally, the patterns of change were compared between the SSS and the *Spontaneous word subtest* of the Carter-Buck; participants 3F, 8M, and 9M looked similar. Overall, this method of comparison did not support a relationship between the tasks on the Carter-Buck and the tasks on the SSS.

Table 3.1

Similarities and differences for each participant in patterns of change for the SSS compared with the subtests on the Carter-Buck.

	Syllables		Imitated Words		Spon. Words	
	Same	Different	Same	Different	Same	Different
3F		X		X	X	
4M	X			X		X
5F		X		X		X
7F		X	X			X
10M	X			X		X
12F		X		X		X
8M	X			X	X	
9M	X		X		X	

The investigator also hypothesized that the percentage of change on the Carter-Buck and the SSS would begin to look more similar across the two tests if the same phonemes were assessed. Because the Carter-Buck includes a much smaller set of target phonemes than on the SSS, the investigator rescored the SSS using only the same targets as the Carter-Buck. The percentage of change for the revised SSS is presented in Figure 3.12. Contrary to the hypothesis, the scores on the SSS for the reduced set of phonemes resulted in very similar patterns of change to the original scores when compared to the original percentages of change for the SSS across participants. The pattern of change on the SSS still did not resemble the pattern of change on the Carter-Buck.

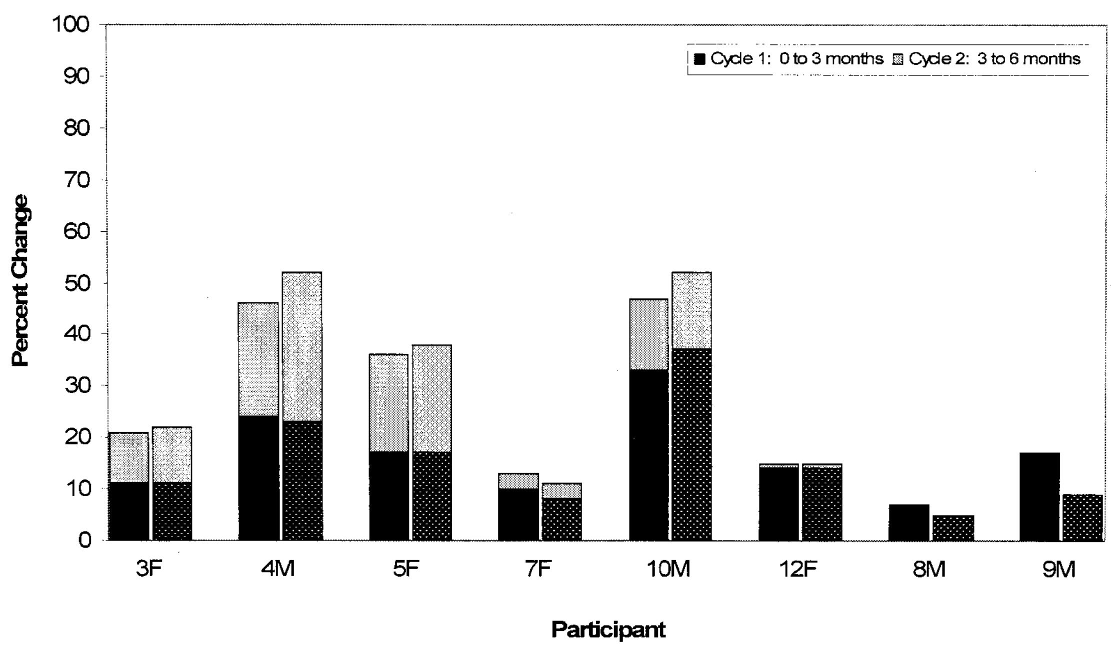

Figure 3.12. Scaffolding Scale of Stimulability: Original percent change scores are solid bars and the revised percent change scores are dotted bars.

Given several considerations, the SSS still showed a greater percentage of change across time than did the Carter-Buck via visual inspection. These results lead the investigator to believe that the SSS assesses a broader range of change across time, is more sensitive to change over time, and is more dynamic than the Carter-Buck. The SSS is quantifying different skills than the Carter-Buck and the two measures do not show concurrent validity in relation to percentage of change across treatment.

Research Question 3

How does accuracy of production on the SSS change in relation to the single-word probe, the Hodson Assessment of Phonological Patterns, or the Percent Consonants in Error in connected speech?

The results were compared between the Scaffolding Scale of Stimulability and the static assessments: the probe, the *Hodson Assessment of Phonological Patterns*, and Percent Consonants in Error to examine predictive validity of the SSS relative to the static assessments. First, the percentages of change were calculated for each child across assessments. Second, the data were grouped together and correlations were calculated between the measures across time periods. Third, the rank orders across the assessments were graphed.

It was hypothesized that the SSS would show the greatest percentage of change followed by the probe, the HAPP-3, and finally, the PCE. However, the patterns of change differed when percent change scores were compared between the SSS and the static assessments (see Figure 3.13). It was expected that all of the children's profiles would look more like that of Participant 10M or 8M, but the results showed different patterns.

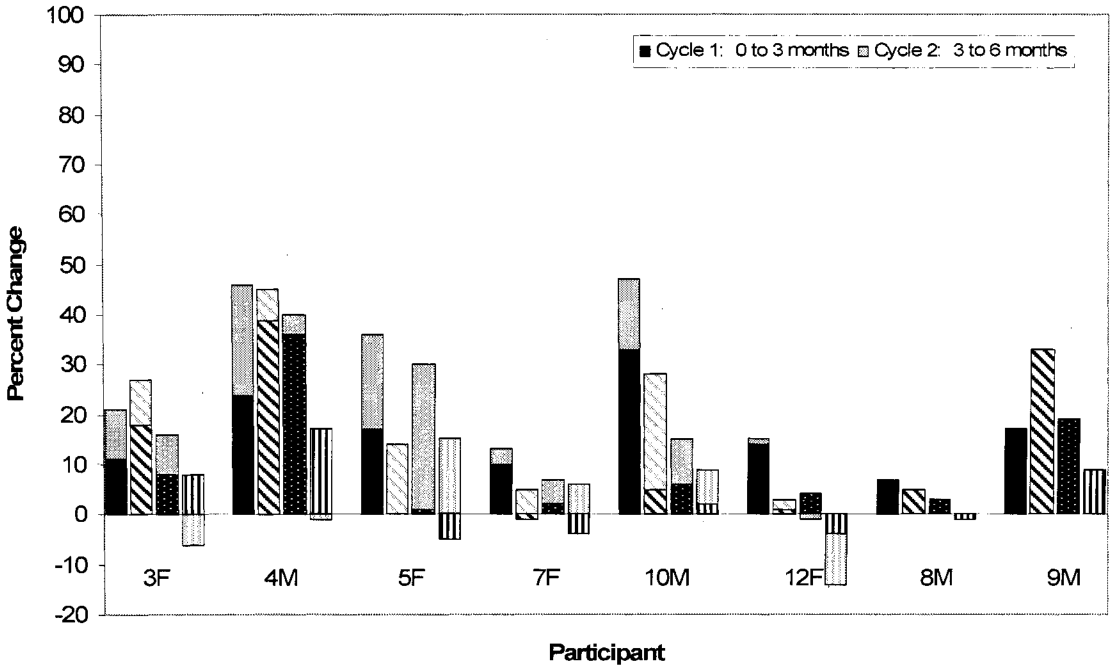

Figure 3.13. Percent change for each participant across the SSS and static assessments. The bars for each participant are in the following order: SSS (solid), Probe (diagonal stripes), HAPP-3 (dots), PCE (vertical stripes).

When two cycles were considered together, the overall percentage of change was the greatest for the SSS when compared to the other measures for five out of six children; only the probe showed a higher percentage of change for participant 3F. In Table 3.2, the rankings across the measures are presented. The probe appeared to be the most unpredictable of the measures; four out of eight children scored much higher on the probe than on the HAPP-3 and the PCE, yet for the other four children the scores were lower than the HAPP-3 and the PCE. If the probe scores are removed from Figure 3.13, the overall pattern across the measures shows a downward "stair-step" with the most change on the SSS, followed by the HAPP-3, and finally the PCE. The alternative rankings are also presented in Table 3.2. Only participant 9 showed similar change on the SSS and HAPP-3, but was only seen for one cycle.

Table 3.2

Rankings of highest overall percentage of change across assessments. The Probe was the most inconsistent of the measures across children. When it was removed (see right), the pattern of change was more consistent.

	SSS	Probe	HAPP-3	PCE	SSS	HAPP-3	PCE
3F	2	1	3	4	1	2	3
4M	1	2	3	4	1	2	3
5F	1	4	2	3	1	2	3
7F	1	4	2	3	1	2	3
10M	1	2	3	4	1	2	3
12F	1	4	2	3	1	2	3
8M	1	4	2	3	1	2	3
9M	2	1	3	4	2	1	3

The pattern for percentages of change across the two cycles differed across the experimental measures. The differences are presented in Table 3.3. The patterns that occurred included an even change across both cycles, a larger change during Cycle 1, a larger change during Cycle 2, and a pattern of regression.

Table 3.3

Patterns of change across measures for six participants treated for two cycles.

	Even change	Greater change for Cycle 1	Greater change for Cycle 2	Regression
SSS	3F, 4M, 5F	7F, 10M, 12F		
Probe	12F	3F, 4M,	5F, 7F, 10M	
HAPP-3	3F, 10M,	4M	5F, 7F,	12F
PCE			10M	3F, 4M, 5F, 7F, 12F

The question still remains, is there a predictive relationship between scores on the SSS and scores on the probe, the HAPP-3, and the PCE? The previous results suggest that three predictions could be made: (1) changes on the SSS are more likely to be evenly distributed across time periods, (2) when a large change occurs it will more likely be during the first cycle of treatment, and (3) the SSS is more likely to show a higher percentage of change than the static measures.

To further assess the relationships between the SSS and the static measures, correlations were calculated between the SSS and each of the static measures at each of the time periods using Kendall's *tau* b. Kendall's *tau* b is a rank order correlation that is being used to compare the probability that participants will likely score in the same rank order across two assessments. The results of the correlations are presented in Table 3.4.

Table 3.4

Kendall's tau b correlations between SSS and static assessments at Time 1, 2, and 3.

	S1	S2	S3	Pr1	Pr2	Pr3	H1	H2	H3	Pc1	Pc2
S2	.571*	--									
S3	.067	-.067	--								
Pr1	.500	.357	-.067	--							
Pr2	.643*	.500	.200	.857**	--						
Pr3	.333	.200	.733*	.200	.467	--					
H1	.593*	.371	-.200	.593*	.445	.067	--				
H2	.929**	.643*	.067	.571*	.714*	.333	.667*	--			
H3	.333	.200	.733*	.200	.467	1.00**	.067	.333	--		
Pc1	.786**	.500	-.067	.429	.429	.200	.667*	.714*	.200	--	
Pc2	1.00**	.571*	.067	.500	.643	.333	.593*	.929**	.333	.786**	--
Pc3	.600	.200	.467	.200	.467	.733*	.333	.600	.733*	.467	.600

Note: *.Correlation is significant at the .05 level (2-tailed). **.Correlation is significant at the .01 level (2-tailed). S=SSS, Pr=Probe, H=HAPP-3, Pc=PCE. The numbers 1, 2, 3 represent time periods that the measures were administered.

Several observations may be made from these scores. One interesting observation is that the scores on the first administration of the SSS are most highly correlated with the scores from the second administrations of the Probe, HAPP-3, and PCE; furthermore, the correlations with the HAPP-3 and PCE are significant at the .01 level. This suggests that the scores on the SSS at an earlier time period may be predictive of later scores on the other measures. This predictive ability could suggest that the SSS is more sensitive to change in the early stages of treatment. Another interesting observation is that by the third administration, the frequency of the correlations increases across the measures. The scores on the SSS at Time 3 correlate with both the HAPP-3 and the PCE; furthermore, five out of six combinations of the measures are correlated at Time 3. One unusual observation is that the scores on the HAPP-3 correlated with all of the other assessments at the same time periods. The highest correlations were with the probe at Time 3 and with the PCE at Time 2.

Scatter plots were used to verify the above correlations and confirm whether outliers were influencing the correlations. The scores on the first administration of the HAPP-3 and the SSS suggested an outlier that could shift the correlations at that time period. Furthermore, the scores between the HAPP-3 and the SSS show that once a child's score approaches 0 on the HAPP-3, the SSS still has 200 points left for potential change. None of the children in the current study scored below a 200 to confirm this potential outcome and ceiling effect on the HAPP-3. Scatter plots comparing the SSS and the Probe showed a cloud in the early administration that gradually shifted to more of a line. In contrast, the PCE appeared more linear with a more cloud-like shift.

The final comparison that is supplemental to the Kendall's *tau* b correlations is a comparison of rank orders in a visual format (see Figure 3.14). The participants were assigned a rank order to further show how different measures characterize children's phonological skills. The rank was labeled from 1 to 8 and reflected the best performance (lowest score) to worst performance (highest score) on each of the assessments across time. Participant 9M was consistently ranked number 1 across all five assessments across the two administrations in which he participated. Participant 4M was consistently ranked across all five measures across administrations one and two with variability across assessments during administration one. Participant 3F was most frequently ranked second or third across the assessments. Participant 5F was most frequently ranked eight or six during administrations one and two, and three or four during administration three. Participant 7F was most frequently ranked 7 during administrations one and two, but ranked highest during administration three across the measures. Participants 8M, 10M, 12F were ranked inconsistently across the measures. Greater consistency occurred for 10M and 12F during administration three when 8M was no longer present. The rankings appear to become more consistent and smooth across time.

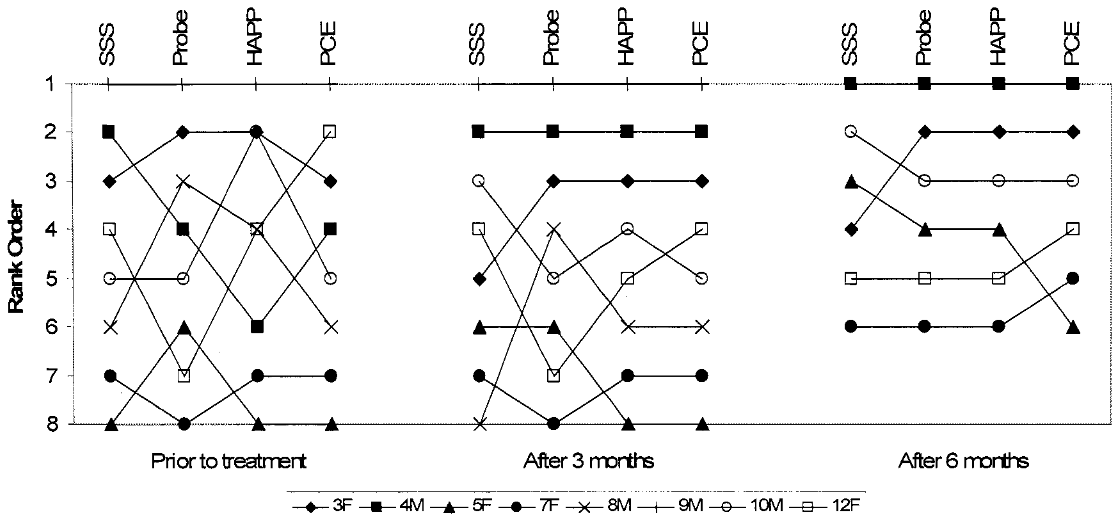

Figure 3.14. Rank order of participants for on each measure prior to treatment, after Cycle 1 (3 months), and after Cycle 2 (6 months).

In summary, it appears that the longer that children are in speech treatment, the more likely they will be consistently ranked across assessments; however, in the earlier stages of treatment, the measures are assessing different skills that will reflect different patterns of change.

CHAPTER FOUR: SUPPLEMENTAL RESULTS

During the process of analysis for the current study, several additional questions arose that were supplemental to the original research questions. In this chapter three related questions are presented with analysis and results. The questions include the following: (1) How do the experimental measures differ in the documentation of change when individual phonemes are analyzed rather than composite scores? (2) How do the combined results across all of the experimental measures inform investigators about the different profiles of each participant? (3) Are all of the environments and cues necessary on the SSS to document change in stimulability?

Supplemental Research Question 1

How do the experimental measures differ in the documentation of change when individual phonemes are analyzed rather than composite scores?

The analysis of the data thus far has only considered the differences across the measures when comparing composite scores. It is possible that data on individual phonemes may also inform documentation of change when they are extracted from the composite; to examine this possibility, samples of data were selected and evaluated by individual phonemes for comparison across the measures.

The first target for comparison was word-initial /k/; performance for initial /k/ was evaluated across all participants and across the five measures at Time 1 and Time 2, i.e., prior to treatment and after three months of treatment. The data for initial /k/ were extracted and rescored individually using the scoring criteria appropriate for each measure. In Table 4.0, the results for initial /k/ at Time 1 are presented. The scores for the SSS ranged between 1 and 21. On the other four experimental measures, the errors are shown as a ratio of errors to opportunities for errors. The Carter-Buck provided three opportunities for initial /k/ on each of the three subtests. The results are presented in sequence for subtest one, two, and three. The probe provided four opportunities to produce initial /k/ and the HAPP-3 provided two. The opportunities on the PCE are variable across the children because the speech sample included spontaneous productions from the child and the vocabulary could not be controlled; in many instances there were no opportunities to produce an initial /k/. Below each of the scores, the types of errors that the children produced are listed; for example, participant 3F substituted /t/ and /d/ for initial /k/ on the Carter-Buck and the probe, but also deleted /k/ on the probe.

Table 4.0
Word-Initial /k/: Scores per assessment at Time 1. SSS ranges from 1-21 with a 1 being best. On the Carter-Buck, probe, HAPP-3, and PCE, the errors are tallied over the number of opportunities. Error types are described below.

	SSS	Carter-Buck	Probe	HAPP-3	PCE
3F	8	3/3, 3/3, 3/3	4/4	2/2	No opp.
		t, d	del, t, d	t, d	
4M	1	0/3, 0/3, 0/3	1/4	2/2	No opp.
			g	f, del	
5F	21	3/3, 3/3, 3/3	4/4	2/2	No opp.
		h, ʔ	h, t	h	
7F	15	3/3, 3/3, 3/3	4/4	2/2	2/2
		del	t/k	t, del	t
10M	15	0/3, 2/3, 2/3	3/4	1/2	0/2
		t	t/k	t	
12F	8	2/3, 3/3, 2/3	4/4	2/2	2/2
		d, g	t, d, del	t	t, d
8M	21	3/3, 3/3, 3/3	4/4	2/2	3/3
		t	t, d, b, del	t, d	t, p
9M	8	0/3, 0/3, 0/3	4/4	2/2	No opp.
			del, t	t	

From this rescoring process, several observations were made. One interesting finding was that at Time 1, four children scored identical scores across all measures except for the SSS. On the Carter-Buck, the HAPP-3, the probe, and the PCE, the children produced /k/ with 0% accuracy; however, on the SSS, the four children were differentiated: participants 5F and 8M scored a 21 and were not stimulable for /k/, which would be expected given the high number of errors on the other tests. Participant 7F scored a 15 for /k/ and could say /k/ in carrier phrases given all cue types. She was stimulable for /k/ and performed better than 5F and 8M for this phoneme. Particpant 3F demonstrated even better stimulability on /k/ with a score of 8, i.e., production of a one-target sentence given verbal model or instructions. For these four children the SSS, but not the other measures, differentiated their production of /k/.

In contrast, some children scored the same SSS score, but performed differently on the other measures. Participants 3F, 12F, and 9M all scored an 8, (i.e., one-target sentence with a verbal model or instruction), yet, participant 3F produced /k/ in error on all of the other assessments; 12F produced /k/ correctly twice on the other assessments, and 9M produced nine targets correctly. Although the children could all produce /k/ in a sentence with some help, the stability of single-word productions varied. The same was true for participants 7F and 10M who both scored a 15, carrier phrase with all cues; 7F produced all single-word productions in error on the other tests, yet 10M produced some /k/ targets correctly.

Another interesting difference in scores occurred for Participant 4M, who scored a 1 (best) on the SSS, yet still made some errors in single-word productions. This finding could be attributed to a number of explanations. The errors could be related to word-specific errors and differences in linguistic and phonetic environment. The target word "comb" on the SSS is selected for maximum stimulability; perhaps, the two and three syllable words such as "cowboy hat" and "candle" or clusters such as "queen" and "clouds" on the HAPP-3 posed greater difficulty. The error on the probe was a voicing problem and was substituted with another velar. Although 4M could produce /k/ in connected speech, the skill was still emerging and was not 100%. In relation to treatment, however, the SSS may inform when a skill has reached a maximum level of performance that will continue to improve on its own (Olswang & Bain, 1986). Although a score of 1 is not equivalent to 100% accuracy in connected speech, continued treatment may not be necessary for the target to continue to improve on its own. Further research must be conducted to document this relationship of the SSS score and treatment withdrawal.

The second data set for initial /k/ from Time 2 (see Table 4.1) also shows how the measures differ in the documentation of change. On the SSS, four participants' scores did not change for initial /k/ from Time 1 to Time 2: 5F and 4M were still not stimulable for /k/, participant 4M who had scored a 1 on the first administration, but had errors on the other assessments, scored 100% accuracy across all of the assessments; participant 3F still scored the same score of 8 on the SSS, yet her single word accuracy showed improvement on the Carter-Buck and the probe.

Table 4.1

Word-Initial /k/: Scores per assessment at Time 2. SSS ranges from 1-21 with a 1 being best, a + indicates that the child may have performed better, but the levels could not be assessed. On the Carter-Buck, probe, HAPP-3, and PCE, the errors are tallied over the number of opportunities. Error types are described below.

	SSS	Carter-Buck	Probe	HAPP-3	PCE
3F	8	1/3, 3/3, 2/3 t	1/4 t	2/2 t	5/5 t, d
4M	1	0/3, 0/3, 0/3	0/4	0/2	0/1
5F	21	3/3, 3/3, 3/3 ʔ^{h}	4/4 tʃ, ʔ^{h}, ʔ	2/2 ʔ^{h}, ʔ	1/1 ʔ
7F	18	3/3, 3/3, 3/3 del, t	4/4 del, t	2/2 del, t	6/6 del
10M	5+	0/3, 0/3 1/3 del	3/4 t	0/2	0/2
12F	6	3/3, 2/3, 2/3 d, g	4/4 d, del	2/2 t	1/1 b
8M	21	3/3, 3/3, 3/3 t, d	4/4 t, d	2/2 t,d	No opp.
9M	5+	0/3, 0/3, 0/3	0/4	0/2	0/3

In contrast, the remaining four participants did show change on the SSS. Two participants of interest are 7F and 10M because they originally had both scored a 15 on the SSS, but performed differently on the spontaneous measures. Participant 7F, who produced /k/ in error on the other measures, showed some regression on the SSS and her score changed from a 15 to an 18, whereas 10M who produced some words correctly on the single word assessments excelled in the second administration and scored a 5+ on the SSS, i.e., two-target productions with a verbal model or instruction with possible higher performance (a plus indicates that the child may produce the target with less support, however it could not be assessed because of language behaviors). In future research on treatment efficacy, the relationship between single-word spontaneous productions and stimulability should be further evaluated; in combination, the two show a strong pattern of change. Additional differences were seen for the remaining two participants, 12F and 9M. Participant 12F showed progress on the SSS, but not on the other measures, whereas participant 9M improved on all of the assessments (except for the Carter-Buck where he already demonstrated 100% accuracy). In summary, some children showed change on the SSS, but not the other measures. Some children showed change on the other measures, but not the SSS. One child showed change on all measures.

The performance for initial /k/ revealed interesting differences across the participants and measures; for comparison, initial /ʃ/ was scored across Time 1 and Time 2 for all participants and all measures (see Tables 4.2 and 4.3). As with initial /k/, four participants produced /ʃ/ in error 100% of the time on all assessments, but demonstrated different scores for the SSS: participants 10M and 8M scored a 20, 5F scored a 19, and 7F scored a 14 (see Table 4.2). Once again the SSS differentiates skills on the same phoneme for children who were not successful unless given some support.

Table 4.2

Word-Initial /ʃ/: Scores per assessment at Time 1. SSS ranges from 1-21 with a 1 being best. On the Carter-Buck, probe, HAPP-3, and PCE, the errors are tallied over the number of opportunities. Error types are described below.

	SSS	Carter-Buck	Probe	HAPP-3	PCE
3F	20	2/2, 3/3, 2/3 t,d	3/3 d, t	1/1 d	No opp.
4M	15	2/2, 2/3, 0/3 del, t	3/3 del	1/1 del	1/1 Del
5F	19	3/3, 3/3, 3/3 t	3/3 t	1/1 t	No opp.
7F	14	3/3, 3/3, 3/3 del	3/3 del	1/1 del	No opp.
10M	20	3/3, 3/3, 3/3 d,g	3/3 t, d	1/1 d	No opp.
12F	15	3/3, 3/3, 3/3 t, d, g	3/3 t,d,g	1/1 t	No opp.
8M	20	2/2, 3/3, 3/3 d	3/3 d, f	1/1 d	No opp.
9M	5+	0/3, 0/3, 0/3	0/3	0/1	No opp.

In contrast, three participants scored the same SSS score, but performed differently on the static assessments. Participants 3F, 10M, and 8M all scored a 20 on the SSS, i.e., production of /ʃ/ in isolation with all cues; however, only 3F scored differently on the other measures with the production of /ʃ/ in a syllable with verbal model on the Carter-Buck. Participants 4M and 12F both scored a 15 on the SSS, yet 4M could also produce /ʃ/ given a verbal model in a word and syllables on the Carter-Buck. The remaining participant, 9M, scored a 5+ on the SSS and produced /ʃ/ with 100% accuracy across the other measures.

Comparisons of change for initial /ʃ/ were also made between Time 1 and Time 2 across participants and measures. Five of the children's scores on the SSS improved on the second administration (3F, 4M, 5F, 10M, 8M), one child's score regressed slightly (7F), and two remained unchanged although one already produced /ʃ/ with 100% accuracy (12F, 9M). Participants 3F, 10M, and 8M, who all scored a 20 at Time 1, showed differences in change at Time 2. Participant 3F had produced an extra syllable correctly on the Carter-Buck at Time 1 only improved by 1 point on the SSS to a 19 at Time 2. It was thought that the extra productions on the Carter-Buck might be an indicator for greater improvement by the second administration, but this was not true. Participants 10M and 8M both changed from a score of 20 at Time 1 to a score of 15 in the second administration; in addition, 10M also improved by one point on the probe. All remaining productions of /ʃ/ remained in error for both participants.

In addition, Participants 4M and 12F, who both scored a 15 on the SSS at Time 1, also progressed differently on the second administration. Participant 4M, who produced /ʃ/ on the Carter-Buck at Time 1, completely acquired /ʃ/ by the second administration; he scored a 3+ on the SSS and produced /ʃ/ with 100% accuracy across the other assessments. Participant 12F, on the other hand, made no change on the SSS or any of the other measures and continued to produce /ʃ/ in error. It was hypothesized that perhaps 12F did not receive treatment for /ʃ/ as part of the cycle, yet she actually did practice /ʃ/ and worked on more stridents than 4M. When looking back at the treatment plans for both children, 4M received treatment for final /ʃ/ for one week of the cycle and 12F received treatment for initial /ʃ/ for one week of the cycle.

Table 4.3

Word-Initial /ʃ/: Scores per assessment at Time 2. SSS ranges from 1-21 with a 1 being best. On the Carter-Buck, probe, HAPP-3, and PCE, the errors are tallied over the number of opportunities. Error types are described below.

	SSS	Carter-Buck	Probe	HAPP-3	PCE
3F	19	2/2, 3/3, 3/3 d, t, dʒ	3/3 d, t	1/1 d	No opp.
4M	3+	0/2, 0/3, 0/3	0/3	0/1	No opp.
5F	11	3/3, 3/3, 3/3 t	3/3 t, tʃ	1/1 t	1/1 ts
7F	15	3/3, 3/3, 3/3 del, del, del	3/3 del	1/1 del	No opp.
10M	15	3/3, 3/3, 3/3 del, t, d	2/3 d	1/1 t	No opp.
12F	15	3/3, 2/2, 3/3 d, t, g	3/3 t,d,g	1/1 t	No opp.
8M	15	2/2, 3/3, 3/3 t	3/3 t,d	1/1 d	No opp.
9M	5+	0/3, 0/3, 0/3	0/3	0/1	0/1

Similar to word-initial /k/, four participants showed change in the production of word-initial /ʃ/ on the SSS, but not on the other experimental measures. Participants 8M, 10M, 5F, and 7F scored 0% accuracy on all the static assessments, but performed differently on the SSS at Time 2; all made progress on the SSS and three continued to remain at 0% accuracy on the other assessments. Only 10M showed a slight improvement on the probe with 1 correct production. Participants 8M and 10M both progressed from 20 to 15, 5F progressed from 19 to 11, and 7F regressed from a 14 to a

15. For three of these four participants, the SSS was the best indicator of change across the two administrations. The remaining participant, 9M, maintained a score of 5+ on the SSS and 100% accuracy across the other measures.

The results from the extracted data for /k/ and /ʃ/ suggest that there are differences in performance at the phoneme level between dynamic and static assessments. For several children, scores were different on the SSS, but the same on the static measures and vice versa, scores were different on the static measures but the same on the SSS. Overall, the combined scores could provide a better profile for the productive capabilities of children with phonological disorders.

Supplemental Research Question 2

How do the combined results across all of the experimental measures inform investigators about the different profiles of each participant?

The primary intent of this study was to evaluate the nature of phonological change by differentiating several experimental measures; however, it is also of interest to consider the performance of each child and how the combined results of the measures provide information about each individual's profile. In addition, the individual variables across children such as language and cognitive skills may influence their scores on these measures. In the following, first the children are presented independently, and second, the differences across the children are considered.

In Figure 4.0, the percentages of change calculated for the experimental measures are separated by participant and plotted together for comparison (see Figure 4.0). The different profiles of these children are summarized in Table 4.4.

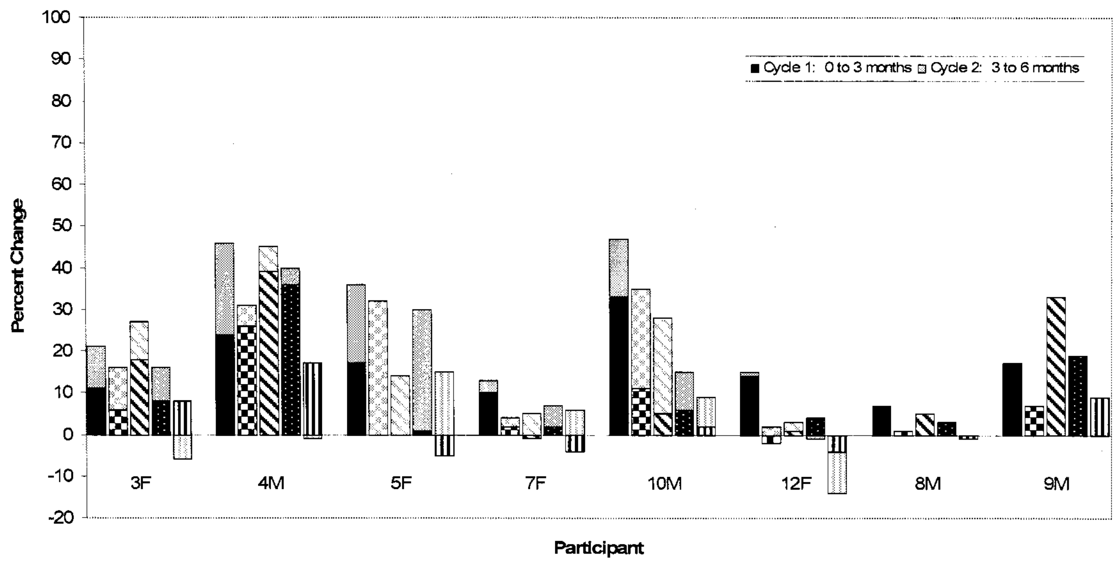

Figure 4.0. Percent change for each participant across all measures and cycles. The bars represent the measures in the following sequence: SSS (solid), Carter-Buck (squares), probe (diagonal stripes), HAPP (dots), PCE (vertical stripes).

Table 4.4

Summary of overall change across the entire treatment period.

	# of Tests that improved on both cycles	Highest Percent Change	Lowest Percent Change	Total Average Change**
3F	4/5	Probe	PCE	16%
4M	4/5	SSS	PCE	36%
5F	2/5	SSS	Probe	24%
7F	3/5	SSS	CB	6%
10M	5/5	SSS	PCE	27%
12F	2/5	SSS	PCE	1%
8M*	4/5*	SSS	PCE	3%
9M*	5/5*	Probe	CB	17%

Note: *8M and 9M were only treated for one cycle. **All total percentages of change were averaged across measures per participant.

The patterns of change for each participant were also evaluated across cycles and the differences for each participant are presented in Table 4.5.

Table 4.5

Patterns of percent change for six participants per cycle who received two cycles of treatment across measures. S=SSS, CB=Carter-Buck, H=HAPP-3, Pr=Probe, and Pc=PCC.

	Even change	Greater change for Cycle 1	Greater change for Cycle 2	Regression
3F	S, CB, H	Pr		Pc
4M	S	CB, Pr, H		Pc
5F	S		CB, Pr, H	Pc
7F	CB	S	H	Pr, Pc
10M	H	S	CB, Pr, Pc	
12F	Pr	S		CB, H, Pc

Overall, the combined assessments provide interesting information about the individual children. The children all met the same entry criteria, received the same amount of treatment, and were treated by the same clinician, yet the amount of change for each child was different. The question then becomes why do these children have such wide ranging differences during a six month period. In Table 4.4, the average percent change for all the assessments is shown for each participant. How is it possible that participant 12F only averaged 1% of change, whereas participant 4M averaged 36% change across the experimental measures? Several variables may have influenced the overall scores, including: age, severity, cognitive skills, language skills, unique phonological characteristics, or unique phonological profile.

One answer is that perhaps the age range was too wide for comparison with such a small group of children. The age of the children ranged from 37 to 54 months with an average age of 45.5 months; however, the age of the child seems to be unrelated to the percentage of change across measures. A Pearson correlation was calculated for age in months and percentage of change across measures, r=.2. The three children who showed the least change included: Participant 12F who was 50 months old, Participant 8M who was 37 months old, and Participant 7F was 49 months old. The three children who showed the most change included: Participant 4M who was 41 months old, 10M who was 48 months old, and 5F who was 54 months old. Correlations were also calculated for each of the original scores by measure and none were related to the age prior to treatment. With this small group of children, age did not appear to be a factor related to the amount of change made in treatment or to the scores on the first assessments.

The level of severity of a child's phonological disorder may also be a factor that influences the amount of change that a child makes in treatment. For the eight participants in this study, a relationship was not evident between the severity at the beginning of treatment and the percentage of change. Participants 5F and 7F were consistently ranked as most severe across the five measures; participant 5F made good progress with 24% overall change, whereas 7F only made 6% change. Even though the two participants started out with relatively similar severity, they did not perform in a similar manner. Participants 12F and 8M made the least overall percentage of change and their rankings of severity were different and inconsistent across the assessments in relation to the other children. Participants 4M and 10M made the greatest amount of change, yet 4M was often scored with a lower severity and 10M was of a mid-range severity. The two participants performed well, but did not have similar levels of severity from the beginning.

Another possible influence on the overall percentage of change may be cognitive skills. Interestingly, participant 4M scored the highest standard score of 110 on the *Peabody Picture Vocabulary Test* (PPVT), which is highly correlated with cognitive abilities, and he made the greatest overall percentage of change across the measures; furthermore, participant 12F scored the lowest standard score of 80 on the PPVT and made the lowest overall percentage of change across the measures. The remaining participants' scores seemed to be mixed throughout; some children scored higher on the PPVT and lower for percentage of change, and some children scored lower on the PPVT and yet made higher percentages of change.

Language skills could also be a factor in the changes made during treatment. Participant 4M again had the highest standard scores on the *Preschool Language Scale-4* with a 119 for receptive skills and a 118 for expressive skills and also made the most progress during treatment with 27% change. Participant 7F had the lowest standard scores with a 95 for receptive skills and an 84 for expressive skills on this assessment and only showed a 6% change. The other children's scores fell in between and were mixed in relation to progress and language abilities.

Unique phonological characteristics may be associated with or have created the differences in changes that the children made during the treatment period. Three children performed poorly overall: 12F, 7F, and 8M. Participant 12F made the smallest overall percentage of change. One aspect of her speech that was different from the other children was that she had a high percent of occurrence of backing. Backing is considered an atypical or deviant phonological process and may be an indicator of a child who will not respond as rapidly to treatment. Participant 8M was unique because his speech patterns were all characterized by a frontal lisp and an overbite. Some of his bilabial stop productions were made with his front teeth and lower lip. He often produced sounds incorrectly because of the forward tongue posture and the distance of his teeth. Participant 7F did not appear to show any specific differences related to sound production. Her only big difference from the other children was in the area of overall language skills.

Although participant 9M shows good change as an overall percentage, it was unusual that he performed better on the HAPP-3 than on the SSS (see Figure 4.0). Dysfluencies developed for this participant over the course of the cycle which may have negatively influenced productions that required sequencing of sentences and immediate repetitions of words as on the SSS, Carter-Buck, and PCE. The participant performed best on measures with spontaneous single-word productions: probe and HAPP-3.

Uncontrolled factors may have also influenced the children's percentage of change. Some parents may have practiced more with their children outside of treatment and given more feedback to their child regarding speech production. In two of the families, new siblings were born that may have disrupted family schedules. In addition, some parents were in the process of changes jobs, new caregivers were more actively involved, and two families moved homes during the treatment period.

A final consideration is that these children may demonstrate different categorical profiles that may require different types of treatment. Certain variables may combine that lead a child to respond better to one treatment type over another. Perhaps Cycles treatment was not the best choice for some children, whereas it was ideal for other children.

The purpose of the current study was not to evaluate the causation of change for these children; however, the scores and differences across the measures may have been influenced by some or all of the above factors. With better understanding of assessment and experimental measures used to document change, perhaps the influence of some of these variables can be experimentally manipulated in the future.

Supplemental Research Question 3

Are all of the environments and cues necessary on the SSS to document change in stimulability?

Supplemental Research Question 3 was the most frequently asked question when the preliminary results of the SSS were presented at conferences and group discussions; therefore it is addressed in this chapter. The SSS comprises a scale of 21 steps; however, are all of the cues and environments necessary to document change across time? In order to answer this question, each step of the hierarchy was counted for the number of times that it was documented as the final score. Next, the number of times each environment was used and the number of times that each cue was used were also counted. Only scores on error productions were evaluated and scores on targets that were already produced correctly by the child in connected speech were not considered. It was expected that the frequency of use would be the greatest for the most supported steps of the hierarchy and that spontaneous levels would be less frequent.

The entire hierarchy was evaluated, followed by assessment of the cues levels and environments. Figure 4.1 illustrates the number of times each level of the hierarchy was used as the final score across all tokens from all participants and administrations. The results showed that scores 2, 4, 6, 7, 9, 10, 12, 13, 16, and 17 were rarely or never used. Most of these scores are associated with the instructional level of cueing and it was interesting that they were rarely a stopping point on the SSS. Scores 2, 4, 7, 10, 12, and 17 in total were only used seven times. The low scores suggest that because the instruction level is not being used, that the level could be combined with verbal model and collapsed to reduce the overall scale by 6.

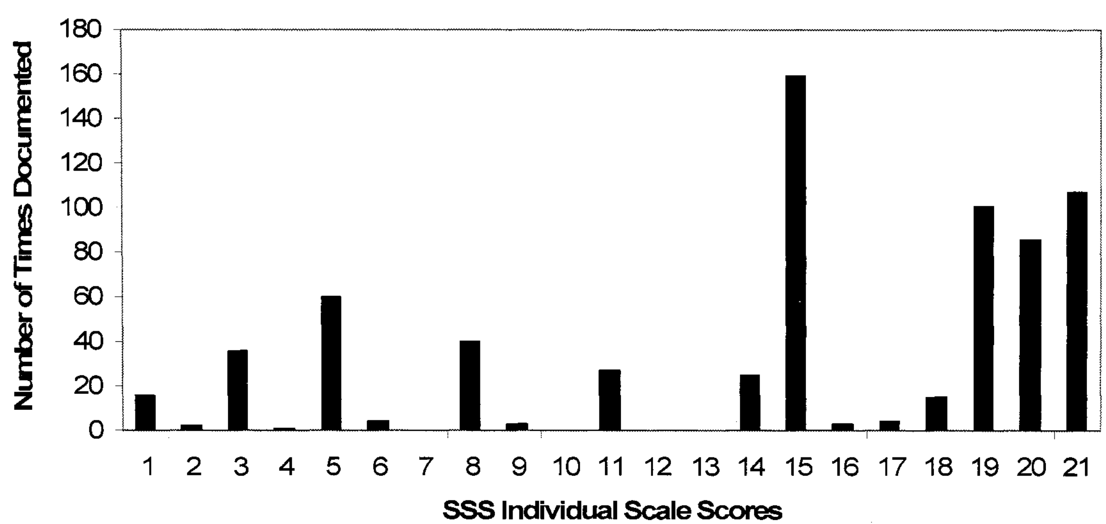

Figure 4.1. Scaffolding Scale of Stimulability: the number of times each level of the hierarchy was documented.

Cues and environments were then considered separately. The levels are supported across the hierarchy based on the number of times they were used. Across the cues presented in Figure 4.2, Cue Level 0 was used 62 times, Cue Level 1 was used 7 times, Cue Level 2 was used 167 times, and Cue Level 3 was used 453 times. The measure may be improved in several ways. First, Cue Level 1 could be removed because it was rarely used. Second, the high usage of Cue Level 3 suggests that the cues could be divided into subgroups. Currently any type of cueing is allowed at the point of Cue Level 3, but perhaps the level could be further divided into a hierarchy of cues. For example, segmentation and prolongation could be grouped together in one cue level, and simultaneous production and tactile cueing into another level. Overall, the hierarchy shows that there were more scores documented at the most supported cues than at the least supported cues (excluding cue level 1) which would be expected for children with phonological disorders.

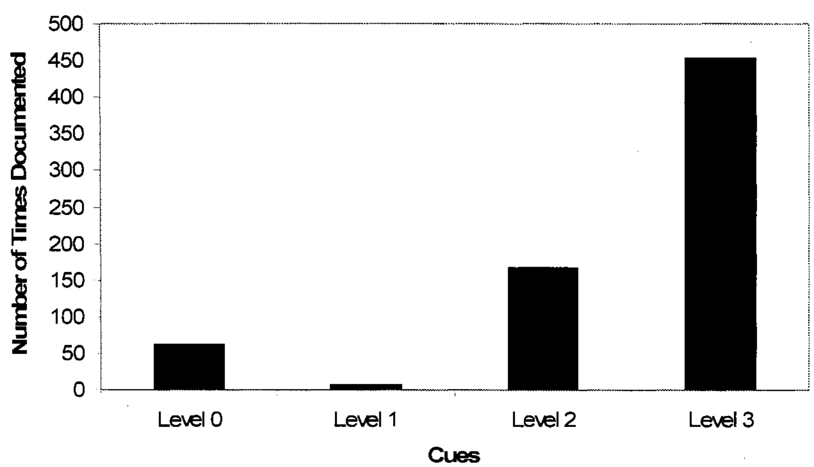

Figure 4.2. Scaffolding Scale of Stimulability: Total number of times that each cue level was scored across all tokens.

The environments were also assessed for the amount of usage for each point along the scale (see Figure 4.3). Each environment was used the following number of times: 107 not stimulable, 86 isolation, 119 word, 184 carrier phrase, 30 novel phrase, 44 one-target sentence, 97 two-target sentences, 18 picture scene. Several observations were made from these results. First, the environment of the carrier phrase was used the most overall. This high score could be related to the fact that there is a stopping rule at score 15 within carrier phrases. Any type of cueing is allowed in carrier phrases, however, cues from Cue Level 3 are not allowed in novel phrases or the above environments and testing is discontinued at this point. Second, the number of times that two-target sentences were used is a little misleading. This environment was used a total of 36 times, however, on 32 scores, it is possible that the child could have scored higher, but had difficulties with the language skills involved in the spontaneous production of connected speech. The targets were not in error, they just couldn't be assessed any higher. This also happened at a score of 5 where 12 times the score was a 5, but 48 times it is possible that the children could have scored higher. The score could be inflated and the environment of connected speech deflated as a result. Overall, the environments that provided the most support: carrier phrase, word, isolation, and not stimulable, were used with a much greater frequency than the environments of less support: connected speech, two-target sentences, one-target sentences, novel phrase.

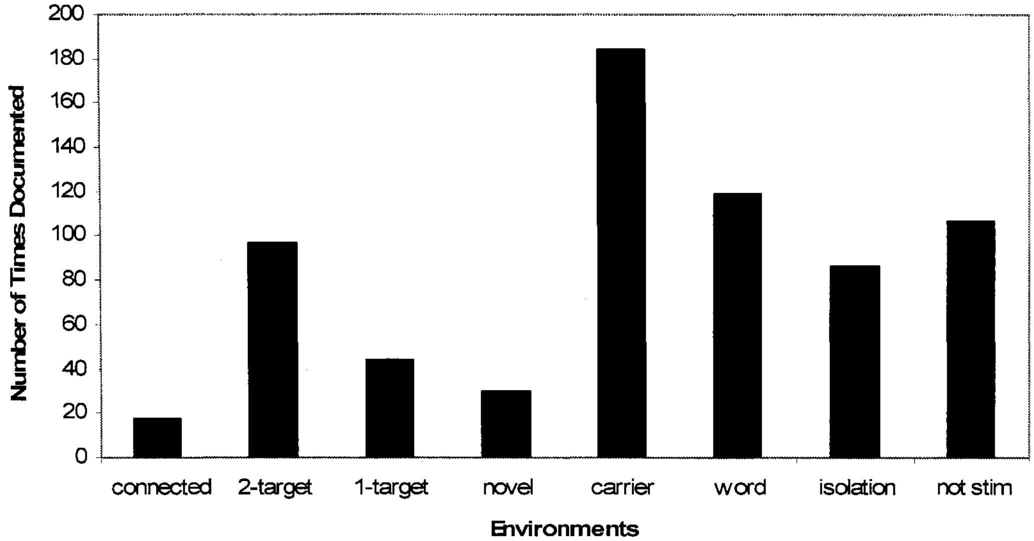

Figure 4.3. Scaffolding Scale of Stimulability: Total number of times that each environment was scored across tokens.

In evaluating all of the environments and cue levels across the SSS, it appears that not all of the items on the scale are necessary to document change across time based on stimulability. The Cue Level 1 could be eliminated or collapsed. The children also did not spontaneously produce carrier phrases, so this item could be removed. The cue levels could be further divided to show differences at the lower levels of the hierarchy that are now combined together. The next possible version of the SSS is represented in Table 4.6.

Table 4.6

Scaffolding Scale of Stimulability: Revised

	Cue Level 0 Spontaneous	**Cue Level 1** Instruction, Verbal Model	**Cue Level 2** Prolongation, Segmentation	**Cue Level 3** Simultaneous, Tactile
Not Stimulable				15
Isolation				14
Word	10	11	12	13
3-word Sentence	6	7	8	9
4-word Sentence	4	5		
2-target Sentence	2	3		
Connected Speech	1			

The new SSS continues to weigh the performance in a more complex environment over the types of cues that are given. For example, sometimes a child can produce a target in a sentence with cues, but cannot spontaneously produce a target in a word. The production of the target in a more complex environment will give the child a better score and outweighs the higher spontaneous score that was in error. One modification is that the carrier phrase and novel phrase environments were collapsed together into a three-word sentence that now includes three-word sentences with the target word as the last word in the sentence. The four-word sentence differs because the target word is embedded in the middle of the sentence. The two-target sentence and the connected speech sample remains the same. Looking at the cueing hierarchy, the instruction cues were collapsed with the verbal model and the segmentation and prolongation cues were separated from simultaneous production and tactile cues.

With these changes, the SSS is reduced to a 15-point hierarchy of stimulability. One benefit of reducing the scale is that it should now take slightly less time to administer. The children will not need to be led through extra and unnecessary cues and levels, yet should still maintain a progression of change over time.

CHAPTER FIVE: SUMMARY AND INTERPRETATION OF FINDINGS

Research Question 1

(1) How does performance change on five different measures of phonology during six months of treatment when each measure is considered independently?

Five measures were compared on three administrations across six months of treatment. Each measure reflected different improvements and patterns of change across the participants. Findings are as follows. (1) On the SSS, all participants' scores improved at every time interval. The general pattern of change was most often a steady, evenly distributed change across the two time periods. When a greater change occurred during a cycle it was always during the first cycle. (2) On the Carter-Buck, only six out of eight participants improved during the first cycle and all participants improved on the second cycle. The pattern of change was mixed between even change, large change during Cycle 1 or 2, and regression. (3) On the Probe, only six out of the eight participants progressed during the first cycle and all participants progressed during the second cycle. The patterns of change also varied between even change, large change during Cycle 1 or 2, and regression, although the uneven pattern of change favoring one cycle or the other was most common. (4) On the HAPP, all eight participants showed improvement during Cycle 1 and five out of six showed improvement during Cycle 2. The pattern of change was most frequently even or a large change favoring Cycle 2, but a large change during Cycle 1 and regression also occurred. (5) Finally, on the PCE, four out of eight children showed improvement during Cycle 1 and only three out of the remaining six children showed improvement during Cycle 2. The pattern of change on the PCE was most often regression, and only one child made improvements on both cycles with a large change during Cycle 2.

Five measures of phonological change were compared across participants and across six months of treatment; overall, the measures showed different patterns of change that suggest differences in sensitivity for each measure.

Research Question 2

(2) How does accuracy of production change on the Scaffolding Scale of Stimulability change in relation to the Carter-Buck Articulation Assessment?

The results of the Scaffolding Scale of Stimulability and the Carter-Buck Articulation Assessment were compared to determine whether two measures that were dynamic in nature assessed change in a similar manner, i.e., to compare concurrent validity. Contrary to the hypothesis, these two measures shared few similarities. The overall percentage of change was greater for all eight participants on the SSS compared to scores on the Carter-Buck. In addition, the patterns of change differed across the two measures. On the SSS, the pattern across two cycles was most likely to be even or show a greater change during Cycle 1. The Carter-Buck, in contrast, showed even change, a greater change during Cycle 1 or Cycle 2, and also regression for one participant. The SSS and the Carter-Buck correlated on the rank order correlation statistic, Kendall's *tau* b, across all three time periods.

Comparison of dynamic assessments to determine similarities and differences revealed that the SSS and the Carter-Buck were quite different. The only similarity between the two measures was that they both ranked the participants in a similar manner at each of the assessment periods as reflected by the Kendall's *tau* b correlations. The differences between the two measures included several factors. The SSS showed a higher percentage of overall change than the Carter-Buck across all of the participants. In addition, the SSS was sensitive enough that change was documented for every child on every cycle. The Carter-Buck was good at showing change, but not with as high of a frequency as the SSS, and even showed regression for one participant.

Another factor indicating that the SSS was a more sensitive measure is that the change on the SSS between two cycles was more likely to be even or favor the first cycle. The Carter-Buck showed large contrasts between the cycles which suggest that a particular skill had to be acquired before a change could occur on the measure and it was less incremental. The larger change during the first cycle on the SSS also suggests that it is more sensitive to changes in the earlier phases of treatment. In fact, the patterns of change on the Carter-Buck looked very similar to the HAPP-3. Thus, although the Carter-Buck provided two different environments and a verbal model cue, this was apparently not enough scaffolding to differentiate change across time on the Carter-Buck.

It was thought that differences between the SSS and the Carter-Buck could be attributed to the fact that the Carter-Buck assesses later developing targets; yet, when the SSS was revised to assess the same targets, the patterns of change resembled the original SSS scores and not the scores on the Carter-Buck. Furthermore, even when subtests of the Carter-Buck were considered, the patterns of change were still unlike those on the SSS. Assessing more environments and different cue levels resulted in differences in the documentation of change. Therefore, it was concluded that the Carter-Buck is not a dynamic assessment and that the SSS and the Carter-Buck are measuring different skills. Other than rankings across the participants, the Carter-Buck and the SSS do not show concurrent validity.

Research Question 3

How does accuracy of production on the SSS change in relation to the probe, the Hodson Assessment of Phonological Patterns, or the Percent Consonants in Error?

The results from the SSS and the probe, the HAPP-3, and the PCE were compared to determine whether a dynamic assessment shared a relationship with static assessments in the documentation of phonological change, i.e., to compare predictive validity. Findings are as follows. The scores on the probe showed the most unpredictable patterns of change of the three static assessments. The probe was the only measure that showed a higher overall percentage of change than the SSS for two children and then showed the lowest percentage of change of any measure for two other children. The relationship between the SSS and the HAPP was more consistent with the SSS showing a higher percentage of overall change for all participants. The SSS also showed a higher overall percentage of change than the PCE across all participants.

The patterns of change were different when comparing the assessments across both cycles. The SSS most often showed an even pattern of change or favored Cycle 1 with a greater change over Cycle 2. The probe most often showed greater change during either the first or the second cycle. The HAPP most often showed even change or a large change during the second cycle. The PCE was most unusual because periods of regression were the most common pattern. Kendall's *tau* b correlations suggested a predictive relationship between the scores on the first administration of the SSS and the second administrations of the probe, HAPP-3, and PCE. Rank orders became more consistent over time.

The results of the probe in comparison to the SSS were quite surprising. The probe captures changes in the production of phonemes at the spontaneous word level. It was thought that the SSS would consistently show more change than the probe because the children are given support; however, the percentage of change on the probe was greater than on the SSS for two children. At the same time, the probe showed the lowest percentage of change for two other children. Perhaps these differences occur because some of the children reached a phase where the production of phonemes in words was finally possible, causing a large jump in the percentage of change for the measure. The inconsistencies for this measure suggest that at times the probe can be more sensitive to change than the SSS, but over an extended period of time and across children, the SSS is more likely to show consistent steady increments of change, especially in the earlier phases of treatment.

The HAPP-3 is another measure that documents phonological change in spontaneous single-word productions. The HAPP-3 is most sensitive to changes from deletions to substitutions and acquisition of sound classes. The HAPP-3 may be less sensitive to change in some children; for example, children may have good HAPP-3 scores, yet still have difficulties producing words in sentences or connected speech, or they may have a poor HAPP-3 score and not show change because they cannot produce sounds in spontaneous words without support. As expected, the HAPP-3 consistently showed a lower percentage of change across administrations than the SSS. The lower percentage of change was expected because of the severity of the children's speech disorders and they were not given any scaffolding on this measure. The patterns of change were similar to those described above for the Carter-Buck. The results suggest that the HAPP-3 can be sensitive to change at times for some children, but does not show as much or as consistent of change as the SSS.

The final assessment for comparison with the SSS was the PCE. The PCE is a measure of children's productions in connected speech during a storytelling activity and has a high range of ecological validity. For change to occur on the PCE, children must be at a level of production where they can independently produce targets spontaneously and in connected speech. It was expected that because of the environmentally complexity of the speech sample that scores would be slow to change on the PCE. This assumption was true, the scores on the PCE were lower than the other measures, but what was most surprising was that periods of regression occurred across all of the children.

The PCE may have extraneous variables in the area of language development and use that influence the scores. Regression across PCE scores could be reflecting growth in language skills and the use of more complex linguistic forms. If a child produces longer phrases with multi-syllabic words, there is a high likelihood that the child will produce more phonemes in error. Another disadvantage of the PCE is that there are not always opportunities for children to produce all sounds. While the PCE holds high ecological validity because it measures accuracy in connected speech, some important information may be missing or overlooked. When these variables are not controlled, it is difficult to compare one sample to the next because the two samples may not be evaluating the same types of productions. Although this measure informs the clinician regarding productions in a natural environment, it is difficult to transfer the information to the treatment context. The PCE may be most beneficial for making diagnostic decisions and maintaining the need for continued treatment rather than documenting specific change across time.

A factor that affects all the measures is the amount of room left for growth should another cycle be administered, i.e., the potential for a ceiling effect across the measures. The SSS was the least likely to show a ceiling effect with well over 200 points of possible improvement, even for the best performing participant. The Carter-Buck, probe, PCE were more likely than the SSS to show ceiling effects, but still had some room for growth for even best performing participants. In contrast, however, the scores on the HAPP for some participants showed very little room to document continued change with less than 10 points remaining for the best score. Continued treatment would likely not differentiate these children although they could still exhibit speech errors.

Floor effects could also explain the patterns of change that were seen across several of the measures. The measures showed periods during which there was little or no change during the first cycle, and for many participants this was followed by a significantly greater change during the second cycle. The small amount of change early in the treatment process suggests presence of a floor effect for some children because they could not perform at an independent level that would be captured by the measures.

The SSS was the least likely to show floor effects because the change was most often even or there was a greater change during the first cycle.

The early phases of treatment are the most difficult in which to document incremental changes because of the persistent floor effects of most of the measures that are currently used. Being able to document change in the early phases of treatment, in particular, could allow for clinicians to make better judgments about treatment planning and better able to document treatment efficacy. In sum, the results suggest that the SSS is more sensitive to change across six months of treatment and that the change is more incremental than other phonological assessments. Furthermore, the SSS is less likely to show ceiling or floor effects that cause insensitivity in other measures.

Supplemental Research Questions

Three supplemental questions were addressed regarding differences in change by phoneme, individual profiles across the children, and the necessity of all 21 levels on the SSS. Supplemental analyses showed that when individual phonemes were compared across assessments, differences in the documentation of change occurred. The SSS differentiated children's performance and showed change by phoneme at times when the other assessments showed no change and 0% accuracy. On the SSS when a score of 1 occurred, some participants still had errors on the unsupported measures, yet when a score of 21 occurred on the SSS, the other measures consistently showed 0% accuracy. At times the scores on the SSS did remain the same across cycles when the static assessments showed improvements in accuracy; however, the SSS score was always at a less supported level than single-word productions of the static assessments that changed.

By analyzing the combined results of the assessments, it was observed that the participants differed in the amount of progress made across treatment. It was speculated that these differences could be related to variables such as age, severity, cognitive skills, language skills, and unique phonological differences. It was found that age and severity did not appear to be related to the amount of change that occurred; however, one child with high cognitive and language skills made the highest overall percentage of change. One child with low cognitive skills and one child with low language skills demonstrated the lowest overall percentage of change across assessments. Unusual phonological processes such as backing could have also accounted for the low score for one of these children. The child with the third lowest percentage of change overall also had structural differences with a forward tongue position and an over-bite that may have influenced the scores.

Finally, supplemental analyses evaluated the usage of the levels of the SSS scale and whether all 21 steps of the hierarchy were necessary. Results revealed that, in fact, not all of the levels were used. Analysis of the cues showed that Cue Level 1, instructional cues, was rarely a stopping point for the scoring of the SSS, whereas Cue Level 3 was quite high. Furthermore, analysis of the environments revealed that novel phrases were low in occurrence. In contrast, carrier phrases were high, possibly because of a stopping rule in combination with the cues. A revised scale was suggested that included 15 levels with a division of Cue Level 3 into two new levels, a combination of carrier phrase and novel phrase levels into 3-word sentences, and deletion of the instructional cue level.

Clinical Applications and Future Directions

One of the strongest advantages of the SSS is the direct applicability of the information gathered during assessment to treatment planning. The clinician may use the data from the SSS to choose targets based on stimulability levels, whether highly stimulable or unstimulable targets are chosen depending on the clinician's theoretical viewpoint. Although static measures may also be used to select treatment targets, they do not necessarily provide information relevant to planning the treatment session. Clinicians are often left uncertain about where to begin instruction, such as the types of cues or environments to use, until they actually interact more with the child. Static assessments only inform the clinician about an independent skill that is often unrelated to the daily activities of the treatment process. In many ways, the SSS is a sampling of treatment.

The SSS shows promise as a measure that is sensitive for documenting phonological change across time and may have potential clinical benefits in the study of treatment efficacy. However, questions remain regarding the feasibility of using the SSS in a "real-world" setting by speech-language pathologists. One issue that needs to be considered is the training process for speech-language pathologists. Currently, the SSS has only been administered by the investigator with 89% procedural reliability. Of concern is how readily other clinicians will learn to administer the SSS and achieve a high rate of reliability. Because the SSS is a dynamic assessment, the skill level of the clinician and the relationship between the clinician and the child will partially determine the child's score. Difficulties may arise as new clinicians administer the SSS, such as the inability to correctly sequence the hierarchy. Because of such difficulties, the SSS may require modification to improve its administration and usefulness. A training manual and assessment protocol must be developed so that other clinicians can accurately administer the SSS including strategies for behavior management, reinforcement, and motivation. One idea is to supplement the manual with a training video with which can learn the administration sequence.

Another issue regarding the SSS is the length of administration, which may take up to one hour. Many clinicians with high caseloads may not find it feasible to dedicate this much time to a phonological assessment. However, it should be noted that, because the SSS is a dynamic assessment, administration time includes both analysis and administration simultaneously; thus, extra time is not required for a detailed analysis. Other phonological assessments only take 10-20 minutes to administer, but may take up to an hour to analyze. The use of the SSS may be considered a more efficient use of time because assessment and analysis are combined and the information gathered may be more directly applied to the treatment process.

The length of administration of the SSS could be reduced by modifying portions of the assessment. If the scale includes fewer cues in the hierarchy as described in Chapter Four, the administration time would be shortened because there would be fewer steps in the overall procedure. Furthermore, the SSS could be reduced to include a core group of phonemes rather than all phonemes in initial and final position. Many of the phonemes elicited from the participants were not in error across any of the participants previously tested. These phonemes includes: /p, b, t, d, w, j, h/ in initial position. Perhaps phonemes that are rarely in error across children could be assessed only in special circumstances, such as with children who have a very limited phonemic inventory.

Future studies should continue to focus on issues relating to the reliability and the validity of the SSS, and subsequent studies may use the SSS to compare treatment outcomes. These issues include the stability of the SSS scores across time, the assessment of the incremental and quantitative aspects of the SSS scale, and the feasibility of documenting developmental norms.

One objective of future studies is to evaluate the stability of the SSS score across time. The SSS has only been administered in three month increments and it is uncertain how stable the scores might be during shorter periods of time, even from one week to the next without treatment. Along with this question of stability is another related question: does the administration of the SSS influence phonological skills? Because administration of the measure requires giving feedback and reinforcement, it is hypothesized that the dynamic nature of the SSS could potentially perform as a treatment if administered frequently enough. It may be beneficial in future studies to increase the frequency of the SSS to determine whether the SSS scores are stable across administrations and whether administration causes change.

A second objective in the evaluation of the SSS is to determine the extent to which items on the scale are incremental in nature. Currently the SSS has been considered an ordinal scale because it is unknown the extent to which a change from a 4 to a 3 is quantitatively the same as a change from a 17 to a 16. Also, because an environment is present in which no scaffolding may occur, the scale could in fact be a ratio scale, which would strengthen the validity of the measure. Statistical assessment using a Rasch analysis on a larger number of tokens and participants may better describe the quantitative attributes of the scale. The Rasch models "provide a foundation for the measurement of quantitative attributes and traits on a continuum, based on categorical data derived from interactions between persons and items (Wikipedia, 2006a)." Furthermore, Guttman scaling may also help assess the implicational hierarchy of the SSS scale, where a score of 15 implies that the child can successfully produce a target at levels 16-21 (Wikipedia, 2006b). Overall, additional analyses across a larger group of participants may further support the SSS as an incremental and quantitative measure.

The third objective for evaluating the SSS is to examine the potential for establishing a developmental continuum of stimulability which might reflect levels of severity. Although the current objective of the SSS has been as a measure of change, with continued study, it may have potential for diagnosis and documentation of the severity of disorder. Scores may be analyzed to determine whether a particular range is associated with severity levels; furthermore, comparison with younger and typically developing children may be documented to better show how stimulability develops over time.

The ultimate objective of future study is to begin to evaluate treatment efficacy. Once the reliability and validity of the SSS are better established, the next step in the research process will be to use the SSS to compare the efficacy and effectiveness of various treatment programs and the components within treatment programs. Clinicians make numerous decisions each day regarding the design of the treatments they administer, yet little is known about the efficacy of their choices. Variables that should be manipulated experimentally include: target selection such as stimulable and non-stimulable targets, the number of treated patterns selected during treatment, and the dosage of treatment. Experimentation with these variables using the SSS to document differences may ultimately lead to improved evidence-based practice.

BIBLIOGRAPHY

Bain, B. (1994). A framework for dynamic assessment in phonology: Stimulability revisited. *Clinics in Communication Disorders, 4,* 12-22.

Bain, B. A., & Olswang, L. B. (1995). Examining readiness for learning two-word utterances by children with specific expressive language impairment: Dynamic assessment validation. *American Journal of Speech-Language Pathology,* 4, 81-92.

Bernthal, J. C., & Bankson, N. W. (2004). *Articulation and phonological disorders* (5th ed.). Boston: Allyn & Bacon.

Bauman-Waengler, J. (2000). *Articulatory and phonological impairments: A clinical focus*. Needham Heights, MA: Allyn & Bacon.

Billingsley, F., White, O. R., & Munson, R. (1980). Procedural reliability: A rationale and example. *Behavioral Assessment, 2,* 229-241.

Campbell, T., Dollaghan, C., Rockette, H., Paradise, J., Feldman, H., Shriberg, L., Sabo, D., & Kurs-Lasky, M. (2003). Risk factors for speech delay of unknown origin in 3-year-old children, *Child Development, 74,* 346-357.

Carter, E., & Buck, M. (1958). Prognostic testing for functional articulation disorders among children in the first grade. *Journal of Speech and Hearing Disorders, 22,* 124-133.

Curry, R., Kennedy, L., Wagner, L., & Wilke, W. (1943). A phonographic scale for the measurement of defective articulation. *The Journal of Speech Disorders, 8,* 123-126.

Diedrich, W. M. (1983). Stimulability and articulation disorders. *Seminars in Speech and Language, 4,* 297-311.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (3rd Ed.). Circle Pines, MN: American Guidance Service.

Farquhar, M. S. (1961). Prognostic value of imitative and auditory discrimination tests. *Journal of Speech and Hearing Disorders, 26,* 342-347.

Gierut, J. A. (1996). Categorization and feature specification in phonological acquisition. *Journal of Child Language, 23,* 397-415.

Gierut, J. A. (1998). Treatment efficacy: Functional phonological disorders in children. *Journal of Speech, Language, and Hearing Research, 41*, S85-S100.

Gierut, J., Elbert, M., & Dinnsen, D. (1987). A functional analysis of phonological knowledge and generalization learning in misarticulating children. *Journal of Speech and Hearing Research, 30,* 462-479.

Gierut, J., Morrisette, M., Hughes, M., & Rowland, S. (1996). Phonological treatment efficacy and developmental norms. *Language, Speech, and Hearing Services in Schools, 27,* 215-230.

Glaspey, A. M., & Stoel-Gammon, C. (2001, November). *Stimulability as a measure of phonological change.* Poster session presented at annual convention of the American Speech-Language-Hearing Association, New Orleans, LA.

Glaspey, A. M., & Stoel-Gammon, C. (2002). *Sensitivity to phonological change: Stimulability, intelligibility, and generalization measures.* Paper presented at the annual convention of the American Speech-Language-Hearing Association, Atlanta, GA.

Glaspey, A. M., & Stoel-Gammon, C. (2005). Dynamic assessment in phonological disorders: The Scaffolding Scale of Stimulability. *Topics in Language Disorders, 25,* 220-230.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe 2 test of articulation.* Circle Pines, MN: American Guidance Service, Inc.

Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. (1999). *The scientist practitioner.* Boston, MA: Allyn and Bacon.

Hodson, B. (1986). *The assessment of phonological processes-revised.* Austin, TX: Pro-Ed/Interstate.

Hodson, B. (2004). *The Hodson assessment of phonological patterns.* Austin, TX: Pro-Ed/Interstate.

Hodson, B., & Paden, E. (1991). *Targeting intelligible speech: A phonological approach to remediation.* Austin, TX: Pro-Ed.

Howell, J., & Dean, E. (1994). *Treating phonological disorders in children: Metaphon—Theory to practice* (2nd ed.). London: Whurr.

Irwin, R. B., West, J., & Trombetta, M. (1996). Effectiveness of speech treatment for second grade children with misarticulations—Predictive factors. *Exceptional Children, 32,* 417-479.

Kisatsky, T. J. (1967). The prognostic value of the Carter-buck Tests in measuring articulation skills of selected kindergarten children. *Exceptional Children, 34,* 81-85.

Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment.* New York: The Guilford Press.

Lippke, B. A., Dickey, S. E., Selmar, J.W., & Soder, A. L. (1997). *The photo articulation test* (3rd. Ed.). Austin, TX: Pro-Ed.

Lof, G. L. (1996). Factors associated with speech-sound stimulability. *Journal of Communication Disorders, 29,* 255-278.

Lowe, R. (1986). *Assessment link between phonology and articulation.* Moline, IL: LinguiSystems, Inc.

Miccio, A. W., Elbert, M., & Forrest, K. (1999). The relationship between stimulability and phonological acquisition in children with normally developing and disordered phonologies. *American Journal of Speech-Language Pathology, 8,* 347-363.

Milisen, R. (1954). A rationale for articulation disorders. *Journal of Speech and Hearing Disorders, Monograph Supplement, 4,* 5-17.

Morrison, J. A., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Research, 35,* 259-273.

Olswang, L. B., & Bain, B. A. (1986). Monitoring phoneme acquisition for making treatment withdrawal decisions. *Applied Linguistics, 6,* 17-37.

Peña, E., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology, 10*, 138-154.

Perrine, S. (1999). *Dynamic assessment of phonological stimulability: Construct validation of a cueing hierarchy.* Unpublished master's thesis, Idaho State University, Pocatello, Idaho.

Powell, T. W., Elbert, M., & Dinnsen, D. (1991). Stimulability as a factor in the phonological generalization of misarticulating preschool children. *Journal of Speech and Hearing Research, 34,* 1318-1328.

Powell, T. W. (1996). Stimulability considerations in the phonological treatment of a child with a persistent disorder of speech-sound production. *Journal of Communication Disorders, 29*, 315-333.

Powell, T. W., Miccio, A. W. (1996). Stimulability: A useful clinical tool. *Journal of Communication Disorders, 29,* 237-253.

Robbins, J., & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders, 52*, 271-277.

Rvachew, S., Rafaat, S., & Martin, M. (1999). Stimulability, speech perception skills, and the treatment of phonological disorders. *American Journal of Speech-Language Pathology*, *8,* 33-43.

Scott, D. A., & Milisen, R. (1954). The effect of visual, auditory, and combined visual-auditory stimulation upon the speech responses of defective speaking children. *Journal of Speech and Hearing Disorders, Monograph Supplement, 4,* 38-43.

Secord, W. A. (1989). The traditional approach to treatment. In N.A. Creaghead, P.W. Newman, & W. A. Secord (Eds.), *Assessment and remediation of articulatory and phonological disorders* (2nd ed., pp. 129-153). Columbus, OH: Merrill.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders,* 47, 256-270.

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.

Snow, K., & Milisen, R. (1954). The influence of oral versus pictorial presentation upon articulation testing results. *Journal of Speech and Hearing Disorders, Monograph Supplement, 4,* 29-36.

Sommers, R. K., Leiss, R. N., Delp, M. A., Gerber, A. J., Fundrella, D., Smith, R. M., et al. (1967). Factors related to the effectiveness of articulation treatment for kindergarten, first, and second grade children. *Journal of Speech and Hearing Research, 10,* 428-437.

Strand, E. A., & McCauley, R. J. (1999). Assessment procedures for treatment planning in children with phonologic and motor speech disorders. In A. J. Caruso & and, E. A. (Eds.), *Clinical management of motor speech disorders in children* (pp.73-107). New York, NY: Thieme Medical Publishers, Inc.

Stoel-Gammon, C. (1985). Phonetic inventories, 15-24 months: A longitudinal study, *Journal of Speech and Hearing Research, 28,* 505-512.

Stoel-Gammon, C., Stone-Goldman, J., & Glaspey, A. (2002). Pattern-based approaches to phonological therapy. *Seminars in Speech and Language, 23,* 3-13.

Travis, L. E. (1931). *Speech pathology.* New York: D. Appleton and Company.

Velleman, S. (1998). *Making phonology functional.* Boston, MA: Butterworth-Heinemann.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1986). *Thought and language.* Cambridge, MA: The MIT Press.

Williams, A. L. (2000). *Speech disorders resource guide for preschool children.* Clifton Park, NY: Singular Publishing Group.

Wikipedia. (2006a). Guttman Scale. *Wikipedia, the free encyclopedia.* Retrieved April 30, 2006, from http://en.wikipedia.org/wiki/Guttman_scale

Wikipedia. (2006b). Rasch Model. *Wikipedia, the free encyclopedia.* Retrieved April 30, 2006, from http://en.wikipedia.org/wiki/Rasch_model

Zimmerman, I. L., Steiner, V.G., Pond, R. E. (2002). *Preschool language scale* (4th ed.). San Antonio, TX: Psychological Corporation.

APPENDIX A

SLP Recruitment Letter

April 30, 2004

Dear Speech-Language Pathologist:

I am writing to ask your help in recruiting children with phonological disorders to participate in a University of Washington research study. The study will examine a new measure that I have developed for assessing phonological change during treatment. The measure is based on cueing the child needs and environmental complexity rather than independent acquisition of sounds at the word, sentence, or conversation level. I will conduct therapy for children whose families choose to participate for six months at the UW Speech and Hearing Clinic. After the study period, the children will be referred back to the clinic waiting list.

In order to recruit these subjects I am asking you to:

1. Distribute information letters to families with children who meet the following criteria:

    1. Moderate, severe, profound phonological disorder
    2. Ages 3:0-5:11
    3. Monolingual English-speaking
    4. Intelligibility judged below 80%
    5. Enrolled in treatment less than 6 months or not yet enrolled
    6. Receptive language within normal limits
    7. No apparent delays in other developmental areas (motor, cognitive, adaptive, . . .)

2. Collect returned letters from families.

3. Once parents have contacted me and given their permission for a Mutual Exchange of Information, submit a summary of therapy and testing completed within the last six months. I will use the summary of test results to help determine the child's eligibility for the study, reduce duplicate testing, and create a new treatment plan.

Attached please find copies of the letter and response forms. If you have any questions, please contact me at 206-931-7309 or aglaspey@u.washington.edu. (Please be aware that I cannot ensure the confidentiality of information sent via e-mail).

Sincerely,

Amy M. Glaspey, Ph.C., CCC-SLP

APPENDIX B

Parent Recruitment Letter

April 30, 2004

Dear Parent:

Your child's speech-language pathologist is forwarding this letter to you because your child may qualify for participation in a University of Washington research study. I am conducting a speech treatment study at the University of Washington Speech and Hearing Clinic. I am a doctoral student from the Department of Speech and Hearing Sciences at the University of Washington. I am studying how children's speech sounds develop when they are in speech therapy. I have made a new test to measure progress and would like to study how well my test works when compared to other tests that are being used by speech-language pathologists.

I am looking for children with the following characteristics:

- Aged three to five years old,
- English speaking,
- Have difficulty making clear speech sounds,
- Understand what is said them,
- Good hearing,
- Not yet enrolled in a speech treatment program or enrolled in speech treatment less than six months.

Families who are eligible and wish to participate in the study would be asked to bring their child to the University of Washington Speech and Hearing Clinic for speech therapy services. Your child would receive individualized speech treatment from me two times per week for 50-minute sessions for six months. The study procedures would be audiotaped and videotaped so that I can document speech development. I am a speech-language pathologist who is certified by the American-Speech-Language and Hearing Association and licensed by the State of Washington Department of Health. The results of the study will help further the development of my measure and may lead to improved treatment programs for children with delayed speech.

If you choose to have your child participate in the study, you would not be charged for treatment during the six months of participation. At the end of the study if your child continues to need speech treatment, you may choose to place your child on the waiting list at the UW Speech and Hearing Clinic for possible continued treatment. Treatment is not guaranteed by UW Speech and Hearing Clinic at the end of the study and is based on the same selection guidelines and equally as any other client. If treatment services are available for your child at that time, then you will be charged regular clinic fees for services. Because the UW Speech and Hearing Clinic is a teaching program, students may observe treatment sessions during the course of the study. For questions regarding enrollment as a client in the Speech and Hearing clinic you may call the clinic office at (206) 543-5440; 4131 15th Ave NE, Seattle, WA 98105.

If you would be interested in gaining more information about this project, please call me at 206-931-7309, email at aglaspey@u.washington.edu, or sign the enclosed letter and return it to your child's speech-language pathologist. (Please be aware that I cannot ensure the confidentiality of information sent via e-mail). I will call you at a convenient time to discuss the details of the project with you. If you do not call or return the enclosed letter, you will not receive further contact from me. Your participation is strictly voluntary. Thank you for your time.

Sincerely,

Amy M. Glaspey, Ph.C., CCC-SLP
Speech-Language Pathologist and Doctoral Student

APPENDIX C

Parental Consent Form

Stimulability as a Measurement of Phonological Change

Researchers: Amy M. Glaspey, Ph.C., CCC-SLP
Doctoral Student, Department of Speech and Hearing Sciences
University of Washington Speech and Hearing Clinic
(206) 543-5440 or (206) 931-7309
aglaspey@u.washington.edu

Carol Stoel-Gammon, Ph.D.
Professor and Chair, Department of Speech and Hearing Sciences
(206) 543-7974
csg@u.washington.edu

We are asking your child to be in a research study. The purpose of this consent form is to give you the information you will need to help you decide whether your child should be in the study or not. Please read the form carefully. You may ask questions about the purpose of the research, what we would ask you to do, the possible risks and benefits, your rights as a volunteer, and anything else about the research or this form that is not clear. When we have answered all your questions, you can decide if you want your child to be in the study or not. This process is called 'informed consent.' We will give you a copy of this form for your records.

PURPOSES AND BENEFITS

The purpose of this study is to find out more about how children learn during their speech treatment. Your child can be in this study because she or he has a phonological delay or disorder. We would like to monitor your child's progress during his or her treatment. If you want your child's individual results, we will provide them to you at the end of each quarter as part of your child's quarterly progress report. We hope that the results of this study will help us develop better ways of treatment to children in the future.

PROCEDURES

Your child will participate in regular treatment practices at the University of Washington Speech and Hearing clinic using a cycles treatment methodology. The investigator, Amy Glaspey, who will work with your child, is a speech-language pathologist who is certified by the American Speech-Language-Hearing Association and licensed by the Washington State Department of Health. The following measures, which are part of regular treatment practices, will be used as part of this study (Estimated administration times are given after each measure. Time may vary depending on the skill level, cooperation, and age of your child):

1. Evaluation of language skills using the Preschool Language Scale IV. (25-45 minutes)
2. Evaluation of understanding of vocabulary using the Peabody Picture Vocabulary Test. (10-15 minutes)
3. Hearing screening. (10-15 minutes)
4. Oral-motor screening. Your child will be asked to make movements with the face and tongue to assess the functioning of the muscles used for talking. (10-15 minutes)
5. Percentage of correct sounds in words calculated by listeners of your child's speech during a connected speech sample. At the beginning and end of each treatment cycle, your child would be given a book without words and asked to tell the story or asked to tell about his/her own favorite book or story. One to two hundred words would be recorded and then unbiased listeners will count how many of the sounds were correct for an overall percentage. (5-10 minutes per administration)
6. Evaluation of speech sounds and patterns using The Hodson Assessment of Phonological Processes-3. This is a measure given at the beginning and end of each treatment cycle to measure long range changes in the speech system. Your child would be shown 50 objects and ask to name them. (15-20 minutes)
7. Evaluation of speech sounds using the Carter-Buck Articulation Assessment. At the beginning and end of each treatment cycle, your child will be asked to say 100-150 words and syllables. When the child says a sound in a word incorrectly, the clinician will say the word and ask the child to repeat it. (20-30 minutes per administration)

With your permission, the results of testing completed within the last six months by other speech-pathologists may be substituted. You will need to sign a Mutual Exchange of Information Form if you would like us to communicate with your child's previous speech-language pathologist(s). This communication may reduce the amount of testing that your child would need to complete and reduce the fatigue that your child may feel from testing.

The following measures are unique to this study and would be administered after your child's treatment sessions if your child chooses to be in this study (with your permission):

8. List of untreated words. Once per week or every other week, your child will be shown 50-70 picture cards and asked to say the pictures. These pictures feature sounds that are being worked on by your child, but in words that are not treated. (10-20 minutes per administration)

9. Rating Scale. At the beginning and end of each treatment cycle, your child will be asked to say all sounds by themselves, in words, or in sentences. Your child will be given verbal modeling and instructions to help him/her say the sounds correctly to the highest level (i.e. sentences without any model) possible. Then as each sound is practiced during treatment, the highest level of practice will be documented. (20-60 minutes per administration)

We would like to videotape and audiotape the study procedures. The video camera and/or audio recorder will be set up in the treatment room. We may also want to use the recordings for future related research or for educational purposes. Although the recordings will not be labeled with your child's name, someone who knows your child may be able to identify her or him from the recordings. We will provide an opportunity for you to review and edit the recordings, and you will provide your written consent before we use the recordings in a public setting.

You will be given any information related to progress of your child's speech development and evaluation as it is part standard treatment procedures.

You will not be charged for treatment or evaluation measures as part of this study; however, if you continue treatment with the UW Speech and Hearing Clinic after the study ends you will be charged standard treatment fees at that time. Treatment services at the UW Speech and Hearing are not guaranteed at the end of the study. You may choose to be placed on the waiting list at the UW Speech and Hearing Clinic and your child will be selected under the same criteria as other clients. Because the UW Speech and Hearing Clinic is a teaching clinic, students who are learning to be speech therapists or taking a course through our department may observe your child from the observation room during treatment. They may ask the investigator questions about how to teach children speech sounds and may hear his or her name. Students have signed a confidentiality statement or been trained in ethics and confidentiality.

We would like to follow your child's treatment for six months. However, if your child improves quickly and no longer needs treatment, we will recommend ending treatment and participation in the study.

RISK, STRESS, AND DISCOMFORT

There are no known risks associated with any of the measures planned for this study. Your child may become tired while saying or repeating words. S/he will be encouraged through play, positive reinforcement, or small rewards such as stickers.

OTHER INFORMATION

Participation in this study is voluntary. Your child may stop at any time. Information about your child is confidential. Study records will be coded with personal information kept in a separate secured location. Only the study investigators will have access to the study records. The link between the study records and your child's name will be kept until all the study information has been analyzed and will be destroyed no later than December 2005. After the study is completed, if you would like to send your child's results from treatment to your child's new therapist, you will need to give your permission by signing a Mutual Exchange of Information form. You can contact Ms. Glaspey and request records to be exchanged until the data links are destroyed in December 2005. The study records and audio/video recordings will be stored indefinitely for potential use during future studies with your permission after you have had an opportunity to review and edit the tapes. The recordings of your child and the study information may be used for purposes not described in this consent form. In that case, we will send our request to the University of Washington Human Subjects Division. This division protects the rights and welfare of research subjects like your child. The board will determine if we need to contact you and ask your consent to do the research. If we publish the results of this study, we will not use your name or your child's name.

We would like to give you the opportunity to meet with a member of the research team to receive your child's individual progress. Please indicate below if you want to receive your child's individual progress results as documented by research.
______ I would like my child's individual progress report.
______ I do NOT want to receive my child's individual progress report.

Amy M. Glaspey, Ph. C., CCC-SLP Date
Doctoral Student/Investigator

Subject's Statement

This study, **Stimulability as a Measurement of Phonological Change,** has been explained to me. I volunteer consent to allow my child ____________________________ to participate. I have had a chance to ask questions. If I have questions later about the research, I can ask one of the researchers listed above. If I have questions about my rights as a research subject, I may call the Human Subjects Division at (206) 543-0098. I give my permission for the research team to audiotape and videotape the study procedures as described above in this consent form. I will receive a copy of this consent form.

Participant's Parent (print): ____________________________________

Participant's Parent Signature: ____________________________________

Date: ________________

APPENDIX D

Assent Form

Stimulability as a Measurement of Phonological Change

Researchers: Amy M. Glaspey, Ph.C., CCC-SLP
Doctoral Student, Department of Speech and Hearing Sciences
University of Washington Speech and Hearing Clinic
(206) 543-5440 or (206) 931-7309
aglaspey@u.washington.edu

Carol Stoel-Gammon, Ph.D.
Professor and Chair, Department of Speech and Hearing Sciences
(206) 543-7974
csg@u.washington.edu

PURPOSE AND BENEFITS

I wanted to check with you and see if you would help me. I want to study about how kids learn to talk clearly and what teachers can do to help them. That way, kids can learn faster and they don't have to work so hard on their speech. Hopefully, this will help me to be a better speech teacher, too.

PROCEDURES

If it's okay with you, you'll stay a little later after your regular speech class, look at some pictures, and tell me what they are called. Then you'll practice saying some words and sentences like you do during your speech class. When it's time for a holiday break or when you come back you'll practice saying a longer list of words and sentences, but that's only once in a long while. We want to take videos and audio recordings of you talking.

RISK, STRESS, AND DISCOMFORT

We'll play some games and you can earn some stickers for saying all of the words and sentences, and maybe even a prize. This will make saying the words more fun so you don't get tired.

OTHER INFORMATION

You don't have to be in our study if you don't want to. We won't tell your name to anyone that you were here saying words, sentences, and working on your speech—but you can tell anyone you would like. We might show the videos of you talking to other speech teachers are learning how to be better teachers. You get to keep a copy of this paper.

____

Amy M. Glaspey, Ph.C., CCC-SLP Date
Doctoral Student/Investigator

Subject's Statement

I have been told about this study. I want to be a part of this study. I have had a chance to ask questions. If I have more questions, I can ask my teacher or his/her "principal."

Participant (print): ______________________________

Participant Signature: ______________________________

Date: ______________

APPENDIX E

Video Consent Form

Department of Speech and Hearing Sciences
University of Washington

Video Consent Form
Stimulability as a Measure of Phonological Change

Investigators:
Amy M. Glaspey, Ph.C., CCC-SLP, Doctoral Student, Department of Speech and Hearing Sciences, College of Arts and Sciences (206) 931-7309
Carol Stoel-Gammon, Ph.D., Professor and Chair, Department of Speech and Hearing Sciences, College of Arts and Sciences (206) 543-7974

Investigator's Statement

PURPOSES AND BENEFITS

This consent form is to allow us to videotape your child for research and education purposes as part of this study. The research purpose is to enable us to view your child's language and social behaviors for coding our data at a later time and by an independent researcher. The education purpose is to enable us to develop a resource library for the Department of Speech & Hearing Sciences. This library will hold recordings/photographs of clinical patients and research participants who have given consent for these recordings and/or photographs to be used for educational purposes. This Video Resource Library will greatly enhance teaching in Speech & Hearing Sciences by providing a wealth of teaching material that cannot be obtained through textbook descriptions of speech and language impairments and intervention. While this Video Resource Library will probably not benefit you (or your child), it will greatly improve the education of speech-language pathology students who are trained through this university.

PROCEDURES

The investigators listed above would like to use recordings that will be made of your child during this study. If you consent, these recordings will be used for the research study in which your child is participating and placed in the Speech & Hearing Sciences Video Resource Library for educational purposes. There will be two separate collections of these recordings in the Video Resource Library: a "departmental" collection and a "public" collection. In general, only faculty, staff, and speech and hearing sciences students will view the Departmental Collection for the purpose of education. Recordings or images in the Public Collection may be used in educational activities for the general public.

We would like your consent to the following specific uses of the video recordings made of your child.

Research Purposes:

- Viewing/listening by University of Washington faculty, students, and staff for research purposes, as described in the Parent Consent Form

Educational Purposes- Video Resource Library Departmental Collection:

- Viewing/listening by University of Washington faculty, students, and staff for education (e.g., course presentations, lectures, etc.)
- Viewing/listening by participants in any educational activities at the discretion of Speech & Hearing Sciences faculty and staff. The educational activities include conferences and workshops attended by students, professionals, and caregivers. The recordings may be shared through any medium, provided it is not available to the general public

Educational Purposes-Video Resource Library Public Collection:

- Viewing/listening by the general public at activities or through media sponsored or licensed by the University of Washington, or its faculty or staff (e.g., Internet/World Wide Web, University of Washington TV).
- Viewing/listening by the general public through licensed commercial enterprises for educational or research purposes, for example a CD-ROM enclosed in a textbook.

If you have any questions or concerns about the recordings or their use, you are free to restrict the uses described above.

If you agree to the use of the recordings for research purposes, please put your initials here _____.

If you agree to the use of the recordings for our Video Resource Library Departmental Collections for educational purposes, please put your initials here _____.

If you agree to the use of the recordings for our Video Resource Library Public Collection, please put your initials here _____.

Restrictions will not affect your child's participation in any current or future research studies or clinical services in the Department of Speech and Hearing Sciences.

Videotaping your child is completely voluntary. Participants are free to withdraw their recordings from this study at any time without penalty and without jeopardizing future services at the University of Washington. After videotaping, you will be free to review the tapes and delete any parts you wish to delete.

RISKS, STRESS DISCOMFORT

Your child may feel an invasion of privacy as he/she is being video recorded. Please remember that these recordings will not be associated with your name or your child's name, or any other identifying information. Your child is free to stop the recording at any time. You are free to withdraw any recording at any time.

OTHER INFORMATION

Research Purposes: The videotapes and the data from this study will be organized by assigning a number to each participant. Assignment of a number will occur at the conclusion of the first session. This number will be used to catalogue all videotapes and the data that we code from them. Only the investigators (listed above) will have access to the names of participants. The Principal Investigators will maintain the master list of participants' names and numbers for five years. This ensures that all information (videotapes and data) will remain confidential. The videotapes and data with numbers as identifiers will be kept in a locked room where they will remain indefinitely. We may use these data for future studies. The final results of the research will be shared with you upon request. If you wish additional information, such as results of tests or copies of videotapes, you may request them as well. Information will only be shared with you.

Educational Purposes: If you give your consent, the recordings from this study will be organized by a number assigned to each individual, so that the identity (i.e., names) of the participants will be available only to project staff. The primary investigator will maintain the master list of participants' names and numbers for five years. This ensures that all information (videotapes and data) will remain confidential. The video recordings will be kept on a secure computer/server, accessible only by the individuals described above. The recordings will be used only for the educational purposes described above and for which you have given consent.

____

Signature of Investigator Date

Participant's Statement

This study and the videotaping have been explained to me. I volunteer to take part in the videotaping. I have had the opportunity to ask questions. If I have questions later on about this videotaping, I can ask one of the investigators listed above. If I have questions about my rights as a research participant, I can call the Human Subjects Divisions at (206) 543-0098. I will receive a copy of this consent form for my records.

____

Signature of Participant Date

Copies to: Participant
Investigator
SPHSC Video Resource Library File

APPENDIX F

Video Assent Form

Department of Speech and Hearing Sciences
University of Washington
Video Assent Form
Stimulability as a Measure of Phonological Change

Investigators:
Amy M. Glaspey, Ph.C., CCC-SLP, Doctoral Student, Department of Speech and Hearing Sciences, College of Arts and Sciences (206) 931-7309
Carol Stoel-Gammon, Ph.D., Professor and Chair, Department of Speech and Hearing Sciences, College of Arts and Sciences (206) 543-7974

Investigator's Statement

We would like to ask you to be part of a study we are doing. If you are part of this study, you may be videotaped and photographed in communication activities. If you agree, we would like to use these tape recordings and/or photographs to help us with the research and for teaching about communication. We would like to have your permission to show your tapes or photographs to the following types of people:

- People who work or study at the University of Washington.
- People who participate in educational meetings where we talk about communication
- People who would watch educational TV programs or would visit a Web site made by us.

You can put an **X** next to the ones that will be okay with you. If you don't want to be recorded, videotaped or photographed let us know. If you decide during the taping, you can let us know then and we will stop. That's okay with us, too. You can still be in the study if you don't want to be videotaped. Do you have any questions?

We will answer your questions at any time. If you change your mind and decide that you don't really want to be videotaped, that's fine. You can tell your parents or someone at school at any time. If you decide not to do this, you can still be in other research studies or come here for therapy if you want to.

____
Signature of Investigator Date

Child/Individual's Statement

The study described above has been explained to me. I have gotten a chance to ask questions. I want to be in the study.

____

Signature of the Child/Individual   Date

Copies to: Parent/Caregiver
Investigator
Clinic and/or Research file
SpHSc Digital Resource Library file

APPENDIX G

HIPAA Form

UNIVERSITY OF WASHINGTON
AUTHORIZATION TO USE, CREATE AND SHARE HEALTH INFORMATION FOR RESEARCH FOR PROJECT ENTITLED,
STIMULABILITY AS A MEASURE OF PHONOLOGICAL CHANGE

PRINICIPAL INVESTIGATOR: AMY M. GLASPEY, Ph.C.
University of Washington, Department of Speech and Hearing Sciences,
206-931-7309, aglaspey@u.washington.edu
INVESTIGATOR: CAROL STOEL-GAMMON, Ph.D.
University of Washington, Department of Speech and Hearing Sciences,
206-543-7974

By law, researchers must protect the privacy of health information about you. In this form the word "you" means both the person who takes part in the research and the person who gives permission for another person to be in the research. Researchers may use, create, or share your health information for research **only if you let them**. This form describes what researchers will do with information about you. Please read it carefully. If you agree with it, please sign your name at the bottom. You will get a copy of this form after you have signed it.

If you sign this form, information will be shared with the people who conduct the research. In this form, all these people together are called "researchers." Their names will appear on the research consent form that you sign.

The researchers will use the health information only as described in the research consent form.

1. What "health information" includes

- Information about you that is created during the research study. This might include the results of tests or exams that become part of the study records; diaries and questionnaires that you might be asked to fill out as part of the study and other records from the study.
- Information in your medical records that is needed for this research study. These might include the results of physical exams, blood tests, x-rays, diagnostic and medical procedures and your medical history.

2. What the researchers may do with health information

The researchers may use and create health information about you for the study. They may also share your health information with certain people and groups. These may include:

- The sponsor of the study, Dr. Carol Stoel-Gammon, and its representatives.

- Government agencies, review boards, and others who watch over the safety, effectiveness, and conduct of the research.
- Other researchers when a review board approves the sharing of the health information.
- Your health insurer if they are paying for care provided as part of the research study.
- Others, if the law requires.

3. **Removing your name from health information**

   The researchers may remove your name (and other information that could identify you) from your health information. No one would know the information was yours.

   If the identifiers are removed, the information may be used, created, and shared by the researchers and sponsor as the law allows. (This includes other research purposes.) This form would no longer limit the way the researchers use, create, and share the information.

4. **How the researchers protect health information**

   The researchers and Dr. Carol Stoel-Gammon will follow the limits in this form. If they publish the research, they will not identify you unless you allow it in writing. These limitations continue even if you take back this permission.

5. **After the researchers learn health information**

   The limits in this form come from a federal law called the Health Insurance Portability and Accountability Act. This law applies to your doctors and other health care providers.

   Once the researchers and others who are not your doctors and health care providers get your health information, this law may no longer apply. But other privacy protections will still apply.

6. **Storing your health information**

   Your health information may be added to a database or data repository. This permission will end when the database or data repository is destroyed.

7. **Please note**

   You do not have to sign this permission ("authorization") form. If you do not, you may not be allowed to join the study. You may change your mind and take back your permission at any time. To take back your permission, write to: Amy M. Glaspey, Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42^{nd} St, Seattle, WA 98105-6246. If you do this, you may no longer be allowed to be in the study. The researchers will keep any information in the study record they already collected.

This form allows the researchers to access health information up to 90 days after you sign this form. If the researchers need access more than 90 days after you sign this form they may ask you to sign this form again. The UW human subjects review board may give the researchers permission to access your records without your written authorization after 90 days.

8. Your signature

I agree to the use, creation, and sharing of my health information for purposes of this research study

______________________________ ____________________

Signature of research subject or subject's legal representative — Date

______________________________ ____________________

Printed name of research subject or subject's legal representative — Representative's relationship to subject

APPENDIX H

Structural-Functional Examination

Adapted from: Robbins, J. & Klee, T. (1987). Clinical Assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders*, 52, 271-277.

Movements and speech are elicited first by verbal instruction. If the child is unsuccessful, the clinician will present the child with a visual and/or auditory model to imitate. The clinician will give a maximum of 3 prompts for each behavior. Administered in 7-10 minutes.

Total Structural score (TSS): Maximum Score 24

24 observations of the child's vocal tract structures and their relationships

Score 1 or 0 for normal or abnormal

Age group	n	M	SD	SE	Min-Max scores	Range
2:6-2:11	10	22.5	0.8	0.3	22-24	2
3:0-3:5	10	23.5	0.8	0.3	22-24	2
3:6-3:11	10	23.0	1.2	0.4	20-24	4
4:0-4:5	10	23.4	0.8	0.3	22-24	2
4:6-4:11	10	22.4	1.7	0.5	20-24	4
5:0-5:5	10	23.3	0.7	0.2	22-24	2
5:6-5:11	10	23.0	1.2	0.4	21-24	3
6:0-6:5	10	22.0	2.0	0.6	18-24	6
6:6-6:11	10	22.4	1.1	0.3	20-24	4
Total Samp.	90	22.8	1.3	0.1	18-24	6

Total Functional Score (TFS): Maximum Score 112

56 observations of the vocal tract functions

Judged with a 3 point rating scale: 2 for adult-like, 1 for emerging, 0 for absent function.

Age group	n	M	SD	SE	Min-Max scores	Range
2:6-2:11	8	97.0	9.2	3.3	78-106	28
3:0-3:5	10	99.3	8.5	2.7	86-109	23
3:6-3:11	8	107.1	3.8	1.3	99-111	12
4:0-4:5	9	108.1	2.1	0.7	104-111	7
4:6-4:11	9	109.3	2.1	0.7	106-112	6
5:0-5:5	8	109.8	2.6	0.9	106-112	6
5:6-5:11	10	109.8	2.6	0.8	103-112	9
6:0-6:5	9	110.6	1.3	0.4	108-112	4
6:6-6:11	10	111.1	1.2	0.4	108-112	4
Total Samp.	81	107.0	6.5	0.7	78-112	34

Rate and Duration measures: Scored individually on a continuum

Instruct child to produce the following as quickly as possible during 3 second intervals.

a) monosyllabic repetition rate (number of reps/second for a three second interval) for /p/, /t/, /k/

b) polysyllabic repetition rate (number/second) for /p ɾ k k/ and patticake

c) maximum phonation time (seconds) for a sustained vowel production

Age Group	/p/		/t/		/k/		ptk1		ptk2		Max	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
2:6-2:11	3.70	0.86	3.70	0.76	3.65	0.49	1.00	0.27	1.26	0.11	5.55	1.87
3:0-3:5	4.66	0.80	4.56	1.02	3.82	1.45	1.06	0.29	1.36	0.32	5.51	1.49
3:6-3:11	4.81	1.20	4.78	1.13	4.83	0.57	1.76	0.90	1.76	0.83	7.79	2.37
4:0-4:5	4.89	1.12	4.77	1.05	4.58	1.08	1.37	0.28	1.56	0.25	8.01	2.11
4:6-4:11	4.64	.082	4.46	0.62	4.29	0.57	1.34	0.27	1.33	0.10	9.22	2.19
5:0-5:5	4.76	0.84	4.82	0.85	4.56	0.94	1.40	0.43	1.58	0.28	8.06	1.97
5:6-5:11	5.09	0.44	5.22	5.22	4.91	0.25	1.58	0.19	1.65	0.21	9.42	1.65
6:0-6:5	5.36	0.64	5.32	0.45	4.94	0.28	1.45	0.25	1.61	0.23	10.9	3.06
6:6-6:11	5.51	0.43	5.37	0.48	4.85	0.71	1.72	0.19	1.64	0.26	11.4	3.02
Total Sample	4.85	0.92	4.80	0.72	4.51	0.88	1.43	0.44	1.54	0.38	8.51	2.91

Structures and Functions

Lips (CN VII)	Structure at rest	Score	Function	Score
	Symmetry	1 0	<u>Oral function:</u>	
			Rounding	2 1 0
	Relationship (open vs. closed)	1 0	Protrusion	2 1 0
			Retraction	2 1 0
			Alternate pucker/smile	2 1 0
			Bite lower lip	2 1 0
			Lip seal	2 1 0
			Puff cheeks	2 1 0
			Open-close lips	2 1 0
			<u>Speech function:</u>	
			Rounding /oU/	2 1 0
			Protrusion /u/	2 1 0
			Retraction /i/	2 1 0
			Alternate /u,i/	2 1 0
			Bite lower lip /f/	2 1 0
			Open-close lips /m/	2 1 0
Mandible (CN V)	**Structure at rest**	**Score**	**Function**	**Score**
	Symmetry	1 0	<u>Oral function:</u> Excursion (click teeth 5x)	2 1 0
	Occlusion	1 0		
	Size (re: facial features)	1 0		

Maxilla	Structure at rest	Score	Function	Score
	Symmetry	1 0		
	Size	1 0		
Teeth	**Structure at rest**	**Score**	**Function**	**Score**
	Decay	1 0		
	Alignment	1 0		
	Gaps	1 0		
	Missing	1 0		
	Occlusion (re: maxillary teeth)	1 0		
Tongue (CN XII)	**Structure at rest**	**Score**	**Function**	**Score**
	Symmetry	1 0	Oral function:	
			Protrusion	2 1 0
	Carriage	1 0	Elevation to alveolar ridge	2 1 0
	Fasciculations	1 0	Anterior-posterior sweep	2 1 0
	Furrowing	1 0	Interdental	2 1 0
	Atrophy	1 0	Speech-function:	
			Elevation to alveolar ridge: /n/, /t/, /l/	2 1 0
	Hypertrophy	1 0	Touch lateral edges of tongue to teeth: /s/ or /S/	2 1 0
			Interdental: /θ/	2 1 0
			Posterior tongue to palate: /k/ or /g/	2 1 0
Velopharynx (CN X)	**Structure at rest**	**Score**	**Function**	**Score**
	Symmetry	1 0	Oral function:	
			Blow on cold mirror	2 1 0
	Uvula	1 0	Suck through straw	2 1 0
	Tonsils	1 0	Speech function:	
			/a/	2 1 0
	Vault height	1 0	/ha.ha.ha/	2 1 0
	Palatal juncture (palpate)	1 0		
Larynx/respiration (CN X)	**Structure at rest**	**Score**	**Function**	**Score**
	Posture during quiet breathing	1 0	Oral function:	
			Cough, laugh, or cry	2 1 0

Coordinated Speech mvmts.	Structure at rest	Score	Function	Score
			Speech function:	
			Maximum phonation time (in seconds): /a/	2 1 0
			Pitch variation	2 1 0
			Loudness variation	2 1 0
			/ha.ha.ha/	2 1 0
			/p/	2 1 0
			/t/	2 1 0
			/k/	2 1 0
			/ptk/ repetitions	2 1 0
			Patticake repetitions	2 1 0
			You	2 1 0
			Top	2 1 0
			Beef	2 1 0
			Fume	2 1 0
			Cowboy	2 1 0
			Band-aid	2 1 0
			Half time	2 1 0
			Banana	2 1 0
			Kitty cat	2 1 0
			Puppy dog	2 1 0
			Communicate	2 1 0
			1950	2 1 0
			Potato head	2 1 0
			Winnie the pooh	2 1 0
Speech Sample				
			Prosody:	
			Rate	2 1 0
			Intonation	2 1 0
			Voice:	
			Pitch	2 1 0
			Loudness	2 1 0
			Quality	2 1 0
			Nasal resonance	2 1 0
Total				
Total Possible		24		112

Rate and Duration measures

/p/	Number of reps/second over 3 second interval	
/t/	Number of reps/second over 3 second interval	
/k/	Number of reps/second over 3 second interval	
/ptk/	Number of reps/second over 3 second interval	
Patticake	Number of reps/second over 3 second interval	
Vowels	Duration of phonation	

Supplemental motor speech assessment:
Strand, E. & McCauley, R. (1999) Clinical Management of Motor Speech Disorders in Children, Edited by A. Caruso and E. Strand. Thieme Medical Publishers, Inc., New York.

Vowels in isolation	/o/, /e/, /aI/	
CV and VC (using various vowels)	me, my, hi, up, on	
CVC (using various vowels)	Mom, pop, cake Hit, cup, ball	
Words of increasing length	Come, compute, computer	
	Base, baseball, baseball player	
Multisyllabic word repetition	Refrigerator Alligator	
Phrase repetition	I want, me too, my turn	
Repetition of sentences of increasing length	I want more, I want more milk, I want more milk please	

APPENDIX I

Scaffolding Scale of Stimulability: Words and Sentences

Word	Carrier phrase	Novel phrase	Embedded sentence	Two target sentence
Nasals				
/m/ mop	It's a **mop**.	Use the **mop**.	The **mop** is dirty.	The **mop** is by the **mat.**
beam	It's a **beam.**	See the **beam.**	The **beam** goes by.	The **beam** is over the **dam**.
/n/ net	It's a **net.**	Grab the **net.**	Her **net** fell off.	I will **knit** a big **net**.
bean	It's a **bean.**	Eat the **bean**.	The **bean** is big.	A **bean** is by the **pen**.
/ŋ/ song	It's a **song**.	Write a **song**.	That **song** is happy.	The **king** heard a happy **song.**
Stops				
/p/ pin	It's a **pin**.	Close the **pin**.	The **pin** is stuck.	My **pet** found an old **pin**.
map	It's a **map.**	Use the **map**.	The **map** is wet.	Put some **tape** on the **map.**
/b/ boat	It's a **boat**.	Row a **boat**.	My **boat** went fast.	I found a **bead** in the **boat**.
web	It's a **web**.	Make a **web**.	The **web** came down.	The **web** is on the **knob**.
/t/ ten	It's a **ten.**	Show me **ten**.	Draw **ten** over here.	I see **ten** by the **top.**
bat	It's a **bat**.	See the **bat**.	The **bat** flew home.	My **bat** landed on my **foot**.
/d/ dad	It's a **dad**.	Meet my **dad**.	My **dad** came here.	The **dad** picked up a **dime**.
bed	It's a **bed**.	Make your **bed**.	His **bed** is new.	I see **mud** on my **bed.**
Velars				
/k/ comb	It's a **comb.**	Use my **comb**.	The **comb** was lost.	My **comb** is in the **cup**.
hook	It's a **hook**.	Bend a **hook.**	The **hook** fell in.	The **hook** is on his **back**.
/g/ gob	It's a **gob**.	Have a **gob.**	A **gob** for you.	Have a **gob** for the **goose**.
hog	It's a **hog**.	See the **hog**.	The **hog** fell asleep.	A **bug** landed on the **hog.**
Stridents				
/θ/ thin	It's **thin.**	He is **thin**.	The **thin** bird flew.	My pen draws **thick** and **thin.**
math	It's **math.**	Do your **math.**	My **math** is done.	A **moth** can not learn **math**.
/ð/ that	It's **that.**	Go over **that.**	Give **that** to me.	Take **this** one and **that** one.
bathe	I can **bathe.**	He should **bathe.**	You **bathe** over here.	You **bathe** and **clothe** your baby.
/f/ fin	It's a **fin**.	Touch the **fin.**	The **fin** is little.	A **phone** can't have a **fin.**
beef	It's **beef.**	Eat your **beef.**	The **beef** is good.	Do not **laugh** about the **beef.**
/v/ van	It's a **van.**	Drive the **van.**	Her **van** was broken.	The **van** stopped at the **vet**.
weave	I can **weave.**	Ready to **weave.**	I **weave** the rug.	You should **weave** in the **cave.**
/s/ sing	I can **sing.**	He will **sing.**	I **sing** all day.	I eat **soup** and I **sing.**
kiss	I can **kiss.**	Give a **kiss**.	A **kiss** for grandma.	A **mouse** gave me a **kiss.**
/z/ zap	I can **zap.**	It will **zap**.	I **zap** my wand.	I **zap** and it will **zoom.**
hose	It's a **hose.**	Pick a **hose.**	A **hose** can leak.	Two **bees** flew on my **hose**.
/ʃ/ chute	It's a **chute.**	Fly a **chute.**	The **chute** blew away.	The **chute** is in the **shop.**
bush	It's a **bush**.	Grow a **bush**.	The **bush** has leaves.	A **dish** fell in the **bush**.
/ʒ/ beige	It's **beige.**	Color it **beige**.	The **beige** pants fit.	Mix the **rouge** and **beige**

				makeup.
/tʃ/ chop	I can **chop.**	You will **chop**.	I **chop** the wood.	I wipe my **chin** and **chop.**
match	It's a **match**.	See a **match.**	A **match** will burn.	I do not **touch** the **match**.
/dʒ/ jam	It's **jam**.	Wipe the **jam.**	The **jam** is sticky.	We ate **jam** in the **jeep.**
badge	It's a **badge.**	Wear a **badge.**	Her **badge** came off.	Her **badge** dropped in the **fudge.**
/h/ hug	I can **hug.**	Mommy will **hug**.	I **hug** my bear.	I like to **hug** and **hop**.
Glides				
/j/ yawn	I can **yawn.**	Make a **yawn**.	I **yawn** in bed.	I **yawn** and eat my **yoke**.
/w/ wood	It's **wood.**	Pick up **wood**.	The **wood** is cut.	I have **one** piece of **wood.**
Liquids				
/l/ lip	It's a **lip.**	Touch a **lip**.	My **lip** is chapped.	The **lid** was on my **lip**.
pool	It's a **pool.**	Open the **pool**.	The **pool** is empty.	I can **fall** in the **pool.**
/r/ rain	I see **rain.**	It may **rain.**	The **rain** came down.	The **rug** is in the **rain**.
bar	It's a **bar.**	Choose a **bar**.	The **bar** tastes good.	Eat the **bar** by the **door.**
Clusters				
/sp/ spin	I can **spin.**	Help me **spin.**	You **spin** a top.	I will **spin** and then **stop.**
/ts/ pits	I see **pits**.	Find two **pits**.	The **pits** are big.	The **pits** are in the **cups.**
/pl/ plum	It's a **plum**.	Eat the **plum.**	The **plum** was eaten.	Put the **plum** in the **glass.**
/tr/ train	It's a **train.**	Catch the **train.**	His **train** goes fast.	A **grape** is on the **train.**
/str/ stripe	It's a **stripe.**	Make a **stripe.**	The **stripe** is long.	Find a **stripe** on a **sprout.**

APPENDIX J

Procedural Reliability for Initial Assessments

PPVT

General Procedure of the Test: The investigator will show pictures to the child and prompt the child to point to the one that is named.

Checklist Instructions: Put a check in the box next to the items that are completed by the investigator. Sum all of the checks and total at the top. Some items are completed by watching the video whereas others will need to be completed by looking at the data sheets.

Boxes:

a) Observe from the video.
b) Observe from the video. "I'm going to show you some pictures,"
c) Look at the data sheet, given the chronological age, was the test initiated at the correct starting point?
d) Look at the data sheet.
e) Look at the data sheet.

PLS-IV

General Procedure of the Test: The investigator will show pictures and objects to the child, give the child instructions, and prompt the child to point or speak objects and pictures.

Checklist Instructions: Complete in the same manner as the PPVT using both the video and the data sheet.

Oral-Motor assessment

General Procedure of the Test: The investigator will assess the structures within the child's mouth and the functioning of the articulators. The investigator will give instructions to the child to perform different tasks including: motor movements, speech-motor movements, and word imitation.

Checklist Instructions: Both the video and the data sheet may be used to complete the checklist.

Boxes:

a) Observe on the video.
b) Use the data sheet to follow the sequence of instructions.
c) Some of the items may be difficult to determine from the video because they are observational. View the data sheet and check that all of the items have been scored.

d) Look at the data sheet to determine whether scoring has occurred.

Initial assessments

Assessment of Procedural Reliability Checklist

Coder ___________________________ Date(s)_________________

PPVT (_____/5)
_____ Participant Code _____ Session

- ☐ Has all test materials: ___picture stimuli, ___protocol
- ☐ Gives directions per test instructions.
- ☐ Begins at the correct chronological age.
- ☐ Discontinues once a ceiling has been met.
- ☐ Scores all test items and calculates standard score.

PLS-IV (_____/5)
_____ Participant Code _____ Session

- ☐ Has all test materials: ___ picture stimuli, ___toy stimuli, ___ protocol
- ☐ Gives directions per test instructions.
- ☐ Begins at the correct chronological age.
- ☐ Discontinues once a ceiling has been met.
- ☐ Scores all test items and calculates standard score.

Oral-Motor Assessment (_____/4)
_____ Participant Code _____ Session

- ☐ Has all test materials: ___flashlight, cup, straw, tongue depressor, ___protocol
- ☐ Gives directions per test instructions.
- ☐ Administers all items.
- ☐ Scores all test items and calculates a score.

Comments

____

____

Instructions:

1) Place a / through the square next to the items that are completed correctly.
2) Add the items for each individual test.

APPENDIX K

Procedural Reliability Protocol for the Scaffolding Scale of Stimulability

Coder__________________________ Participant Code ______ Coding Date(s) _________________

Target	Envmnt.	Type	Picture only		Instruction		Verbal Model		Seg, ProLn, Sim, Tactile	
			Adult	Child	Adult	Child	Adult	Child	Adult	Child 21
Time	Isolation								____ d + -	ir 20 + + + - - - -
	Word		____ d + -	ir 16 + + + - - - -	____ d + -	ir 17 + + + - - - -	____ d + -	ir 18 + + + - - - -	____ d + -	ir 19 + + + - - - -
ir tally 1 2 3	Sentence	Carrier Phrase	____ d + -	ir 12 + + + - - - -	____ d + -	ir 13 + + + - - - -	____ d + -	ir 14 + + + - - - -	____ d + -	ir 15 + + + - - - -
	Sentence	Novel	____ d + -	ir 9 + + + - - - -	____ d + -	ir 10 + + + - - - -	____ d + -	ir 11 + + + - - - -		
	Sentence	Embed within	____ d + -	ir 6 + + + - - - -	____ d + -	ir 7 + + + - - - -	____ d + -	ir 8 + + + - - - -		
	Sentence	2 Target Sentence	____ d + -	ir 3 + + + - - - -	____ d + -	ir 4 + + + - - - -	____ d + -	ir 5 + + + - - - -		
	Picture Scene		____ d + -	ir 1 + + + - - - -	____ d + -	ir 2 + + + - - - -				

APPENDIX L

Procedural Reliability: Scaffolding Scale of Stimulability Instructions

Definitions

Environment

The environment is the linguistic context in which a sound is produced. The clinician changes the environment in response to the child's production. If the child's production is correct, then the environment is made more complex; conversely, if the child's production is in error, then the environment is made less complex.

The following environments are listed from least complex to most complex.

1) Isolation: The target is elicited alone or in a syllable with "uh" (CV or VC). The clinician gives a verbal model of the sound with or without cues. The cues are individualized to the child's errors and any type of cue may be used.

- Example—mmmm

2) Word: The target sound is elicited in a CVC word with the exception of targets for multiple syllables and clusters. The words are listed on the summary sheet. Dynamic assessment of each target begins at the word level. The child is shown a picture and asked to name the picture, "What's this?"

- Example—mop

3) Sentence: The target sound is elicited in a sentence. The clinician may elicit four types of sentences: carrier phrase, novel phrase, one-target sentence, or two-target sentence. The progression is from least to most complex sentences. The child is shown a picture and instructed, "Tell me about this" or "Tell me about these."

a) Carrier phrase: a repetitive phrase that starts with "He can …" or "It's a …" depending on whether the target is in a word that is a noun or a verb.
  - Example—It's a mop.

b) Novel phrase: a three word phrase with the target sound in a word at the end of the phrase. The phrase change from target to target.
  - Example—Use a mop.

c) One-target sentence: A 4 word sentence with the target sound in a word in the middle of the sentence. The word with the target may be the second or third word in the sentence, but may not occur at the end or beginning of the sentence.
  - Example—The mop is dirty.

d) Two-target sentence: A 4-5 word sentence with the target sound in two words. The words with target sounds may occur anywhere in the sentence.
  - Example—The mop is by the map.

4) Picture scene: a spontaneous production of the target sound in two different words while describing a picture scene. The productions must be in at least two connected words. Single word production such as labeling do not count. The picture scene has a several opportunities for productions, some scenes more than others. The child is shown a scene and instructed, "Tell me a story about this picture."

Cue Levels

The following cues are in sequence and are additive. Once a picture has been presented, it remains on the table until the dynamic assessment for the target is complete. The clinician adds cues in response to the child's error productions.

0) Picture: The clinician shows the child a picture that represents a word and asks the child to name it.

1) Instruction: The clinician gives an instruction that my help the child produce the target accurately. The nature of the instruction is dependent on the child's previous error response. The instruction may include feedback about the error response and then a suggestion about how to alter the articulators to improve production.

Example—"Not quite, pull your tongue back. Try again."

2) Verbal Model: The clinician says the target in its environment immediately before the child. The clinician also cues the child to watch.

Example—"Look at me. Say, mop."

3) Segmentation, Prolongation, Simultaneous Production, Tactile. At this level the clinician may choose a variety of strategies depending on the child's error and the characteristics of the target. Either segmentation or prolongation is modeled to emphasize the target sound. For example, stops are segmented and fricatives are prolonged. If this does not work, then the clinician may try another cue. Simultaneous production may be selected for continuants. Tactile cues may include manipulation of the articulators or a tactile representation of the sound such as a tap on the hand for a stop or a slide down the arm for a fricative. All cues at this level are scored the same.

Sequencing Rules:

The dynamic assessment begins at word level (SSS 16). The clinician shows a picture to the child and says, "What's this?" The child then names the target word. Once the child responds, the clinician must make an immediate judgment about the production. If the child makes an **error response**, the clinician moves to the *right* on the hierarchy and adds instruction cues. The clinician will give an instruction regarding the participant's production which may include feedback such as, "You said that very close, try again," or

"Pull your tongue back, try again." If the child continues to make errors, the clinician will continue to move right along the hierarchy to Verbal model and *up* to Isolation. In contrast, if the child makes a **correct response**, then the clinician will change the environment and move *down* on the hierarchy. The clinician will prompt the child, "Tell me about that," and elicit a sentence production. If the child produces a correct sentence then the clinician will move to the two-target sentence level. The child is scored on the type of sentence produced, but is not required to produce all sentences types at the spontaneous level. The clinician turns the page that includes pictures of the two targets. First, the clinician makes sure that the child can label both words and then says, "Tell me about these." The child must produce both of the target words correctly to receive credit.

Potential difficulties:

The child doesn't know the target word.
Ideally, the child will respond with the correct word; however, when the child does not know the word, then the clinician may give a **delayed model.** The clinician will name the picture, say a filler sentence, and then prompt the child again.

The child says the target correctly at the word level, but does not spontaneously create a sentence when prompted.
Often younger children will not generate their own sentences without a verbal model. If the child does not respond to the prompt, "Tell me about that," at the sentence level then the clinician circles "ir" on the data sheet and the clinician will move immediately to the right and Verbal Model cueing.

The child continues to not respond to the prompt to produce a sentence.
If a child demonstrates an "ir" response at the sentence level across three different phonemes, then on the fourth phoneme the clinician is allowed to skip the spontaneous and instruction cues and go straight to a verbal model on all subsequent elicitations. The clinician will not be in violation of the procedural sequence; however, if the clinician does not wait at least three times for sentences productions, then the clinician is scored in error.

The clinician keeps repeating the same cueing level.
Sometimes a child may be inattentive or unresponsive. The clinician is allowed to repeat a cue up to three times at the Instruction or Verbal model levels. The clinician is allowed to repeat cues up to four times at the Segmentation/prolongation/simultaneous production/tactile level. If the clinician repeats the same cue more than three times on I or VM levels, or four times on S/P/SP/T level, then the clinician is scored in error.

The clinician skipped a level or went the wrong way.

The clinician may at times make errors in the presentation of the cues. Each step is judged in relation to the step before it. The clinician is marked in error for the step that is missed, but not for the entire elicitation task.

The child made an error at the spontaneous sentence level, but the clinician gave a verbal model in a less complex environment rather than at the same sentence level. When a child makes an error at the sentence, two-target, or picture scene level, and needs a verbal model cue, the clinician is allowed to give verbal models at lower levels.

Coding:

Adult Codes
____ = The coder will number the sequence that the clinician uses along the hierarchy of the dynamic assessment.
d = The coder will circle "d" when the adult has given a delayed model.
\+ = The coder will circle a plus when the clinician uses a correct step in the hierarchy.
\- = The coder will circle a minus when the clinician used an incorrect step in the hierarchy.
Total = The coder will then add the total of correct procedures over the total number of steps on the hierarchy attempted.

Child Codes
\+ = The coder will circle a plus each time the child makes a correct production in response to the clinician's prompt.
\- = The coder will circle a minus each time the child makes an error production in response to the clinician's prompt.
ir = The coder will circle "ir" each time the child makes an inappropriate response. An "ir" may include no response, something off topic, or the wrong word even after it has been modeled.
\# = The coder will circle the last correct response that the child has given. This will be the child's score.

Other
Target = Fill in the target being elicited.
Time = The time on the tape may be recorded here (optional).
ir tally = When the child produces an "ir" response at the word level, circle the number appropriate number. Once the child has scored three "ir's," then the procedure rules change.

APPENDIX M

Procedural Reliability: Carter-Buck

Instructions

General procedure of the test:

The Carter-Buck has three subtests. One Subtest 1, the child is asked to name pictures that are seen in the book. The Clinician is not allowed to say any of the pictures in the book even if the child doesn't know the word. On Subtest 2, the child is given a verbal model of the words and the child is asked to repeat the words. The child should say the word right after the clinician. On Subtest 3, the child is given a verbal model of syllables and asked to repeat the syllables. The child should say the syllables right after the clinician.

Checklist instructions:

Watch the video and put a slash in the box when the clinician correctly completes the procedure. You will watch three different videos.

Boxes:

a) The protocol is a piece of paper with word lists. The picture stimuli include two small white picture books. Give credit when you see these items.

b) The clinician should say something similar to the following instructions:
Test 1: I'll show you some pictures and I want you to name them for me. What is this?
Test 2: Look at me and listen carefully. Then you say the word.
Test 3: Look at me and listen carefully. Then you say what I say.

c) Count the mistakes. Count how many times the clinician accidentally says a word when she should not. Put the number on the line. Divide by 112 to get a percentage. These should occur infrequently, so watch carefully.

d) Count the mistakes. Count how many times the clinician accidentally talks between saying the word and the child's response. Add the mistakes from both subtests 2 and 3. Put the number on the line. Divide by 229. Subtract this number from 100 and put it on the line. These should occur infrequently, so watch carefully.

e) Count the cues. The clinician should prompt the child to look at her face at least two times during subtest 2 and during subtest 3. For each subtest, stop counting once you have seen two prompts.

f) Look at the data sheet and make sure that there are no circles in Columns T2. stim and stim that say missing.

g) You should see the child naming pictures first, then the clinician saying the words with the child repeating, and finally the clinician saying syllables.

h) You are watching the video, so check the box.

i) Look at the data sheet and make sure that something is written in every box

Carter-Buck Procedural Reliability Checklist

Coder____________________________ Date(s)__________________

1)______ Participant Code _______ Session (_______**/9)**

- ☐ Has all test materials: ____ Protocol, _____ Picture stimuli
- ☐ Gives directions per test instructions
- ☐ On Test 1, the clinician gives a verbal model less then 10% of the time.
  (the instructions require that the clinician not say the target words).
  Total number of verbal models: ______/112 = ______%
- ☐ On Test 2, 3, the clinician gives a verbal model more than 90% of the time.
  (the instructions require that the clinician says the target words).
  Total number of missed models: ______/229 = ______%
- ☐ On Tests 2, 3, the clinician cues the child to look at the clinician's face at least 2 times per subtest (if the child looks away from the clinician) 2 = _____, 3 = _____
- ☐ Administers all test items.
- ☐ Administers subtests in correct order.
- ☐ Video-tapes child's productions.
- ☐ Scores all items.

2)______ Participant Code _______ Session (_______**/9)**

- ☐ Has all test materials: ____ Protocol, _____ Picture stimuli
- ☐ Gives directions per test instructions

- ☐ On Test 1, the clinician gives a verbal model less then 10% of the time.
  (the instructions require that the clinician not say the target words).
  Total number of verbal models: ______/112 = ______%
- ☐ On Test 2, 3, the clinician gives a verbal model more than 90% of the time.
  (the instructions require that the clinician says the target words).
  Total number of missed models: ______/229 = ______%
- ☐ On Tests 2, 3, the clinician cues the child to look at the clinician's face at least 2 times per subtest (if the child looks away from the clinician) 2 = _____, 3 = _____
- ☐ Administers all test items.
- ☐ Administers subtests in correct order.
- ☐ Video-tapes child's productions.
- ☐ Scores all items.

3)______ Participant Code _______ Session **(_______/9)**

- ☐ Has all test materials: ____ Protocol, _____ Picture stimuli
- ☐ Gives directions per test instructions
- ☐ On Test 1, the clinician gives a verbal model less then 10% of the time.
  (the instructions require that the clinician not say the target words).
  Total number of verbal models: ______/112 = ______%
- ☐ On Test 2, 3, the clinician gives a verbal model more than 90% of the time.
  (the instructions require that the clinician says the target words).
  Total number of missed models: ______/229 = ______%
- ☐ On Tests 2, 3, the clinician cues the child to look at the clinician's face at least 2 times per subtest (if the child looks away from the clinician) 2 = _____, 3 = _____
- ☐ Administers all test items.
- ☐ Administers subtests in correct order.
- ☐ Video-tapes child's productions.
- ☐ Scores all items.

APPENDIX N

Procedural Reliability: Probe

Instructions

General Procedure of the Test:

The child is asked to name pictures that are seen on a laptop. The child should name the words without a model from the clinician. However, if the child does not know the word, the clinician is allowed to say the word, talk a little, and prompt the child again.

Checklist Instructions:

Watch the video and put a slash in the box when the clinician correctly completes the procedure. You will watch three different videos.

Boxes:

a) The test material is the laptop which you may or may not able to see in the video. You will know it is there by the way the child looks and names words.

b) The clinician should give an approximation of the following instructions:

I'm going to show you some pictures and I want you to say them for me. Mommy says, I…

c) Look over the data sheet and make sure that something is written in each square.

d) Count the mistakes. Listen to the video and count the number of times that the clinician says the work and the child imitates immediately after the clinician. Sometimes the child won't know the work and the clinician will say it; however, the clinician should talk a little in-between and ask the child to say the word.

e) You are watching the video that was taped

f) Check the data sheet and make sure that the transcription is filled in for each item and then make sure that the test was scored.

Probe Procedural Reliability Checklist

Coder ______________________________ Date(s) ____________________________

1) ____ Participant Code ____ Session **(_______/6)**

- □ Has all the test materials: ____ laptop
- □ Gives all directions per test instructions
- □ Administers all items.
- □ Presents items without a direct verbal model at least 90% of the time. ____/50
- □ Video-tapes child's productions.
- □ Transcribes and scores all items.

2) ____ Participant Code ____ Session **(_______/6)**

- □ Has all the test materials: ____ laptop
- □ Gives all directions per test instructions
- □ Administers all items.
- □ Presents items without a direct verbal model at least 90% of the time. ____/50
- □ Video-tapes child's productions.
- □ Transcribes and scores all items.

3) ____ Participant Code ____ Session **(_______/6)**

- □ Has all the test materials: ____ laptop
- □ Gives all directions per test instructions
- □ Administers all items.
- □ Presents items without a direct verbal model at least 90% of the time. ____/50
- □ Video-tapes child's productions.
- □ Transcribes and scores all items.

APPENDIX O

Procedural Reliability: HAPP-3

Instructions

General Procedure of the Test:

Children are asked to name objects. The child should name the words without a model from the clinician. However, if the child does not know the word, the clinician is allowed to say the word, talk a little, and prompt the child again.

Checklist Instructions:

Watch the video and put a slash in the box when the clinician correctly completes the procedure. You will watch three different videos.

Boxes:

a) The protocol is a paper with all the target words listed on it. You may not see the form directly, but may see the clinician glancing over to it.

b) The clinician should give an approximation of the following instructions:

Here are some toys. I want you to say their names for me. What's this?

c) Look over the data sheet and make sure that something is written is each square.

d) Count the mistakes. Listen to the video and count the number of times that the clinician says the word and the child imitates immediately after the clinician. Sometimes the child won't know the word and the clinician will say it; however, the clinician should talk a little in-between and ask the child to say the word.

e) You are watching the video that was taped.

f) Check the data sheet and make sure that the transcription is filled in for each item and then make sure the test was scored.

Hodson Assessment of Phonological Processes

Coder______________________________ Dates(s)____________________________

1) ____ Participant Code ____ Session **(_______/6)**

- □ Has all the test materials: ____Protocol, ____Toy stimuli
- □ Gives all directions per test instructions
- □ Administers all items.
- □ Presents items without a direct verbal model at least 90% of the time. ____/50
- □ Video-tapes child's productions.
- □ Transcribes and scores all items.

2) _____ Participant Code ____ Session **(________/6)**

- □ Has all the test materials: ____Protocol, ____Toy stimuli
- □ Gives all directions per test instructions
- □ Administers all items.
- □ Presents items without a direct verbal model at least 90% of the time. ____/50
- □ Video-tapes child's productions.
- □ Transcribes and scores all items.

3) _____ Participant Code ____ Session **(________/6)**

- □ Has all the test materials: ____Protocol, ____Toy stimuli
- □ Gives all directions per test instructions
- □ Administers all items.
- □ Presents items without a direct verbal model at least 90% of the time. ____/50
- □ Video-tapes child's productions.
- □ Transcribes and scores all items.

APPENDIX P

Procedural Reliability: Percent Consonants in Error

Instructions

General Procedure of the Test:

The clinician gives the child a wordless book and asks the child to tell the story. The clinician may need to prompt the child to engage her/him in the story line. Prompts may include a story starter such as, "One day…" or comments, "Tell me about…" The child needs to talk in connected speech for at least two minutes.

Checklist Instructions:

Watch the videos and put a slash in the box when the clinician correctly completes the procedure. You will watch three different videos.

Boxes:
a) The picture book is the only material.

b) The clinician should give an approximation of the following instructions:
I'm going to give you a story. This story doesn't have any words. So you are going to make them up. You are going to tell the story.

c) The child may have difficulty starting a story, so the clinician may give a story starter such as, "One day…" or "Tell me about this picture." The clinician may even start telling the story for a couple pages.

d) Watch the counter and make sure the child is talking in connected speech for at least 2 minutes. The clinician is allowed to make short comments, repeat the child's phrase, or give and instruction to continue. The clinician should not be telling the entire story during the 2 minutes.

e) You are watching the video that was taped

f) Look at the data sheets. Make sure that you see the gloss (how an adult would say the words), transcription of the gloss, and transcription of the child's speech. Check that the test has been scored, which will be at the top of the data sheet.

Percent Consonants in Error

Coder ______________________________ Date(s)________________________

1) ____ Participant Code ____ Session **(______/6)**

- □ Has all the test materials: ____Picture book
- □ Gives all directions per test instructions
- □ Gives a story starter or modeling as needed.
- □ Child speaks in connected speech for at least 2 minutes.
- □ Video-tapes child's productions.
- □ Speech is glossed, transcribed, and scored for PCC.

2) ____ Participant Code ____ Session **(______/6)**

- □ Has all the test materials: ____Picture book
- □ Gives all directions per test instructions
- □ Gives a story starter or modeling as needed.
- □ Child speaks in connected speech for at least 2 minutes.
- □ Video-tapes child's productions.
- □ Speech is glossed, transcribed, and scored for PCC.

3) ____ Participant Code ____ Session **(______/6)**

- □ Has all the test materials: ____Picture book
- □ Gives all directions per test instructions
- □ Gives a story starter or modeling as needed.
- □ Child speaks in connected speech for at least 2 minutes.
- □ Video-tapes child's productions.
- □ Speech is glossed, transcribed, and scored for PCC.

APPENDIX Q

SSS: Individual Token Scores

Participants are listed across by administration and targets are listed in the first column going down. A ".01" denotes levels that the child may have been able to go higher on the scale, but this could not be assessed. A dash indicates sounds that were used correctly in connected speech. Participants 8M and 9M were only evaluated two times.

	3F. 1	3F. 2	3F. 3	4M. 1	4M. 2	4M. 3	5F. 1	5F. 2	5F. 3	7F. 1	7F. 2	7F. 3
/m-/	5.01	-	-	-	-	-	-	-	-	-	-	-
/-m/	18	14	11	5.01	-	1	15	11	5.01	5	-	-
/n-/	5.01	-	-	-	-	-	-	-	-	-	-	-
/-n/	5.01	5.01	5.01	-	-	-	-	-	-	-	-	-
/-ŋ/	5.01	5.01	5.01	-	21	21	21	20	15	19	20	19
/p-/	5.01	-	-	-	-	-	-	-	-	-	-	-
/-p/	5.01	5.01	5.01	-	-	-	15	11	3.01	1	-	-
/b-/	5.01	-	-	-	-	-	-	-	-	-	-	-
/-b/	8	5.01	5.01	5.01	5.01	3.01	15	8	3.01	5.01	-	-
/t-/	5.01	-	-	-	-	-	-	-	-	-	-	-
/-t/	19	15	11	-	-	-	15	8	1	-	-	-
/d-/	5.01	-	-	-	-	-	-	-	-	-	-	-
/-d/	5.01	15	5.01	5.01	5.01	3.01	15	8	1	-	-	-
/k-/	8	8	8	1	1	3.01	21	21	21	15	18	14
/-k/	11	5.01	5.01	1	5.01	1	21	21	21	19	15	19
/g-/	8	5.01	5.01	1	5.01	1	21	21	21	15	15	15
/-g/	5.01	5.01	5.01	19	5.01	1	21	21	20	19	15	19
/θ-/	20	19	19	20	21	5	19	15	15	20	19	19
/-θ/	20	19	19	20	21	3.01	19	14	8	20	15	11
/ð-/	21	20	20	19	21	5.01	19	15	14	15	11	8
/-ð/	21	20	20	20	21	3.01	20	15	11	21	20	19
/f-/	20	19	5.01	1	5.01	1	20	15	3	15	15	15
/-f/	21	20	19	3.01	3.01	6	20	11	3.01	18	19	15
/v-/	20	19	14	3.01	5.01	1	19	15	15	19	8	15
/-v/	20	20	14	5	5.01	3.01	20	15	6	20	15	15
/s-/	19	15	15	19	5.01	1	19	15	2	18	11	11
/-s/	20	17	15	19	5.01	3.01	18	11	1	19	19	15
/z-/	19	14	5.01	19	5.01	1	15	15	8	19	15	15
/-z/	20	15	15	20	5.01	1	19	15	3.01	19	19	18

/ʃ-/	20	15	15	15	1	1	19	15	9	14	15	15
/-ʃ/	20	19	19	18	3.01	9	15	11	15	19	15	19
/-ʒ/	20	15	15	14	8	5	19	15	15	19	19	19
/tʃ-/	18	19	15	20	20	5	19	5.01	3.01	19	20	20
/-tʃ/	15	11	14	20	20	15	15	5.01	3.01	15	15	15
/dʒ-/	14	15	5.01	21	21	20	5.01	3.01	3.01	21	20	19
/-dʒ/	19	15	11	21	21	20	15	15	8	21	20	20
/h-/	5.01	5.01	5.01	-	-	-	-	-	-	15	5.01	5.01
/j-/	18	-	18	-	-	-	-	-	-	5.01	-	-
/w-/	5.01	-	5.01	-	-	-	-	-	-	-	-	-
/l-/	5.01	5.01	5.01	5.01	-	-	20	19	15	5	5.01	1
/-l/	20	16	15	20	20	1	15	19	15	16	20	19
/r-/	5.01	5.01	5.01	21	21	21	18	15	15	21	21	15
/-r/	8	5.01	5.01	21	21	21	-	5.01	-	21	21	20
/sp-/	15	15	14	11	3.01	1	15	15	15	19	11	15
/-ts/	15	14	8	15	3.01	3	15	3.01	3.01	19	15	15
/pl-/	8	8	5.01	19	5.01	3.01	21	21	19	21	15	5.01
/tr-/	5	5.01	5.01	21	21	21	21	21	19	21	21	21
/str-/	15	11	14	21	21	21	15	19	19	21	21	21

	8M.1	8M.2	9M.1	9M.2	10 M.1	10 M.2	10 M.3	12F .1	12F .2	12F .3
/m-/	-	-	-	-	-	-	-	-	-	-
/-m/	-	-	-	-	-	-	-	-	-	-
/n-/	-	-	-	-	-	-	-	-	-	-
/-n/	-	-	-	-	-	-	-	-	-	-
/-ŋ/	21	20	8	5.01	3.01	3.01	3.01	-	-	-
/p-/	-	-	5.01	5.01	-	-	-	-	-	-
/-p/	5.01	-	5.01	5.01	-	-	-	5.01	3	3.01
/b-/	-	-	-	-	-	-	-	-	-	-
/-b/	8	5.01	8	5.01	-	-	-	5.01	3.01	3.01
/t-/	-	-	-	-	-	-	-	-	-	-
/-t/	-	-	-	-	-	-	-	-	-	-
/d-/	-	-	-	-	-	-	-	-	-	-
/-d/	-	-	-	-	-	-	-	-	-	-
/k-/	21	21	8	5.01	15	5.01	1	8	6	5
/-k/	21	21	-	-	5	3.01	1	5.01	1	1
/g-/	21	21	5.01	5.01	8	3.01	1	18	5	18
/-g/	21	21	-	-	1	3.01	1	15	5	2
/θ-/	19	19	19	15	19	14	8	19	19	15

/-θ/	19	19	19	15	19	11	11	19	15	15
/ð-/	20	15	18	15	18	5.01	8	17		15
/-ð/	20	20	19	15	20	15	14	19	15	15
/f-/	8	15	-	5.01	20	5	1	15	15	15
/-f/	14	5.01	11	11	15	14	3.01	15	15	15
/v-/	3.01	15	8	5.01	19	3.01	3	20	15	14
/-v/	11	11	8	5.01	19	15	15	20	15	15
/s-/	20	19	5.01	5.01	20	15	3.01	19	15	15
/-s/	17	8	-	5.01	19	8	8	19	15	15
/z-/	19	19	15		20	8	3.01	19	15	15
/-z/	19	15	8	5.01	19	15	11	19	14	15
/ʃ-/	20	19	5.01	5.01	20	15	3.01	20	19	15
/-ʃ/	20	15	5.01	5.01	20	15	8	15	15	15
/-ʒ/	21	15	14	8	20	11	4.01	20	19	15
/tʃ-/	20	20	5.01	8	20	5.01	3.01	20	15	15
/-tʃ/	20	19	5.01	5.01	19	3.01	3.01	20	19	15
/dʒ-/	21	20	-	-	21	3.01	3.01	20	21	21
/-dʒ/	21	19	8	8	21	15	5.01	20	21	19
/h-/	-	-	-	-	-	-	-	-	-	-
/j-/	-	-	-	-	-	-	-	-	-	-
/w-/	-	-	-	-	-	-	-	-	-	-
/l-/	21	20	-	-	20	14	15	5.01	-	-
/-l/	21	20	-	-	20	15	15	20	15	15
/r-/	20	19	21	20	21	21	19	21	21	21
/-r/	5.01	-	21	20	21	21	20	21	21	21
/sp-/	19	15	19	5.01	19	15	8	15	14	15
/-ts/	21	15	5.01	5.01	19	9	15	15	11	15
/pl-/	21	21	15	15	21	19	19	15	6	3.01
/tr-/	19	20	21	5.01	21	21	21	21	21	21
/str-/	21	20	21	14	21	21	21	21	21	21

CURRICULUM VITAE

Amy M. Glaspey

Education

Ph.D. (2006) Speech and Hearing Sciences
University of Washington; Seattle, WA

- Dissertation title: Dynamic Assessment of Phonological Disorders: The Scaffolding Scale of Stimulability
- Dissertation advisor: Carol Stoel-Gammon, Ph.D.

M.S. (1995) Communication Disorders & Sciences
University of Oregon; Eugene, OR

M.S. (1995) Early Intervention/Childhood Special Education
University of Oregon; Eugene, OR

B.S. (1993) Communication Disorders & Sciences
University of Oregon; Eugene, OR

Employment History

2000-2005 Speech-Language Pathologist. Private Practice, Issaquah, WA. Treated preschool and elementary age children with phonological disorders in their homes.

2000-2002 Speech-Language Pathologist. Pediatric Speech and Language Associates. Conducted speech/language evaluations and therapy for preschool and elementary age children in a clinical setting.

1999-2000 Speech-Language Pathologist. Issaquah School District, Issaquah, WA. Conducted speech/language evaluations and therapy with preschool and elementary age children, participated in the Northwest Initiative for Teaching and Learning project and teacher study teams.

1995-1999 Speech-Language Pathologist. Umatilla-Morrow Education Service District. Conducted speech/language therapy within a rural and culturally diverse population. Traveled as an itinerant therapist and served children in their homes, consulted with early intervention/early childhood special education and Head Start teachers, and provided direct service in group settings with parent participation. Used interpreters during treatment with children from Spanish speaking families. Supervised SLP assistants. Documented units of service, log notes, and Medicaid billing with computerized data base system.

- 1997-1999. Participated in action research project to design and use a curriculum to teach articulation and phonology skills in Spanish. Awarded grant to develop project.
- 1995-1997. Participated in language sample analysis research group and provided inservices to other speech/language/hearing department staff.

Publications and Presentations

Journal Articles

Glaspey, A. & Stoel-Gammon, C. (2005). Dynamic Assessment in Phonological Disorders: The Scaffolding Scale of Stimulability. *Topics in Language Disorders.*

Stoel-Gammon, C, Stone-Goldman, J., & **Glaspey, A.** (2002). Pattern-based approaches to phonological therapy. *Seminars in Speech and Language, 23,* 3-13.

Contributed Posters and Presentations

Glaspey, A. & Stoel-Gammon, C. (2005). Stimulability as a measure for Evidence Based Practice. Paper presentation. Annual Convention of the American Speech-Language-Hearing Association. San Diego, CA.

Glaspey, A. & Stoel-Gammon, C. (2005). Patterns of Phonological Change with the Scaffolding Scale of Stimulability. Paper presentation. International Child Phonology Conference. Fort Worth, TX.

Glaspey, A. & Stoel-Gammon, C. (2004). Scaffolding Scale of Stimulability: Comparing treatment measures. Poster presentation. Annual Convention of the American Speech-Language-Hearing Association, Philadelphia, PN.

Glaspey, A. & Stoel-Gammon, C. (2004). The Scaffolding Scale of Stimulability. Paper presentation. International Child Phonology Conference. Tempe, AZ.

Glaspey, A. & Stoel-Gammon, C. (2003). Efficacy in cycles treatment: Evidence from untreated single-word probes. Poster Presentation. Annual Convention of the American Speech-Language-Hearing Association. Chicago, IL.

Glaspey, A. & Stoel-Gammon, C. (2003). Generalization probes as evidence of treatment efficacy with cycles treatment. Poster Presentation. International Child Phonology Conference, Vancouver, BC, Canada.

Glaspey, A. & Stoel-Gammon, C. (2002). Sensitivity to phonological change: Stimulability, intelligibility, and generalization measures. Paper Presentation. Annual Convention of the American Speech-Language-Hearing Association, Atlanta, GA.

Glaspey, A. & Stoel-Gammon, C. (2002). Stimulability: Changes across one cycle of treatment. Poster Presentation. International Symposium for Child Language Development and the International Association and Study of Child Language, Madison, WI.

Glaspey, A. & Stoel-Gammon, C. (2001). Stimulability as a measure of phonological change. Poster Presentation. Annual Convention of the American Speech-Language-Hearing Association, New Orleans, LA.

Invited Presentations

Glaspey, A. & Newcom, M. (1998). *Rural SLP's and Spanish Speaking Students: Phonology and Language Intervention Issues*. Oregon Speech-Language-Hearing Association Regional Conference. Pendleton, Oregon.

Traylor, D. & **Glaspey, A.** (1997). *Creating a Data System.* Oregon Department of Education Training Program. LaGrande, Oregon.

Teaching Experience

Independent Instruction

Instructor. *Articulation and Phonological Disorders;* University of Washington graduate course (SPHSC 539); 4 credits; Winter quarter 2005 (20 students)

Instructor. *Proseminar: Instructional Development Forum;* University of Washington doctoral course (SPHSC 563); 1 credit; Fall quarter 2004, Winter 2005, Spring 2005 (12 students)

Instructor. *American English Sound System;* University of Washington undergraduate course (SPHSC 111); 2 credits; Winter quarter 2002 (70 students); Fall quarter 2003 (65 students); Spring quarter 2006 (61 students)

Acting Undergraduate Academic Advisor. University of Washington; Department of Speech and Hearing Sciences; Summer quarter 2003

Clinical Supervisor. University of Washington Speech and Hearing Clinic; Supervised students' treatment of clients with pediatric speech and language disorders. Fall quarter