TREATMENT OF CHILDHOOD APRAXIA OF SPEECH: A SINGLE-CASE EXPERIMENTAL DESIGN STUDY OF INTENSITY OF TREATMENT

A Master's Thesis
Submitted to
the Temple University Graduate Board

In Partial Fulfillment

of the Requirements for the Degree

MASTER OF ARTS

---

by:

Nicolette Kovacs

August 2017

Examining Committee Members:

Edwin Maas, PhD, Advisory Chair, Communication Sciences and Disorders, Temple University

Jamie Reilly, PhD, Communication Sciences and Disorders, Temple University

Francine Kohen, M.S. CCC-SLP, Communication Sciences and Disorders, Temple University

ABSTRACT

Childhood Apraxia of Speech (CAS) is a pediatric motor-speech disorder which has been controversial due to its difficulty to diagnose and little progress in treatment. The purpose of the present study was to examine a principle of motor learning (PML) within the context of an evidence-based treatment for this disorder, as a way to improve outcomes for children with CAS. In particular, this study examines the role of intensity, specifically, massed versus distributed practice, when treating CAS using a modified form of Dynamic Temporal Tactile Cueing (DTTC; Strand et al., 2006). Two participants with CAS between the ages of 5 and 11 received massed and distributed practice on individualized targets in an single-case alternating treatments design with multiple baselines. Accuracy of speech targets on probe tasks was judged by blinded listeners. Results were interpreted through inspection of graphs and calculation of effect sizes. The results of the study showed that massed practice had a marginal benefit over distributed practice. Implications from this study suggest the importance of continued research examining the role of PML in CAS treatment and the value of using a massed-treatment approach when treating CAS.

ACKNOWLEGEMENTS

First, I want to thank Dr. Edwin Maas for his endless support and encouragement from the very beginning of the project. Without his time, energy, dedication, and knowledge-base, the completion of this thesis would have been impossible. I would also like to thank the members of the Speech, Language and Brain Lab for their hard work throughout the many stages of this project. This project was also made possible by Anne van Zelst, the speech-language pathologist who provided the treatment to participants in the study. I appreciate her time-commitment, patience, and dedication during treatment sessions. I also would like to thank my other committee members, Dr. Jamie Reilly and Francine Kohen for their time, support and words of encouragement throughout this project. Finally, I would like to thank my family and close friends for their endless support during this process.

TABLE OF CONTENTS

Page

CHAPTER

TABLE OF CONTENTS

(CONTINUED)

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Background

Childhood Apraxia of Speech (CAS) is a motor speech disorder characterized by difficulty with planning movements for speech (Strand et al., 2006). CAS has caused controversy within the field of speech-language pathology due to its difficulty to diagnose and little or slow progress in treatment. The American Speech-Language Hearing Association (ASHA) Technical Report defines CAS as a distinct type of pediatric speech sound disorder in which precision of movements are compromised with no signs of neuromuscular deficits (ASHA, 2007a). The ASHA report indicates that diagnosis and treatment for CAS are challenging and not standardized. The current best practice for diagnosis of CAS is considered to be expert opinion based on three hallmark features of CAS: inconsistent productions on vowels and consonants, prosodic errors and lengthened and disrupted co-articulatory transitions (ASHA, 2007a). However, once CAS is diagnosed, the speech-language pathologist faces more challenges in terms of treatment approaches and prognostic indicators. Since there is limited evidence to support treatment approaches (e.g., ASHA, 2007a; Murray, McCabe, & Ballard, 2014), there are no clear prognostic indicators to provide children and their families. Nevertheless, ASHA (2007) notes that intensive and individualized treatment and the principles of motor learning appear to be important in treatment. The ASHA Position Statement (2007b) mentions that a research demand for CAS is to determine the most effective ways to treat children with CAS. This question requires studies that examine different components of treatment programs in order to determine the best course of action for CAS. The present study aims to examine the role of intensity of treatment for CAS to provide evidence for speech-language pathologists, parents, and children who have been diagnosed with this disorder.

Principles of Motor Learning and CAS

Since speech production is a motor skill, the principles of motor learning (PML) are potentially useful to determine effective ways in which clinician can provide treatment for children with CAS. The principles of motor learning refer to variables that facilitate learning, such as more practice resulting in greater learning than less practice. Learning is defined as improved performance on retention and transfer tests (maintenance and generalization), rather than performance improvements observed during practice, because such improvements during practice may be temporary (e.g., Maas et al., 2008). The principles of motor learning have been traditionally studied in non-speech motor tasks, but researchers have applied them to speech motor tasks. Available evidence suggests that certain important aspects of motor learning, such as precursors to learning, conditions of practice and feedback, can be applied to speech motor tasks to improve speech (Edeal & Gildersleeve-Neumann, 2011).

Edeal and Gildersleeve-Neumann (2011) describe four important components of motor learning: precursors to learning, feedback, rate and conditions of practice. Precursors to motor learning are necessary elements that must be established between the client and clinician prior to onset of treatment. Edeal and Gildersleeve-Neumann (2011) highlight trust and established treatment goals as necessary precursors to motor learning.

The second component, feedback, is important as intrinsic and extrinsic feedback have different influences on treatment. Extrinsic feedback includes knowledge of results and performance on therapy tasks; however, equally as important is intrinsic feedback which requires the client to evaluate their own performance which is critical for self-monitoring (Edeal & Gildersleeve-Neumann, 2011). Rate is an important component to consider as it has an influence on learning. Reduction in rate decreases the mental load on the client to focus on the task at hand, as they improve, the clinician can encourage them to increase their rate. The final component, conditions of practice, describe the way in which targets will be elicited. A variety of practice conditions have been described as influential for learning, such as practice distribution (massed versus distributed practice) and practice schedule (blocked versus random practice) (Edeal & Gildersleeve-Neumann, 2011). For example, in blocked practice, one specific target is practiced at a time, whereas in random practice, more than one target is practiced at a time. Research has shown that blocked practice results in better performance during practice compared to random practice, but that random practice results in greater learning (maintenance and generalization) than blocked practice (see Maas et al., 2008, for review). In other words, gains observed for blocked practice are temporary and do not necessarily lead to robust *learning,* whereas gains from random practice are maintained over time.

Another example of a potentially useful practice condition is intensity, or practice distribution, the focus of the present study. Intensity describes the duration of sessions and the length of time between sessions, i.e. how practice is distributed over time (Edeal & Gildersleeve-Neumann, 2011). In the motor learning literature, a distinction is often made between massed and distributed practice. Massed practice refers to the practice on a given number of trials or sessions in a small period of time (e.g., 100 trials per session for 4 sessions yielding 400 productions). Distributed practice is given a number of trials or sessions over longer periods of time (e.g., 50 trials per session for 8 sessions yielding 400 productions). Discussion of massed versus distributed practice is common amongst speech-language pathologists. However, prior to discussing distribution of practice in terms of speech motor learning, it is important to examine non-speech motor learning as a base for developing a thorough hypothesis.

The non-speech motor learning literature tends to favor distributed practice over massed practice. Baddeley and Longman (1978) tested massed versus distributed practice while training postmen to type. The results indicated that distributed practice was more beneficial to learning and retaining the necessary motor skills. In a meta-analysis of motor learning experiments, Lee and Genovese (1988) found distributed practice yielded better performance consistently. Moreover, Lee and Genovese (1988) reported that retention appeared to be stronger in those studies which used distributed practice instead of massed practice. Shea, Lai, Black and Park (2000) examined the effects of practice distribution over days compared to within days for learning motor skills. They reported that distributed practice provided the greatest enhancement of motor skills. Certain studies have postulated hypotheses to explain the advantages of distributed practice over massed practice. Bahrick and Hall (2005) theorize the advantage of distributed practice to be due to retrieval failures as an individual learns over a longer period of time. Others have suggested that distributed practice may afford greater opportunity for memory consolidation, which depends on rest or periods of inactivity (e.g., Shea et al., 2000).

Despite the non-speech motor learning literature favoring distributed practice, findings in the neuroplasticity literature suggests massed practice may be better than distributed practice. Kleim and Jones (2008) discuss various findings within the literature which suggest distributed practice can cause weakening of synaptic responses whereas massed practice may cause long-term potentiation. For example, in studies of motor learning in rats, some studies have found greater learning and extent of neuroplasticity with 400 practice trials per day for 10 days (Kleim et al., 2002) than studies in which rats only performed approximately 63 trials per day for 20 days (Luke, Allred, & Jones, 2004). Thus, the neuroplasticity literature would predict greater learning with massed practice than with distributed practice. However, much of this evidence from the neuroplasticity literature has come from animal studies, making it unclear whether or to what extent these findings apply to treatment of children with CAS.

Practice Distribution (Intensity) in Speech Motor Learning

The opposing findings in the non-speech motor and neuroplasticity literature suggests that in reality, researchers do not know whether massed or distributed practice is more beneficial for learning. With respect to speech treatment specifically, Baker (2012) describes the fundamental question of intervention intensity stating that too much or too little could cause harm. Studies of intensity provide critical information for the speech-language pathologist as they can indicate how many sessions are necessary for progress and provide evidence to advocate for services. Baker (2012) identifies relevant concepts when examining intensity of sessions. First, it is important to consider the total number of sessions and their duration. Second, the number of teaching episodes in a given session are important to consider as it provides information about the level of support provided to the client. Finally, the number of client responses in a session, especially when discussing motor learning, informs the clinician about how many repetitions of a certain phrase are necessary to see progress in treatment.

Relevant terms are discussed by Warren, Fey, and Yoder (2007). First, *dose form* refers to the particular activity that the clinician provides, that is, what counts as a teaching episode. Second, *dose* indicates the number of teaching episodes that occur in a given session. Third, *dose frequency* is the number of intervention sessions per unit of time. Fourth, *total intervention duration* is the total period of time that a given intervention takes place. Finally, *cumulative intervention intensity* is dose x dose frequency x total intervention duration and represents the total amount of treatment. When discussing the present study and the CAS literature, all of the terms identified by Warren et al. (2007) are relevant as they indicate how much more evidence is needed when discussing treating children with CAS. Other studies have provided useful information for *dose form* by identifying what sort of treatment seems to work for CAS (see Murray et al., 2014, for systematic review), but relatively few studies have examined other aspects of intensity in treatment for CAS.

The literature has not provided a clear insight into how the speech-language pathologist should distribute practice across sessions to provide the most gains for the child. Studies have yielded results contrary to the motor learning literature when comparing massed versus distributed practice (dose frequency), in that several studies have suggested an advantage for massed over distributed practice. There has been research that examined *dose frequency* by comparing efficacy of treatment in massed versus distributed practice. Allen (2013) examined intensity of intervention for children with non-motor-based speech sound disorders (SSD) by comparing a multiple oppositions treatment provided three times per week for 8 weeks (massed practice) to once per week for 24 weeks (distributed practice). Results indicated that a massed practice was more beneficial for this approach.

In another study, Namasivayam et al. (2015) examined intensity in treatment of CAS. Children were divided into a 'higher intensity' group (2 sessions per week for 10 weeks) and a 'lower intensity' group (1 session per week for 10 weeks). Namasivayam et al. reported greater gains for the higher intensity group than for the lower intensity group. While this study examined intensity, it is limited because the lower intensity group received half as many sessions as the higher intensity group (i.e. also had a lower cumulative intervention intensity) suggesting that treatment progress cannot provide a clean comparison regarding practice distribution (dose frequency), as practice distribution was confounded with practice amount. Thus, future studies should control for amount of practice between groups.

A different study conducted by Thomas, McCabe, and Ballard (2014) found a slight benefit for massed practice over distributed practice in a comparison across studies. The study compared the efficacy of Rapid Syllable Transition treatment (ReST) for children with CAS provided twice-weekly for six weeks (distributed set) to ReST provided four times-weekly over three weeks (massed set) from a prior study (Murray, McCabe, & Ballard, 2015). In this way, the amount of practice (cumulative intervention intensity) was the same between studies. Results indicated that children in both groups (massed and distributed treatment) had maintained their skills at four months post-treatment. However, Thomas et al. (2014) found that children in the massed treatment condition had ongoing improvements after the cessation of treatment. Researchers concluded that a massed implementation of ReST treatment provides slight gains over a distributed implementation.

Thus, some initial evidence suggests that massed practice may be more beneficial than distributed practice in treatment for CAS specifically. However, these initial studies were either confounded by practice amount or rested on comparisons between different studies. Further research is needed regarding the various aspects of intensity in order to have a better understanding of treating CAS. In the present study, we kept constant the dose form, total intervention duration, and cumulative intervention intensity, and varied only dose and dose frequency at the level of individual targets, as explained further below.

Intensity and CAS - The Present Study

No studies have examined intensity while controlling for amount. The present study explores the impact of intensity of treatment while controlling for amount. An integral stimulation based approach adapted from Dynamic Temporal and Tactile Cueing (DTTC; Strand et al., 2006) was used due to its independently replicated evidence-base and because its focus on motor learning provides a logical vehicle to examine conditions of practice related to principles of motor learning. DTTC incorporates principles of motor learning and integral stimulation to emphasize shaping of movements for speech production slowly for positioning accuracy and to maximize proprioceptive processing (Strand et al., 2006). In a single subject multiple baseline design treatment efficacy study, Strand et al. (2006) found that DTTC resulted in gains for young children with CAS. While this study provided evidence for the efficacy of DTTC, it did not compare different practice conditions. In addition, this study used a very massed regimen (2 times a day, 5 times a week for 6 consecutive weeks), which is an atypical regimen and unrealistic for most settings. Therefore, it is not clear to what extent the improvements seen in Strand et al. (2006) are due to the DTTC approach or the intensive practice.

The present study directly examined practice distribution (dose frequency as defined above) in a single subject alternating treatments design with two children with CAS. As explained further below, practice distribution was defined over speech targets (a set of targets practiced over eight weeks or four weeks), in order to keep other variables related to intensity constant. The motor learning literature and the neuroplasticity literature suggest opposite predictions, as reviewed above. Specifically, if speech motor learning in CAS follows principles of motor learning, we predict greater learning for distributed practice items than for massed practice items. Conversely, if speech motor learning in CAS follows principles of neuroplasticity, we predict greater learning for massed practice items than for distributed practice items.

CHAPTER 2

METHODOLOGY

Participants

Participant Criteria and Evaluation

For the purpose of the study, two participants were recruited who met several inclusion criteria, including age between 4 and 12 years old, a verbal output of over 50 words, communicative intent, a score below the 10th percentile on the Goldman-Fristoe Test of Articulation (GFTA; Goldman and Fristoe, 2000), CAS as a primary diagnosis (see below for details), and normal hearing. Specific exclusionary criteria included concomitant diagnosis of neurobehavioral disorders that impact communication (such as autism), impaired vision, and significant impairments of oral structures (e.g., cleft palate). A full evaluation of the children was conducted using formal and informal measures. Information pertaining to case history was obtained through parent questionnaires.

In order to be included, the children were required to meet diagnostic criteria for CAS. Specifically, evidence of inconsistent errors on vowels and consonants, difficulty transitioning between sounds and syllables, and prosodic abnormalities were examined for diagnostic criteria of CAS (ASHA, 2007). Since there is no validated standardized diagnostic test for CAS, expert opinion is considered to be the best level of evidence to diagnose this disorder (McCauley & Strand, 2008). A primary diagnosis of CAS had to be agreed upon by four expert speech-language pathologists (SLP). Initial diagnosis was made by the evaluating SLP, who has extensive experience in diagnosis of pediatric speech disorders. Three additional expert SLPs observed a video recording of the evaluation and independently assigned a score of 0 (no CAS), 1 (possible CAS), or 2 (definitely CAS). To be included in the study, children must have received an average rating across all four SLPs ≥ 1 and could not receive a score of 0 from any SLP. These ratings were based on a variety of speech samples throughout the evaluation, including those from the Dynamic Evaluation of Motor Speech Skill (DEMSS; Strand, McCauley, Weigand, Stoeckel, & Baas, 2013), a test to assess motor planning and programming skills. In the DEMSS, the child is asked to repeat speech movement gestures which vary in length and complexity. Scoring evaluates vowel accuracy, utterance accuracy, consistency and stress accuracy. Since the DEMSS is not yet standardized and distributed for wide-spread use, the DEMSS was used here only to gather qualitative observations about speech motor skills; results were not compared to a normative sample.

A second criterion was that children must obtain a CAS Score of 2 on the maximum performance task (MaxPT) protocol developed by Thoonen et al. (1996, 1999) and manualized by Rvachew, Hodge, and Ohberg (2005). This protocol consists of nine tasks used to examine skills of the motor speech system, such as prolonging vowels and fricatives as long as possible and producing syllable sequences (e.g., ‘pataka’) at maximal rates. These tasks are scored and used to yield two independent scores: a dysarthria score and a CAS score. Each score has three levels: 0 = disorder not present, 1 = undetermined, 2 = disorder present). Thoonen et al. (1996, 1999) demonstrated that scores from this protocol differentiate CAS from other speech-sound disorders with 100% sensitivity and 91% specificity.

A final criterion for the diagnosis of CAS for the purpose of this study was that children must have demonstrated inconsistent productions on multiple attempts of the same word, as determined by a score of greater than 40% on the Inconsistency subtest of the Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd, Hua, Crosbie, Holm, & Ozanne, 2006). On this portion of the DEAP, children must repeat a list of 25 words three times and each word is scored as inconsistent (different production on attempts) or consistent (same production on all attempts).

In addition to the measures described below to determine diagnosis of CAS, a number of descriptive measures were collected, including measures of receptive language (Semel, Wiig, & Secord, 2003), expressive vocabulary and receptive vocabulary (Dunn & Dunn, 2007; Williams, 2007), and nonverbal cognitive skills (Reynolds & Kamphaus, 2003). Oral structure and function was examined using an oral motor exam developed for children (Robbins & Klee, 1987).

Several parent-rated measures were also obtained to characterize functional communication. These included the Focus on the Outcomes of Communication Under Six (FOCUS-34; Thomas-Stonell, Oddson, Roberston, Walker, & Rosenbaum, 2013) and the Intelligibility in Context Scale (ICS; McLeod, Harrison, & McCormack, 2012). The FOCUS-34 measures the child's current communication participation skills through presentation of situations that may occur in the home, school and within the community.

The parent rates statements on a 7-point scale where 1 indicates 'not at all like my child' and 7 indicates 'exactly like my child.' An example statement is: My child is understood the first time when talking with adults who do not know my child well. Scores on the FOCUS-34 are converted to represent the average rating on the 7-point scale. The ICS requires the parent to rate the degree to which their child's speech is understood by various communication partners on a 5-point scale ranging from 1 (never) to 5 (always). A sample item is 'Do your child's teachers understand your child?'. An average score is computed to represent the average rating.

Participant 1

Participant 1 was a male aged 7;11. Based on parent report, he had a positive family history of speech sound disorders, no reported feeding/swallowing issues, and normal hearing and vision. Participant 1 was diagnosed with CAS at age 3. The Goldman Fristoe Test of Articulation-2 (GFTA-2; Goldman & Fristoe, 2000) was used to assess the participant's articulation skills. Administration of the GFTA-2 allows for spontaneous elicitations of sounds at the single word level. On this test, Participant 1 received a standard score of 55, which was considered to be below average (3rd percentile).

With respect to CAS diagnosis, Participant 1 received an inconsistency score of 80% on the DEAP (Dodd et al., 2006), suggesting his productions were inconsistent. On the MaxPT protocol (Rvachew et al., 2005; Thoonen et al., 1996, 1999), Participant 1 received a CAS score of 2 suggesting the disorder was present. On the DEMSS, Participant 1 was found to have decreased articulatory accuracy, distorted vowels, vowel substitutions, inconsistent productions, and atypical prosodic errors. Finally, the average CAS rating by the four independent SLP experts was 1.88, and none of the SLPs assigned a score of 0 (1.5, 2, 2, 2). Thus, Participant 1 met all our stringent criteria for CAS diagnosis.

An oral motor exam (Robbins & Klee, 1987) revealed intact oral structure and some difficulty with precision, range, and coordination of sequenced movements (e.g., alternating lip rounding and protrusion). His vocal quality, loudness, and resonance were considered in the normal range. Thus, there was no evidence of structural abnormalities (e.g., cleft palate) or dysarthria.

In addition to testing the oral and speech motor system, several other measures were obtained. The Clinical Evaluation of Language Fundamentals - 4 (CELF-4; Semel et al., 2003) is a standardized measure used to assess receptive and expressive language skills. Subtests administered provide a scaled score with a mean of 10, average scores range from 7 to 13. For this study, only the three core receptive language subtests were administered. Participant 1 received an average score on all three subtests. The Peabody Picture Vocabulary Test - 4 (PPVT-4; Dunn & Dunn, 2007) was administered to measure the participant's receptive vocabulary. Participant 1 received a PPVT-4 standard score of 89, suggesting he is within the lower-average range. The Expressive Vocabulary Test - 2 (EVT-2; Williams, 2007) was administered to assess expressive vocabulary and word retrieval. Participant 1 received an EVT-2 score of 88 which is considered to be within the lower end of the average range. Finally, Participant scored in the normal range on the nonverbal subtests of the Reynolds Intellectual Assessment Scales (RIAS; Reynolds &

Kamphaus, 2003). Thus, Participant 1 did not present with co-occurring language impairments or nonverbal cognitive impairments.

In terms of functional communication, Participant 1 received a score of 2.0 (out of 7) on the FOCUS-34 (Thomas-Stonell et al., 2013), suggesting a significant limitation in communicative participation. On the Intelligibility in Context Scale (McLeod et al., 2012), Participant 1 received an average score of 3.2, suggesting that he was usually understood by immediate family but infrequently by unfamiliar individuals.

Participant 2

Participant 2 was a male aged 6;0 with no family history of speech sound disorders. His hearing was determined to be within normal limits but had a history of recurrent ear infections. Participant 2 received a CAS diagnosis at age 3. He received a GFTA-2 score of 45 which was considered to be below average (below the 1st percentile).

With respect to CAS diagnosis, he obtained an Inconsistency score of 64% on the DEAP (Dodd et al., 2006), suggesting inconsistent productions across repeated attempts. On MaxPT tasks, he received a CAS score of 2 and a Dysarthria score of 2, suggesting a diagnosis of dysarthria and CAS. Results of the DEMSS (Strand et al., 2013) revealed that Participant 2 had decreased articulatory accuracy, distorted vowels, vowel substitution errors, inconsistent productions. Finally, his average CAS rating across the four expert SLPs was 1.63, with no SLPs providing a rating of 0 (0.5, 2, 2, 2). Thus, Participant 2 met criteria for CAS diagnosis and inclusion in the study.

The oral mechanism exam (Robbins & Klee, 1987) revealed intact oral structures but some difficulty maintaining a strong lip seal, reduced range of movement for the tongue on protrusion and elevation. Occasional drooling was noted. Reduced vocal loudness and hypernasality were noted consistently on all speech tasks. Taken together, these observations suggest presence of dysarthria, characterized primarily by weakness. This is consistent with his MaxPT Dysarthria score.

In addition to speech impairments (CAS and dysarthria), Participant 2 also presented with expressive and receptive language impairments. On the CELF-4 (Semel et al., 2003), he was unable to complete the Concepts and Following Directions subtest despite multiple attempts. On the Word Classes subtest, he received a score of 7 which is considered to be low average, and on the Sentence Structure subtest, Participant 2 received a standard score of 1 which was well below average. His standard score on the PPVT-4 (Dunn & Dunn, 2007) was 75, which was below average. On the EVT-2 (Williams, 2007), Participant 2 received a standard score of 60, which was below average. Finally, Participant 2 scored in the normal range for his age on measures of nonverbal cognition.

In terms of functional communication, Participant 2 received a score of 3.15 on the FOCUS-34 (Thomas-Stonell et al., 2013), indicating a significant limitation in communicative participation. On the ICS (McLeod et al., 2012), Participant 2 received a score of 3.2, indicating that he is usually understood by immediate family and teachers but rarely by unfamiliar listeners.

Table 1.

Participant Information

Testing	Participant 1: Age: 7;11	Participant 2 Age: 6;0
FOCUS-34 (max. 7)	2.0	3.15
ICS (max. 5)	3.2	3.2
Word Inconsistencies on DEAP Criterion >40%	80%	64%
GFTA-2 SS (%ile) Criterion: Below 10th percentile	55 (< 3rd)	45 (< 1st)
MaxPT 0 = disorder not present 1 = undetermined 2 = disorder present	CAS Score: 2 Dysarthria Score: 0	CAS Score: 2 Dysarthria Score: 2
CELF-4 Average Score = 7-13	Concepts and Following Directions: 9 Word Classes: 12 Sentence Structure: 12	Concepts and Following Directions: no score[a] Word Classes: 7 Sentence Structure: 1
EVT SS (%ile) Average Score > 75	88 (21st)	60 (0.4th)
PPVT SS (%ile) Average Score > 75	89 (23rd)	75 (5th)
RIAS	Normal Range	Normal Range

[a] Participant 2 had significant difficulty with the Concepts and Following Directions subtest despite multiple attempts.

Design

An alternating treatments design with multiple baselines across behaviors and an untreated control condition was used. The treatment was divided into 2 phases, Phase I and Phase II, each involving eight sessions over four weeks (see Figure 1). Each session was divided into halves and split between conditions (massed practice, distributed practice). The order of conditions was pseudo-randomized for each session (session 1 order determined randomly by roll of a die, session 2 the reverse order of conditions, session 3 order determined by roll of a die again, etc.).

As a means of comparing massed practice versus distributed practice, the critical variable of interest, three set of ten targets were generated for each condition (Massed, Distributed , Control). The Massed sets were divided into two subsets of five items; each subset was practiced only in one phase. The ten Distributed items were practiced in both phases. Control items were never practiced and served as a basis for comparison of treatment effects. Target selection procedures are described in the next section.

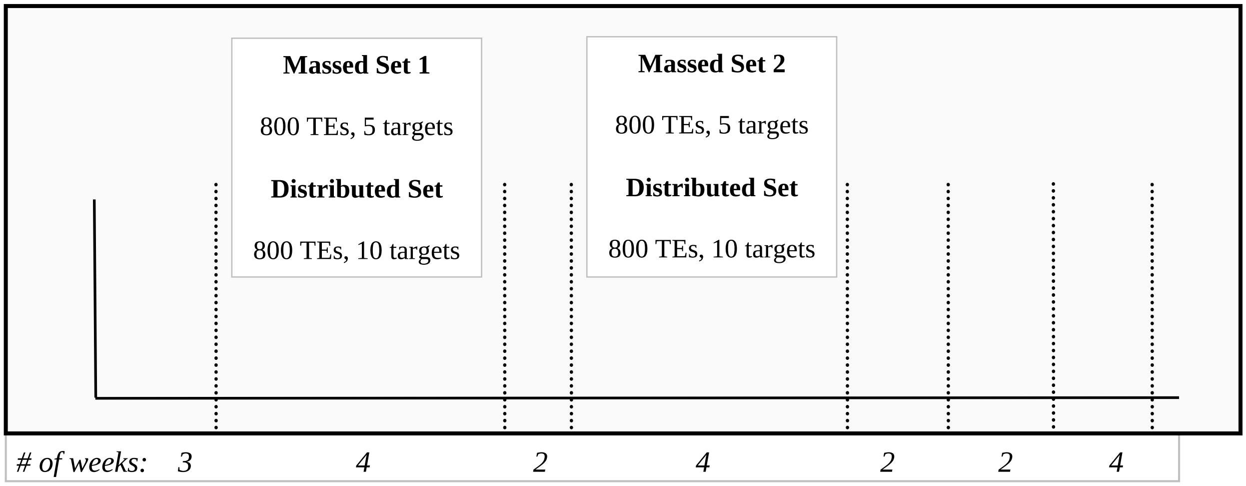


Figure 1. Overview of the design. TEs = Teaching Episodes. See text for further detail.

Target Selection

For each participant, a list of potential targets (words or phrases) was generated. In order to identify appropriate targets, the SLP worked in conjunction with participants and their parents to develop targets based on personal relevance and interest as well as baseline speech skills. Target suggestions were developed from structured questionnaires given to the parents in order to increase functionality of targets for the child (Wilson & Gildersleeve-Neumann, 2014).

From this larger list, three sets of ten items were created that each included items from the same interest categories (e.g., favorite food items, names of loved ones). This was done in an attempt to minimize potential differences in motivation. In addition, the sets were balanced in terms of syllable length and baseline accuracy as scored by a trained research assistant (two-sampled two-tailed t-test p-values > 0.35 for both participants). Matching of sets was completed before assigning a list of targets to a treatment condition in order to minimize risks of bias. When the sets were matched, they were randomly assigned to a condition: Control set, Distributed practice, Massed practice. Finally, the Massed set was divided into two subsets of five items, each matched to each other and to the other sets, and then randomly assigned to Phase I or II. Item sets for each participant are provided in Tables 2 and 3 below.

The ten items in the Distributed set were practiced over eight weeks of treatment (i.e., in both phases), and the ten items in the Massed set were practiced over four weeks of treatment (i.e., 5 items in Phase I, 5 items in Phase II). In this way, given a fixed amount of time in each session (25 minutes per condition), massed items were practiced more frequently in a shorter amount of time. At the end of the eight weeks of treatment both distributed and massed sets were practiced an equal number of times, whereas the distribution of practice differed. In other words, we varied only dose (number of trials per session). Dose form (treatment approach), dose frequency (number of sessions per week), total intervention duration (8 weeks), and cumulative intervention intensity (total amount of practice) were the same between conditions. Massed items for Phase II were probed during Phase I, representing the multiple baselines across behaviors component of the design, to ensure that there was no carryover between Phase I and Phase II.

Table 2.

Item Sets for Partcipant 1

DISTRIBUTED	**MASSED**	**CONTROL**
I want a cereal bar	breakfast (1)	I like chocolate milk
XXX is my speech teacher	My dad's name is XX (1)	X is my brother
XXX (friend’s name)	When are we going? (1)	it's almost 6 o'clock
I like playdough	5 minutes later (1)	XXXX XXX Lane (address)
I live in XXX	When is the movie? (1)	Where's the bathroom?
What time is it?	Minecraft (2)	Do you want to play?
memory	XX is my friend (2)	I'm sorry
Talk to you later	My name is XX XXX (2)	kitchen
Beauty and the Beast	Can you help me please? (2)	after 2 hours
What's your name?	Have a nice day (2)	Where are my shoes?

Note: Identifying information is redacted (each syllable replaced by X). For massed items, number in parenthesis indicate treatment phase.

Table 3.

Item Sets for Participant 2

DISTRIBUTED	**MASSED**	**CONTROL**
XX XX (name)	excuse me (1)	Pepper (dog)
thank you	yesterday (1)	hello
teacher	Mickey (1)	towel
doing	What's your name? (1)	tomorrow
Disney World	sister (1)	Phanatic
XX (friend)	XXX (sister) (2)	XX (friend)
please	XX (friend) (2)	see you later
hamburger	good morning (2)	guacamole
Where is the bathroom?	pizza (2)	How are you?
I'm tired	I'm hungry (2)	I'm thirsty

Note: Identifying information is redacted (each syllable replaced by X). For massed items, number in parenthesis indicate treatment phase.

Procedures

Probe Procedures

In order to assess speech accuracy, productions were elicited on a probe task. Probes included 30 items (10 items for *distributed practice*, 10 items for *massed practice,* and 10 control items). To assess independent production, delayed reading (Participant 1) or delayed imitation (Participant 2) were utilized to elicit responses. Items on each probe were presented in a new random order each time. No feedback was provided on probe performance.

Probes were presented throughout the study, beginning in baseline. During treatment, probes were presented weekly at the beginning of the session in order to measure learning from previous sessions. During the 2-week maintenance periods following Phase I and Phase II, two probes were administered, once per week. Finally, to measure long term retention, two follow-up probes were administered following the final treatment session at 4 weeks and 8 weeks.

Treatment procedures

This study utilized an integral stimulation based approach adapted from DTTC (e.g., Strand et al., 2006). Treatment used principles of motor learning and decreasing levels of support as appropriate. At the beginning, as necessary, the child had high support and treatment moved towards more independent productions. First, the speech-language pathologist encouraged production through imitation and as productions became more accurate, the child was encouraged to produce targets independently. To increase motor planning time and maximize awareness and proprioceptive feedback, the clinician encouraged slow productions of targets. As productions improved, the rate of productions increased to that of a conversational rate. Cues were varied and faded throughout the treatment process. Feedback was provided on a fading feedback schedule (as in Ballard, Maas, & Robin, 2007), and included verbal knowledge of results and performance. Targets were practiced in random order in each condition.

A teaching episode was defined as a sequence of events that began and ended with elicitation of a given target using a specific elicitation method, with clinician feedback and supported practice (‘teaching’) in between these two elicitations. If the child’s response was incorrect at the start of a teaching episode, the SLP provided verbal feedback (if the teaching episode was marked for feedback), and then used cues and instructions to elicit up to five additional supported attempts. Techniques depended on the child’s error and responsiveness to particular cues, but included visual and tactile cues, slow simultaneous production, and verbal descriptions (e.g., “that’s too choppy”, “round your lips more”). After several supported production attempts, the clinician elicited another production using the same elicitation method as at the start of the teaching episode and provided feedback, before starting the next teaching episode for the next target. If a child’s response was correct at the initial attempt of the teaching episode, the clinician provided feedback (if the teaching episode was marked for feedback) and immediately elicited another attempt (e.g., “can you say that again?”).

The level of support on initial elicitation varied depending on the child’s response. Each condition in each session began with immediate imitation at a normal, conversational rate (“watch me, listen to me, say what I say” – the integral stimulation component of the treatment). If the child produced incorrect responses on the first attempt of two consecutive teaching episodes, the elicitation method changed to immediate imitation at a slower rate. If the child produced correct responses on the first attempt of two consecutive teaching episodes, the SLP elicited the next teaching episode at the next level (e.g., delayed imitation at conversational rate, independent production). A detailed protocol is provided in Appendix A and B.

Data Analysis

Since this project is a single-case experimental design study, the primary comparison was between massed and distributed sets for each child. This comparison provided information regarding which intensity of treatment (*dose frequency)* results in greater learning (retention) in children with CAS. The dependent measure was perceptual accuracy of speech on probes, and was analyzed from audio recordings by independent raters who were blinded both to treatment status of targets (massed, distributed, control) and to time of probe (e.g., baseline, follow-up), to eliminate analyst bias. Any clues regarding treatment status or time of probe were removed from audio recordings before analysis, and probes were scored in randomized order to minimize systematic effects of perceptual drift.

Perceptual accuracy was captured using a scoring system that was adapted from previous studies using similar treatment approaches (Maas & Farinella, 2012; Maas et al., 2012; Strand et al., 2006). Detailed information regarding the scoring system is provided in Appendix C. Briefly, raters scored a child's productions on two components, segmental and prosodic accuracy. For segmental scoring, each syllable in a target was judged for accuracy, and the number of syllables with errors was used to compute a segmental accuracy score with a maximum of 2, as follows: 2 = all syllables correct (“correct”), 1 =

at least 50% of syllables correct (“minor error”), 0 = less than 50% of syllables correct (“major error”). Errors included substitutions, omissions, additions, distortions, and intrasyllabic metatheses, but specific error types were not tallied. For prosody scoring, prosodic accuracy of the entire target was rated on a binary scale (correct/incorrect). Prosodic errors included syllable segmentation, stress errors, and omissions and additions of syllables. Self-corrections were also included for the prosody scoring as these affect the entire utterance and reflect some level of difficulty with production of the target. Segmental and prosodic scores were then combined so that each attempt at a target has 3 possible points (2 for segmental accuracy and 1 for prosodic accuracy). Inter-rater reliability was assessed by having a second blinded independent rater score 16.2% (6/37) of the probes. The average percentage of scores within 1 scale point was 93% (SD = 4%, range 88-98%).

Scores were averaged across targets within a set and converted to a percent accuracy score for each probe. Data was analyzed via visual inspection of graphs, supplemented with standardized effect sizes to compare conditions. Effect sizes were computed as follows: (performance post treatment) – (average of baseline performance) / (standard deviation of baseline performance) (e.g., Beeson & Robey, 2006; Busk & Serlin, 1992). Effect sizes for each condition were calculated separately for the immediate post-treatment maintenance phase (M2), the follow-up phase (FU), and the entire post-treatment phase (M2 and FU combined). We operationally defined an effect to be present if the effect size was > 1 (as in Maas et al., 2012).

If speech motor learning in CAS follows the motor learning literature, we would expect to see larger effect sizes in the Distributed set than in the Massed set. In contrast, if speech motor learning in CAS follows the neuroplasticity literature, we would expect larger effect sizes in the Massed sets than in the Distributed sets. Finally, treated items (Distributed and Massed) were expected to show larger effect sizes than untreated items (Control).

CHAPTER 3

RESULTS

Data for each child can be found in Figure 2 (Participant 1) and Figure 3 (Participant 2). Each participant is described in detail below. Effect sizes for both participants are reported in Table 4, and absolute percent change from baseline is presented in Table 5.

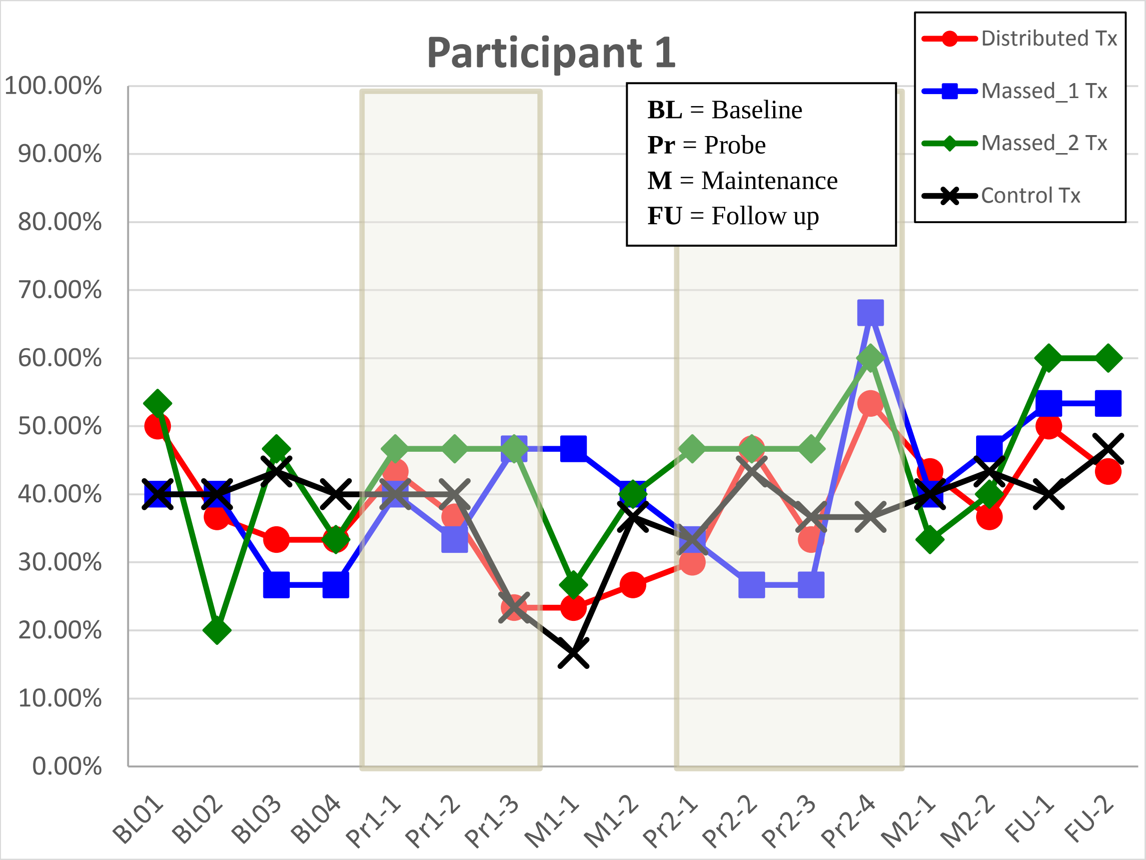


Figure 2. Data for Participant 1.

Participant 1

As the graph demonstrates, Participant 1 demonstrated a relatively stable or slightly declining baseline (data points from BL01-BL04). Following onset of treatment, the first massed set was in treatment and improvements were noted on this set. The distributed set demonstrated a decrease in accuracy in the first treatment phase, as did the control set, whereas the Massed_2 set showed no change. In the second treatment phase, the distributed set did demonstrate an increase in accuracy, as did the Massed_2 set. With the exception of the last probe in phase II, the Massed_1 set and control set did not show clear increases in accuracy during phase II. Inspection of the graph suggests that the control set demonstrated a slight increase in accuracy across the study period. At follow-up, both massed sets demonstrated marginally higher accuracy than the distributed set.

As can be seen in Table 4, effect sizes for Participant 1 revealed no evident change during the immediate post-treatment maintenance phase in any of the three conditions. However, during the follow-up phase, Participant 1 showed improvement on all sets, with a notably larger effect size for the Massed condition (2.65) than the other two conditions, which did not differ substantially (Distributed = 1.05, Control = 1.50). This pattern is also reflected in Figure 2: both Massed sets have greater accuracy than the other sets during follow-up. When combining data across the entire post-treatment phase, there was improvement from baseline for the Massed condition items (1.59), but no improvement for the Distributed items (0.63) and a borderline improvement for Control items (1.00). As the absolute percent change scores (Table 5) show, the overall gains are modest in all conditions.

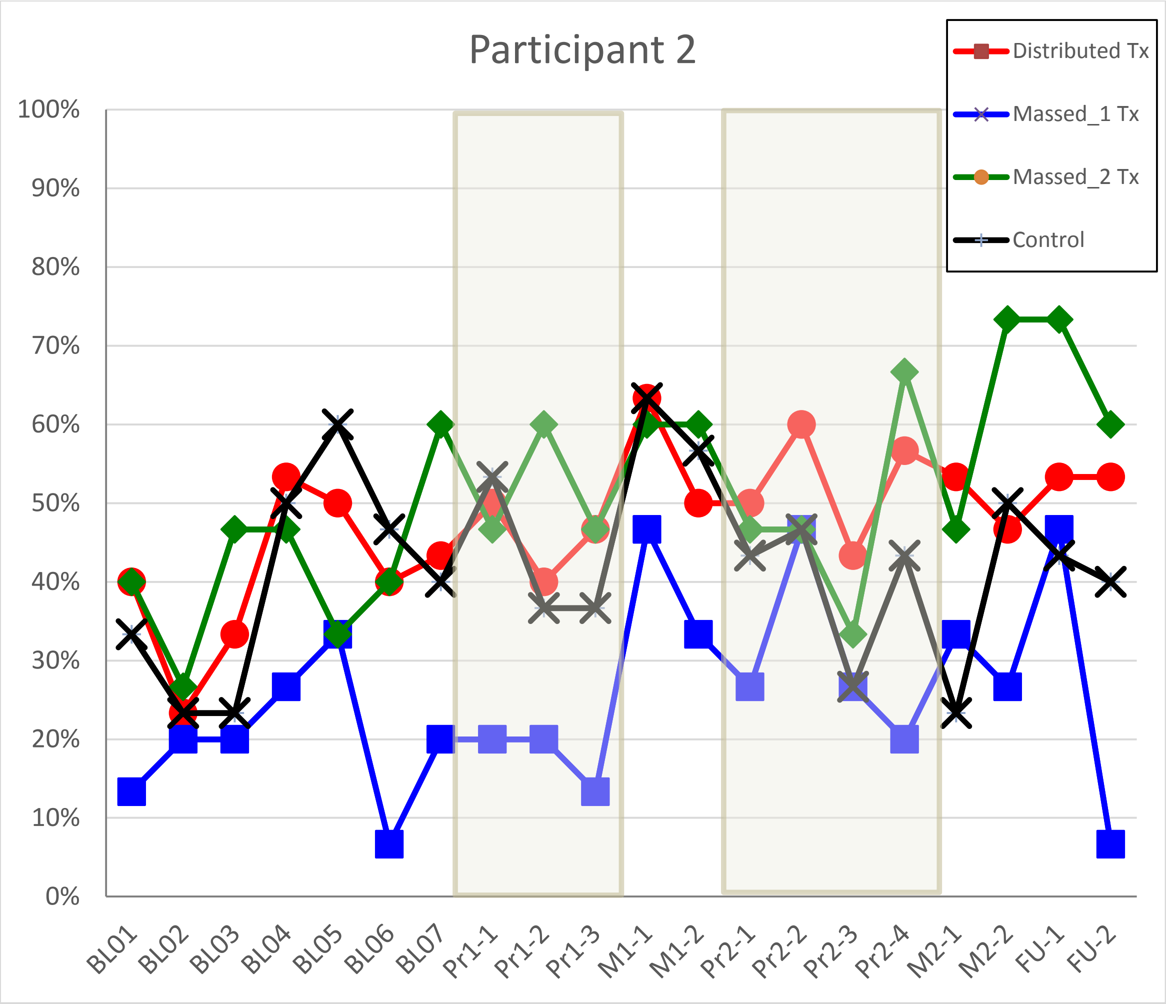


Figure 3. Data for Participant 2.

Participant 2

Participant 2 presented a slightly increasing baseline (data points from BL01-BL05) but stabilized before initiation of treatment (BL06-BL07). There was no clear improvement during treatment phase 1, but accuracy was higher than baseline in the maintenance phase for all sets, although these improvements were modest. In the second treatment phase, the Massed_2 set initially showed a decrease in accuracy but improved by the final treatment probe session (Pr2-4). The distributed set remained stable throughout this second phase of treatment and did not show further improvement. The control set and Massed_1 were relatively stable or slightly declining. Following all treatment, little change was observed in any set except for the Massed_2 set, which showed continued improvement.

In terms of effect sizes (Table 4), the combined Massed condition showed a clear positive effect, both in the maintenance phase (2.13) and at follow-up (2.39). The Distributed condition showed no effect during the maintenance phase (0.94) but a positive effect emerged during the follow-up phase (1.28). There was no effect for the Control set in either phase (M2: -0.21, FU: 0.16). Absolute percent change also favored the Massed condition but changes were modest, as were differences between conditions (Table 5).

Table 4: Effect sizes by condition and phase for both participants

	Participant 1			**Participant 2**		
	M2	**FU**	**M2 + FU**	**M2**	**FU**	**M2 + FU**
Distributed	0.21	1.05	0.63	0.94	1.28	1.11
Massed	0.53	2.65	1.59	2.13	2.39	2.26
Control	0.50	1.50	1.00	-0.21	0.16	-0.03

Table 5: Absolute percent change from baseline by condition and phase for both participants.

	Participant 1			**Participant 2**		
	M2	**FU**	**M2 + FU**	**M2**	**FU**	**M2 + FU**
Distributed	2.0%	8.0%	5.0%	9.5%	12.9%	11.2%
Massed	4.0%	21.0%	12.5%	14.1%	15.7%	14.9%
Control	1.0%	3.0%	1.7%	-2.9%	2.1%	-0.3%

CHAPTER 4

DISCUSSION

The present study examined the role of intensity (massed versus distributed) of practice when treating CAS using a modified version of Dynamic Temporal Tactile Cueing (DTTC) (Strand et al., 2006), an integral stimulation based treatment approach. A single-case experimental design was utilized with multiple baselines. The findings obtained from the two participants in this study suggest that massed practice was marginally more beneficial than distributed practice. Both participants showed larger standardized effect sizes for maintenance and follow-up for the Massed condition items than for the Distributed condition items or the Control condition items. This pattern of greater improvement for items in massed practice was also reflected in the absolute percent change data, although it should be noted that gains and condition differences were relatively modest. Taken together, the findings obtained are consistent with the neuroplasticity literature, which suggests that learning is optimized with massed practice. These findings are consistent with those of previous studies (Namasivayam et al., 2015; Thomas et al., 2014), but provide more direct evidence for a massed practice benefit because the present study involved a within-study and within-participant comparison (unlike Thomas et al., 2014) and did not confound distribution and amount of practice (as in Namasivayam et al., 2015).

A further observation is that improvements were relatively modest, in terms of effect size and absolute percent change, in particular for the Distributed condition. Gains in the present study were less clear and strong than in previous studies using similar integral stimulation based treatment approaches for CAS (e.g., Edeal & Gildersleeve-Neumann, 2011; Strand et al., 2006). A number of factors may have impacted results of this study, including motivation, the nature of targets, and length and intensity of treatment.

It is possible that the participants' level of motivation impacted how they performed in treatment. Motivation is an important precursor to treatment (Maas et al., 2008). According to Maas et al. (2008), motivation encompasses the participant's understanding of the purpose of the task and knowledge of their goals. Strand et al. (2006) also suggested that the child must be motivated and desire change. For example, if participants did not fully understand the purpose of their participation in treatment (i.e. to improve intelligibility and prevent communication breakdown), they may not have participated in treatment sessions to the best of their ability. Although motivation could have impacted results obtained in the study, motivation was not formally measured, as there are no measures of motivation that are appropriate for or validated with children with speech disorders. Anecdotally, both children understood the purpose of treatment and were motivated, although it is possible that the purpose of the probe task was less clear, and thus motivation for this task may have been suboptimal. Future research should examine the role of motivation in treatment and develop ways to increase motivation during treatment as well as probe outcome tasks.

The present study sought to facilitate motivation by choosing treatment targets that were personally meaningful to each child. This approach was based in part on findings from Maas and Farinella (2012) and Maas et al. (2012), who used a similar treatment approach with children with CAS but chose words and phrases that were not specifically selected for personal functional relevance but rather for particular properties (e.g., consonant clusters, stress patterns). In those studies, improvements were observed for treated items, but no generalization to untreated items was evident for any of the children. Thus, Maas et al. (2012) recommended that this general treatment approach be applied to functionally meaningful targets, and this was the approach adopted here. In the present study, we selected targets based on functional relevance with little consideration of whether targets were within capacity in terms of phonetic inventory. For example, if a child's name includes the sound 'r', it was included even if that sound was not yet in the phonetic inventory. This was deemed appropriate on motivational grounds (a child is likely motivated to be able to say his/her name) and because children with CAS do not necessarily follow typical developmental trajectories for sound production (ASHA, 2007a). However, it is possible that this resulted in targets that were too complex, affecting both potential for improvement and possibly motivation. Future studies should examine the role of target complexity in treatment for CAS.

The results may also have been impacted by the length and intensity of treatment. Other studies have utilized different amounts and distributions of practice and reported greater gains (e.g., Edeal & Gildersleeve-Neumann, 2011; Strand et al., 2006). For example, Edeal and Gildersleeve-Neumann included two children with CAS, one receiving three 40-minute sessions per week for 10 consecutive weeks and the other receiving two 50-minute sessions for 5 consecutive weeks. Strand et al. included four children who received two 30-minute sessions per day, five days per week, for six consecutive weeks. In the present study, children received two 50-minute treatment sessions per week for two 4-week blocks, with only half of each session devoted to each condition. This was done to improve the adaptability of this treatment approach to a more realistic time schedule for a clinical setting. One possible interpretation of the discrepancy between these prior studies and the present study is that indeed practice amount and intensity are important parameters of treatment, and that there may be a critical minimum amount, distribution, and duration of treatment to achieve gains. However, there were other important methodological differences between these studies that may account for the differences in outcomes (e.g., target selection, data analysis). Therefore, future studies should examine these parameters further while keeping other methodological variables the same.

Finally, although it is possible that improvements observed in the present study are attributable to maturation or other factors rather than treatment, the consistent benefit of massed practice across both participants suggests that this is unlikely to be the complete story. Massed practice does appear to have a benefit above and beyond the improvements seen in the untreated control items, whereas distributed practice may not be sufficient to induce clear gains beyond no treatment. This notion has important implications for clinical practice. In particular, it appears that more practice, in a massed distribution, may be necessary for large, meaningful gains to emerge. Thus, these findings may lend further credence to the utility of intensive treatment programs for children with CAS that have begun to emerge in recent years, both in clinical practice and in research contexts (e.g., Preston, Leece, & Maas, 2016; Strand et al., 2006). Future research should continue to study various aspects of intensity of treatment for CAS, as well as other principles of motor learning, in order to develop maximally effective treatment approaches for children with CAS.

CHAPTER 5

CONCLUSION

The results from this study suggested a slight benefit of massed practice over distributed practice in a single case experimental design study. However, caution must be exercised when interpreting results, as the condition differences were small. Gains were relatively modest compared to previous studies. Future research should continue to examine aspects of treatment intensity in order to devise optimal treatment approaches and regimens for children with CAS.

REFERENCES

Allen, M. M. (2013). Intervention efficacy and intensity for children with speech sound disorder. *Journal of Speech, Language, and Hearing Research, 56,* 865-877.

American Speech-Language-Hearing Association. (2007a). Childhood Apraxia of Speech [Technical Report]. Available at www.asha.org/policy.

American Speech-Language-Hearing Association. (2007b). Childhood Apraxia of Speech [Position Statement]. Available at www.asha.org/policy.

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics, 21*, 627–635.

Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory & Language, 52*, 566-577.

Baker, E. (2012). Optimal intervention intensity. *International Journal of Speech-Language Pathology, 14*, 401–409

Ballard, K. J., Maas, E., & Robin, D. A. (2007). Treating control of voicing in apraxia of speech with variable practice. *Aphasiology, 21*, 1195–1217.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessonslearned from the aphasia literature. *Neuropsychology Review, 16*, 161-169.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-Case Research Design and Analysis: New directions for psychology and education.* Hillsdale, N.J.: Lawrence Erlbaum Associates

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2006). *Diagnostic Evaluation of Articulation and Phonology*. San Antonio, TX: PsychCorp.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th Edition). AGS/Pearson.

Edeal, D. M., & Gildersleeve-Neumann, C. E. (2011). The importance of production frequency in therapy for childhood apraxia of speech. *American Journal of Speech-Language Pathology, 20,* 95-110.

Forrest, K. (2003). Diagnostic criteria of developmental apraxia of speech used by clinical speech-language pathologists. *American Journal of Speech-Language Pathology,* 12, 376–380.

Goldman, R., & Fristoe, M. (2000). *The Goldman–Fristoe Test of Articulation* (2nd Edition). Austin, TX: Pro-Ed.

Kleim J. A., Barbay S., Cooper N. R., Hogg T. M., Reidel C. N., Remple M. S., et al. (2002). Motor learning-dependent synaptogenesis is localized to functionally reorganized motor cortex. *Neurobiology of Learning and Memory, 77,* 63–77.

Kleim, J. A., & Jones, T. A. (2008). Principles of experience-dependent neural plasticity: Implications for rehabilitation after brain damage. *Journal of Speech, Language, and Hearing Research, 51*, S225-S239.

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport, 59,* 277-287.

Luke, L. M., Allred, R. P., & Jones, T. A. (2004). Unilateral ischemic sensorimotor cortical damage induces contralesional synaptogenesis and enhances skilled reaching with the ipsilateral forelimb in adult male rats. *Synapse, 54,* 187–199.

Maas, E., Butalla, C. E., & Farinella, K. A. (2012). Feedback frequency in treatment for childhood apraxia of speech. *American Journal of Speech-Language Pathology, 21,* 239-257.

Maas, E., & Farinella, K. A. (2012). Random versus blocked practice in treatment for childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 55,* 561-578.

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17,* 277-298.

McCauley, R. J., & Strand, E. A. (2008). A review of standardized tests of nonverbal oral and speech motor performance in children. *American Journal of Speech-Language Pathology, 17,* 81-91.

McLeod, S., Harrison, L. J., & McCormack, J. (2012). The Intelligibility in Context Scale: Validity and reliability of a subjective rating measure. *Journal of Speech, Language, and Hearing Research, 55,* 648-656.

Murray, E., McCabe, P., & Ballard, K. J. (2014). A systematic review of treatment outcomes for children with Childhood Apraxia of Speech. *American Journal of Speech-Language Pathology, 23,* 486-504.

Murray, E., McCabe, P., & Ballard, K. J. (2015). A randomized controlled trials for children with childhood apraxia of speech comparing Rapid Syllable Transition treatment and the Nuffield Dyspraxia Programme–Third Edition. *Journal of Speech, Language, and Hearing Research, 58,* 669-686.

Namasivayam, A. K., Pukonen, M., Goshulak, D., Hard, J., Rudzicz, F., Rietveld, T., Maassen, B., Kroll, R., & van Lieshout, P. (2015). Treatment intensity and childhood apraxia of speech. *International Journal of Language and Communication Disorders, 50,* 529-546.

Powers, C., Gildersleeve-Neumann, C. E., Jakielski, K. J., Maas, E., & Stoeckel, R. (2015). Investigating amount and intensity of practice in treatment of childhood apraxia of speech. Poster presented at the American Speech-Language-Hearing Association Convention (Denver, Colorado, November 2015).

Preston, J. L., Leece, M. C., & Maas, E. (2016). Intensive treatment with ultrasound visual feedback for speech sound errors in childhood apraxia. *Frontiers in Human Neuroscience, 10*:440. doi: 10.3389/fnhum.2016.00440

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales* (RIAS). Lutz, FL: Psychological Assessment Resources, Inc.

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., Hodge, M., & Ohberg, A. (2005). Obtaining and interpreting maximum performance tasks from children: A tutorial. *Journal of Speech-Language Pathology and Audiology, 29,* 146-157.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals®* (4th Edition). San Antonio, TX: Pearson.

Shea, C. H., Lai, Q., Black, C., & Park, J. H. (2000). Spacing practice sessions across days benefits the learning of motor skills. *Human Movement Science, 19,* 737–760.

Strand, E. A., Stoeckel, R., & Baas, B. (2006). Treatment of severe childhood apraxia of speech: A treatment efficacy study. *Journal of Medical Speech-Language Pathology, 14,* 297-307.

Strand, E. A., McCauley, R. J., Weigand, S. D., Stoeckel, R. E., & Baas, B. S. (2013). A motor speech assessment for children with severe speech disorders: Reliability and validity evidence. *Journal of Speech, Language, and Hearing Research*, *56*(2), 505-520.

Thomas, D., McCabe, P., & Ballard, K. (2014). Rapid Syllable Transitions (ReST) treatment for Childhood Apraxia of Speech: The effect of lower dose frequency. *Journal of Communication Disorders, 51,* 29–42.

Thomas-Stonell, N., Washington, K., Oddson, B., Robertson, B. & Rosenbaum, P. (2013). Measuring communicative participation using the FOCUS: Focus on the Outcomes of Communication Under Six. *Child: Care, Health, and Development*, *39,* 474–480.

Thoonen, G., Maassen, B., Gabreëls, F., & Schreuder, R. (1999). Validity of maximum performance tasks to diagnose motor speech disorders in children. *Clinical Linguistics & Phonetics, 13,* 1-23.

Thoonen, G., Maassen, B., Wit, J., Gabreëls, F., & Schreuder, R. (1996). The integrated use of maximum performance tasks in differential diagnostic evaluations among children with motor speech disorders. *Clinical Linguistics & Phonetics, 10,* 311-336.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews, 13,* 70-77.

Williams, K. (2007). *Expressive Vocabulary Test* (2nd Edition). Circle Pines, MN: American Guidance Services.

Wilson, A., & Gildersleeve-Neumann, C. E. (2014). Functional Communication Parent Questionnaire. Unpublished instrument, Portland State University.

APPENDIX A - TREATMENT PROTOCOL

Teaching episodes will be presented in random order for both conditions. There will be 5 targets for the massed condition and 10 targets for the distributed condition in each treatment phase. A card deck will contain 5 stimulus cards for each target word per condition. Random order will be implemented by shuffling the cards before the treatment session and reshuffling the deck as needed during the session once every card has been used. Steps for each teaching episode are as follows: (see also Protocol Flowchart below)

1. Start teaching episode by taking stimulus card from deck.

2. Elicit first attempt using one of the following elicitation methods (see below for criterion to switch between elicitation methods):

- Immediate imitation, slow rate ("watch me, listen carefully, and say ___.")
- Immediate imitation, normal rate ("watch me, listen carefully, and say ___.")
- Delayed imitation, normal rate ("watch me, listen carefully, and say ___." Point to child after 3 seconds to cue delayed imitation)
- Spontaneous production (Show picture associated with target and say "What is this?")

3. Child responds (***count as attempt #1***)

If production is <u>correct</u>:

4. Wait 2-3 seconds

a. provide feedback, if indicated for this trial (e.g., "good one!" "that was right")

b. don't provide feedback, if not indicated for this trial. Instead provide general encouragement (e.g., "Keep doing the best you can")

5. Elicit another attempt (e.g., "can you say that again?")

6. Child responds (count as attempt #2)

7. Wait 2-3 seconds

a. provide feedback, if indicated for this trial.

b. don't provide feedback, if not indicated for this trial; general encouragement.

8. Place a check mark on the feedback tracking sheet (see below)

9. Go to Step 1 with next target to start next teaching episode (e.g., "Let's do another one").

APPENDIX A: TREATMENT PROTOCOL (CONTINUED)

If production is incorrect:

4. Wait 2-3 seconds

    a. provide feedback, if indicated for this trial (e.g., “good try, but not quite”, “close!”, “your mouth was too open”)

    b. don’t provide feedback, if not indicated for this trial. Instead provide general encouragement (e.g., “Keep doing the best you can”)

5. Provide appropriate support (e.g., slow simultaneous production, tactile/gestural cue)

6. Elicit another attempt using same elicitation method as for attempt #1

7. Child responds (count as attempt #2).

8. Wait 2-3 seconds

    a. provide feedback, if indicated for this trial.

    b. don’t provide feedback, if not indicated for this trial; general encouragement.

9. Place a check mark on the feedback tracking sheet (see below)

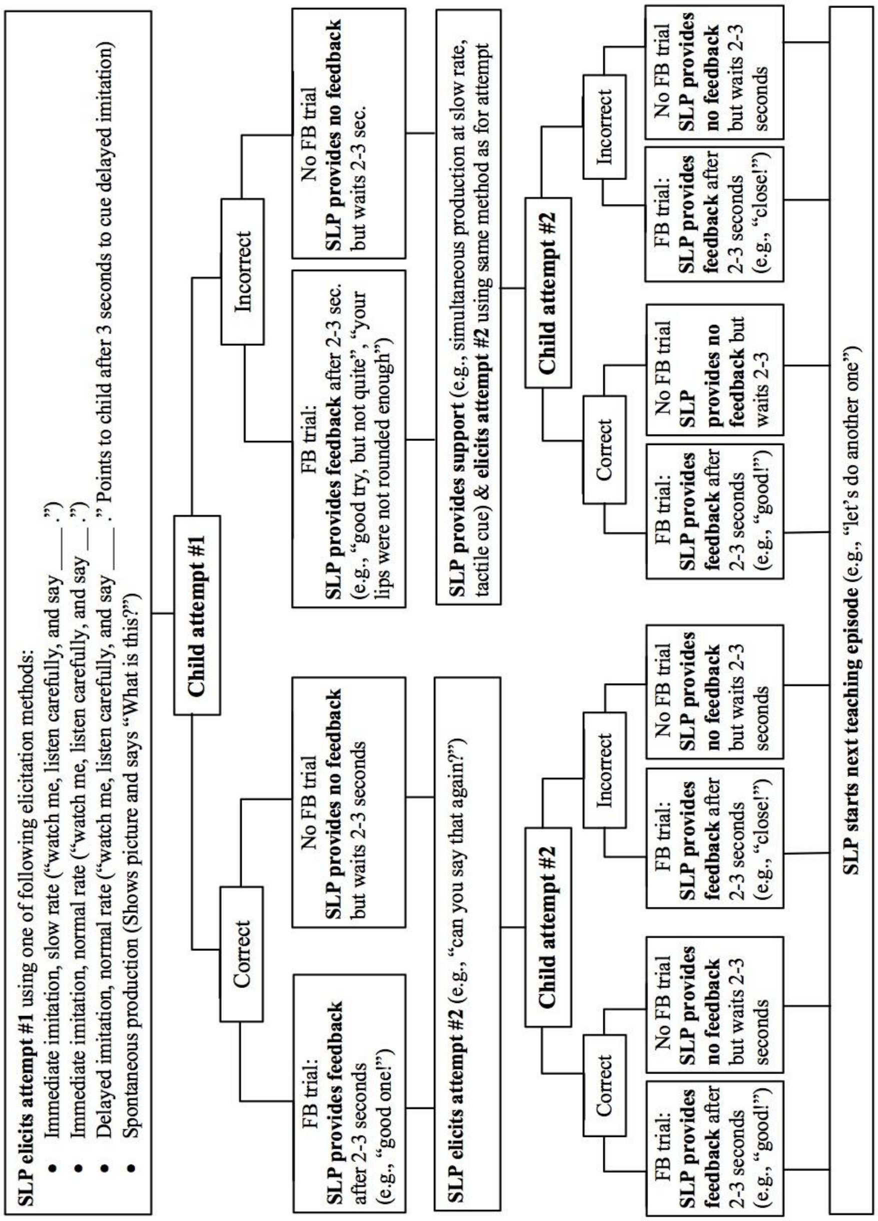
10. Go to Step 1 with next target to start next teaching episode (e.g., “Let’s do another one”).

Continue until a total of 100 trials have been completed or one half of the session time has elapsed. After a brief break, begin the second treatment condition and continue until a total of 100 trials have been completed or the session is ended.

Feedback is given with decreasing frequency during each condition. Feedback is provided for 100% (10) of the first 10 teaching episodes, 90% (9) of the next 10 teaching episodes, and so on until feedback is given on 10% (1) of the last 10 teaching episodes. To facilitate keeping track of feedback schedules, the SLP will have a feedback tracking sheet with 100 slots with the fading feedback schedule marked (cf. Ballard et al., 2007). By placing a check mark in a slot for each teaching episode, the SLP can easily see whether the next trial requires feedback or not.

Start with immediate imitation at normal rate. The criterion to increase difficulty level of elicitation condition is 2/2 consecutive correct on the first attempt for a given target within a session. That is, if the child produces target A correctly after the first elicitation on two consecutive teaching episodes of target A (separated by other targets or not), then the third teaching episode of target A will be elicited with the next level elicitation method. Similarly, if the child produces a target incorrectly on 2/2 consecutive teaching episodes of target A with an elicitation method, the next teaching episode of target A will revert to the previous difficulty level of elicitation method (or stay in immediate imitation at slow rate).

APPENDIX B - PROTOCOL FLOWCHART



APPENDIX C: SCORING SYSTEM

Background & Overview

For our purposes, the scoring system has to meet several goals:

- It has to be valid. For our purposes that involves two things: (1) it has to be able to differentiate minor from major errors, weighted to some extent by target complexity, and (2) it has to be sensitive to change – i.e. It has to be able to detect changes (e.g., due to treatment), including for targets of different complexity levels. For example, a minor error on a 5-syllable phrase is less serious than a minor error on a 1-syllable word.

- It has to be reliable. Different raters should be able to agree, and the same rater should be consistent with her/himself. This also relates to replicability of research findings. The detailed instructions below are intended to ensure reliability. We will check this of course.

- It has to be (at least somewhat) clinically relevant. We are using perceptual judgments rather than acoustic measures, for example. This also means that it has to be manageable. We don't want to do narrow transcription but rather use a bit more of a holistic judgment of accuracy.

- The system should be able to distinguish between segmental and suprasegmental accuracy.

The scoring system described below meets these criteria. This scoring system is adapted from Strand et al. (2006), and was also used in Maas et al. (2012) and Maas and Farinella (2012), but adapted to account for different target lengths/complexities (operationalized here in terms of number of syllables). NOTE that we judge accuracy (not intelligibility) of the child's production.

The system includes segmental and prosody scoring components, both of which involve a binary scale (correct/incorrect). Briefly, for Segmental scoring, raters judge the accuracy of each syllable in a target, and for Prosody scoring, raters judge the prosodic accuracy of the entire target word or phrase. Segmental accuracy of the entire target word or phrase is then scaled to a 3-point scale (correct, minor errors, major errors) according to the criteria outlined in Table 1 below; this is accomplished automatically in Excel using a formula so raters do not need to apply those criteria. Combining the Segmental and Prosodic scores, each target has a maximum possible score of 3 (2 points for segmental accuracy + 1 point for prosodic accuracy). Each scoring component is explained in further detail below, followed by some general guidelines and examples.

APPENDIX C: SCORING SYSTEM CONTINUED

NOTE: This version of the scoring system was adapted from a longer version intended to train treatment scorers.

Table 6. Segmental Scoring

Score	**Description**	**Criteria**
2	Correct	No segmental errors in any syllable in any word
1	Minor error	≥ 50% but <100% of syllables correct
0	Major error	<50% of syllables correct

For Segmental Scoring, we derive a score based on the proportion of syllables with (one or more) segmental errors (defined in Table 2 below). The criteria for the Segmental Score are given in Table 1 above; these criteria are designed to be applicable to both single-word targets and multi-word targets (but note that monosyllabic targets can only receive a 0 or 2). A further rationale for aspects of this system is provided below.

Raters do not need to apply the criteria in Table 1; these scores are calculated automatically in the Excel sheet. Instead, raters simply indicate for each syllable whether it was correct or not (in terms of segmental accuracy). We do not distinguish between major and minor segmental errors, nor do we classify error types. The judgment is only whether the syllable contains one or more of the listed errors or not. If a syllable is correct, enter 1 in the relevant cell in the spreadsheet; if a syllable has any of the listed errors, enter 0 (see example in Figure 1). Based on these judgments, a formula in the Excel file then automatically applies the score criteria given in Table 1, as well as a percentage of syllables correct (see cell U2 in Figure 1).

Table 7. Segmental Error Types and Explanations

Error Type	**Description**	**Notes**
Substitution	Consonant or vowel that sounds like an example of a different phoneme.	May also be distorted but still sounds like a phoneme that is not the target.
Omission	Missing sound(s).	
Distortion	Consonant or vowel that is recognizable as the target sound but is not a clear, well-articulated example of that sound.	May sound like a sound between the target and some other sound, but not always (e.g., a lateral lisp, a frictionalized stop).

		Includes excessive aspiration of plosives, derhoticization of rhotic vowels/diphthongs and consonants, excessive lengthening of consonants, and any other distortions. Vowel lengthening is not scored for the Segmental score but affects the Prosody score if it affects the perception of stress (Score 0 if it leads to perception of exaggerated, reduced or incorrect stress).
Addition	A consonant is added to a syllable.	Vowel/syllable additions are scored in *Prosody Scoring.*
Metathesis	Two consonants are misordered within the syllable	Metathesis between syllables is counted as 2 syllables in error (e.g., /kato/ for ‘taco’ is 2 syllables with errors), whereas metathesis within a syllable is counted as only 1 syllable in error (e.g., /supɚnæm/ for ‘Superman’ is 1 syllable with errors). Whole-syllable or whole-word metatheses are scored a 0 in Prosody Scoring. For Segmental Scoring, treat each syllable as though it appeared in the correct position.
Unintelligible	Syllable is unrecognizable	
Note: Error types are judged for each syllable of the target		

Prosody Scoring

For Prosody Scoring, only score correct (1) vs. incorrect (0), based on presence of the errors and abnormalities listed in Table 3 below. Basically, raters judge stress patterns, correct number of syllables, and fluency (syllable repeats, syllable segmentation). In addition, we include self-corrections here as well. A further rationale for this system is provided below.

Table 8. Prosodic error types and explanations.

Error Type	**Description**	**Notes**
Segmentation	Pauses between sounds or syllables.	Score 0 regardless whether all or only some syllables are segmented.
Omitted syllable(s)	One or more syllables are missing.	Missing syllables affect prosody. BUT normal contractions that result in an omitted vowel/syllable should not be penalized because those are not considered true omissions. Score both syllables correct unless there are errors
Added syllable(s)	One or more syllables are added.	Vowel additions result in added syllables, which affect prosody. This also includes partial or whole-syllable repetitions (disfluencies). For Segmental Scoring: If the added syllable is in the middle of the word, give the child the benefit of the doubt by scoring the syllable that is segmentally accurate as the target, regardless of position
Syllable reversals	Two or more syllables are switched.	For Segmental Scoring, score each syllable as though it appeared in correct position.
Reduced/equalized stress	Reduced differentiation between stressed and unstressed syllables.	This can be the result of vowel lengthening or shortening, alterations of loudness, and/or changes in pitch.
Exaggerated stress	Correct stress pattern but exaggerated.	This can be the result of vowel lengthening or shortening, alterations of loudness, and/or

		changes in pitch.
Incorrect stress	Reversed stress pattern – not simply equalized or exaggerated. (stressed becoming unstressed and vice versa).	This can be the result of vowel lengthening or shortening, alterations of loudness, and/or changes in pitch.
Self-corrections	Child initiates another attempt without prompting.	If the clinician elicits another attempt, it is likely due to inattention or background noise, etc. In this case, score the newly-elicited attempt.
Note: If any of these are present, the Prosody Score = 0.		

General Scoring Rules & Procedures

- Generally, judge the first attempt by a child, unless the child produces a spontaneous selfcorrection (i.e. not prompted by the clinician). Spontaneous self-corrections are scored incorrect for Prosody Scoring but are scored the same as first attempts with respect to segmental accuracy (you only score segmental accuracy for self-corrections because the prosody score is already 0).
- If targets are elicited via a repetition task, then judgment is relative to the model presented by the clinician. In other words, if the clinician uses abnormal prosody and the child imitates that, this is not an error on the child’s part. Be sure to make a note if this happens.
- Dialectal variations or normal speech phenomena are not considered errors.
- For Segmental Scoring, use the Sonority Sequencing Principle and Onset Maximization Principle to determine syllable boundaries (and thus whether one or two syllables contain errors).