JSLHR

Research Article

Bang for Your Buck: A Single-Case Experimental Design Study of Practice Amount and Distribution in Treatment for Childhood Apraxia of Speech

Edwin Maas,[a] Christina Gildersleeve-Neumann,[b] Kathy Jakielski,[c] Nicolette Kovacs,[a] Ruth Stoeckel,[d] Helen Vradelis,[a] and Mackenzie Welsh[a]

Purpose: The aim of this study was to examine 2 aspects of treatment intensity in treatment for childhood apraxia of speech (CAS): practice amount and practice distribution.
Method: Using an alternating-treatments single-subject design with multiple baselines, we compared high versus low amount of practice, and massed versus distributed practice, in 6 children with CAS. Conditions were manipulated in the context of integral stimulation treatment. Changes in perceptual accuracy, scored by blinded analysts, were quantified with effect sizes.
Results: Four children showed an advantage for high amount of practice, 1 showed an opposite effect, and 1 showed no condition difference. For distribution, 4 children showed a clear advantage for massed over distributed practice post treatment; 1 showed an opposite pattern, and 1 showed no clear difference. Follow-up revealed a similar pattern. All children demonstrated treatment effects (larger gains for treated than untreated items).
Conclusions: High practice amount and massed practice were associated with more robust speech motor learning in most children with CAS, compared to low amount and distributed practice, respectively. Variation in effects across children warrants further research to determine factors that predict optimal treatment conditions. Finally, this study adds to the evidence base supporting the efficacy of integral stimulation treatment for CAS.
Supplemental Material: https://doi.org/10.23641/asha.9630599

Parents of children with childhood apraxia of speech (CAS), and speech-language pathologists (SLPs) who work with these children, often ask how much treatment is needed for meaningful improvement and how best to distribute treatment. Children with CAS often show little or slow progress in standard therapy (American Speech-Language-Hearing Association [ASHA], 2007; Campbell, 1999; Hall, 2000; Shriberg, Aram, & Kwiatkowski, 1997), leading to recommendations for intensive intervention to optimize outcomes (ASHA, 2007; Hall, 2000; Strand & Skinder, 1999). Given that resources (money, personnel, time) are limited, it is important to maximize use and impact of these limited resources. The current study continues and extends a systematic research program that examines effects of various practice conditions in treatment for CAS to optimize outcomes (Edeal & Gildersleeve-Neumann, 2011; Maas, Butalla, & Farinella, 2012; Maas & Farinella, 2012). In particular, we examine two critical aspects of intensity, namely, *amount* and *distribution* of practice, in six children with CAS, in the context of an integral stimulation–based intervention.

Our approach to optimizing treatment efficacy for children with CAS is guided by the motor learning literature, which offers insight into factors that facilitate learning (retention and transfer) of motor skills (Schmidt & Lee, 2005). This literature indicates that some practice conditions enhance learning relative to others, effects sometimes referred to as principles of motor learning (e.g., Strand, Stoeckel, &

[a]Department of Communication Sciences and Disorders, Temple University, Philadelphia, PA
[b]Department of Speech and Hearing Sciences, Portland State University, OR
[c]Department of Communication Sciences and Disorders, Augustana College, Rock Island, IL
[d]Department of Speech-Language Pathology, Mayo Clinic, Rochester, MN

Correspondence to Edwin Maas: emaas@temple.edu

Helen Vradelis is now at the Department of Communication Sciences and Disorders, University of Wisconsin, Madison.

Editor-in-Chief: Julie Liss
Editor: Maria Grigos

Received May 29, 2018
Revision received November 30, 2018
Accepted May 2, 2019
https://doi.org/10.1044/2019_JSLHR-S-18-0212

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Baas, 2006; see Maas et al., 2008, for a review). Incorporation of principles of motor learning in treatment for CAS has been recommended (e.g., Hall, 2000; Robin, Maas, Sandberg, & Schmidt, 2007) because practice conditions exert their effect primarily on motor planning and programming (e.g., Wright, Black, Immink, Brueckner, & Magnuson, 2004), the presumed locus of deficit in CAS (ASHA, 2007; Maas & Mailend, 2017; Nijland et al., 2002; Strand et al., 2006). The effects of these practice conditions are thought to arise at the level of motor planning and programming because reaction time evidence suggests that (for example) random practice, but not blocked practice, of multicomponent movement sequences results in recoding and consolidation of the sequence into a single motor plan (Wright et al., 2004). The ability to form stable motor plans for multipart sequences (such as multisyllabic words and phrases) is important for developing accurate, fluent, and efficient speech, and therefore, conditions that facilitate such recoding would seem particularly relevant for children with CAS. Several treatment approaches incorporate principles of motor learning (e.g., Murray, McCabe, & Ballard, 2015; Skelton & Hagopian, 2014; Strand et al., 2006). However, the evidence for such recommendations remains limited because few studies have directly compared conditions in CAS treatment (Edeal & Gildersleeve-Neumann, 2011; Maas et al., 2012; Maas & Farinella, 2012; Namasivayam et al., 2015; Preston, Leece, McNamara, & Maas, 2017), and although findings from the motor learning literature can provide “best guesses” about optimal conditions, ultimately these principles of motor learning must be studied in the populations of interest (e.g., CAS) because principles of motor learning are largely based on studies of adults with intact motor systems learning nonspeech motor tasks (see Maas et al., 2008, for further discussion).

This study focused on two critical aspects of treatment intensity (amount and distribution), because intensity is frequently cited as a critical treatment variable (Edeal & Gildersleeve-Neumann, 2011; Hall, 2000; Maas, Gildersleeve-Neumann, Jakielski, & Stoeckel, 2014; Warren, Fey, & Yoder, 2007). In fact, intensive intervention programs for CAS have emerged in various places across the United States in recent years, and although several studies have examined outcomes of such intensive programs (Murray, McCabe, & Ballard, 2015; Namasivayam et al., 2015; Preston, Leece, & Maas, 2016; Strand et al., 2006), intensity has received little systematic study in treatment for CAS. Only Edeal and Gildersleeve-Neumann (2011) and Namasivayam et al. (2015) have directly compared high and low amounts of practice, and no studies have examined intensity while controlling for practice amount. Thus, the primary purpose of this study was to provide direct, controlled evidence regarding the role of amount and distribution of practice in treatment for CAS. A secondary goal was to further examine the efficacy of integral stimulation treatment approaches for CAS; these approaches incorporate integral stimulation as a core element (use of multimodal cues to elicit and support a child’s attempt, e.g., simultaneous visual and auditory models; see This Study and Method sections). Integral stimulation–based approaches have among the strongest evidence to date (Maas et al., 2014; Murray, McCabe, & Ballard, 2014), with several studies from independent research groups reporting treatment effects, despite differences in protocols (e.g., Edeal & Gildersleeve-Neumann, 2011; Maas & Farinella, 2012; Strand et al., 2006). To address this secondary goal, we used integral stimulation–based treatment as the vehicle to study intensity.

Practice Amount

Practice amount refers to the number of practice trials and sessions provided throughout the treatment period (cumulative intervention intensity; Warren et al., 2007). The motor learning literature indicates that more practice results in greater retention (Shea, Kohl, & Indermill, 1990), and similar effects have been reported for typical speakers (Kim, LaPointe, & Stierwalt, 2012). More practice trials provide more opportunity for the learner to determine the relationship between motor commands (pattern of muscle contractions), initial starting configuration (current location of articulator), and movement outcome (whether target was achieved). Motor learning, including speech motor learning, involves establishing such complex relationships over multiple attempts at a movement (Guenther, Ghosh, & Tourville, 2006; Schmidt & Lee, 2005; Wolpert, Ghahramani, & Flanagan, 2001). The learner can then use this knowledge to perform this and similar movements in the future by deriving motor commands for the intended goal given the current position of articulators. This view of motor learning accounts for the fact that people can perform a given motor task despite variability in context. Thus, a larger number of practice trials results in more accurate and robust knowledge of the relationships that support performance on future attempts.

In the CAS literature, two studies examined practice amount (Edeal & Gildersleeve-Neumann, 2011; Namasivayam et al., 2015). In an alternating-treatments single-subject design with two children, Edeal and Gildersleeve-Neumann randomly assigned target sounds to two conditions: moderate (30–40 trials/session) and high production frequency (100+ trials/session). Sounds were embedded in functional words with simple syllable shapes (CV, VC, CVC, CVCV) and practiced using an integral stimulation approach. Both children showed greater retention and transfer for more frequently practiced sounds on probes at the end of treatment sessions. While these results are encouraging, the study included only two children who received different practice amounts (cumulative intervention intensity) and different practice distributions. Furthermore, probe scores only took into account accuracy of the target sounds rather than the entire word.

Namasivayam et al. (2015) used a pre/post design to compare two groups of children with CAS receiving a motor-based speech intervention (which also included integral stimulation methods) with different practice amounts. The more-practice group ($N = 21$) received 20 sessions over 10 weeks; the less-practice group ($N = 12$) received 10 sessions over 10 weeks. SLPs ($N = 46$) were assigned to one of these conditions using stratified random assignment to account for clinician skill level. Namasivayam et al. compared performance before and after practice in each group separately for speech sound accuracy, intelligibility, and functional communication. Significant changes for speech sound accuracy and functional communication were found only for the more-practice group, although both groups showed comparable effect sizes, and no direct group comparisons were performed. Sentence intelligibility did not improve for either group.

Thus, available evidence suggests that more practice results in greater learning, including for children with CAS. However, to date, this evidence is limited to two published studies, one of which (Namasivayam et al., 2015) confounded practice amount and distribution: The more-practice group received twice as many sessions per week as the less-practice group (i.e., sessions were more closely spaced). This study was designed to conduct a within-study comparison of different practice amounts while controlling for practice distribution.

Practice Distribution

While clinicians and researchers generally agree that more practice leads to better outcomes, it is not clear how to distribute practice across time to maximize learning. Studies using similar treatment approaches report different magnitudes of improvement (e.g., Maas & Farinella, 2012; Strand et al., 2006), which may be due in part to different practice distributions (Edeal & Gildersleeve-Neumann, 2011; Maas et al., 2012, 2014). Practice distribution (dose frequency; Warren et al., 2007) refers to how a given practice amount is divided over time and typically involves a distinction between massed and distributed practice (Baddeley & Longman, 1978; Shea, Lai, Black, & Park, 2000). In massed practice, many trials occur in a short period, whereas in distributed practice, the same number of trials is divided over a longer period.

The motor learning literature shows greater retention for distributed practice than for massed practice (Baddeley & Longman, 1978; Shea et al., 2000; see Lee & Genovese, 1988, for a meta-analysis). This effect is not specific to motor learning but is also observed in other learning domains and is sometimes referred to as the *spacing effect* (e.g., verbal learning: Bahrick & Hall, 2005; Bjork & Allen, 1970). With respect to motor learning, in a seminal study, Baddeley and Longman trained four groups of British postal workers on a keyboarding task for mail sorting. All participants received 60 hr of training and were randomly assigned to one of four schedules applied to a 5-day workweek. All groups received practice 5 days per week, but the length and number of sessions were systematically varied as follows: (a) one session of 1 hr per day (12 weeks of training: $12 \times 5 \times 1 = 60$ hr); (b) two sessions of 1 hr per day, separated by at least 2 hr (6 weeks: $6 \times 5 \times 2 = 60$ hr); (c) one session of 2 hr per day (6 weeks: $6 \times 5 \times 2 = 60$ hr); and (d) two sessions of 2 hr per day, separated by at least 2 hr (3 weeks: $3 \times 5 \times 4 = 60$ hr). Retention was examined at 1, 3, and 9 months following training, using subgroups matched on performance at the end of training. The two groups receiving 1-hr sessions (Groups 1 and 2) had significantly faster rates of correct key strokes than groups receiving 2-hr sessions (Groups 3 and 4), with the group of one 1-hr session per day (most distributed practice, Group 1) showing the fastest performance. The group receiving two 2-hr sessions per day (most massed practice, Group 4) was not only the slowest but also made significantly more errors than other groups. Although all groups showed some decrease in speed of correct keystrokes at follow-up, this decrease stabilized after 3 months for all groups except for the most massed practice group, which declined further at 9 months.

Various explanations have been proposed for the spacing effect. One hypothesis is that long-term memory formation depends on strengthening (consolidation) of neural and cognitive representations (Brashers-Krug, Shadmehr, & Bizzi, 1996; Shea et al., 2000) and that successful consolidation requires some degree of uninterrupted time ("downtime") after practice. Massed practice can be viewed as a disruption of the consolidation process because the relative amount of "downtime" is less than when practice trials or sessions are spaced further apart. Another explanation is that distributed practice results in more detailed representations, because it results in greater forgetting than massed practice (Bjork & Allen, 1970; Lee & Weeks, 1987). The idea is that, in massed practice, the action plan representation is still available in memory and thus requires less reconstructive processing to retrieve and perform again. By contrast, in distributed practice, the learner has forgotten more of the task and has to engage in more processing to recreate the action plan. It is this additional practice in reconstructing the action plan that results in more elaborate action plans and ability to retain and transfer the skill (Lee & Weeks, 1987).

Based on this literature, we might predict a benefit of distributed over massed practice in treatment for CAS. However, whereas the neuroplasticity literature is in agreement with the motor learning literature regarding benefits of practice amount, evidence from the neuroplasticity literature suggests the opposite for practice distribution: Massed practice results in greater learning than distributed practice (Kleim & Jones, 2008). For example, Kleim et al. (2002) found improved reaching behaviors in rats trained on a regimen of 400 trials/day over 10 days, whereas Luke, Allred, and Jones (2004) found no improvement in reaching skills in rats trained on a regimen of 60 trials/day over 20 days. This difference led Luke et al. to suggest that massed (more intense) practice may be needed to induce changes (Kleim & Jones, 2008). However, this interpretation rests on a between-study comparison (with a confound between amount and distribution) and on an animal model. A randomized controlled trial with human learners examined the effect of treatment intensity on recovery of arm and leg function in stroke patients with limb paresis (Kwakkel, Wagenaar, Twisk, Lankhorst, & Koetsier, 1999). Improvement was greater with intense training than with less intense training at 6, 12, and 20 weeks post stroke. However, this study and others (e.g., Askim et al., 2010; Størvold, Jahnsen, Evensen, & Bratberg,

2018) also conflated amount and distribution. Overall, recommendations based on the neuroplasticity literature suggest that massed practice results in better learning than distributed practice, although amount and distribution are often conflated.

To our knowledge, no studies have examined practice distribution in typical speakers, but several studies have explored the role of practice distribution in speech intervention research for various clinical populations (Allen, 2013; Spielman, Ramig, Mahler, Halpern, & Gavin, 2007; Wambaugh, Nessler, Cameron, & Mauszycki, 2013; see Kaipa & Peterson, 2016, for a systematic review), including for those with CAS (Namasivayam et al., 2015; Thomas, McCabe, & Ballard, 2014). In Namasivayam et al. (2015; reviewed above), children with CAS received treatment twice per week (massed) or once per week (distributed) over 10 weeks. The massed group, but not the distributed group, showed significant improvement post treatment. Thus, massed practice appeared to benefit learning compared to distributed practice. However, as noted above, this study confounded distribution with amount of practice (massed practice also involved more practice), making it impossible to separate the contributions of each factor to the overall effect. Thomas et al. (2014) examined practice distribution of rapid syllable transition (ReST) treatment (Ballard, Robin, McCabe, & McDonald, 2010; Murray, McCabe, & Ballard, 2015). They compared four children with CAS who received distributed ReST (two sessions/week for 6 weeks) to massed (standard) ReST (four sessions/week for 3 weeks) published in a prior study (Murray, McCabe, & Ballard, 2015). Thomas et al. found comparable gains immediately post treatment but continued gains only for standard ReST, suggesting that massed practice is more beneficial than distributed practice.

Taken together, evidence from various literatures results in two competing hypotheses about optimal practice distribution for CAS: One hypothesis (speech motor learning in CAS follows principles of motor learning) predicts greater learning with distributed practice; the other (speech motor learning in CAS follows principles of neuroplasticity) predicts greater learning with massed practice. Both these literatures have replicated support for their predictions, and it is unclear a priori which of these is most likely to apply to speech motor learning in CAS. This uncertainty poses a challenge for SLPs and their clinical decisions regarding how to optimally structure treatment for children with CAS. Empirical study is needed to resolve this uncertainty and adjudicate between two plausible but contradictory hypotheses. Some research has begun to address this question. However, the number of studies with children with CAS is small, and given the noted confounds and reliance on cross-study comparisons, further research is warranted to determine optimal treatment distribution. The current study was designed to examine practice distribution in a single study while controlling for practice amount.

This Study

This study examined practice amount and distribution in the context of an integral stimulation treatment ("watch me, listen carefully, say what I say") that incorporates principles of motor learning, and tactile cues and reduced speech rate as needed (Strand et al., 2006). We chose this approach because it is considered "probably effective" (Murray et al., 2014) and the only approach with replicated evidence from independent research groups (Maas et al., 2014). Although details of implementation have varied across these studies (e.g., Edeal & Gildersleeve-Neumann, 2011; Maas & Farinella, 2012; Strand et al., 2006), all involved a core reliance on integral stimulation as support and elicitation technique, tactile cues, gradual fading of cues, and a focus on whole-word movement accuracy during treatment. If treatments involving the same components show efficacy despite variations in implementation (e.g., with respect to frequency of sessions, target selection criteria), the support for those shared components is strengthened by showing generalizability across variations of implementation.

We examined effects of practice amount and distribution on retention of practiced targets (words/phrases) using a single-case, experimental, alternating-treatments design with six children with CAS. Practice amount and distribution were manipulated by creating matched sets of targets for different conditions. The research questions and hypotheses were the following:

1. Does practice amount influence speech motor learning in treatment for children with CAS?
   a. **Hypothesis 1** (*more-is-better hypothesis*) is that more practice offers more opportunities for learning and predicts greater retention for targets that receive more practice.
2. Does distribution influence speech motor learning in treatment for children with CAS?
   a. **Hypothesis 2a** (*motor learning hypothesis*) is that speech motor learning in CAS follows motor learning principles and predicts greater retention for targets in distributed practice.
   b. **Hypothesis 2b** (*neuroplasticity hypothesis*) is that speech motor learning in CAS follows neuroplasticity principles and predicts greater retention for targets in massed practice.

Method

Participants

Six children with CAS (see Table 1) were recruited via postings on the Childhood Apraxia of Speech Association of North America website (apraxia-KIDS.org) and at local outreach events. Children were evaluated by an experienced SLP with expertise in CAS. Evaluation included a case history and formal and informal measures to evaluate speech, language, and cognitive skills. Four additional children were evaluated but did not meet our stringent inclusion criteria for CAS diagnosis (see below). All testing and treatment took place in a quiet, child-friendly room in the Speech, Language, and Brain Lab at Temple University. The number of sessions

Table 1. Participant information.

ID	Sex	Age	GFTA[a]	DEAP[b]	EVT[c]	PPVT[d]	CELF[e]	RIAS[f]	MaxPT Dys[g]	MaxPT CAS[h]	CAS Dx[i]
001	M	11;3	< 40	68%	73	74	61	48	2	2	1.50 (0.58)
002	M	7;11	55	80%	88	89	105	42	0	2	1.75 (0.50)
003	M	5;11	47	52%	88	92	98	43	0	2	1.75 (0.50)
004	M	5;11	41	72%	84	82	n/c	48	0	2	1.50 (0.58)
005	M	6;0	45	64%	60	75	n/c	45	2	2	1.75 (0.50)
008	M	4;7	75	52%	96	116	n/a	61	0	2	1.25 (0.50)

Note. M = male; n/c = not computed (see text); n/a = not administered.

[a]Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000) standard score. [b]Diagnostic Evaluation of Articulation and Phonology (Dodd et al., 2006) Word Inconsistency Subtest percent inconsistent. [c]Expressive Vocabulary Test–Second Edition (Williams, 2007) standard score. [d]Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007) standard score. [e]Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel et al., 2003) Receptive Language Index standard score. [f]Reynolds Intellectual Assessment Scales (Reynolds & Kamphaus, 2003) nonverbal cognition composite *t* score. [g]Maximum Performance Test Protocol (Rvachew et al., 2005; Thoonen et al., 1999) dysarthria score. [h]Maximum Performance Test Protocol (Rvachew et al., 2005; Thoonen et al., 1999) apraxia of speech score. [i]Mean (standard deviation) of four expert speech-language pathologist ratings (0 = *no CAS*, 1 = *possible CAS*, and 2 = *CAS*).

required to determine eligibility varied depending on how many sessions were needed to complete the tasks used to diagnose CAS; sessions typically lasted 60 min or less. Parent permission was obtained for all children, and all children provided assent. All study procedures were approved by the Temple University Institutional Review Board.

Inclusionary criteria were as follows: (a) age between 4 and 12 years; (b) from homes and educational settings where the primary language spoken was English; (c) verbal output (50+ words) and communicative intent as determined by the SLP and parent report; (d) presence of a speech sound disorder (SSD), based on a score below the 10th percentile on the Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); (e) CAS as the primary speech diagnosis (see below); (f) normal hearing based on parent report; and (g) typical nonverbal cognition as determined by a *t* score within 1.5 *SD*s of the mean on nonverbal subtests of the Reynolds Intellectual Assessment Scales (Reynolds & Kamphaus, 2003). Exclusionary criteria were as follows: (a) diagnosis of disorders that significantly affect communication and/or social interactions (e.g., autism), as per referral diagnosis; (b) uncorrected vision impairments (per parent report) that would likely interfere with the ability to take advantage of visual cues; (c) significant impairments of oral structure (e.g., cleft palate) as judged by the SLP based on an oral motor examination (Robbins & Klee, 1987); and (d) a primary diagnosis of dysarthria, as judged by the SLP. In addition, children were administered the Peabody Picture Vocabulary Test–Fourth Edition (Dunn & Dunn, 2007), the Expressive Vocabulary Test–Second Edition (Williams, 2007), and the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003). These measures were used for descriptive purposes only. The Dynamic Evaluation of Motor Speech Skill (DEMSS; Strand, McCauley, Weigand, Stoeckel, & Baas et al., 2013) was administered to obtain further information about motor speech skill from a dynamic assessment developed specifically for this purpose (Strand et al., 2013). The speech samples elicited from this test were used for clinical judgment; the total score on the DEMSS was not used to diagnose CAS. None of the children had a history of recurrent ear infections, feeding or swallowing problems, learning disability, neurological disease, or neurodevelopmental condition, except where noted.

Children had to meet stringent diagnostic criteria for CAS to qualify for the study. First, expert SLP judgment is currently the gold standard for CAS (ASHA, 2007; Murray, McCabe, Heard, & Ballard, 2015). To strengthen confidence in diagnosis, each child was required to be judged as having CAS by four independent expert SLPs, including the SLP who administered the assessment and three offsite experts (second, third, and fifth authors) who independently rated children from video recordings of the assessment (including DEMSS, Goldman-Fristoe Test of Articulation, Diagnostic Evaluation of Articulation and Phonology, conversational speech, and any other portion deemed relevant by the rater). Each SLP independently rated each child on a 3-point scale (0 = *no CAS*, 1 = *possible CAS*, and 2 = *CAS*) based on signature perceptual speech features of CAS (inconsistent vowel and consonant errors on repeated productions, difficulties achieving and transitioning into articulatory configurations, abnormal prosody). Inclusion required an average rating > 1 and no rating of 0 from any SLP (averages provided in Table 1).

Second, children were required to receive an apraxia score of 1 or 2 on the Maximum Performance Task (MaxPT) protocol developed and prospectively validated by Thoonen, Maassen, Wit, Gabreëls, and Schreuder (1996) and Thoonen, Maassen, Gabreëls, and Schreuder (1999) and manualized by Rvachew, Hodge, and Ohberg (2005). In this protocol, children sustain vowels and fricatives as long as possible and perform diadochokinetic tasks (e.g., "say /pʌtʌkʌ/ as fast and as long as you can"). Thoonen et al. (1999) showed that the combination of scores differentiates CAS from other SSDs with 100% sensitivity and 91% specificity.

Third, children had to exhibit inconsistent productions on repeated attempts of the same word, based on a score > 40% on the Word Inconsistency subtest of the Diagnostic Evaluation of Articulation and Phonology (Dodd, Hua, Crosbie, Holm, & Ozanne, 2006). In this test, children produce 25 words three times (with other tasks interposed), and each word is scored as consistent (same all three times) or inconsistent (two or more different productions).

CAS001. CAS001 was a right-handed boy aged 11;3 (years;months) who was adopted from South Korea at the age of 1 year and lived with his White adoptive parents and sister. He was exposed to Korean as an infant, but the language at home was English. Birth history and family history of speech disorders were unknown. Developmental milestones were slightly delayed, and at the age of 3 years, he was diagnosed with severe CAS. His speech errors were characterized by distortion, omission, and substitution errors and included gliding of liquids and cluster reduction. He showed reduced articulatory accuracy, vowel distortions and substitutions, atypical prosody, and inconsistent productions on repeated trials on the DEMSS. He performed below norms on all MaxPT tasks; he produced repetitions of monosyllables but produced no more than two trisyllable sequences in a row and made several sequencing errors. Oral mechanism examination revealed reduced range of motion during nonspeech oral motor sequences and limited tongue–jaw dissociation on tongue protrusion and elevation despite normal isolated nonspeech oral movements. He had occasional pitch breaks and difficulty sustaining phonation on repeated /hɑ.hɑ.hɑ/ productions. Loudness was at times reduced, and resonance was hypernasal, suggesting possible mild dysarthria. Receptive language scores were well below age expectations, although his scores must be considered with caution, as he exhibited marked apprehension of making errors, especially when items required a verbal response, and was often reluctant to respond and rehearsed silently before responding. Nonverbal cognition was in the normal range, and phonological awareness skills were below the normal range (Phonological Awareness Composite standard score = 67) based on the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999).

CAS002. CAS002 was a 7;11 left-handed non-Hispanic White boy who lived with his parents and older brother. English was the only language spoken at home. His mother reported a family history of speech disorders (his older brother had speech-language delays but normalized quickly with treatment). He was diagnosed with CAS at the age of 3 years. His speech errors were primarily omission and substitution errors and included initial voicing, final devoicing, cluster reduction, and gliding. He also exhibited an unusual but relatively consistent substitution of retroflex [ɭ] for /ɹ/. On the DEMSS, he demonstrated notably reduced articulatory accuracy, vowel distortions and substitutions, atypical prosody, and inconsistent productions on repeated attempts. Although his maximum phonation rate and maximum fricative duration rates were below age norms (Thoonen et al., 1999), his monosyllable repetition rate was in the normal range. He exhibited great difficulty with the trisyllable: He did not maintain the sequence, producing both vowel errors and consonant additions (e.g., [pɑptokɑ] for /pʌtʌkʌ/). The oral mechanism exam revealed normal range of motion and symmetry for isolated movements but some reduced range of movement, imprecision, and incoordination when alternating movements. Vocal quality, resonance, and loudness were within normal limits. His language and nonverbal cognition were in the normal range, and his phonological awareness was in the low average range (Comprehensive Test of Phonological Processing Phonological Awareness Composite standard score = 88).

CAS003. CAS003 was a 5;11 non-Hispanic right-handed White boy who lived with his parents and two brothers (including an identical twin, CAS004 below). English was the only language spoken at home. His father reported a medical history of premature birth (30.5 weeks) and delayed speech and language development. Other developmental milestones were achieved within expected time frames. His father also reported a family history of speech disorders: His father did not start fully talking until the age of 4 years, and his identical twin brother also presented with a speech disorder. CAS003's speech errors were characterized primarily by inconsistent substitution and omission errors, including voicing errors, stopping, deaffrication, gliding, and cluster reduction. Atypical errors were also noted, including substitutions of later-developing sounds for earlier-developing ones (e.g., [ʧ] for /f/) and clusters for singletons (e.g., [sn] for /n/). On the DEMSS, he exhibited moderately decreased articulatory accuracy, vowel distortions and substitutions, atypical prosody, and inconsistent productions. His maximum phonation duration was slightly below the normative range, but his monosyllable repetition rate was in the normal range. Despite producing each monosyllable repeatedly, he did not produce the trisyllable correctly, with inconsistent consonant substitutions and sequencing errors (e.g., [pakaba] for /pʌtʌkʌ/). Oral mechanism evaluation revealed no difficulty with isolated oral movements but reduced range of movement, slow rate, and limited tongue–jaw dissociation during repeated alternating movements. Vocal quality, loudness, and resonance were within normal limits. CAS003 demonstrated age-appropriate language and nonverbal cognitive skills.

CAS004. CAS004 was a 5;11 non-Hispanic right-handed White boy who lived with his parents and two brothers (including his identical twin brother, CAS003 above). His developmental and family histories were the same as described for CAS003 above. CAS004 mostly made omission and substitution errors, as well as distortions; errors included initial voicing, gliding, and cluster reduction. He also occasionally substituted later-developing sounds for earlier-developing sounds (e.g., [ʃ] for /k/). The DEMSS revealed moderately decreased articulatory accuracy, vowel distortions and substitutions, atypical prosody, and inconsistent errors. He performed below the normative range for maximum phonation duration on MaxPT, but monosyllable repetition rate was in the low-average range. He had difficulty maintaining /kʌ/ for more than seven repetitions (reverting to [tʌ] thereafter). Despite producing each of the syllables repeatedly, he did not produce the trisyllable more than twice in a row, switching to and alternating with [pʌtʌtʌ]. Inconsistent vowel distortions and substitutions were noted (e.g., [pʌtitʌ] for /pʌtʌkʌ/), as were occasional consonant additions (e.g., [pʌtiptʌ]). Oral structures were intact and symmetrical, with possible malocclusion of dentition. Some lateral movement was noted on jaw opening, but he had no difficulty with other isolated movements of the lips or tongue. However, reduced range of motion and incoordination were observed in sequenced and alternating movements. Vocal quality, loudness, and resonance were within normal limits. His receptive language skills according to CELF subtests were in the low-average range. Due to increased distractibility, the Concepts and Directions subtest was administered over multiple sessions; given this nonstandardized administration, the subtest score and the Receptive Language Index were not computed. However, for the Word Classes subtest, CAS004 obtained a scaled score of 10 (percentile = 50), and for the Sentence Structure subtest, he received a scaled score of 6 (percentile = 16). Finally, his nonverbal cognitive skills were in the normal range.

CAS005. CAS005 was a 6;0 right-handed non-Hispanic White boy who lived with his parents and older sister. English was the only language spoken at home, and there was no family history of SSD. His mother reported a history of "severe" recurrent ear infections with a slight hearing loss and myringotomy tubes at the age of 2 years, but intact hearing at the time of the study. His medical history revealed supraventricular tachycardia in utero at 33 weeks. He was diagnosed with CAS at the age of 2;9. Testing at intake revealed a severe SSD characterized primarily by omissions, distortions, and atypical substitution errors (e.g., [n] for /ʤ/, [bw] for /l/, [gl] for /f/). The DEMSS revealed moderately decreased articulatory accuracy, vowel distortions and substitutions, and inconsistent productions. Although atypical prosody was not observed on the DEMSS, he did exhibit frequent difficulty with transitioning between sounds and syllables in spontaneous connected speech. His spontaneous speech was slow and effortful. He scored below normative values on all MaxPT tasks. He produced repetitions of single syllables, although in most cases fewer than 10 syllables, and often voiced the consonants. He did not produce the trisyllable sequence correctly. His slow monosyllable repetition rate and reduced maximum phonation duration resulted in a dysarthria score of 2. On the oral mechanism exam, he inconsistently exhibited an open mouth rest posture, and frequently, his tongue rested on his lower lip. Drooling and pooling of saliva were noted intermittently during the assessment. He had no difficulty with isolated lip movements; however, incoordination and reduced range of motion were noted on repeated alternating movements (protrude–retract). Lip seal was slightly inadequate, and he did not bite his lower lip or elevate his tongue with either verbal directive or visual model. He sustained phonation on /ɑ/ but had difficulty producing /hɑ.hɑ.hɑ./. Vocal quality was slightly breathy, with reduced loudness. Hypernasality was also noted throughout the assessment. His reduced oral motor skills in the absence of structural deficits suggest that his SSD includes a contribution of dysarthria, characterized primarily by weakness. Clinical judgment suggested that the relative contribution of CAS to the overall SSD outweighed that of dysarthria. CAS005 also exhibited a receptive language disorder. He could not complete the CELF subtest Concepts & Following Directions despite repeated attempts; thus, the Receptive Language Index could not be computed. He scored in the low-average range for Word Classes (scaled score = 7, percentile = 16) and well below average for Sentence Structure (scaled score = 1, percentile = 0.1). Nonverbal cognitive skills were age appropriate.

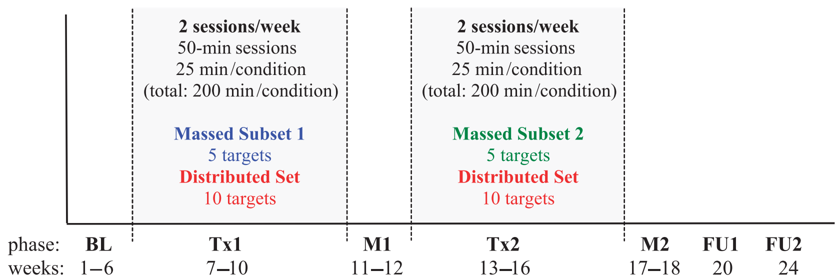
CAS008. CAS008 was a 4;7 right-handed non-Hispanic White/African American boy who was an only child living with his parents. English was the primary language spoken at home, with occasional Spanish exposure from grandmother and at school. His father reported family history of speech disorder, with the father receiving speech therapy until the age of 12 years. CAS008's speech errors consisted primarily of substitution and omission errors, with occasional distortions; errors included final consonant deletion, stopping, final devoicing, deaffrication, gliding, and cluster reduction. Atypical errors included substituting clusters for singletons (e.g., [bl] for /n/, [ts] for /s/) and later-developing sounds for earlier ones (e.g., [v] for /w/). The DEMSS revealed decreased articulatory accuracy, vowel distortions and substitutions, atypical prosody, and inconsistent productions. Final schwa additions and schwa insertions were also noted. The MaxPT protocol was discontinued on the first attempt due to difficulty and was completed in a subsequent session. He exhibited difficulty sustaining phonation and frication, most likely due to difficulty coordinating respiration with phonation onset. His fricatives were distorted and unstable, and he did not sustain /z/ for more than a few seconds, after which he sustained a vowel instead. Although maximum phonation duration was below the normative range, his monosyllabic repetition rate was in the normal range (albeit with occasional voicing alternations). He had significant difficulty with the trisyllabic sequence and did not produce the correct sequence more than twice and only at a slow rate. His errors were mainly sequencing errors, with occasional voicing alternations and vowel distortions (e.g., [pʌkʌtʌ], [tʌkʌpʌ], [pɑgɪbʌ], for /pʌtʌkʌ/). Oral structures were intact and symmetrical, with possible malocclusion (overbite). He had no difficulty with isolated oral movement tasks but missequenced repeating alternating movements. Vocal quality, loudness, and resonance were within normal limits. Mild hypernasality was occasionally noted during the coordinated speech tasks. His vocabulary was at or above average, and nonverbal cognitive skills were above average.

Design

Design Overview

To compare conditions, we used an alternating-treatments design (Kearns, 1986) with different treatment target sets for each condition, combined with a withdrawal/maintenance phase and multiple baselines across behaviors (target sets) and participants, with a minimum of three baseline sessions. Total study duration (excluding initial assessment) was approximately 24 weeks, including 3–6 weeks of initial baseline; 8 weeks (16 sessions) of treatment divided into two 4-week phases (eight sessions), each followed by a 2-week withdrawal/maintenance phase; and two follow-up points, at 4 and 8 weeks posttreatment (see Figure 1). Due to scheduling conflicts, actual study duration ranged from 23 (CAS003 and CAS004) to 27 (CAS008) weeks.

Figure 1. Overview of study design. BL = initial baseline phase; Tx1 = Treatment Phase 1; M1 = Maintenance Phase 1; Tx2 = Treatment Phase 2; M2 = Maintenance Phase 2; FU = follow-up.



Both massed and distributed practice conditions included 10 targets (words/phrases; see Target Selection). All 10 distributed practice items were trained in both treatment phases (16 sessions), whereas massed practice items were divided into two 5-item subsets, with each subset practiced in only one treatment phase (eight sessions). Treatment sessions included both conditions for an equal amount of time (25 min/condition). Because massed subsets contained half as many targets as the distributed set, and given equal time per session, targets in massed subsets received twice as much practice per phase as items in the distributed set. Therefore, comparison of Massed Subset 1 to the distributed set after Phase 1 addressed practice amount. Because each massed subset was practiced only in one phase (vs. two phases for the larger distributed set), the practice amount for the complete massed set and the distributed set was equal after Phase 2. Therefore, comparison of massed and distributed sets after Phase 2 addressed practice distribution. Thus, distribution was manipulated by varying set size with a given amount of time, rather than by varying number of sessions per unit time. To clarify, the 10 massed practice targets were practiced over 4 weeks (eight sessions of ~25 min), five in each phase, and distributed practice targets, all 10 items, were practiced over 8 weeks (16 sessions of ~25 min). By varying target set size (5 vs. 10) and keeping session duration the same, targets in massed practice received twice as much practice per session, and per phase, as distributed items. However, because distributed items were targeted in two phases (as opposed to one phase for massed items), the overall amount of practice for massed items was the same as that for distributed items. For example, if 20 teaching episodes (see below) can be completed in 25 min, then each of the five massed practice targets receives, on average, four teaching episodes per session. In contrast, the 10 distributed practice items would receive, on average, two teaching episodes per session. Thus, to achieve the same amount of practice, the distributed practice items received practice over 16 sessions (8 weeks) versus eight sessions (4 weeks) for massed practice targets. At the level of the complete set, the overall treatment period was also the same (8 weeks to achieve the same amount of practice for all items in each set).

Order of conditions within each session was pseudorandomized as follows: The child rolled a die before the first weekly session to determine which condition would be presented first in that session; the following session would have the reverse order. Thus, the order of conditions was counterbalanced by week but randomized across weeks, and each condition was presented an equal number of times in the first and second half of a session (8/16 first, 8/16 second).

Target Selection

Target words/phrases were tailored to each child based on personal relevance to enhance motivation. Possible targets were generated from a list provided by child and parents based on a questionnaire (Wilson & Gildersleeve-Neumann, 2014) and were administered several times in direct imitation prior to baseline in order to narrow and refine the list (e.g., changing names to preferred nicknames) and to gauge pretreatment accuracy and stability and rule out items that were consistently correct or improving during baseline (based on judgment by a research assistant not blinded to the time of data collection).[1]

From each child's list, three sets of 10 words/phrases were created as potential treatment targets for each child. Effort was made to balance sets for interest area (e.g., friends' names, food items), and sets were matched for length (number of syllables), complexity (Index of Phonetic Complexity; Jakielski, 2002, 2017), and accuracy on the first two baseline sessions and average baseline accuracy, as scored by a research assistant (all *p*s > .05, two-tailed

[1] This analysis was performed in an effort to gauge performance and establish stable baselines prior to initiation of treatment and did not constitute our primary outcome. The primary outcome data reported below were based on blinded analysis after collection of the final follow-up data point, so that data could be analyzed blinded to both treatment status and data collection time point (e.g., pre vs. post treatment). See Data Analysis section below. We note that one consequence of this decision to rely on blinded outcome assessors for our primary outcome is that stable baselines cannot be guaranteed prior to initiation of treatment. Establishing stable baselines before initiating treatment typically involves nonblinded analysis with respect to pre- versus posttreatment status (including in this study) and thus controls neither for assessor bias nor for perceptual drift. As a result, stability of baselines in such cases may be overestimated. The blinded analyses are more credible and more rigorous methodologically.

two-sample t tests). To eliminate bias, matching of sets was completed prior to assignment to condition.

Once sets were matched, they were randomly assigned (via a random number generator) to one of three conditions: (a) Distributed Practice, (b) Massed Practice, and (c) Control. Next, the Massed Practice set was divided into two subsets of five items, matched for length, complexity, and baseline accuracy to each other and to Distributed Practice and Control sets. After matching, the two subsets were randomly assigned to Phase 1 or 2. Item set information is given in Supplemental Material S1.

Procedure

The study took place in a quiet, child-friendly room at Temple University. A parent and a research assistant were often present in addition to the child and the SLP. All sessions were both video-recorded (using a Canon Vixia HF10 recorder) and audio-recorded (at 44.1 kHz and 16 bits using an M-Audio Aries condenser microphone connected to a Marantz CDR420 CD recorder).

Treatment Procedure

Treatment phases involved two weekly 1-hr sessions to approximate realistic logistic constraints in typical settings. Both conditions were administered in each session; sessions were divided into two halves (25 min/condition, controlled by egg timer), separated by a 10-min break. Treatment was provided by the same SLP who conducted the assessments.

Our integral stimulation–based approach was derived from, and similar to, Dynamic Temporal and Tactile Cueing (DTTC; Strand et al., 2006) but differs in target selection, less intensive session frequency, and less dynamic implementation. To facilitate replicability, experimental control, and consistency, we developed a protocol that systematically controls and varies treatment parameters (see Appendix A). For this reason, we refer to this approach as ASSIST (Apraxia of Speech Systematic Integral Stimulation Treatment). Like other integral stimulation treatments, ASSIST incorporates motor learning principles, modeling, and various cues. The SLP initially facilitated speech through immediate imitation (integral stimulation: "watch me, listen to me, say what I say") and worked toward independent productions. Articulation was shaped through multimodal (tactile, visual, auditory; simultaneous production) cueing techniques to promote accurate movement gestures. Cues were individualized and varied dynamically depending on the child's response and motivation. The SLP gradually faded these cues and varied the interval between her and the child's production. The goal was for children to produce the entire target word or phrase correctly, with correctly articulated segments and accurate and fluent prosody, in order to provide them with a functional, personally meaningful set of utterances that might be useful in their daily life.

A critical element of the protocol for present purposes is the *teaching episode*. A teaching episode refers to the sequence of events from initial elicitation of a target through the child's final attempt before switching to another target. If the child's initial attempt is incorrect, the SLP provides feedback and cues (e.g., tactile, simultaneous production at a slower rate) and up to five additional partial or complete attempts by the child before eliciting the target again with the same elicitation method as at the start of the teaching episode. If the child's initial attempt is correct, the SLP requests another production (e.g., "can you say that again?"), to reinforce the speech motor pattern and to maintain a roughly equal number of production attempts across conditions.

For each condition, targets were written on two index cards (20 for distributed practice, 10 for massed practice). Before each session, decks were shuffled to implement random practice; targets appeared on two cards to decrease predictability of upcoming targets. Given that children produced multiple attempts per teaching episode, this practice schedule is considered randomized block practice (small blocks, presented in random order). Once all cards had been presented, the SLP reshuffled the deck and continued. To implement reduced feedback frequency (Maas et al., 2012), verbal feedback on initial and final attempts in a teaching episode was faded, with feedback on all of the first 10 teaching episodes in a session, nine of 10 for the next 10, and so forth (Ballard, Maas, & Robin, 2007). Feedback included knowledge of results (accuracy, e.g., "That was good!") and knowledge of performance (nature of the response, e.g., "Your lips were not closed."). The SLP used her clinical judgment to decide the focus of feedback for each attempt, depending on the error produced, the child's response and ability to implement the change, and so forth. For instance, if a child was unable after several attempts to approximate a particular segment, the SLP would focus on prosodic errors or fluency (e.g., "Let's focus on keeping the sounds together this time") before returning to the difficult segment, in order to minimize frustration and maintain motivation and attention. Thus, while ASSIST is less dynamic than DTTC, it still incorporates allowances for clinical judgment in treatment delivery for reasons of ecological validity.

To systematically work toward independence in speech production, four different elicitation methods were used, in order from more to less support (less to more independent): immediate imitation at a slowed rate, immediate imitation at a normal rate, delayed imitation at a normal rate, and independent (e.g., in response to a question). Each session half started with immediate imitation at a normal rate, and the elicitation method changed depending on the child's performance during the session as judged by the treating SLP. Elicitation method changed after 2/2 consecutive correct or incorrect first attempts of a teaching episode (see Appendix A).

Probe Procedure

The primary data to assess treatment effects were productions elicited on a probe task that included all 30 items (distributed, massed, control). Items were presented in a different random order each time by the treating SLP using either delayed reading (CAS001, CAS002) or delayed imitation (CAS003–CAS008). Delayed production tasks were used to reduce reliance on the SLP's auditory and visual model. For delayed reading, children were shown an index card with the written item, read it silently, and produced the item only after the SLP removed the card (to minimize effects of reading out loud on prosody). No feedback on accuracy or performance was provided. Probes were administered weekly, except during follow-up (2-week interval and 4-week interval). During treatment phases, probes were administered at the beginning of a session.

Treatment Fidelity Procedure

To ensure treatment fidelity (Kaderavek & Justice, 2010), the treating SLP followed an operationalized, manualized protocol (see Appendix A). An independent analyst also assessed fidelity for a minimum of two randomly selected sessions from video recordings; this analyst was not involved in treatment or probe data analysis. Rated fidelity aspects included the numbers and proportions of teaching episodes with reduced feedback frequency (< 90%), feedback delay (1–3 s), feedback on both initial and final attempts, same elicitation method for both initial and final attempts, and the number and proportion of teaching episodes with an intended initial elicitation attempt. It became clear during the fidelity analysis that judgments of speaking rate (slow vs. normal rate) were not reliable, and therefore we combined these elicitation methods for fidelity purposes. We also report the number of teaching episodes, the number of whole-target production attempts per teaching episode, and the amount of time per condition (see Appendix B).

Data Analysis

The dependent measure was perceptual accuracy of words/phrases on probes. Accuracy was judged from audio recordings by independent raters blinded to condition (massed, distributed, control) and time point (e.g., baseline, follow-up). To enable blinding to time point and control for perceptual drift, recordings were presented in random order (and thus analysis could only begin after collection of the final follow-up probe for each child). Clues about time point or treatment status were removed from files by a research assistant not involved in the analysis.

Analysts judged accuracy on a syllable-by-syllable basis by judging whether each syllable contained error(s) or not. Errors included substitutions, omissions, distortions, additions, metatheses, and unintelligible syllables.[2] Accuracy was judged against the SLP's model; casual speech and dialect variations were not considered errors (even if different from the SLP's model). For example, if the SLP produced the item *How are you today?* (CAS003) with a full vowel in *today* ([tuˈdeɪ]) but the child produced a reduced vowel ([təˈdeɪ]), this was considered correct because vowel reduction is an acceptable variation in this context. Based on these syllable-by-syllable accuracy judgments, each item was awarded a score of 0, 1, or 2 to represent major errors (0), minor errors (1), and correct responses (2; similar to Maas et al., 2012; Strand et al., 2006), in order to account for differences in length and complexity of items. A score of 0 (major error) indicated that fewer than 50% of the syllables were correct, a score of 1 (minor error) indicated that 50% or more (but not 100%) of the syllables were correct, and a score of 2 (correct) indicated that all syllables were error free.[3] For each probe, item scores were averaged by set and divided by 2 to provide the condition score in percent accuracy. Average interrater reliability based on 19 randomly selected probes (16% of probes) indicated exact agreement on syllable-by-syllable judgments for 79% of the syllables scored by both raters and exact agreements for 68% of item scores (with 96% within 1 scale point); intrarater reliability based on one randomly selected probe indicated exact agreement on syllable-by-syllable judgments for 91% of the syllables double-scored and exact agreement for 83% of item scores (100% within 1 scale point).

Analysis involved visual inspection of graphs, supplemented with standardized effect sizes (*d* statistic; Beeson & Robey, 2006; Busk & Serlin, 1992) to compare and quantify changes relative to baseline in each condition for each child. Effect size *d* was computed as [(mean $\text{score}_{\text{post}}$ − mean $\text{score}_{\text{pre}}$) / SD_{pre}]. We operationally define $d > 1$ as improvement (pre–post gains exceed baseline standard deviation; Maas et al., 2012; McAllister Byun, Hitchcock, & Swartz, 2014; McAllister Byun, Swartz, Halpin, Szeredi, & Maas, 2016). Note that a larger standard deviation during baseline (more variable, less stable baseline) has the effect of reducing the ability to detect an effect, as it increases the denominator. Thus, increased variability during baseline sets a higher bar for obtaining an effect. For completeness, we also report unstandardized gains (percent change). We expected larger effect sizes for treated than untreated control items after each treatment phase. In addition to individual analyses, we also compared conditions across the six children using nonparametric Cochran–Mantel–Haenszel chi-square (χ^2) tests to detect overall condition differences and nonparametric Wilcoxon signed-ranks tests to identify significant pairwise differences ($\alpha = .05$).

To address the role of practice amount, we compared effect sizes for Massed Practice Subset 1 to the distributed practice set during Maintenance Phase 1, averaged across both probes in this phase. Hypothesis 1 (more-is-better hypothesis) predicts a larger effect size for Massed Practice Subset 1 than for the distributed practice set, because after Phase 1, massed practice subset items had received twice as much practice as distributed practice items.

To address the role of practice distribution, we compared effect sizes for massed practice items (Subsets 1+2 combined) and the distributed practice set in Maintenance Phase 2 and follow-up. Hypothesis 2a (motor learning hypothesis) predicts larger effect sizes for the distributed than

[2]Initially, prosody was also included in the scores, but these judgments were unreliable and are not reported here.

[3]Although this scoring system reduces sensitivity for monosyllabic targets (which can receive only segmental scores of 0 or 2), this is unlikely to impact the findings because there were very few monosyllabic targets for any child.

the massed practice set; Hypothesis 2b (neuroplasticity hypothesis) predicts the reverse.

Results

Individual Analyses

Summary data for all children and group means are presented in Tables 2 and 3. Data for children CAS001–CAS003 are plotted in Figure 2; and data for CAS004–CAS008, in Figure 3.

CAS001. CAS001 demonstrated somewhat variable but nonrising baselines prior to Treatment Phase 1. On initiation of treatment, Massed Subset 1 showed an increase in accuracy. Distributed items showed a more gradual improvement toward the end of the phase, but this improvement was comparable to the Control set, and performance for both these sets declined when treatment was withdrawn. As-yet-untreated Massed Subset 2 did not change until Treatment Phase 2, although a slight increase was noted on the probe immediately prior to treatment. The Distributed set again showed a modest increase during Phase 2.

Effect sizes for Phase 1 indicated treatment effects for both treated conditions, with an advantage for the low-amount set (d = 2.31, 20.0% gain) compared to the high-amount set (d = 1.09, 16.7% gain). Control items did not reveal a significant effect (d = 0.57, 10.0% gain).

At Maintenance Phase 2, CAS001 showed a minimal advantage for massed over distributed practice in terms of effect size but not in terms of percent gain (massed: d = 2.74, 15.8% gain; distributed: d = 2.31, 20.0% gain). At follow-up, both effect size and percent gain revealed an advantage for distributed practice (d = 3.46, 30.0% gain) compared to massed practice (d = 2.74, 15.8% gain). Untreated items did not improve either at maintenance (d = 0.24, 4.2% gain) or at follow-up (d = 0.52, 9.2% gain).

CAS002. CAS002 showed stable or slightly declining baselines prior to Phase 1 treatment. High-amount items (Massed 1) improved during Treatment Phase 1. During Maintenance Phase 1, all sets declined except for the high-amount set. During Treatment Phase 2, the distributed set showed improvement. At follow-up, accuracy of the massed set was slightly higher than accuracy for distributed items.

Table 2. Effect sizes (*d*) and percent change (%change) by condition for amount (Treatment Phase 1).

	High amount		Low amount		Control	
Child	*d*	%change	*d*	%change	*d*	%change
001	1.09	16.7	2.31	20.0	0.57	10.0
002	1.83	17.5	–1.79	–11.3	–2.98	–18.8
003	2.35	22.5	1.13	5.4	1.12	9.6
004	2.83	14.2	–0.41	–1.7	–0.60	–3.8
005	3.18	28.6	2.47	13.2	1.39	15.0
008	0.21	1.8	0.89	9.4	1.38	9.0
M	1.45	13.5	0.69	6.6	0.26	5.3
SD	1.22	9.9	1.45	11.4	1.77	14.1

Table 3. Effect sizes (*d*) and percent change (%change) by condition for distribution.

		Massed		Distributed		Control	
Period	Child	*d*	%change	*d*	%change	*d*	%change
M2	001	2.74	15.8	2.31	20.0	0.24	4.2
	002	2.53	11.3	0.60	3.8	0.60	3.7
	003	2.58	16.3	1.83	8.8	0.73	6.3
	004	3.50	17.5	1.84	7.5	0.20	1.3
	005	2.32	18.6	1.07	5.7	–0.23	–2.5
	008	0.48	4.8	2.06	21.9	0.61	4.0
	M	2.36	14.1	1.62	11.3	0.36	2.8
	SD	1.01	5.2	0.65	7.7	0.36	3.1
FU	001	2.74	15.8	3.46	30.0	0.52	9.2
	002	2.87	13.8	0.99	6.3	–0.20	–1.3
	003	2.19	13.8	2.35	11.3	1.02	8.8
	004	3.50	17.5	0.61	2.5	0.20	1.3
	005	2.63	21.1	2.00	10.7	0.69	7.5
	008	1.47	14.8	0.89	9.4	1.00	6.5
	M	2.57	16.1	1.72	11.7	0.54	5.3
	SD	0.68	2.8	1.09	9.5	0.47	4.3

Note. M2 = Maintenance Phase 2 (Probes M2-1 and M2-2); FU = follow-up period (4- and 8-week follow-up combined).

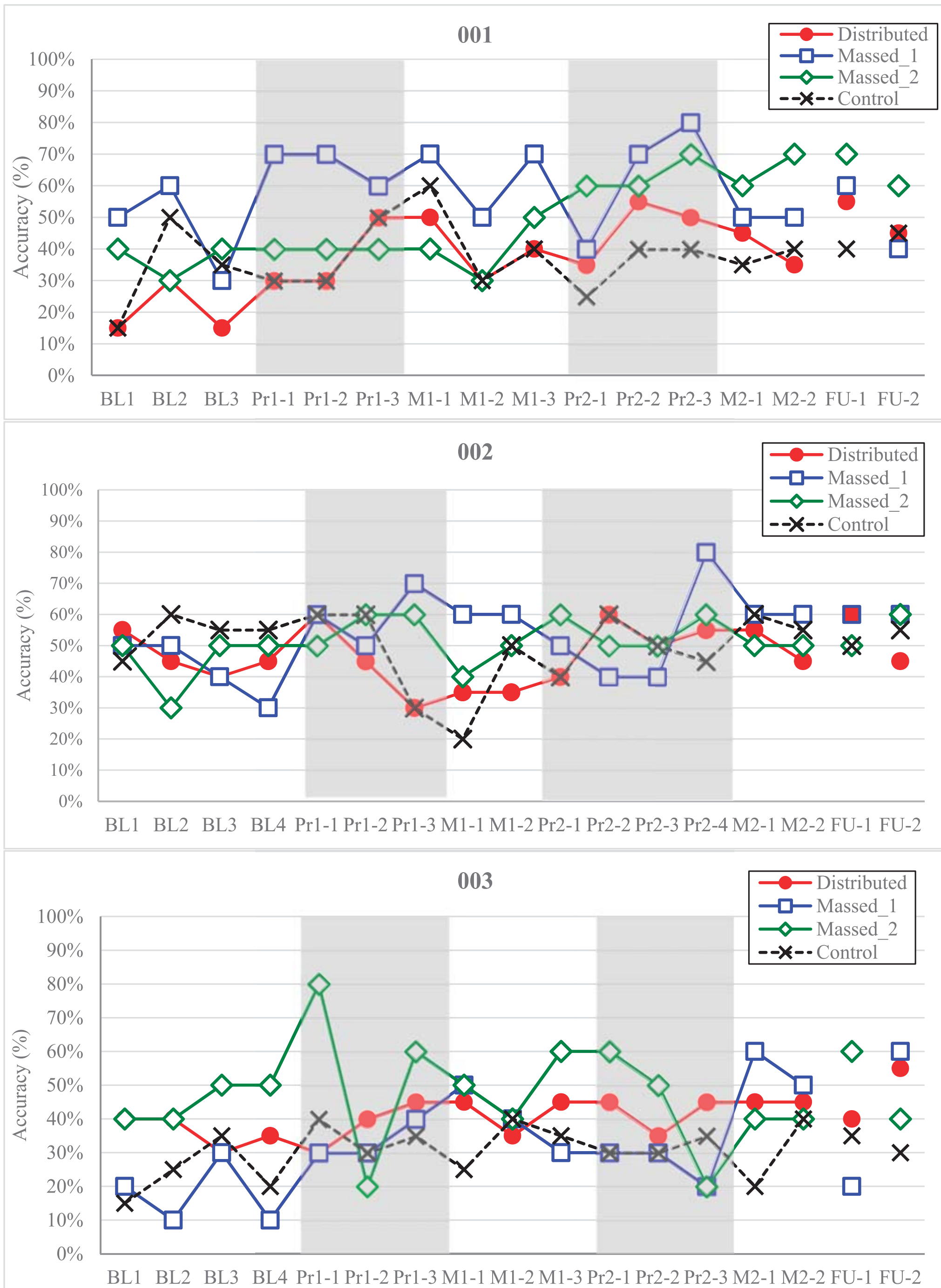
Both effect sizes and percent gains after Phase 1 indicated a benefit for high-amount items (d = 1.83, 17.5% gain) compared to the low-amount items and the Control items, both of which showed a decline (low amount: d = −1.79, 11.3% loss; Control: d = −2.98, 18.8% loss). Thus, only the high-amount set showed a gain.

In the immediate 2-week posttreatment phase, CAS002 showed a clear advantage for Massed items (d = 2.53, 11.3% gain), with no effects for Distributed (d = 0.60, 3.8% gain) or Control (d = 0.60, 3.7% gain) items. This advantage was maintained during the follow-up phase, at which point the Distributed items approached an effect (d = 0.99, 6.3% gain) while the Control items declined (d = −0.20, 1.3% loss). Thus, only Massed items showed a robust treatment effect.

CAS003. CAS003 showed largely stable baselines before treatment, with a slight increase for untreated items. At the end of Treatment Phase 1, both treated sets showed improvement. During Treatment Phase 2, a slight decline was noted for the massed sets. After both treatment phases, performance was generally maintained at levels comparable to those after Treatment Phase 1 and above baseline for Massed Subset 1 and Distributed items. After Phase 1, CAS003 showed a larger effect size for high-amount items (d = 2.35, 22.5% gain) than for low-amount (d = 1.13, 5.4% gain) and Control (d = 1.12, 9.6% gain) items.

During Maintenance Phase 2, CAS003 showed an advantage for Massed (d = 2.58, 16.3% gain) over Distributed (d = 1.83, 8.8% gain) items; both outperformed Control items (d = 0.73, 6.3% gain). During follow-up, a small advantage for Distributed over Massed items emerged for effect size (d = 2.35 vs. d = 2.19) but not for percent gain (13.8% vs. 11.3%). Although control items also improved at

Figure 2. Data plots for Children 001, 002, and 003. Shading indicates treatment phases. Note that, due to scheduling constraints, the number of probes during Maintenance Phase 1 varied between children.

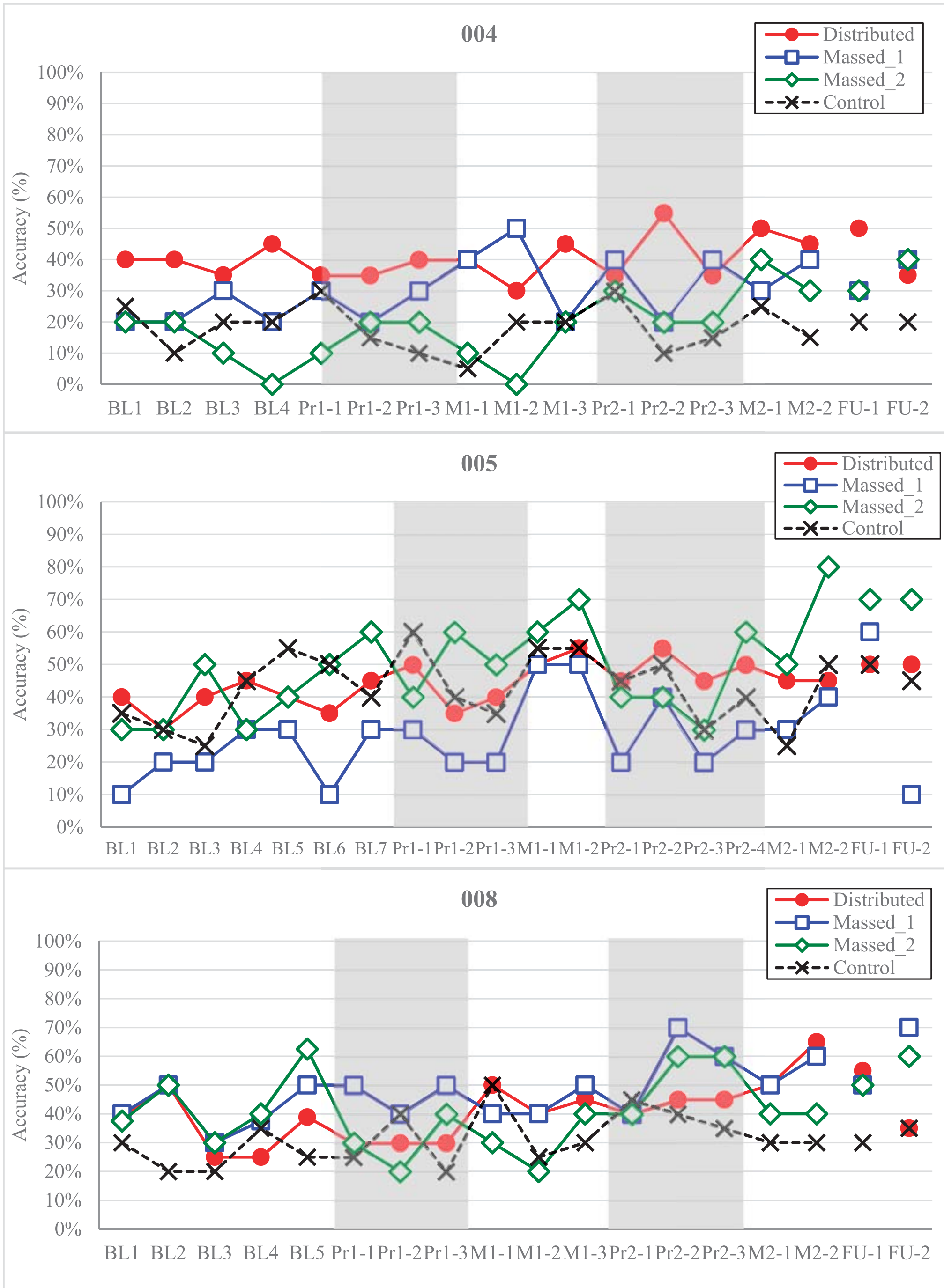


follow-up, gains were smaller than for treated items (d = 1.02, 8.8% gain).

CAS004. CAS004 demonstrated stable baselines before initiation of treatment. Following Treatment Phase 1, Massed 1 items showed a small improvement that dissipated prior to Treatment Phase 2. After Phase 2, Distributed items and Massed 2 items showed a small increase.

After Phase 1, CAS004 improved only on high-amount items (d = 2.83, 14.2% gain), with slight declines for low-amount (d = −0.41, −1.7%) and Control (d = −0.60, −3.8%) items.

Figure 3. Data plots for Children 004, 005, and 008. Shading indicates treatment phases. Note that, due to scheduling constraints, the number of probes during Maintenance Phase 1 varied between children.



CAS004 showed a robust advantage for Massed Practice items compared to Distributed Practice items, both at Maintenance 2 (Distributed: d = 3.50, 17.5% gain; Massed: d = 1.84, 7.5% gain) and during follow-up (Distributed: d = 3.50, 17.5% gain; Massed: d = 0.61, 2.5% gain). Untreated items did not show improvement at either point (d = 0.20, 1.3% gain at both points).

CAS005. CAS005 demonstrated a rise around the third baseline probe for Control items and Massed Subset 1 items, which subsequently stabilized before treatment onset.

Massed Subset 2 did show a rising baseline during this initial baseline, which largely stabilized prior to its being treated. No clear response to treatment was evident during Treatment Phase 1, but treated items did show higher accuracy immediately following Phase 1. Distributed items remained stable through Treatment Phase 2, and Massed Subset 2 showed an increase toward the end of Treatment Phase 2, whereas Massed 1 and Control items declined during Phase 2 treatment. Effect sizes for practice amount revealed greater gains for high-amount items (d = 3.18, 28.6% gain) than for low-amount (d = 2.47, 13.2%) or Control (d = 1.39, 15.0% gain) items.

CAS005 showed a consistent advantage for Massed Practice compared to Distributed Practice both in the immediate posttreatment phase (Massed: d = 2.32, 18.6% gain; Distributed: d = 1.07, 5.7% gain) and during the follow-up phase (Massed: d = 2.63, 21.1% gain; Distributed: d = 2.00, 10.7% gain). Both treated conditions also consistently outperformed untreated Control items (Maintenance 2: d = −0.23, 2.5% loss; follow-up: d = 0.69, 7.5% gain).

CAS008. CAS008 demonstrated somewhat variable baselines, with indications of rising trends immediately prior to treatment for both massed subsets; Massed Subset 2 stabilized prior to Phase 2 treatment. Following Treatment Phase 1, improvement was noted only for the Distributed items and, to some extent, for the Control items (which returned to baseline levels after the first Maintenance 1 probe). In Treatment Phase 2, both massed subsets improved while Distributed items remained stable. In the Maintenance 2 phase, Distributed items showed some additional gains, which largely dissipated during follow-up, unlike massed items that retained gains.

Effect sizes for practice amount indicated no treatment effects; unexpectedly, only Control items showed an effect (d = 1.38, 9.0% gain). Although percent gain for low-amount items was of similar magnitude (9.4%), greater baseline variability resulted in a nonsignificant effect size (d = 0.89); high-amount items showed no improvement (d = 0.21, 1.8% gain).

Regarding distribution, CAS008 showed inconsistent effects: At Maintenance 2, he showed an advantage for Distributed (d = 2.06, 21.9% gain) over Massed (d = 0.48, 4.8% gain) Practice, whereas at follow-up, this pattern was reversed (Distributed: d = 0.89, 9.4% gain; Massed: d = 1.47, 14.8% gain). Control items did not show clear improvement (Maintenance 2: d = 0.61, 4.0% gain; follow-up: d = 1.00, 6.5% gain).

Group Analyses

Group means by condition are provided in Tables 2 and 3. At Maintenance Phase 1, there were no significant effects for practice amount for either effect sizes or percent gains, although there was a trend (p < .10) in the comparison between high-amount and Control sets (S = 8.5, p = .094, for both d and percent gain; all other ps > .10; see Appendix C, Table C1).

For practice distribution, Cochran–Mantel–Haenszel tests indicated significant condition differences for both effect size and percent change at both M2 and FU (see Appendix C, Table C2). Pairwise comparisons revealed no significant group differences between Massed and Distributed conditions (ps > .20). At M2, both treated sets showed a significant advantage compared to control items for percent change (S = 10.5, p = .031 for both) and a trend for effect size (Distributed vs. Control: S = 7.5, p = .063; Massed vs. Control: S = 9.5, p = .063). At follow-up, both Massed and Distributed sets showed significantly greater gains for percent change (S = 10.5, p = .031 for both), but only Massed items differed significantly from the Control items in terms of effect size (S = 10.5, p = .031; Distributed: S = 9.5, p = .063).

Discussion

This study was designed to examine two critical aspects of treatment intensity, amount and distribution, in treatment for CAS. This study also examined the overall impact of integral stimulation–based treatment for CAS. Below, we discuss findings relative to these goals.

Practice Amount

At the individual level, four of the six children showed an advantage of more practice (CAS002, CAS003, CAS004, and CAS005). One child (CAS001) showed an unexpected advantage for the low-amount condition, and one child (CAS008) did not show effects for either treated set but did improve on control items. These unexpected effects most likely reflect item-specific effects. Despite our best efforts at matching item sets for difficulty and personal interest, random assignment to conditions may have led to low amount (CAS001) and control sets (CAS008) containing more items that were more motivating or used more frequently outside treatment.

Averaged across all six children, there was a numerical advantage for more practice compared to less practice after Treatment Phase 1, in terms of both effect sizes and percent gain (see Table 2). Item sets receiving a low amount of practice (the Distributed set) did not differ notably from untreated sets, and neither of these sets showed an effect on average (average d < 1), unlike high-amount items (average d = 1.45). The numerical advantage for high amount of practice was not significant in the nonparametric group analyses, however, possibly due to the small sample.

Taken together, the individual data and the group data are suggestive of an advantage for more practice, with four of six children showing greater gains and effect sizes than items receiving less practice and an average effect size across children who exceeded baseline standard deviations. These findings are consistent with both the motor learning literature (e.g., Shea et al., 1990) and the CAS treatment literature (Edeal & Gildersleeve-Neumann, 2011; Namasivayam et al., 2015) and add support to the notion that children with CAS benefit from more practice. However, this interpretation must be taken with caution given that the differences were generally modest and the group analysis failed to detect significant condition differences. One factor that may have contributed to these modest effects is that the overall amount of practice during Phase 1 was small, with only eight 25-min sessions per condition. This amount of practice may not have been sufficient for condition differences to emerge. Some support for this view comes from the observations that more robust treatment effects were evident after the second treatment phase and that only two children showed positive effects and a clear advantage over the control items for the low-amount condition (CAS001 and CAS005), compared to five children for the high-amount condition (CAS001, CAS002, CAS003, CAS004, and CAS005). Thus, it appears that, despite the relatively few and short sessions, reducing the set size (thus increasing the number of practice opportunities per item) resulted in speech motor learning. Clearly, further research is needed with larger sample sizes and greater differences between conditions. Nevertheless, our findings suggest that more practice is likely to result in greater gains, consistent with Hypothesis 1 and with previous literature (Edeal & Gildersleeve-Neumann, 2011; Namasivayam et al., 2015).

Practice Distribution

For practice distribution, five of the six children showed a numerical benefit in terms of effect size for massed practice over distributed practice during the 2-week maintenance period immediately following Treatment Phase 2, whereas only one (CAS008) showed an advantage for distributed practice (see Table 3). For percent change, the pattern was the same except that CAS001 showed a slight advantage for distributed rather than massed practice. At follow-up, four of the six children showed an advantage for massed practice compared to distributed practice for effect size (CAS002, CAS004, CAS005, and CAS008) and two showed the opposite (CAS003 and CAS008); Participant CAS003 did show a numerical massed practice advantage for percent change. Such ambiguous effects (inconsistent between *d* and percent change) are difficult to interpret and always involved small differences (*d* differences < 0.5, %change differences < 5%). To our knowledge, there are no established criteria for determining condition differences in alternating-treatments designs; hence, we rely primarily on numerical differences in standardized effect sizes (our primary outcome measure). However, to exercise reasonable caution in interpretation, we only consider conditions to differ if both effect size and percent change are in the same direction and consider ambiguous effects to be ties. With this more conservative criterion, four of the six children showed a massed practice advantage at maintenance; and four at follow-up, with one reverse pattern and one tie at each period.

Of the two children with a distributed practice advantage that was consistent between effect size and percent change (CAS008 at M2 and CAS001 at FU), one (CAS008) showed a reverse pattern at maintenance. For him, the data indicate a steady decline for distributed practice items over the follow-up period and a steady increase for massed practice items (see Figure 3). While his data must be interpreted with caution given that he also showed improvements for untreated items in Treatment Phase 1, this pattern across the follow-up phase does suggest that massed practice may have resulted in more robust speech motor learning (longer term retention) than distributed practice. Unfortunately, we do not have longer term follow-up data to determine whether this trend would have continued. In contrast, for CAS001, there appears to be a more robust advantage for distributed practice: He showed a tie during maintenance (distributed practice advantage for percent change), and massed practice items declined during the follow-up period, while distributed practice items improved. It is interesting to note that CAS001 was considerably older than the other children, because as noted in the introduction, most principles of motor learning are based on studies with adults. Thus, the observation that the oldest child was the only one to show a relatively clear advantage for distributed practice (consistent with the motor learning literature) raises the intriguing possibility that effects of practice conditions may vary with age, such that massed practice may be optimal for younger children and benefits of distributed practice emerge for older children. At this point, given these limited data, this possibility remains necessarily speculative, but it deserves attention in future research.

When considered as a group, there was a numerical advantage for massed practice compared to distributed practice in terms of both effect size and percent change, at both maintenance and follow-up. Although the two treatment conditions did not differ significantly, likely due to the small sample size, only the massed practice condition differed significantly from the control condition for effect size at maintenance. In addition, massed practice showed an effect ($d > 1$) for five of six children at maintenance and all six children at follow-up, compared to five of six and three of six for distributed practice (and zero of six and one of six for control items), respectively.

On the whole, the weight of the evidence of our findings suggests a benefit for massed over distributed practice for children with CAS. These findings are consistent with the neuroplasticity literature (Hypothesis 2b), in which massed practice is considered an important learning principle (Kleim & Jones, 2008), as well as with the CAS treatment literature (Namasivayam et al., 2015; Thomas et al., 2014). As in Thomas et al., we showed that the massed practice advantage became more evident at longer retention intervals (with the exception of CAS001). Thus, our findings provide further support from a well-controlled within-study comparison for the use of intensive, massed practice in CAS treatment. It should be reiterated that the manner in which practice distribution was manipulated here differs from that in previous studies, which varied the period and/or practice amount (Namasivayam et al., 2015; Thomas et al., 2014). Nevertheless, the findings provide converging evidence from divergent methods to suggest that massed practice likely leads to greater, more robust improvement for most, though not all, children with CAS. Although speculative, according to the neuroplasticity literature (which is largely based on animal models of acquired brain damage), massed practice results in strengthened synaptic connections between neurons in pertinent neural circuits (Kleim & Jones, 2008). By extension, it is conceivable that, in children with CAS, massed practice strengthens synaptic connections in neural circuits underlying speech motor planning, for example, connections between auditory target and corresponding motor command (Terband, Maassen, Guenther, & Brumberg, 2009) or between somatosensory input and motor command (Terband & Maassen, 2010; Terband, Maassen, Guenther, & Brumberg, 2014). Future studies using neuroimaging and computational modeling simulations may be able to shed further light on these possibilities.

ASSIST and Integral Stimulation Treatment Efficacy

All children demonstrated treatment effects for at least one condition after completing both phases of treatment, with larger gains and effect sizes for treated items than for untreated control items, for which effect sizes in most cases did not exceed 1. In addition to demonstrating treatment effects for all children individually, group averages and statistical analysis of group data confirmed that treated items showed greater gains than untreated items. Thus, this study also adds another replication to the literature supporting the efficacy of integral stimulation treatment for CAS. Despite differences in participants (e.g., DTTC was primarily developed for, and studied in, younger children with more severe CAS) and differences in implementation (discussed below), our protocol included the core aspects of integral stimulation treatment (e.g., integral stimulation, tactile cues, slowed rate, gradual fading of cues, focus on whole-target movement accuracy). As such, the fact that we observed treatment effects strengthens the evidence base for integral stimulation–based treatments, by demonstrating treatment effects for this combination of shared components despite variations in other treatment parameters. Such convergence illustrates that the effects in prior studies were not specific to those participants or clinicians, or specific target selection criteria. Rather, the growing evidence base suggests that it is the integral stimulation–based elements that are responsible for the emergence of treatment effects.

Nevertheless, despite this replicated evidence, the degree of improvement varies across studies. In this study, gains were modest compared to some previous studies using integral stimulation–based approaches (Strand & Debertine, 2000; Strand et al., 2006). While the presence of treatment effects across studies supports the efficacy of integral stimulation–based treatment regardless of variation in treatment parameters, the magnitude of treatment effects is likely influenced by variations in treatment protocol and by methodological differences. Below, we briefly discuss some of these differences to inform design of future treatment studies.

One of the most important differences in approach relates to target selection. In this study, targets were selected primarily based on personal functional relevance. Previous studies have shown gains for treated items but little to no generalization to untrained items (Maas et al., 2012; Maas & Farinella, 2012; Strand et al., 2006). Following recommendations from our earlier work (Maas et al., 2012; Maas & Farinella, 2012), we chose personally meaningful targets to enhance motivation and potential for carryover. This choice was also consistent with the neuroplasticity principle of salience (Kleim & Jones, 2008), which recommends practice on relevant, meaningful behaviors. However, the relative complexity of our targets likely resulted in smaller gains compared to studies in which targets were functionally relevant but less complex (e.g., Strand et al., 2006). Selecting targets that are personally relevant and build on movement gestures and syllable shapes already mastered may lead to greater gains on those items and may enhance motivation due to increased success during treatment (Jakielski, 2017). However, a drawback of this strategy is that it often excludes targets with high personal relevance for a child (e.g., their name includes a sound not yet mastered). Anecdotal clinical observations also suggest that high motivation may help a child overcome the difficulty of a complex target. Nevertheless, selection of challenging targets in this study likely contributed to the relatively modest overall effect sizes compared to studies using DTTC. As discussed elsewhere (Maas et al., 2014), target selection is a critical element of any treatment approach, and to our knowledge, there has not been a systematic comparison of different methods of target selection in treatment for CAS.

Another important difference in approach with previous studies relates to overall amount and distribution of practice. For example, studies using DTTC typically involved 1 hr of treatment per day, 5 days per week, for 6 weeks. Instead, in our work here and elsewhere, treatment was typically administered in two or three weekly 1-hr sessions over 5–8 weeks. Thus, studies using DTTC typically involved more, and more massed, treatment. If massed practice is indeed more beneficial for children with CAS, then our more distributed treatment here (in terms of sessions over weeks) may have led to more modest effect sizes.

Finally, ASSIST is less dynamic than DTTC, because in order to adequately study the intervention and enable replication, we opted to create a more structured treatment protocol, which necessarily limited some of the dynamic elements of the treatment. For instance, the current protocol did not include a prepractice phase, and the choice of elicitation method was prespecified (though still based on the child's performance as judged by the SLP). Similarly, each teaching episode began and ended with an elicitation of the target, practice was always random, and a verbal feedback schedule was prespecified. Thus, while ASSIST retained some dynamic elements of treatment (e.g., type of cues and support provided in a teaching episode, choice of elicitation method based on a child's performance), ASSIST did have fewer dynamic aspects than DTTC. Greater gains might have emerged with a less structured protocol.

Several methodological differences may also have contributed to the more modest gains in this study. One such difference is that we randomly assigned targets to conditions, after careful matching of target sets. This strategy was important to establish experimental control and avoided any potential bias (conscious or not) in selecting the easiest or most motivating targets, which may have contributed to larger gains in previous studies.

Another difference is that we administered probes at the beginning rather than at the end of a session (as in Edeal & Gildersleeve-Neumann, 2011; Strand & Debertine, 2000). Moreover, probes involved delayed production tasks (as in Edeal & Gildersleeve-Neumann, 2011), rather than immediate imitation (as in Maas et al., 2012; Strand et al., 2006). Not all children reliably achieved delayed elicitation methods during treatment; thus, the probe task represented a more challenging level (a higher bar) than that experienced during treatment, which may have depressed performance on probes but reflects a higher (desired) level of skill improvement.

Finally, our data were analyzed by blinded analysts, rather than by the treating clinician (e.g., Edeal & Gildersleeve-Neumann, 2011; Strand et al., 2006). This blinding eliminated bias in analysis and controlled for potential perceptual drift across the treatment period. Some portion of the gains in previous studies may have been attributable to potential bias or perceptual drift; by controlling those factors, here, we may have a more accurate estimate of the true (more modest) change attributable to the child's improvements in speech motor skill.

Clinical Implications

The findings have several clinical implications for the treatment for CAS. These implications must be considered preliminary given the caveats noted throughout the discussion above. First, the replication of the efficacy of integral stimulation–based intervention for six new children with CAS supports the clinical use of such interventions. Although the evidence base remains small, integral stimulation approaches have been supported by seven peer-reviewed studies to date, with a total of 17 children (Baas, Strand, Elmer, & Barbaresi, 2008; Edeal & Gildersleeve-Neumann, 2011; Gildersleeve-Neumann & Goldstein, 2015; Maas et al., 2012; Maas & Farinella, 2012; Strand & Debertine, 2000; Strand et al., 2006). This study expands this to 23 children.

Second, children with CAS show greater speech motor learning with more practice. Treatment Phase 1 findings suggest that eight sessions of 25 min may not result in robust gains when targeting 10 items, at least when using ASSIST. Whereas five of six children demonstrated larger and positive effects for the five-item set than for untreated items, only two of six showed such effects for the 10-item set. Thus, when the limiting factor is number and duration of sessions, clinicians can increase practice amount and expect gains by decreasing the number of targets. In a 25-min session, children completed on average 25 teaching episodes and 71 whole-word production attempts, regardless of condition (see Appendix B). For a set of five items, this means approximately five teaching episodes and about 14 whole-word production attempts per item, compared to 2.5 teaching episodes and seven whole-word attempts per item for a set of 10 items.

Third, most children with CAS may show greater, more robust improvements for items practiced in a massed regimen. In this study, distribution of a fixed amount of practice was manipulated by varying set size and practicing items either in two separate "bursts" of 4 weeks or over a period of 8 weeks. Thus, for a child with a larger set of target items, our findings suggest that selecting subsets for more intensive practice may lead to greater gains over a given period (here, the 10 massed items generally outperformed the 10 distributed items after 8 weeks). To the extent that we observed a massed practice advantage, our findings also provide some support for intensive service delivery models (e.g., summer camps) in which distribution is defined in terms of number of sessions over time (e.g., Preston et al., 2016). However, direct study is needed to determine the benefits of such intervention models relative to more distributed service delivery models (e.g., same number of sessions throughout a semester or school year).

Future Directions

Future studies should replicate and extend this work, and investigate other aspects of intensity in treatment for CAS. Given the interindividual variability in treatment response and pattern, future research should also examine participant factors (e.g., age, severity) and target selection factors (e.g., complexity) that may impact treatment factors, in order to move toward an evidence-based approach to providing optimal and personalized care for children with CAS.

To date, the support for efficacy of integral stimulation treatment has been based exclusively on speech accuracy measures. However, the success and clinical relevance of a treatment must not solely be based on impairment-level measures but also on measures of activity and participation. Very few studies have used activity- and participation-level measures in treatment for motor-based pediatric speech disorders, including CAS (Kearney et al., 2015). Future studies should examine other, more functional outcome measures, such as parent ratings of intelligibility and participation, as well as direct intelligibility measures obtained from unfamiliar listeners (as in Namasivayam et al., 2015). A study examining such measures for the children in this study is currently in progress and will be reported in a separate report.

Conclusions

In conclusion, this study showed greater gains with more practice for most children with CAS. Not all children showed this benefit, possibly due to the limited amount of overall practice and some unanticipated item-specific effects. In addition, five of the six children revealed a massed practice advantage, contrary to the motor learning literature but consistent with the neuroplasticity literature. The massed practice benefit emerged especially at longer term retention intervals. The study also provided another well-controlled replication of the efficacy of integral stimulation–based interventions (here, ASSIST). Overall gains were modest, however, and several important considerations were discussed to help optimize treatment and research designs for future studies. Future studies should also include more functional outcome measures.

Acknowledgments

This work was supported by a generous grant from the Childhood Apraxia of Speech Association of North America (currently known as Apraxia Kids; principal investigator: Maas). The content is solely the responsibility of the authors and does not necessarily represent the official views of Apraxia Kids. We thank Apraxia Kids (formerly known as Childhood Apraxia of Speech Association of North America) for funding. We also thank Carolina Echeverri, Talia Irgangladen, Leanne Long, Brian Kulsik, and Sarah Rosenberg for their help with data and fidelity analyses; the SLP for evaluating the children and administering the treatment; and Kyra Skoog for computing complexity scores for target stimuli. Most importantly, we thank the children and their families for their time and participation. Portions of this research were conducted as part of two master's theses (by Nicolette Kovacs and Mackenzie Welsh). Preliminary data were presented at the American Speech-Language-Hearing Association Convention (Los Angeles, CA, November 2017).

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. (2007). *Childhood apraxia of speech* [Tech. Rep.]. Retrieved from http://www.asha.org/policy

Askim, T., Mørkved, S., Engen, A., Roos, K., Aas, T., & Indredavik, B. (2010). Effects of a community-based intensive motor training program combined with early supported discharge after treatment in a comprehensive stroke unit. A randomized, controlled trial. *Stroke, 41,* 1697–1703.

Baas, B. S., Strand, E. A., Elmer, L. M., & Barbaresi, W. J. (2008). Treatment of severe childhood apraxia of speech in a 12-year-old male with CHARGE association. *Journal of Medical Speech-Language Pathology, 16,* 181–191.

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 and Language, 52,* 566–577.

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

Ballard, K. J., Robin, D. A., McCabe, P., & McDonald, J. (2010). A treatment for dysprosody in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 53,* 1227–1245.

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

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior, 9,* 567–572.

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature, 382*(6588), 252–255.

Busk, P. L., & Serlin, R. (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, NJ: Lawrence Erlbaum Associates.

Campbell, T. F. (1999). In A. J. Caruso & E. A. Strand (Eds.), Functional treatment outcomes in young children with motor speech disorders, *Clinical management of motor speech disorders in children* (pp. 385–396). New York, NY: Thieme.

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–Fourth Edition (PPVT-4)*. Circle Pines, MN: AGS.

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.

Gildersleeve-Neumann, C. E., & Goldstein, B. A. (2015). Cross-linguistic generalization in the treatment of two sequential Spanish–English bilingual children with speech sound disorders. *International Journal of Speech-Language Pathology, 17,* 26–40.

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

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96,* 280–301.

Hall, P. K. (2000). A letter to the parent(s) of a child with developmental apraxia of speech. Part IV: Treatment of DAS. *Language, Speech, and Hearing Services in Schools, 31,* 179–181.

Jakielski, K. J. (2002). *The index of phonetic complexity: Preliminary findings*. Paper presented at the Annual Child Phonology Conference, Wichita, KS.

Jakielski, K. J. (2017). *Building speech and quantifying complexity. The manual*. Pittsburgh, PA: Childhood Apraxia of Speech Association of North America.

Kaderavek, J. N., & Justice, L. M. (2010). Fidelity: An essential component of evidence-based practice in speech-language pathology. *American Journal of Speech-Language Pathology, 19,* 369–379.

Kaipa, R., & Peterson, A. M. (2016). A systematic review of treatment intensity in speech disorders. *International Journal of Speech-Language Pathology, 18,* 507–520.

Kearney, E., Granata, F., Yunusova, Y., van Lieshout, P., Hayden, D., & Namasivayam, A. (2015). Outcome measures in developmental speech sound disorders with a motor basis. *Current Developmental Disorders Reports, 2,* 253–272.

Kearns, K. P. (1986). Flexibility of single-subject experimental designs. Part II: Design selection and arrangement of experimental phases. *Journal of Speech and Hearing Disorders, 51,* 204–214.

Kim, I., LaPointe, L. L., & Stierwalt, J. A. G. (2012). The effect of feedback and practice on the acquisition of novel speech behaviors. *American Journal of Speech-Language Pathology, 21,* 89–100.

Kleim, J. A., Barbay, S., Cooper, N. R., Hogg, T. M., Reidel, C. N., Remple, M. S., & Nudo, R. J. (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.

Kwakkel, G., Wagenaar, R. C., Twisk, J. W. R., Lankhorst, G. J., & Koetsier, J. C. (1999). Intensity of leg and arm training after primary middle-cerebral-artery stroke: A randomised trial. *The Lancet, 354,* 191–196.

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.

Lee, T. D., & Weeks, D. J. (1987). The beneficial influence of forgetting on short-term retention of movement information. *Human Movement Science, 6,* 233–245.

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., Gildersleeve-Neumann, C. E., Jakielski, K. J., & Stoeckel, R. (2014). Motor-based intervention protocols in treatment of childhood apraxia of speech (CAS). *Current Developmental Disorders Reports, 1*(3), 197–206.

Maas, E., & Mailend, M.-L. (2017). Fricative contrast and coarticulation in children with and without speech sound disorders. *American Journal of Speech-Language Pathology, 26,* 649–663.

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

McAllister Byun, T., Hitchcock, E. R., & Swartz, M. T. (2014). Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research, 57,* 2116–2130.

McAllister Byun, T., Swartz, M. T., Halpin, P. F., Szeredi, D., & Maas, E. (2016). Direction of attentional focus in biofeedback treatment for /r/ misarticulation. *International Journal of Language and Communication Disorders, 51,* 384–401.

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 trial 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.

Murray, E., McCabe, P., Heard, R., & Ballard, K. J. (2015). Differential diagnosis of children with suspected childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 58,* 43–60.

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

Nijland, L., Maassen, B., van der Meulen, S., Gabreëls, F., Kraaimaat, F. W., & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics & Phonetics, 16,* 461–483.

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.

Preston, J. L., Leece, M. C., McNamara, K., & Maas, E. (2017). Variable practice to enhance speech learning in ultrasound biofeedback treatment for childhood apraxia of speech: A single case experimental study. *American Journal of Speech-Language Pathology, 26,* 840–852.

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

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

Robin, D. A., Maas, E., Sandberg, Y., & Schmidt, R. A. (2007). Motor control and learning and childhood apraxia of speech. In P. Hall, L. Jordan, & D. Robin (Eds.), *Developmental apraxia of speech: Theory and clinical practice* (2nd ed., pp. 67–86). Austin, TX: Pro-Ed.

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.

Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning: A behavioral emphasis* (4th ed.). Champaign, IL: Human Kinetics.

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

Shea, C. H., Kohl, R. M., & Indermill, C. (1990). Contextual interference: Contributions of practice. *Acta Psychologica, 73,* 145–157.

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.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997). Developmental apraxia of speech: I. Descriptive and theoretical perspectives. *Journal of Speech, Language, and Hearing Research, 40,* 273–285.

Skelton, S. L., & Hagopian, A. L. (2014). Using randomized variable practice in the treatment of childhood apraxia of speech. *American Journal of Speech-Language Pathology, 23,* 599–611.

Spielman, J., Ramig, L. O., Mahler, L., Halpern, A., & Gavin, W. J. (2007). Effects of an extended version of the Lee Silverman Voice Treatment on voice and speech in Parkinson's disease. *American Journal of Speech-Language Pathology, 16,* 95–107.

Størvold, G. V., Jahnsen, R. B., Evensen, K. A. I., & Bratberg, G. H. (2018). Is more frequent physical therapy associated with increased gross motor improvement in children with cerebral palsy? A national prospective cohort study. *Disability and Rehabilitation*. Advance online publication. https://doi.org/10.1080/09638288.2018.1528635

Strand, E. A., & Debertine, P. (2000). The efficacy of integral stimulation intervention with developmental apraxia of speech. *Journal of Medical Speech-Language Pathology, 8,* 295–300.

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,* 505–520.

Strand, E. A., & Skinder, A. (1999). Treatment of developmental apraxia of speech: Integral stimulation methods (pp. 109–148). In A. J. Caruso & E. A. Strand (Eds.), *Clinical management of motor speech disorders in children*. New York, NY: Thieme.

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.

Terband, H., & Maassen, B. (2010). Speech motor development in childhood apraxia of speech: Generating testable hypotheses by neurocomputational modeling. *Folia Phoniatrica et Logopaedica, 62,* 134–142.

Terband, H., Maassen, B., Guenther, F. H., & Brumberg, J. (2009). Computational neural modeling of speech motor control in childhood apraxia of speech (CAS). *Journal of Speech, Language, and Hearing Research, 52,* 1595–1609.

Terband, H., Maassen, B., Guenther, F. H., & Brumberg, J. (2014). Auditory–motor interactions in pediatric motor speech disorders: Neurocomputational modeling of disordered development. *Journal of Communication Disorders, 47,* 17–33.

Thomas, D. C., McCabe, P., & Ballard, K. J. (2014). Rapid Syllable Transitions (ReST) treatment for childhood apraxia of speech: The effect of lower dose-frequency. *Journal of Communication Disorders, 51,* 29–42.

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.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: Pro-Ed.

Wambaugh, J. L., Nessler, C., Cameron, R., & Mauszycki, S. C. (2013). Treatment for acquired apraxia of speech: Examination of treatment intensity and practice schedule. *American Journal of Speech-Language Pathology, 22,* 84–102.

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–Second Edition*. Circle Pines, MN: AGS.

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

Wolpert, D. M., Ghahramani, Z., & Flanagan, J. R. (2001). Perspectives and problems in motor learning. *Trends in Cognitive Sciences, 5*(11), 487–494.

Wright, D. L., Black, C. B., Immink, M. A., Brueckner, S., & Magnuson, C. (2004). Long-term motor programming improvements occur via concatenation of movement sequences during random but not during blocked practice. *Journal of Motor Behavior, 36,* 39–50.

Appendix A

Treatment Protocol

Steps for each teaching episode are outlined in the Protocol Flowchart below.

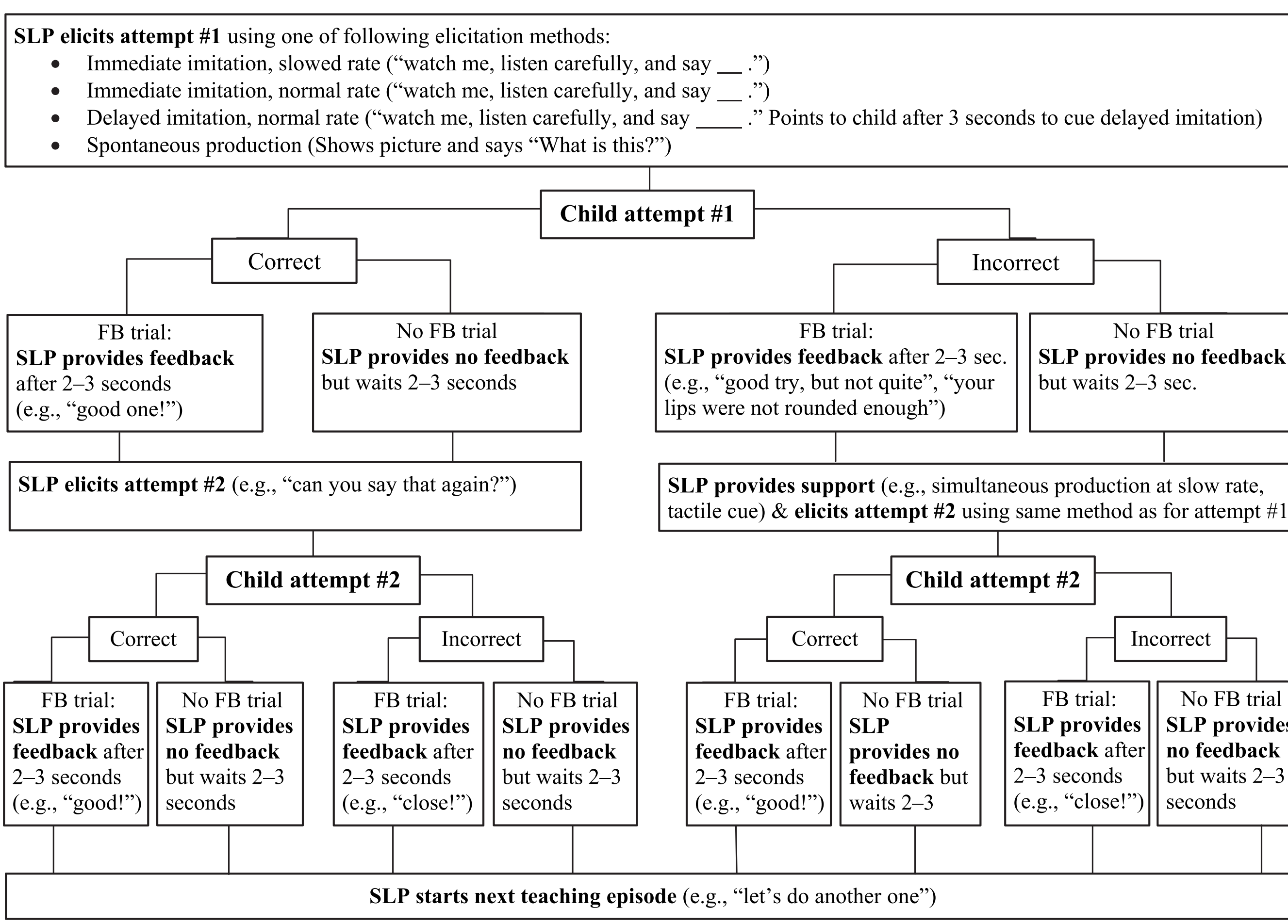
In each treatment phase, there are five targets for the massed condition and 10 targets for the distributed condition. A card deck contains two stimulus cards for each target per condition.

Teaching episodes are presented in random order for both conditions. Random order of teaching episodes is implemented by shuffling the cards before the treatment session and reshuffling the deck as needed during the session once every card has been used.

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 speech-language pathologist has a feedback tracking sheet with 100 slots with the fading feedback schedule marked (Ballard et al., 2007).

Start each condition in each session with immediate imitation at a normal rate. The criterion to increase difficulty level of elicitation condition is 2/2 consecutive correct Attempts #1. That is, if the child produces a correct response on Attempt #1 on two successive teaching episodes (whether these are the same target or not), then the third teaching episode begins with the next level elicitation method. Similarly, if the child produces 2/2 consecutive incorrect Attempts #1 with an elicitation method, the next teaching episode reverts to the previous difficulty level of elicitation method (or stay in immediate imitation at a slow rate).

Continue until session time has elapsed (determined by an egg timer set to 25 min per condition). After a brief break, begin the second treatment condition and continue until session time is up.



Appendix B

Treatment Fidelity Data

Table B1. Fidelity measures.

Child	001		002		003		004		005		008	
Fidelity aspect	**M**	**D**	**M**	**D**	**M**	**D**	**M**	**D**	**M**	**D**	**M**	**D**
No. (%) sessions reviewed	4/16 (25)		3/16 (19)		2/16 (12.5)		2/16 (12.5)		3/16 (19)		2/16 (12.5)	
Duration (minutes)[a]	23	24	23	23	24	23	21	21	22	23	23	22
Number of TEs[b]	26	28	25	26	27	28	16	15	34	31	22	27
Total whole-word attempts[c]	64	78	63	75	67	70	41	38	108	92	74	78
Whole-word attempts/TE[d]	2.48	2.72	2.53	2.88	2.49	2.48	2.53	2.58	3.30	2.96	3.40	2.89
FB frequency[e]	74%	64%	75%	58%	63%	53%	73%	72%	74%	80%	84%	58%
FB on the first and last attempts[f]	58%	43%	59%	34%	42%	28%	53%	54%	59%	67%	70%	43%
FB delay[g]												
<1 s	68%	71%	70%	82%	87%	80%	59%	84%	79%	76%	89%	86%
1–3 s	29%	27%	27%	16%	12%	15%	28%	14%	19%	23%	9%	14%
4–6 s	3%	1%	3%	3%	1%	5%	12%	0%	2%	1%	2%	0%
> 6 s	0%	1%	1%	0%	0%	0%	2%	3%	0%	0%	0%	0%
Elicitation per protocol[h]	96%	99%	97%	98%	96%	100%	100%	96%	85%	84%	87%	89%
Same elicitation: first and last[i]	92%	88%	88%	86%	92%	97%	98%	91%	90%	84%	78%	74%

Note. M = massed; D = distributed.

[a]Treatment duration per session (time from start of the first teaching episode to start of the last teaching episode; rounded to the nearest minute). [b]Number of teaching episodes per session. [c]Number of whole-word target attempts per session. [d]Number of whole-word target attempts per teaching episode (including first and last attempts). [e]Proportion of total number of attempts (initial and final) with FB (number of initial and final attempts with FB / total number of initial and final attempts). [f]Proportion of teaching episodes with the same presence/absence of FB on initial and final attempts. [g]Proportion of initial and final attempts with feedback delays in each given range. [h]Proportion of teaching episodes in which the initial elicitation method adhered to the protocol. [i]Proportion of teaching episodes with the same elicitation method for initial and final attempts (based only on TEs with an incorrect initial attempt).

Appendix C

Results From Nonparametric Statistical Group Analyses

Table C1. Results of nonparametric group analyses for Research Question 1 (practice amount).

Dependent variable	Test	Comparison	Statistic	p	Advantage
d	CMH	High vs. low vs. control	$\chi^2 = 4.33$	.115	
	Wilcoxon	High vs. low	$S = 6$	.250	—
		High vs. control	$S = 8.5$	.094	—
		Low vs. control	$S = 7.5$	.156	—
% change	CMH	High vs. low vs. control	$\chi^2 = 3.00$	.223	
	Wilcoxon	High vs. low	$S = 7.5$	.156	—
		High vs. control	$S = 8.5$	.094	—
		Low vs. control	$S = 4.5$	.4375	—

Note. Significant effects ($p < .05$) are bolded, and the direction of effect is indicated (Advantage column). Em dashes mean "no statistically significant difference." CMH = Cochran–Mantel–Haenszel test; Wilcoxon = signed-ranks test.

Table C2. Results of nonparametric group analyses for Research Question 2 (practice distribution).

Phase	Dependent variable	Test	Comparison	Statistic	p	Advantage
M2	*d*	CMH	Distributed vs. Massed vs. Control	$\chi^2 = 6.35$	**.042**	
		Wilcoxon	Distributed vs. Massed	$S = 6.5$	.219	—
			Distributed vs. Control	$S = 7.5$	.063	—
			Massed vs. Control	$S = 9.5$	.063	—
	% change	CMH	Distributed vs. Massed vs. Control	$\chi^2 = 9.33$	**.009**	
		Wilcoxon	Distributed vs. Massed	$S = 3.5$	.500	—
			Distributed vs. Control	$S = 10.5$	**.031**	Distributed > Control
			Massed vs. Control	$S = 10.5$	**.031**	Massed > Control
FU	*d*	CMH	Distributed vs. Massed vs. Control	$\chi^2 = 7.00$	**.030**	
		Wilcoxon	Distributed vs. Massed	$S = 5.5$	.313	—
			Distributed vs. Control	$S = 9.5$	.063	—
			Massed vs. Control	$S = 10.5$	**.031**	Massed > Control
	% change	CMH	Distributed vs. Massed vs. Control	$\chi^2 = 10.33$	**.006**	
		Wilcoxon	Distributed vs. Massed	$S = 5.5$	.313	—
			Distributed vs. Control	$S = 10.5$	**.031**	Distributed > Control
			Massed vs. Control	$S = 10.5$	**.031**	Massed > Control

Note. Significant effects ($p < .05$) are bolded, and the direction of effect is indicated (Advantage column). Em dashes mean "no statistically significant difference." M2 = Maintenance Phase 2; CMH = Cochran–Mantel–Haenszel test; Wilcoxon = signed-ranks test; FU = follow-up.