JSLHR

Research Article

An N-of-1 Randomized Controlled Trial of Interventions for Children With Inconsistent Speech Sound Errors

Susan Rvachew[a] and Tanya Matthews[a]

Purpose: The aim of this study was to test the hypothesis that children with inconsistent speech errors would respond differentially to 1 of 3 specific interventions depending on their primary underlying impairment: Children with deficient motor planning were expected to respond best to an auditory–motor integration (AMI) intervention, and children with deficient phonological planning were expected to respond best to a phonological memory and planning (PMP) intervention.
Method: Twelve participants were diagnosed with a motor planning (n = 7) or phonological planning (n = 5) deficit based on a comprehensive assessment, which included the Syllable Repetition Task as an important source of diagnostic evidence. An N-of-1 randomized controlled trial was used. Each child experienced all 3 interventions: AMI, PMP, and control (CTL); however, these interventions were randomly allocated to sessions within weeks (3 sessions per week × 6 weeks for 18 sessions). The AMI intervention procedures targeted knowledge of the acoustic–phonetic target and integration of auditory and somatosensory feedback during speech practice. The PMP intervention procedures targeted segmenting and recompiling the phonological plan for each word. The CTL intervention was standard drill practice. The child was taught 5 pseudowords in a meaningful context in each intervention condition.
Results: Same-day (SD) probes assessed transfer from taught pseudowords to untaught real words, and next-day (ND) probes assessed retention of that learning. Nonparametric resampling tests with pooling of p values across children with the same diagnosis were used to assess the results. Pooled p values indicated a significant benefit of AMI over PMP for the group with a motor planning deficit (p = 2.01E-04 for SD probes and 2.97E-03 for ND probes) and a significant benefit of PMP over AMI for the group with a phonological planning deficit (p = 1.22E-02 for SD probes and 1.32E-02 for ND probes). Response to the CTL intervention was variable within groups.
Conclusion: In this study, the child's underlying psycholinguistic deficit helped to predict response to intervention.

Children with a speech sound disorder (SSD) tend to produce errors that are consistent and predictable in their patterning (Dodd et al., 2017), usually typical phonological error patterns or common clinical distortions. It is known that these error types can be successfully treated with phonological and/or traditional treatment approaches (Hesketh, Adams, Nightingale, & Hall, 2000; Law, Garrett, & Nye, 2003). Furthermore, treatment of one member of a sound class can generalize to other segments in that class (McReynolds & Bennett, 1972), and treatment of consistent error patterns in one word position can generalize to another word position where the error appears (Olswang & Bain, 1985; Powell & McReynolds, 1969). In contrast, those few children who produce inconsistent and unpredictable errors are often difficult to treat in that they may not learn the phonemes that are targeted in speech therapy and, even when they do, new phonemes may not generalize to untaught contexts (Forrest, Dinnsen, & Elbert, 1997; Forrest, Elbert, & Dinnsen, 2000). Therefore, children with inconsistent errors form a unique subgroup of SSD that requires a particular and nonphonological approach to speech therapy in order to achieve a satisfactory outcome.

The specific nature of the deficit that is associated with inconsistent speech sound errors and the specific type of treatment that should be employed is, however, controversial. On the one hand, Forrest (2003) found that inconsistent speech errors are viewed as the primary marker of childhood apraxia of speech (CAS) by North American speech-language pathologists (SLPs). On the other hand,

[a]School of Communication Sciences and Disorders, McGill University, Montréal, Québec, Canada

Correspondence to Susan Rvachew: susan.rvachew@mcgill.ca

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

Received July 17, 2018
Revision received January 6, 2019
Accepted April 23, 2019
https://doi.org/10.1044/2019_JSLHR-S-18-0288

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

Dodd (2014) describes inconsistency in speech production as a primary marker of inconsistent phonological disorder (IPD). Given that CAS is a motor planning disorder whereas IPD is a phonological planning disorder, it is reasonable to presume that these two putative subtypes of SSD will respond most favorably to different approaches to intervention. Herein, we describe a study in which this hypothesis was assessed by using a single-subject randomization design (Rvachew & Matthews, 2017a) to describe individual children's response to interventions that addressed deficits in motor versus phonological planning. Although the children met criteria for CAS, the hypothesized pattern of treatment response for each child was determined by the primary level of breakdown in planning, either motor planning or phonological planning. Each child's diagnosis was determined by their profile of deficits across a battery of assessments in which the Syllable Repetition Task (SRT) was primary, because this test was designed by Shriberg et al. (2009) to identify deficits in planning at these different levels independently of inconsistency in speech errors during the production of real words. Each of these subtypes of SSD will be discussed in turn, in relation to diagnostic signs and intervention procedures.

Motor Planning Disorder

Diagnostic Signs

CAS is defined above all as a disorder of speech motor control in which the child is unable to coordinate the temporal and spatial parameters of multiple articulatory gestures (American Speech-Language-Hearing Association, 2007). Therefore, inconsistency in phoneme production arises when the movements of multiple articulators are mistimed during the production of a particular phoneme or syllable. Similarly, miscoordination of pitch and duration cues may result in dysprosody, perceived as inaccurate lexical stress or phrase-level prosodic contours. Finally, the challenge of coarticulating segments within and across syllables may lead to unusual segregation of phonemes, especially at syllable boundaries. These signs, taken individually, may not be reliable indicators of a CAS diagnosis, however. Inconsistency in the production of words or phonemes is not a reliable marker of CAS (Murray, McCabe, Heard, & Ballard, 2015) because it occurs frequently in the speech of children with language disorders or speech delay as well (Iuzzini-Seigel, Hogan, & Green, 2017). Unusual prosody may be a more promising diagnostic marker, but identifying a reliable and objective indicator of dysprosody for this population has proven difficult (Velleman & Shriberg, 1999), and some children's SSD is so severe they are not yet producing enough connected speech for a clear analysis of prosody. Shriberg, Lohmeier, Strand, and Jakielski (2012) claimed that the only specific sign of apraxia is a transcoding deficit, operationalized in their test battery as addition errors when repeating nonsense syllables on the SRT (Shriberg et al., 2009). These addition errors occur at syllable junctions (e.g., /bada/ → [banda] as described in Rvachew & Matthews, 2017b) and reflect the inability to transition between these elements, even though the nonsense words on this test are made up of a restricted set of only four simple syllables ([ma], [na], [ba], and [da]). Other researchers have reported that an inability to complete maximum performance tasks and, in particular, sequence [pataka] rapidly is a characteristic of CAS (Thoonen, Maassen, Gabreels, & Schreuder, 1999). A complete inability to sequence these syllables, even at a slow rate, might be another indicator of impairment in "transcoding processes that plan and programme the representations for the motoric gestures of manifest speech" (p. 447), as defined by Shriberg et al. (2012).

Intervention

Many different approaches to the treatment of CAS have been reported in the scientific literature (for a review, see Murray, McCabe, & Ballard, 2014; Ballard et al., 2015). Nonetheless, these approaches share certain common features. First, the specific goals when treating CAS are universally concerned with movement patterns rather than specific phoneme accuracy (Strand & Skinder, 1999). Therefore, when treating CAS, a small set of targets will be practiced during each session, chosen to focus attention on the coordination and temporal patterning of particular articulators within a specified prosodic context; accurate and automatic production of the whole target is expected. Furthermore, a phonological approach will necessarily focus on meaningful words and meaningful phonological contrasts, whereas interventions for the treatment of a motor planning disorder often employ pseudoword stimuli (Murray, McCabe, & Ballard, 2015). A primary advantage of practicing such stimuli is that the child undergoes the process of learning a new "word" while avoiding interference from mislearned motor plans that have been stored in the lexicon. An additional advantage is that pseudowords can be designed to focus on specific movement patterns individualized to the child's needs. Despite these advantages, many interventions for CAS focus on functional words and phrases in order to motivate practice and facilitate generalization (for further discussion, see Maas, Gildersleeve-Neumann, Jakielski, & Stoeckel, 2014). The intervention approaches used in the studies reported herein taught pseudowords but in meaningful drill-play contexts.

Second, Murray et al. (2014) reported that 90% of studies reporting treatment approaches for CAS made use of the principles of motor learning that define learning as transfer and retention of a new skill to untaught contexts as differentiated from practice performance (Maas et al., 2008). Maas and colleagues (Maas et al., 2014, 2008) recommend that treatment sessions be composed of a prepractice portion during which procedures are implemented to prepare the learner for successful practice and a practice portion during which principles of motor learning are employed to ensure response and/or stimulus generalization. Important principles include *distributed practice* (more time between practice sessions supports retention), *variable practice* (practicing multiple targets in varied contexts supports transfer), *feedback frequency* (reduced frequency supports learning), and *feedback timing* (immediate feedback from the SLP inhibits processing of intrinsic feedback and therefore delayed feedback is desirable).

Third, children who have motor planning disorders must receive intense treatment in order to achieve a good outcome (Campbell, 1999; Namasivayam et al., 2015). Murray et al. (2014) found that the median frequency of treatment sessions was three times per week. Another way to increase cumulative intervention intensity is to increase dose—that is, the number of practice trials per session (Warren, Fey, & Yoder, 2007). Studies have found that a minimum of 100 practice trials per session is feasible and effective (Edeal & Gildersleeve-Neumann, 2011; A. L. Williams, 2012).

Finally, treatments for CAS typically involve the provision of multimodal stimulation to support accurate production on each practice trial (Strand, Stoeckel, & Baas, 2006). The amount of stimulation provided is graded on each trial to correspond to the child's needs as indexed by performance on preceding trials. One reason that children with CAS might need enhanced sensory information prior to, during, and after each speech practice attempt is a hypothesized impairment in the feedforward control mechanism that underlies the development of speech motor control. This mechanism is impaired due to poor mapping between articulatory gestures and their sensory effects, so that predictions about the outcome of implementing selected motor plans are often inaccurate. Terband and Maassen (2010) proposed that CAS is caused by "weak forward models" as a consequence of poor oral sensitivity so that the mappings between articulatory gestures and acoustic outputs would be imprecise; consequently, the ability to plan and adjust motor movements in real time is impaired while speaking due to degraded information about the current somatosensory state of the vocal system. As would be expected, these children subsequently develop an overreliance on feedback control while speaking. This prediction was supported by Iuzzini-Seigel, Hogan, Guarino, and Green (2015) who found that children with CAS produced less accurate voice onset time when masking noise was present, whereas children with SSD or typical speech were able to maintain accuracy under this condition. Terband, van Brenk, and van Doornik-van der Zee (2014) perturbed auditory feedback when children were producing vowel sounds; they observed that children with typical speech changed their vowel production to compensate for and adapt to the altered feedback, whereas children with motor speech impairments did not. When listeners with strong forward models are presented with systematically perturbed auditory feedback, they continuously update their motor plans in an effort to achieve a stable acoustic output over time; when children who are overreliant on auditory feedback experience this perturbation, motor plans are selected to achieve a continuously updated auditory target, hence "target drift" as observed by Shiller, Sato, Gracco, and Baum (2009). Therefore, Terband et al. concluded that auditory–motor integration (AMI) plays a key role in severe SSDs; specifically, children must develop a strong correspondence between articulatory gestures and their acoustic outcomes in order to compensate for the imprecise mapping between articulatory and somatosensory states (see also Liégeois et al., 2019).

Given this theoretical perspective and previous findings that perceptual training improves sensory–motor adaptation in children (Shiller & Rochon, 2014), the treatment approach adopted to improve motor planning in this study focused on procedures designed to sharpen knowledge of the auditory target and improve the children's ability to monitor their own acoustic output in relation to that target. This strengthened internal model would form the foundation for the acquisition of feedforward control with repeated practice. Learning from prepractice activities was carried over into the practice portion of the session during which the child was encouraged to integrate knowledge of the target with self-produced feedback during high-intensity speech practice.

Phonological Planning Disorder

Diagnostic Signs

Historically, CAS has been conceived as a continuum of planning deficits that might encompass linguistic elements, motor elements, or both (see Crary, 1993; Shriberg, Aram, & Kwaitkowski, 1997). According to this view, a phonological planning disorder might be viewed as a subtype of CAS (as described in Ozanne, 2005); in contrast, Bradford and Dodd (1996) argued that children with an underlying breakdown in phonological planning constituted a distinct diagnosis, not overlapping with CAS and requiring a unique approach to intervention. Specifically, these children were diagnosed with IPD, defined as inconsistent production of the same words during repeated administration of a picture naming task (Dodd et al., 2017). Dodd and Bradford (2000) suggested that the children who produce unpredictable and atypical speech errors have a "deficit at the level of constructing, storing and/or retrieving the phonological output plan" (p. 190). These inconsistencies are observed at the level of the whole word because the errors arise at the point of phonological assembly: The number of "slots" for syllables or segments within syllables might be misplanned, and the slots might be filled with the wrong phonemes or in the wrong order. Relying on token-to-token inconsistency as the sole diagnostic sign to identify children with a phonological planning deficit can be problematic, however. Often, children with a phonological planning deficit will have word finding problems and possibly weak vocabulary skills in general, and it may be necessary to elicit the picture naming items with imitative prompts. These children demonstrate excellent imitative abilities, and therefore their accuracy under these elicitation conditions will provide an overestimate of their consistency. Furthermore, because children with CAS often have difficulties at the phonological and motor planning levels (Shriberg et al., 2012), it is necessary to rule out motor planning difficulties in order to diagnose a specific breakdown in phonological planning. Phonological planning is a prearticulatory planning step (i.e., it is a psycholinguistic process that precedes motor planning). The neural network for prearticulatory planning overlaps neural structures in Broca's area that are responsible for short-term memory (Flinker et al., 2015; Hickok et al., 2014). Therefore, differential diagnosis of a phonological planning disorder versus a motor planning disorder can be supported by the use of the SRT, in this case examining for evidence of a phonological memory deficit in the absence of a transcoding deficit, but coincident with within-word inconsistency as revealed by the Word Inconsistency Assessment (Dodd, Zhu, Crosbie, Holm, & Ozanne, 2006).

Intervention

It is not necessary to take a position on the question of whether a phonological planning deficit is part of the continuum of planning deficits that comprises CAS, or a distinct subtype of phonological disorder, to hypothesize that motor versus phonological planning deficits might benefit from unique intervention procedures. One intervention has been recommended and assessed for use with children who have IPD: The Core Vocabulary Approach is designed to increase consistency in the production of functional words with a focus on stimulus generalization, that is, transfer across multiple speaking environments (Dodd & Bradford, 2000). A focus on functional vocabulary is not unique in speech therapy (Murray et al., 2014). The treatment goal of achieving consistency—as opposed to accuracy—is novel, but it has been shown that the approach can be combined with stimulability training to ensure both consistency and accuracy (Iuzzini & Forrest, 2010). The distinctive aspects of this intervention lie in treatment procedures that are used to help the child achieve consistent productions of the target words. During prepractice, the child is taught to produce the words by segmenting them, syllable by syllable and phoneme by phoneme, learning the segments individually in order with the provision of visual cues. The visual cues are critical because, during practice, the child is encouraged to use them to form a phonological plan independently, recombining the segments to produce the target words. Importantly, the child practices the words without hearing a prior auditory model of the target, as is typical in virtually every other approach to speech therapy. Withholding of a prior auditory model forces the child to assemble their own phonological plan; in the event that the child cannot do this accurately, the visual cues that were introduced during prepractice can be used by the child to achieve a successful production, with backward or forward chaining as additional support as required. This approach has been shown to be successful for children diagnosed with IPD in case studies and multiple baseline single-case experiments (Crosbie, Holm, & Dodd, 2005; Dodd & Bradford, 2000; McIntosh & Dodd, 2008). No studies have determined whether children with inconsistent speech errors due to a motor planning disorder, as opposed to a phonological planning disorder, will respond to this approach to intervention.

In this study, the teaching procedures that are central to the Core Vocabulary Approach were adopted in our phonological memory and planning (PMP) intervention. Children were taught new words via segmentation procedures, introducing visual cues and implementing forward and backward chaining to sequence the requisite phonemes in each target word during the prepractice portion of the treatment session. In a departure from the Core Vocabulary Approach, the children were taught pseudowords but in a meaningful context. In this way, the goals and practice part of each treatment session could be parallel to those for the AMI intervention. The rationale for teaching pseudowords during PMP sessions was roughly the same as in the case of the AMI sessions: Specifically, the words could be constructed to meet the individual needs of the child, and the children would be experiencing the process of learning new words in all treatment sessions.

Summary and Hypotheses

In this study, 12 children participated after being diagnosed with a motor planning disorder or a phonological planning disorder. Each child received 18 treatment sessions, with each treatment session targeting one of three goals selected individually for that child. Treatment goals were randomly assigned to one of three different experimental conditions, and these conditions, in turn, were randomly allocated to treatment sessions using block randomization so that the outcomes could be assessed in the context of a single-subject randomized alternation design (Edgington, 1987; Rvachew & Matthews, 2017a). The three treatment conditions were as follows: (a) a usual care control (CTL) condition consisting of intense practice alone (using the principles of motor learning and the integral stimulation hierarchy); (b) AMI condition in which prepractice procedures focused on knowledge of the auditory target and error monitoring, followed by intense practice; and (c) PMP condition in which prepractice procedures focused on segmenting the words to help the child develop their own phonological output plan before proceeding to intense practice. Practice involved pseudowords, whereas learning was assessed with real words that had similar phonetic characteristics to the target structures. Resampling tests were used to assess transfer and retention in relation to the following hypotheses:

1. Each child was expected to show a differential treatment response with a pattern of treatment response associated with the child's diagnosis.
2. Children with a diagnosis of motor planning disorder were expected to obtain higher probe scores in the AMI condition in comparison to the CTL and PMP conditions.
3. Children with a diagnosis of phonological planning disorder were expected to obtain higher probe scores in the PMP condition in comparison to the CTL and AMI conditions.

Method

Participants

Recruitment and Identification

The trial protocol was approved by the institutional review boards of McGill University and the Centre de

Recherche Interdisciplinaire en Réadaptation du Montréal Métropolitain. Brochures were distributed to SLPs and through the Facebook page for Apraxia Kids–Canada requesting referrals of English-speaking children aged 4–6 years with a primary SSD and suspected or confirmed CAS (as indicated by producing many more speech errors than would be expected for the child's age and especially inconsistent speech errors). The age requirement was relaxed to allow younger and older children to participate in the study. Recruitment and intake began in Winter 2013 and was terminated in Spring 2015 when the research funds were exhausted.

Intake Procedure

Parents and referring SLPs were interviewed, and written documents (case history form, past assessment reports, signed consent form) were collected to ensure that (a) the child did not present with certain exclusionary conditions, specifically primary sensory disorders (e.g., sensorineural hearing loss, significant visual impairments) or syndromes (e.g., Down syndrome, Fragile X); (b) the family and the child, in the opinion of the family and referring SLP, could comply with the trial protocol; and (c) the parent would provide written consent for their child's participation. Subsequently, a 2-hr assessment was scheduled during which tests were administered to confirm eligibility for trial inclusion and to select treatment targets, specifically the complete Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd et al., 2006), the SRT (Shriberg et al., 2009), the Peabody Picture Vocabulary Test–III (PPVT-III; Dunn & Dunn, 1997), and the Matrices subtest of the Kaufman Brief Intelligence Test–Second Edition (KBIT-2; A. S. Kaufman & Kaufman, 2004). Children were retained for participation in the study[1] if their percent consonants correct (PCC) score on the DEAP Articulation Assessment was no more than 85 and their SRT Competency Score was below normal limits (i.e., $z < -1.00$), and inconsistent speech errors were observed. Children with milder forms of speech delay were excluded. Furthermore, the child was required to demonstrate connected speech during spontaneous conversation with the examiner and the ability to comply with the 2-hr assessment procedure, that is, name pictures, follow instructions, imitate actions and speech sounds, and persist with formal tasks for a length of time.

Final Sample

Twenty-three children were assessed for inclusion in this protocol (see supplemental materials for details: Supplemental Material S1 for intake data, Supplemental Material S2 for history and final status information, and Supplemental Material S3 for evidence of CAS inclusion criteria). Seven were excluded because their profiles were consistent with phonological delay, articulation disorder, or language impairment. This intake process resulted in 17 children who met common checklist criteria for CAS diagnosis (e.g., Shriberg et al., 2012; as shown in Supplemental Material S3). Of these 17 who were invited to participate in the study, four withdrew before finishing enough probe assessments for analysis (in two cases, family emergencies precluded further participation; one child was absent from school too frequently to attend all the sessions; one child did not enjoy the therapy and was excluded due to lack of assent). The 12 children who were accepted as participants[2] and who completed the study are described with respect to their intake assessments in Table 1.

The children ranged in age from 3.06 to 8.09 years. The one girl obtained a relatively high PCC on the DEAP Articulation Assessment but, like all the other children, had significant difficulty repeating sequences of syllables correctly on the SRT. On the PPVT, one child achieved a standard score below 70 and seven children achieved standard scores below 85. Cognitive delays in the nonverbal domain were also observed, with one child obtaining a nonverbal IQ below 70 and four children obtaining a nonverbal IQ below 85. One child was too young to complete the KBIT, and another refused to complete this test; therefore, there is no indication of their nonverbal intelligence. The broad variability in performance across these intake measures was not considered to be problematic for this study because each child served as his or her control given the single subject randomization design.

Procedure

Diagnostic Assessments

Certain formal assessments were administered in order to ensure eligibility for trial participation and to determine whether each child presented with a motor planning disorder, a phonological planning disorder, or another diagnosis that would preclude trial participation. As mentioned previously, the PPVT and the KBIT were administered for intake purposes. For diagnostic and treatment planning purposes, the complete DEAP (Dodd et al., 2006) was administered, including the Articulation Assessment, Oral Motor, Phonology, and Word Inconsistency Assessments.[3] An oral peripheral examination was conducted including maximum performance tasks (Rvachew, Hodge, & Ohberg, 2005; Thoonen et al., 1999). Speech perception and implicit phonological awareness skills were assessed using procedures described in Rvachew and Grawburg (2006), but the results are not formally reported here because many children were

[1]The children were treated in a university clinic in which the student clinicians obtained a standard number of practice hours by treating all the children who were referred. However, data are reported here for only those children who met the intake criteria for this study. A full list of all children who were referred with intake information and an explanation for all exclusions are provided in Supplemental Materials S1–S3.

[2]Some of these children were treated more than once in this clinic, but only one data set for each child is presented here so that all p values reported and pooled are independent of each other.

[3]During the first cohort, the assessment battery included the Kaufman Speech Praxis Test for Children (N. R. Kaufman, 1995), but this assessment was abandoned in subsequent cohorts to permit a broader age range of participants and to ensure completion of the Word Inconsistency Assessment.

Table 1. Description of participants at intake by diagnostic group.

Participant	Age	Sex	Single words PCC	Single words PVC	SRT-competency PCC	SRT-competency z Score	PPVT SS	KBIT NVIQ
Motor planning group								
TASC02	8;09	M	60	95	74	–2.40	79	56
TASC04	3;07	M	34	65	66	–1.70	110	nd
TASC16	4;01	M	49	69	54	–1.60	78	71
TASC23	9;06	M	43	64	22	–12.58	74	67
TASC24	4;09	M	66	92	46	–2.12	83	100
TASC30	4;06	M	37	83	36	–2.76	79	nd
TASC34	8;09	M	49	83	66	–4.58	117	105
Phonological planning group								
TASC05	4;07	M	61	95	60	–1.22	85	90
TASC09	4;09	M	52	98	40	–2.50	89	107
TASC20	4;05	F	82	94	42	–2.37	109	112
TASC21	5;05	M	61	86	36	–4.41	69	81
TASC26	4;10	M	55	86	40	–4.05	70	100

Note. Sex indicated as male (M) or female (F). Age is shown as years;months. Single-word naming performance was scored for percent consonants correct (PCC) and percent vowels correct (PVC). Raw competency score on the Syllable Repetition Task (SRT) is also PCC. Peabody Picture Vocabulary Test–III (PPVT) and Kaufman Brief Intelligence Test (KBIT) scores are reported as standard scores with a mean of 100 and an *SD* of 15 except where there were no data (nd). SS = Standard score; NVIQ = Nonverbal IQ.

not within the age range of the normative data for these tests (see Supplemental Material S1).

Diagnostic Criteria

Further to inclusion in the trial on the basis of the criteria described above and shown in Supplemental Material S3, each child was further classified as having a primary deficit in motor planning or phonological planning. Each child was required to listen to two-, three-, and four-syllable sequences of [na], [ma], [da], and [ba] syllables as presented by the SRT and then repeat them. The number of correct consonants produced by syllable length was scored to yield a raw score and a *z* score that is provided in Table 2. The raw score can be interpreted with respect to cutoff scores provided in Shriberg et al. (2012) as being indicative of CAS on the basis of empirical data. The *z* scores are derived from reference scores presented in Lohmeier and Shriberg (2011); however, the samples sizes for each age interval in this source are small, and the standard deviations can be large, especially for younger children. Therefore, we have found that it was important to consider the raw scores. The children's SRT performance is shown in Table 2, with bolding used to highlight *z* scores that are below age expectations and raw scores that are below the cutoff for CAS.

A deficit in phonological planning was associated with an SRT-memory score of less than 67.5, indicating significantly poorer performance on three-syllable items in comparison to two-syllable items (one child, TASC09, was assessed against the reference data for the ratio of performance for four-syllable items vs. three-syllable items because his two-syllable performance was unusually poor). Memory score is lower as performance declines with item length. A phonological planning deficit is also associated with a very high degree of inconsistency on the Word Inconsistency Assessment (at least 40%). Children who demonstrated high token-to-token inconsistency on this test also produced atypical and inconsistent errors in their connected speech samples and on the phonological assessment at the level of segment accuracy and syllable structure matching. Children who were determined to have a phonological planning deficit were observed to score below the cutoff on the SRT-memory subtask but above the cutoff on the SRT-transcoding subtask. Their performance on tests of oral–motor function was varied among children, with one child in this group even presenting with some signs of subclinical dysarthria (TASC21). Groping was not observed in any of the children with a phonological planning deficit in the absence of a transcoding deficit. Three children with a phonological planning deficit failed the DEAP Oral–Motor Screen because they were unable to sequence [pataka] accurately. In this group, three children also had significant difficulty with syllable matching. The children with phonological planning deficits produced a small number of syllable segregation errors.

A deficit in motor planning was associated with significant transcoding difficulties, specifically four or more addition errors while repeating the strings of nonsense syllables and a raw score of 80 or less on the SRT. Performance on the DEAP Oral–Motor Screen and on maximum performance tasks was also examined, with a dyspraxia score of 2 considered to be definitive (Rvachew et al., 2005; Thoonen et al., 1999). The children in this study obtained a dyspraxia score of 2 because they were unable to achieve an accurate sequence of [pataka] consistently and accurately, with one exception, TASC21, who repeated monosyllable and trisyllable sequences accurately but at a very slow rate. Some

Table 2. Indicators of encoding, phonological memory, and transcoding deficits by participant and diagnostic group.

Participant	Memory		WIA	Transcoding			OMS	MPTs		Groping	Lexical stress	Syllable segregation
	Raw	*z*		Raw	#Add	*z*		Dys	Apr			
Motor planning group												
TASC02	72	**–2.57**	**87%**	**67**	**6**	**–4.95**	**F**	0	**2**	No	40%	0%
TASC04	82	0.81	nd	**78**	**4**	–0.51	**F**	nd	**2**	No	70%	20%
TASC16	**44**	**–1.10**	**88%**	**80**	**4**	–0.90	**F**	**2**	**2**	Yes	100%	100%
TASC23	100	0.89	**64%**	**72**	**5**	**–3.94**	**F**	0	**2**	No	86%	43%
TASC24	**1**	**–2.86**	**64%**	**78**	**4**	–0.89	**F**	0	**2**	Yes	46%	8%
TASC30	73	0.06	**100%**	**16**	**11**	**–7.25**	**F**	nd	**2**	Yes	100%	50%[a]
TASC34	72	**–2.57**	**76%**	**78**	**4**	**–2.93**	**F**	**2**	**2**	Yes	17%	8%
Phonological planning group												
TASC05	**52**	–0.8	nd	89	2	0.16	P	0	**0**	No	10%	10%
TASC09	71	**–2.28**	nd	89	2	–0.16	**F**	0	**2**	No	50%	20%
TASC20	**66**	–0.22	**63%**	89	2	0.16	P	0	1	No	0%	9%
TASC21	**41**	**–4.29**	**87%**	94	1	0.27	**F**	**2**	**2**	No	92%	11%[a]
TASC26	**60**	**–2.66**	**52%**	89	2	–0.39	F	0	2	No	43%	7%

Note. Transcoding and memory scores are from the Syllable Repetition Task. Word Inconsistency Assessment (WIA) scores indicate the percent words out of 50 that were produced inconsistently over three independent trials. The Oral–Motor Screen (OMS) is from the Diagnostic Evaluation of Articulation and Phonology (DEAP). Maximum performance tasks (MPTs) were scored to indicate dysarthria (Dys) or apraxia (Apr), with a score of 2 indicating a greater likelihood of the diagnosis than 1 and 0 meaning no indicators of the diagnosis (see Thoonen et al., 1999). Groping is scored as yes when observed during the OMS. Lexical stress and syllable segregation errors were scored as percent occurrences on multisyllabic words elicited on the DEAP except in those participants who did not produce any multisyllabic words;
[a]in which case, syllable segregation was scored from the SRT. Bolding indicates scores that are below age expectations. nd = no data.

additional information about signs that are common to CAS is also recorded in Table 2. The children who obtained low SRT-trancoding scores also failed the DEAP Oral–Motor Screen. These children all demonstrated high word inconsistency, and four of them obtained low SRT-memory scores. Groping was observed in four of these children. These children demonstrated many errors involving inappropriate syllable matching or inappropriate lexical stress. Three of the children also produced syllable segregation errors frequently during single-word naming. TASC02 was observed to have unusually long stop gaps and syllable segregation errors in connected speech as well as difficulties maintaining an appropriate speaking rate (usually too fast). TASC16 was highly disfluent. Note that these children also had phonological memory deficits, as shown in Table 2.

Selection of Goals and Target Words

Intermediate goals were selected by considering all the single-word and connected speech data available from the intake assessment and conducting a multilinear analysis to take strengths and needs into account at the prosodic and segmental tiers of the prosodic hierarchy (Bernhardt, Stemberger, & Major, 2006; Rvachew & Brosseau-Lapré, 2018). In all cases, goals were selected to stabilize structures that were present but inconsistent in the child's speech. Intermediate goals were selected to represent structures at three of the following levels of the phonological hierarchy: (a) higher level prosodic structures such as prosodic phrase, word, or foot, with these targets reflecting difficulties with complexity of longer utterances including poor assignment of syllable stress, weak syllable deletion or weakening through deletion of word-internal consonants, consonant harmony or other assimilation patterns, or increasing substitution errors with greater utterance length; (b) subsyllabic units, with the most common being deletions from complex onsets or codas; (c) major sound class features, reflecting absence of a major sound class category or alternations among them (children often demonstrated confusion among major sound classes, such as liquids ↔ glides, nasals ↔ liquids, or fricatives ↔ affricates being substituted in both directions); and (d) stabilization of a specific place feature or phoneme that was emerging in the child's system. Most children demonstrated some inventory constraints, but these missing phonemes were avoided as specific targets unless there were no other choices. The three intermediate goals for each child were listed in order from the highest to lowest with respect to the phonological hierarchy. Goals were chosen to ensure that the complete set comprised stimulable targets at similar levels of consistency to the extent that this was possible.

Subsequently, http://www.random.org/lists was used to randomly assign targets to treatment conditions, independently for each child, in order to control for any possible differences in target difficultly. The result is shown in Supplemental Material S4, which shows the three intermediate targets for each child and the randomly assigned condition for each goal. The first participant, TASC02, will be discussed in more detail as an example. This child was able to produce most phonemes accurately with the exception of /s, z, θ, ð/. Errors increased markedly in three-syllable words and in connected speech; however, as speaking rate increased, omission of word-internal and word-final coda consonants was common alongside weak syllable deletion ("look out the window" /lʊk aʊt əv ðə wɪndoʊ/ → [lʊ aʊ də wɪndʊ]). Accuracy might be maintained as utterances lengthened if he slowed his rate, but phoneme and syllable segregation increased ("a sponge" /əspʌndʒ/ → [ʌs.pʰʌntʃ]), and his prosody became atypical ("mirror" /ˈmɪˌɹɚ/ → [ˈmɪˈwʌ]). Therefore, the first goal was to provide a frame for the production of three syllables with a strong–weak–strong or weak–strong–weak stress pattern with inclusion of the coda consonants. Pseudowords such as "Biftenope" were assigned as names to monsters who engaged in unusual activities such as "upteening." Moving down the phonological hierarchy to subsyllabic units, all types of clusters were reduced or simplified on an inconsistent basis in single words and in connected speech ("spray her with water" /spɹeɪ hɝ wɪθ wɑtɚ/ → [fe hɝː wɑtə]; "climb out of her bed" /klaɪm aʊt ʌv hɚ bɛd/ → [twaɪm aʊʔ ʌv hɚ bɛd]; "climb off this" /klaɪm ɑf ðɪs/ → [tɑm ɑf dɪst]). The /l/ clusters were selected as the next intermediate target because his sibilants were consistently distorted but the /l/ phoneme was produced correctly on an inconsistent basis. Pseudowords such as "blees" were paired with "alien" flowers and gardening tools. Finally, moving to the segmental tiers of the phonological hierarchy, the place feature [Dorsal] was selected because this feature was matched 25% of the time in his speech, whereas other potential segment targets were consistently in error ("how is that gonna come out?" /haʊ ɪz ðæt ɡanʌ kʌm aʊt/ → [haʊ dæt dɑː tʌm aʊt]). Pseudowords such as "Kip" were selected as names for dogs and their favorite toys and foods. Random assignment resulted in the prosodic level target (three-syllable words) being treated in the AMI condition, the subsyllabic unit target (/l/ clusters) being treated in the CTL condition, and the place feature target ([Dorsal] → /k, ɡ/) being treated in the PMP condition.

Probes

A set of probe items was created uniquely for each intermediate target and each child, containing 30 items. Whereas the specific treatment targets were pseudowords, the probe items were real words or phrases, selected to sample the targeted prosodic and segmental structures. An effort was made to construct the probe items from phones that were in the child's inventory as much as was possible. Length of probe items was typically two to four syllables long but constrained to be within the child's utterance length in syllables on average. Although phrase-level probe items were preferred, single-syllable probes might be constructed to sample difficult phonemes for those children who typically produced shorter utterances or who produced only very simple forms in longer utterances. Although treated pseudowords were expected to be produced with complete accuracy during treatment sessions, probe items were not expected to be produced completely correct during probes; rather, the child was expected to demonstrate use of the targeted structure. Therefore, TASC02 was taught to name his monster "Biftenope" → [ˈbɪftəˈnop] with all segments and the stress pattern produced correctly. When asked to say "garbage man" during probe administration, he might say [ˈdʌbədʒˈmæn], and this would be scored as correct because the word internal coda [dʒ] and the word final coda [n] were both produced accurately and the word was also produced with correct syllable marking and stress pattern.

With respect to probe administration, the set of 30 probe items was used to create sets of 12 unique probes for administration as same-day (SD) probes and next-day (ND) probes for each intermediate goal. These probes were created by selecting 10 items at random from the list of 30 items 12 times and assigning two of these lists to each week, one as an SD probe and one as an ND probe. At the end of the session during which the intermediate goal was treated, the SD probe would be administered; when the child returned for the next treatment session, the ND probe for that same goal would be administered. Therefore, SD probes were always administered at the end of the session targeting that goal and ND probes were administered at the beginning of the next session, regardless of which goal was treated on that day. Recall that SD probes assess transfer of learning from the pseudowords taught during the therapy session to the real words produced during the probe; ND probes assess retention of that learning until the next treatment session.

Probe administration was always by imitation with no visual (picture) stimulus provided. The treating therapist read the probe items live in front of the child and asked the child to repeat the item. No feedback regarding performance was provided to the child. Probe responses were recorded with a SONY HDR-XR150 handy camera and a Zoom handy recorder H1 with settings of 24 bit/96 kHz in Wave format. The child's responses were phonetically transcribed after listening to the audio recording and viewing the associated waveform display provided by Audacity 3.1 or PHON software, Version 2.1.8 (2016). Probe responses were transcribed by research assistants with graduate level training in clinical phonetics. After transcription, the first author scored the probe items as correct or incorrect according to the scoring protocol that was individualized for each child and goal at the time of goal selection. For example, as explained above, the scoring protocol indicated that correct productions of probe items by TASC02 for the first goal must include the word-internal and word-final coda, all three syllables, and the correct stress pattern. Therefore, during probe administration, the productions from this child that received scores of 1 were "garbageman" → [ˈdaɚbədʒˈmæn], "empty box" → [ˈɛmtiˈbotsʲ], and "patchwork" → [ˈpætʃˌwɝt], whereas these productions received scores of 0: "unpack it" → [ˈʌˈpʌnˌtɪt], "batmobile" → [ˈbæˈmoˈbi], and "computer" → [ˈtuˈbudə].

Common Intervention Procedure

Treatment sessions were conducted by student SLPs who were completing a second-year practicum or their final internship at McGill University. The students were assigned to this research clinic 1 or 2 days per week, but the children attended the clinic 3 days per week, and the treatment conditions were assigned at random to those days. Therefore, the children experienced two or three different student SLPs during the 6-week intervention, and the student SLPs provided all three treatment conditions according to the random schedule. All of the children with two exceptions were treated at a rehabilitation hospital, which also contained a school from which four of the participants were recruited. In this setting, all members of the research team were in the treatment room together: The student SLP and the child worked at a small table or in front of a mirror; the clinical educator (usually the second author) observed alongside an undergraduate student research assistant who recorded the session with a SONY HDR-XR150 handy camera; and the child's parent or parents, if attending, observed the session from inside the room as well. TASC30 and TASC34 were treated at our McGill University clinic site in a treatment room that was equipped with a two-way mirror. In this case, the student SLP, the child, and the research assistant who was video-recording the session were in the treatment room together. The clinical educator and the child's family observed from behind the two-way mirror. The room was equipped with a two-way sound system so that the clinical educator and student SLP could communicate with each other.

Each treatment session consisted of the same four events: (a) ND probes, 2–5 min; (b) prepractice, 20 min; (c) practice, 20 min; and (4) SD probes, 2–5 min. At the beginning of each session, the child was presented with a unique pathway picture that described for the child and the student SLP the assigned events for that day; for example, a puppy, four paw prints, and a dog house might be shown on the page with each paw print annotated to indicate the four activities for that session. During TASC02's second session, the pathway was annotated to indicate that the four activities would be an /l/-cluster ND probe, PMP prepractice for the dog names "Kip" and "Goop," practice /k, g/ target words, and, finally, /k, g/ SD probe. The child would stamp each paw print after each activity was completed.

The student SLP would create lesson plans for each session to be consistent with the assigned target and treatment approach for that day and implement the plan after approval from the second author. The prepractice and practice portions were planned to be each 20 min in duration. The target number of practice trials was intended to be at least 100 trials during all sessions. The practice procedures were roughly similar during every session. The prepractice procedures were determined by the random allocation procedure, which was conducted by the first author prior to the child's first session at the time that the treatment goals were selected and assigned to conditions. Specifically, www.randomizer.org was used to generate six blocks of three numbers in an independent random order for each block. These numbers determined the sessions that would be assigned to treatment conditions: For example, this procedure resulted in the following order of treatment assignments for TASC02 by week: CTL PMP AMI | AMI PMP CTL | CTL AMI PMP | PMP CTL AMI | CTL PMP AMI | PMP AMI CTL. These random sequences were generated independently for each child so that the intervention protocol for each child constitutes an independent experiment.

During sessions assigned to the CTL condition, no prepractice activities were assigned to prepare the child for successful practice of that day's target words. Rather, certain assessment procedures were assigned to take place during the first 20 min of the session, in order by week as follows for each child: free speech sample with a picture book (this is in addition to the connected speech sample that is part of the DEAP conducted during the intake assessment); SAILS speech perception test; phonological awareness test; maximum performance tests; KBIT if the child did not complete this during the intake assessment and if the child was old enough to undertake this test; and, finally, an oral-mechanism examination. If these assessment activities did not require the full 20 min, the extra time would be used with an activity not requiring speech such as a puzzle or drawing on an iPad.

In all three interventions, the procedures were designed to teach the child five pseudowords that were paired with objects or cutout laminated pictures, grouped into themes for each goal. For example, in the monster theme, "Biftenope" and "Hapnidreem" are monsters who like "to upteen/upteening" (walk on the ceiling); in the flower theme, "blees" and "ploon" are alien flowers that can be dug up with a "plinter" (a special tool); and in the dog theme, "Goop" and "Kip" play with a "koot" (favorite toy). The student SLPs were encouraged to teach two words the first week and two more the second week and introduce a fifth word during the third week while moving the children to phrase level material as soon as possible. Given the goal of moving the children toward the production of phrases, the word sets were designed to include subject nouns, object nouns, verbs, and, sometimes, adjectives. Some word sets included real but rare words in order to support authentic theme-relevant conversation, for example, "stake" in relation to planting the flowers.

AMI Prepractice Procedure

In the AMI condition, four procedures were implemented with amplification of the student SLP's speech to the child who was wearing headphones. These activities had the goal of improving the child's acoustic–phonetic representation of each target word and helping the child integrate auditory and somatosensory feedback while practicing production of that target. Auditory bombardment involved repeated production of the target words, presented in the context of a story that the student SLP narrated with the support of props (i.e., the objects that would be used later during practice). Subsequently, the student SLP taught the child to engage in an error detection task involving the target words for that session. These error detection tasks were designed so that the SLP would present the words at an appropriate level (i.e., single word or phrase, blocked or variable), and the child would be instructed to perform unique actions to signal correct versus incorrect productions. For example, if the student SLP called "Biftenope" → [ˈbɪftəˈnop], the child should move the dog to the food bowl; however, if the student SLP called "Biftenope" →

[ˈbɪdəˈnop], the child should make the dog lie down on its bed. The next activity would be focused on stimulation providing the child an opportunity to attempt the target words and receive appropriate feedback (recast, imitation, or expansion) from the student SLP. The child would be explicitly taught to monitor and correct his or her own productions after he had successfully learned to identify errors produced by the student SLP.

PMP Prepractice Procedure

In the PMP prepractice condition, four procedures were implemented to help the child construct a phonological plan and implement the plan independently during production practice without receiving prior models from the SLP. First, the student SLP provided a visual cue for each phoneme in the target word, consisting of an index card illustrated with a Bigmouth sound figure (www.voxaux.com) and the appropriate orthographic referent. For example, when introducing the word "blees," it would be explained that the word consisted of four sounds, namely, "b," "l," "ee," and "s." The card for the last sound would be illustrated with a snake-like figure and the letter "s." The Bigmouth sound figures provide information about articulation cues and/or sound cues for each English phoneme. For example, the figure for [p] shows lips closed while holding a popped balloon. The student SLP would then demonstrate the articulatory features of each sound to the child in a mirror while placing each card on the mirror in order and asking the child to imitate the sounds one at a time, using placement cues as necessary to elicit a correct production. Then, the student SLP would teach the child to chain the sounds sequentially to recreate the target word: [s], [is], [lis], and [blis], to recreate "blees" all the while moving the appropriate visual cues together in the right sequence and with the picture of the flower referent for "blees" visible on the mirror throughout the activity. Finally, the clinician and child would move to a small table to engage in a drill-play activity that involved spontaneous use of the day's target words, for example, naming the two flowers, "blees" and "ploon," correctly before planting them in a garden. This naming activity would proceed at a slow pace with the child using the visual cue cards to achieve the correct response if he or she had forgotten the name or was producing it incorrectly. In keeping with the principles of motor learning, feedback was largely about performance during the prepractice portion of the section, with knowledge of results feedback preferred during the practice portion.

Practice Procedure

Practice procedures were roughly the same in all three conditions except that, in the two experimental conditions, the student SLP would be able to choose a starting point for therapy based on prepractice performance and the child would be encouraged to use strategies learned during prepractice to support effective practice. In all three conditions, the clinicians were asked to help the child complete approximately 100 practice trials at the challenge point according to the principles described in Rvachew and Brosseau-Lapré (2018). Specifically, target words were practiced in five trial increments, and then a decision was made whether to change a practice parameter based on the child's performance with a correct response scored on each trial when the child's production completely matched the target. If the child's performance was 80% or better, a practice parameter was changed to increase difficulty. If the child's performance was worse than 80%, a practice parameter was changed to decrease difficulty. During the first few weeks, the student SLP would consult with the clinical educator (i.e., the second author) frequently about which practice parameter to change, but as the clinician gained experience and confidence, he or she would make decisions independently. The practice parameters that could be changed included (a) pretrial stimulation, according to the integral stimulation hierarchy (Strand & Debertine, 2000; Strand et al., 2006)[4]; (b) the target itself, meaning that it could be simplified to a part word or segmented target or it could be made more complex by inserting it into a phrase or sentence frame; (c) variability or consistency in the practice schedule (Preston, Leece, McNamara, & Maas, 2017); or (d) feedback frequency (Maas, Butalla, & Farinella, 2012). The decision was left open to the treatment team to decide which parameter to modify, keeping in mind that the goal was to keep the child at the challenge point while increasing the complexity of the practice material to the extent possible. For example, during TASC02's first session, the word "blees" was introduced as the practice target at the delayed imitation level yielding one of five correct responses prompting the team to change the pretrial stimulation parameter to imitative model plus coproduction, which in turn yielded three of five correct responses and then five of five correct responses prompting a return to delayed imitation. Later in the session, a second word was introduced, permitting a variable (rather than blocked) practice schedule, and subsequently, practice with phrase level stimuli was achieved.

Practice procedures were somewhat more structured during CTL practice sessions, although with similar goals, that is, maintaining a high rate of practice at the challenge point and moving the child to more complex practice stimuli as soon as possible. The student SLPs were provided with 120 picture cards to support drill practice, with target complexity level determined prior to each practice session. Feedback rate was maintained at a 60% rate by marking the backs of the cards to indicate whether the clinician should provide feedback or not after each given trial. Feedback was verbal for three out of five randomly selected trials, and then tangible reinforcement was provided after every fifth trial (e.g., puzzle piece to put in puzzle). The child

[4]The highest step in the integral stimulation hierarchy was an imitative prompt followed by co-production by child and therapist. Strand et al. (2006) describe adding tactile and gestural cues above this level if necessary to support accurate production of target items by the child. In our study, target items were constructed of stimulable phonemes, and therefore we did not find it necessary to add tactile cues as a means of stimulating a correct response.

was led through the steps: (a) imitative variable words, (b) spontaneous variable words, (c) imitative variable phrases, (d) spontaneous variable phrases, (e) imitative variable sentences, and (f) spontaneous variable sentences, with movement from one step to another based on criterion level performance of at least 80% correct averaged over 120 trials at each step. If the child's performance fell below 40% correct, responding substepping to a blocked practice schedule was implemented. This structured procedure was meant to be roughly similar to what should happen during the AMI and PMP practice sessions, but more flexibility was allowed during those sessions in order to achieve high-intensity practice at the challenge point while encouraging the child to use strategies that proved successful during prepractice. Procedures are also described in the Procedure Manual that is included in Supplemental Material S9.

Follow-Up

Children returned to the clinic at a time convenient to the family during the subsequent academic term for a follow-up assessment. At that time, the probe assessments for the three treatment goals were repeated, and the single-word naming test of articulation was repeated.

Blinding

Separating the treatment team from the research team served to blind the treatment team and the data analysis teams to the child's diagnosis and thus the specific hypothesis for each child. The treatment team, coordinated by the second author, was responsible for collecting assessment and probe data, providing the interventions, and writing clinical reports. The first author selected the targets, randomly assigned them to treatment conditions, and randomly assigned treatment conditions to sessions. These tasks occurred prior to the analysis of the assessment data (other than clinician transcription of DEAP responses), and therefore there was no diagnosis to bias these tasks. The first author coordinated teams of research assistants who transcribed all data that were recorded by the treatment team, ultimately deriving the diagnosis and probe scores for each child. All probe item transcriptions were completed by research assistants who were blind to the treatment condition and the child's diagnosis. Blinding was accomplished by labeling all recordings by participant number and session number and keeping all clinical files and research data in separate locations (the research clinic was in a rehabilitation hospital, and the laboratory where the research data were coded by a separate team was at the university some distance away). Because the process of transcribing data and deriving the relevant summary data stretched one to two academic terms beyond the actual treatment of the child, it was possible to complete these tasks in parallel but independently of each other and independent of the treatment process.

Reliability

The interrater reliability for transcription of the session and ND probes was determined by comparing point-by-point agreement for transcription of all the consonants in 10% of all the probes, scored by two independent transcribers who were blind to the child's diagnostic grouping, as well as session number and treatment condition associated with the probe. Additionally, 10% of probes were rescored independently for correctness of each item according to the scoring protocol by research assistants who were blind to treatment assignment and child diagnosis. Transcription reliability proved to be 81% agreement, which was similar to other treatment studies involving children with CAS (e.g., Ballard, Robin, McCabe, & McDonald, 2010; Maas et al., 2012; Preston, Brick, & Landi, 2013). Item-by-item agreement for scoring of probe correctness was 95%.

Fidelity

Treatment fidelity components included (a) a 3-day training workshop for interns and clinical educator that preceded each new cohort of participants, consisting of lectures, video demonstrations, and role play of all treatment and assessment procedures; (b) application of prepractice and practice procedures according to a detailed procedure manual; (c) close supervision of student SLPs to ensure that lesson plans conformed to the study protocol prior to implementation as well as close monitoring of treatment sessions to ensure that lesson plans were implemented in accordance with the study protocol; and (d) video recording of 100% of treatment sessions, with (e) follow-up coding of 10% of prepractice video recordings to document compliance with the randomly assigned treatment protocol and (f) coding of 100% of practice video recordings to document student SLP manipulation of practice parameters.

With respect to treatment fidelity during the experimental prepractice sessions, video review by the second author of the AMI condition revealed the use of auditory bombardment, focused stimulation, error detection, and self-monitoring tasks in 86%, 59%, 82%, and 23% of videos, respectively. Focused stimulation was sometimes omitted due to time constraints, whereas activities to teach self-monitoring were often omitted during the early weeks of the study because the child was struggling to achieve correct identification of errors produced by the student SLP. Nonetheless, across the six sessions implemented in the AMI condition, the combination of monitoring lesson plans, live sessions, and video recordings confirmed that all children experienced all four procedures. Furthermore, procedures from the PMP intervention protocol were never implemented during AMI prepractice sessions (note, however, that practice of part-word stimuli, i.e., segmenting and chaining targets, was allowed as a strategy to ensure practice at the challenge point during the practice portion of the session). In the PMP condition, video review revealed the use of visual cues with segmentation of the target word, cued articulation of each phoneme in the word, chaining to reconstruct the word, and spontaneous naming tasks in 90%, 81%, 81%, and 40% of videos, respectively. Procedures from the AMI intervention protocol were never implemented during PMP prepractice sessions. However, the student SLPs sometimes had difficulty refraining from providing imitative models when asking the children to name the target items. Sessions were closely monitored, and the clinical educator intervened quickly to ensure that the naming procedure was implemented correctly during prepractice sessions.

Regarding the practice portion of the session, there are two indicators of treatment fidelity: First, the child was expected to achieve at least 100 practice trials, and second, the SLP was expected to help the child maintain practice at the "challenge point," meaning a performance level that was neither too high nor too low. The mean number of practice trials per session overall was 106 (SD = 23), and the mean percent correct practice trials per session was 57 (SD = 11). For those children with a motor planning deficit, the number of trials and percent correct responses on average per session were 104 and 61% in the AMI condition, 96 and 38% in the CTL condition, and 94 and 58% in the PMP condition, respectively. For those children with a phonological planning deficit, the number of trials and percent correct responses on average per session were 118 and 61% in the AMI condition, 141 and 61% in the CTL condition, and 97 and 78% in the PMP condition, respectively. Regarding the use of specific strategies by treatment condition, it is notable that the student SLPs were 2.5 times more likely to elicit spontaneous productions from the children in the PMP condition than they were in the AMI condition, indicating that they made an effort to avoid imitative models in this condition when possible while maintaining the child's performance level at challenge point. In contrast, the student SLPs used co-production four times more often in the AMI condition than they did in the PMP condition. Variable practice (in which different words were practiced within the same block of five trials, as opposed to constant practice in which the same word was produced five times in succession) was twice as frequent in CTL and PMP sessions compared to AMI sessions. Session level data are presented in the supplemental materials (Supplemental Materials S5 and S6) for each participant and indicate that the student SLPs successfully achieved a high intensity of practice; furthermore, the children were practicing almost at challenge point on average. The exceptions were two children who achieved very few correct productions of specific phoneme targets (TASC23 [ʃ] and TASC30 [f, v]) throughout the 6-week treatment protocol resulting in a low percent trials correct in the CTL condition.

Results

The results will be presented separately for each diagnostic group, but the analysis strategy was the same for the two groups of children. Recall that the first hypothesis was that each child would achieve a differential response to treatment. This hypothesis was tested by calculating the sums of squares of the differences between the means in each treatment condition and the grand mean: This is the observed test statistic. A distribution of this statistic $\sum(\text{M-GM})^2$ was created by random shuffling of the observed data within blocks (weeks) 10,000 times. The proportion of statistics (i.e., $\sum(\text{M-GM})^2$ calculated from the reshuffled data in this distribution) that were equal to or greater than the observed test statistic was taken as the p value and evaluated against alpha, .05/2 = .025, to take into account two tests (one for the SD probe scores and another for the ND probe scores). A significant result for any given child suggests that the child obtained a superior result in at least one treatment condition relative to the other(s). These p values were then pooled using the formula provided by Edgington (1972) to assess whether the effect was reliable when the experiment was replicated across all the participants with the same diagnosis.

The second and third hypotheses were that children in the two diagnostic subgroups would show a different pattern of results: Specifically, children with a motor planning deficit as a group should respond best to the AMI intervention, and children with a phonological planning deficit as a group should respond best to the PMP intervention. Therefore, two planned comparisons were conducted using the same resampling method, but in this case, the test statistic was the difference between means. Specifically, for the group with a motor planning deficit, the tests were AMI M – CTL M and AMI M – PMP M; for the group with a phonological planning deficit, the tests were PMP M – CTL M and PMP M – AMI M. The multiple comparisons have less power than the overall test of an effect because there are fewer permutations for each test (i.e., $(3!)^6$ = 46, 656 permutations for the overall test vs. $(2!)^6$ = 64 for the multiple comparisons). Pooling the p values across the replications of the experiment improves the power. Therefore, each hypothesis was assessed for each child and then by pooling p values across all the children with the same diagnosis. The pooled p values will always be smaller for the phonological planning deficit group than the motor planning deficit group because there are fewer p values to pool. The formula for pooling the p values is strongly impacted by the number of p values in the pool. The result is also strongly impacted by the sum of the p values. The test statistic (difference between means) and the resampling test that was conducted were one sided so that, when the difference between means was not in the expected direction, the individual p value would be very large, increasing the size of the pooled p value. In summary, the pooled p values reflect the size of the effects, consistency of the direction of the effects with the hypothesis, and the number of p values in the pool. The p values cannot be compared across the motor planning and phonological planning groups because the size of these groups is not the same. Note that the raw data are shown in Supplemental Materials S7 and S8.

Motor Planning Group

Table 3 shows the mean and standard deviation of the probe scores by treatment condition for each child in this group along with the resulting test statistic and p value. A significant result was obtained for five of seven children when examining SD probe scores and for all seven children when examining ND probe scores. The pooled p for the SD probes is 3.70E-07, and the pooled p for the ND probes is

Table 3. Mean (*M*) and standard deviation (*SD*) of same-day and next-day probe scores by treatment condition for children with a motor planning deficit.

	AMI		CTL		PMP		Test statistic	
Participant	*M*	*SD*	*M*	*SD*	*M*	*SD*	$\Sigma(M\text{-}GM)^2$	*p*
Same-day probe scores								
TASC02	7.50	1.04	5.83	1.47	6.33	2.87	1.46	.3493
TASC04	3.50	1.38	0.33	0.51	2.33	1.86	5.13	**.0068**
TASC16	4.33	0.66	3.00	0.77	0.83	0.47	6.24	**.0186**
TASC23	7.33	1.21	1.50	1.76	6.50	1.38	19.91	**.0022**
TASC24	5.00	1.26	3.00	1.90	0.33	0.82	10.96	**.0006**
TASC30	3.17	1.08	0.00	0.00	2.83	1.47	6.06	**.0060**
TASC34	4.00	1.79	2.83	1.17	1.17	1.33	4.06	**.0239**
Pooled *p* values from the same-day probes								**.0000**
Next-day probe scores								
TASC02	7.17	1.19	5.67	2.25	0.50	0.83	24.46	**.0018**
TASC04	4.00	1.50	0.00	0.00	1.50	1.50	8.17	**.0003**
TASC16	4.33	0.61	2.50	0.71	0.00	0.00	9.46	**.0020**
TASC23	5.17	1.94	1.33	1.86	6.50	1.38	14.39	**.0067**
TASC24	6.17	1.17	3.33	1.63	0.33	0.52	17.02	**.0003**
TASC30	4.42	1.31	0.50	0.83	3.67	1.75	8.64	**.0072**
TASC34	1.33	1.03	3.00	1.26	0.00	0.00	4.52	**.0104**
Pooled *p* values from the next-day probes								**.0000**

Note. Significant *p* values in bold. AMI = auditory–motor integration condition; CTL = usual care control condition; PMP = phonological memory and planning condition.

3.18E-15. Therefore, it cannot be concluded that probe scores are independent of the randomly allocated treatment conditions.

Planned comparisons for children with a motor planning deficit are shown in Table 4. For each comparison, the raw effect size (i.e., mean difference) is shown along with the confidence interval (CI), followed by the associated *p* value and standardized effect size. The effect sizes for paired samples and the CIs were calculated using the ESCI calculator provided by Cumming and Finch (2005; https://

Table 4. Results of multiple comparisons for children with motor planning deficits.

Participant	AMI–CTL	CI	*p*	*d*	AMI–PMP	CI	*p*	*d*
Same-day probe scores								
TASC02	1.67	[–0.29, 3.62]	.06	1.30	1.17	[–1.35, 3.69]	.26	0.54
TASC04	**3.17**	[1.94, 4.40]	**.02**	3.04	1.17	[–0.98, 3.30]	**.02**	1.17
TASC16	1.33	[–1.03, 3.70]	.14	0.75	**3.50**	[1.54, 5.46]	**.02**	2.47
TASC23	**5.83**	[4.03, 7.64]	**.02**	3.86	0.83	[–1.59, 3.26]	.31	0.64
TASC24	**2.00**	[0.85–3.15]	**.02**	1.24	**4.67**	[2.83, 6.50]	**.01**	4.38
TASC30	**3.17**	[2.03, 4.30]	**.02**	4.15	0.33	[–2.16, 2.83]	.38	0.26
TASC34	1.17	[–1.44, 3.77]	.19	0.77	**2.83**	[0.59, 5.07]	**.01**	1.79
Pooled *p* values AMI–CTL: **.0000**					Pooled *p* values AMI–PMP: **.0002**			
Next-day probe scores								
TASC02	1.50	[–2.06, 5.06]	.22	0.71	**6.67**	[4.72, 8.62]	**.02**	4.46
TASC04	**4.00**	[2.37, 5.63]	**.02**	3.65	**2.50**	[0.77, 4.22]	.15	1.89
TASC16	1.83	[–0.69, 4.35]	.09	1.12	**4.33**	[2.75, 5.91]	**.02**	4.07
TASC23	**3.83**	[0.23, 7.43]	.03	2.02	–1.33	[–3.39, 0.73]	.97	–0.79
TASC24	**2.83**	[0.50–5.17]	.03	2.00	**5.83**	[4.29, 7.38]	**.02**	6.46
TASC30	**3.92**	[3.01, 4.82]	**.01**	3.55	0.75	[–2.15, 3.66]	.28	0.48
TASC34	–1.67	[–3.93, 0.60]	.97	–1.44	**1.33**	[0.24, 2.41]	.03	1.82
Pooled *p* values AMI–CTL: **.0018**					Pooled *p* values AMI–PMP: **.0030**			

Note. Bolding indicates raw effect sizes that are associated with a CI that does not contain 0 and significant *p* values. AMI = auditory–motor integration condition; CTL = usual care control condition; CI = confidence interval; *d* = standardized effect size; PMP = phonological memory/planning condition.

thenewstatistics.com/itns/esci/). Raw effect sizes that are associated with a CI that does not contain 0 are bolded as are *p* values that are smaller than .0125. The results indicate that four children obtained higher probe scores in the AMI condition compared to the CTL condition when examining both SD and ND probe scores, although the *p* values for three of these individual participants are not significant on the next day. Both pooled *p* values for this comparison are significant, however, at 7.44E-07 for SD probes and 1.80E-03 for ND probes. When comparing probe scores in the AMI condition to those obtained in the PMP condition, effect sizes are relatively large for three children's SD probes and five children's ND probes, yielding significant pooled *p* values, specifically 2.01E-04 for SD probes and 2.97E-03 for ND probes.

The effects are in the expected direction with two exceptions: TASC23 obtained similarly large probe scores in the AMI and PMP conditions for both SD and ND probes; TASC34 showed a decline in performance on AMI probes between the SD and ND probe interval, resulting in a negligible advantage to the CTL condition on ND probes. TASC30 probe scores were highest in the AMI condition, but the difference was significant in comparison to the CTL condition only. The remaining children obtained noticeably larger AMI SD and ND probe scores in comparison to the CTL and PMP probe scores. In fact, the advantage to the AMI condition over the PMP condition was often larger when the children were assessed at the beginning of the next treatment session.

Phonological Planning Group

Table 5 shows the mean and standard deviation of the probe scores by treatment condition for each child in this group along with the resulting test statistic and *p* value. A significant result was obtained for four of five children when examining SD probe scores and for all five children when examining ND probe scores, resulting in pooled *p* values of 1.10E-09 and 1.12E-12, respectively. Therefore, it cannot be concluded that probe scores are independent of the randomly allocated treatment conditions.

Planned comparisons for children with a phonological planning deficit are shown in Table 6. When examining SD and ND probe results for the PMP versus CTL comparison, it can be seen that the direction of effect across children was unpredictable: Three children obtained the expected benefit to the PMP condition, whereas two children achieved better scores in the CTL condition. Given these inconsistent effects, the pooled *p* values for this comparison are not significant, specifically 3.03E-01 and 3.02E-01 for the SD and ND probes, respectively. When examining the SD and ND probe results for the PMP versus AMI condition comparison, four children show the expected effect of a large comparative benefit to the PMP condition, whereas one child achieved a better score in the AMI condition. Overall, the pooled *p* values were significant for this comparison: 1.22E-02 for SD probes and 1.32E-02 for ND probes. These results suggest that the null hypothesis can be rejected for the PMP versus AMI comparison; that is, it is not likely that the probe scores are independent of the randomly allocated treatment conditions.

Follow-Up Data

Follow-up assessments were obtained for 10 of the 12 participants. Four children increased their probe scores when compared to probe performance during the treatment phase of the experiment. Substantial increases in PCC during single-word picture naming were observed for all but

Table 5. Mean (*M*) and standard deviation (*SD*) of same-day and next-day probe scores by treatment condition for children with a phonological planning deficit.

	AMI		CTL		PMP		Test statistic	
Participant	*M*	*SD*	*M*	*SD*	*M*	*SD*	$\sum(M\text{-}GM)^2$	*p*
Same-day probe scores								
TASC05	4.67	2.16	7.33	2.42	9.67	0.82	12.52	**.0004**
TASC09	2.83	1.16	8.00	1.67	7.83	1.16	17.24	**.0135**
TASC20	6.00	2.28	3.50	1.97	8.67	1.21	13.35	.0257
TASC21	2.33	1.63	6.00	1.89	1.33	1.36	12.07	**.0018**
TASC26	0.17	0.40	0.83	0.75	8.50	1.37	42.89	**.0007**
Pooled *p* values from the same-day probes								**.0000**
Next-day probe scores								
TASC05	3.00	1.55	8.17	1.17	9.67	0.52	24.46	**.0027**
TASC09	3.67	0.33	9.67	0.21	7.67	0.81	18.67	**.0001**
TASC20	5.83	2.78	5.00	1.54	9.00	0.89	8.91	**.0073**
TASC21	2.67	1.96	6.00	1.54	0.50	0.83	15.35	**.0007**
TASC26	0.00	0.00	1.17	0.75	7.00	1.41	28.13	**.0001**
Pooled *p* values from the next-day probes								**.0000**

Note. Significant *p* values in bold. AMI = auditory–motor integration condition; CTL = usual care control condition; PMP = phonological memory and planning condition.

Table 6. Results of multiple comparisons for children with phonological planning deficits.

Participant	PMP–CTL	CI	*p*	*d*	PMP–AMI	CI	*p*	*d*
Same-day probe scores								
TASC05	**2.33**	[0.50, 4.17]	.03	1.29	**5.00**	[3.12, 6.88]	**.02**	3.06
TASC09	–0.17	[–1.71, 1.38]	1.00	–0.12	**5.00**	[3.67, 6.33]	**.02**	4.28
TASC20	**5.17**	[3.03, 7.31]	**.02**	3.15	**2.67**	[–1.67, 7.00]	.11	1.08
TASC21	–4.67	[–5.52, 3.81]	1.00	–2.82	–1.00	[–2.88, 0.88]	.92	–0.66
TASC26	**7.67**	[5.83, 9.50]	**.01**	6.90	**8.33**	[7.06, 9.60]	**.02**	8.20
Pooled *p* values PMP–CTL: .3028					Pooled *p* values PMP–AMI: **.0122**			
Next-day probe scores								
TASC05	**1.50**	[0.05, 2.95]	.03	1.66	**6.67**	[5.40, 7.94]	**.02**	5.77
TASC09	–2.00	[–2.66, 1.34]	1.00	–2.93	**4.00**	[4.00, 4.00]	**.08**	—
TASC20	**4.00**	[2.85, 5.15]	**.02**	3.16	**3.17**	[0.32, 6.02]	.05	1.53
TASC21	–5.50	[–6.95, 4.05]	1.00	–4.42	–2.17	[–4.09, 0.24]	1.00	–1.43
TASC26	**5.83**	[3.80, 7.87]	**.02**	5.15	**7.00**	[5.52, 8.48]	**.02**	7.00
Pooled *p* values PMP–CTL: .3022					Pooled *p* values PMP–AMI: **.0132**			

Note. Bolding indicates raw effect sizes that are associated with a CI that does not contain 0 and significant *p* values. PMP = phonological memory/planning condition; CTL = usual care control condition; CI = confidence interval; *d* = standardized effect size; AMI = auditory–motor integration condition.

two children who maintained their overall PCC scores (see Table 7).

Discussion

Children were recruited to this study by asking SLPs to refer children who presented with a severe SSD characterized by inconsistent speech sound errors. Ultimately, seven children were excluded from participation because their profile was consistent with phonological or articulation disorder, leaving 12 children who produced inconsistent errors and scored below age expectations on the SRT. Of these children who were admitted for participation in this study, seven were observed to produce transcoding errors on the SRT, whereas the remaining five had difficulties with PMP but did not produce four or more transcoding errors on the SRT. These two groups of children were judged to have motor planning deficits in the former case and phonological planning deficits in the latter case, using the diagnostic process described in Rvachew and Matthews (2017b) and the rationale described by Shriberg et al. (2012). It was hypothesized that children with a motor planning deficit would respond best to an AMI intervention, whereas children with a phonological planning deficit would respond

Table 7. Follow-up probe scores and percent consonants correct (PCC) in single words by participant and group.

	Follow-up probes			PCC-single words	
Participant	AMI *M*	CTL *M*	PMP *M*	Intake	Follow-up
Motor planning group					
TASC02	1.00	0.00	1.00	60.00	*58.00*
TASC04	*4.00*	0.00	**5.00**	34.00	**50.00**
TASC16	nd	nd	nd	nd	nd
TASC23	nd	nd	nd	nd	nd
TASC24	**9.00**	1.00	**8.00**	66.00	**94.00**
TASC30	**6.50**	*1.00*	2.00	37.00	*41.00*
TASC34	0.00	*3.00*	0.00	49.00	**57.00**
Phonological planning group					
TASC02	*4.00*	0.00	0.00	61.00	**85.00**
TASC04	*4.00*	*10.00*	*8.00*	51.00	**76.00**
TASC16	*6.00*	*5.00*	*8.00*	82.00	**92.00**
TASC23	*3.00*	*6.00*	0.00	61.00	**68.00**
TASC24	0.00	*2.00*	*8.00*	55.00	*56.00*

Note. Those in standard font indicate decline or floor level performance, those in italics indicate maintenance, and those in bold indicate improvement of more than 1 *SE* interval. AMI = auditory–motor integration condition; CTL = usual care control condition; PMP = phonological memory/planning condition; nd = no data.

best to an intervention that focused on phonological planning. The hypotheses were supported for the motor planning group in full and for the phonological planning group in part, as will be discussed further below.

Motor Planning Group

In keeping with the hypothesis of weak forward models (Terband et al., 2014), the prepractice procedures used in the AMI treatment condition were intended to strengthen the children's knowledge of the auditory target and their ability to monitor and respond to intrinsic feedback while engaged in speech practice. Previous studies have shown that speech perception training to ensure knowledge of the auditory target supports appropriate responses to auditory feedback during speech practice by children (Shiller & Rochon, 2014). In accordance with our first hypothesis, every child judged to have a motor planning deficit showed a significant response overall to the treatment conditions, with the same pattern of mean SD and/or ND probe scores: Specifically, the mean probe scores (as shown in Table 3) were higher in the AMI condition when compared to the CTL and PMP conditions. The size and significance of this pattern of results (response to the AMI treatment in comparison to the PMP and CTL conditions) were assessed via planned comparisons for the group of children with a motor planning deficit, as shown in Table 4, revealing statistically significant results via the pooled p values. The effect sizes showed that the advantage to the AMI condition relative to the PMP condition was large for five of the seven children upon next-day testing. The remaining two children showed a large response to the AMI intervention and the PMP intervention in comparison to the CTL condition. These unexpected results may have occurred because difficulties with PMP overlap with motor planning deficits, as shown in Table 2. On the other hand, some departures from expectations are expected in the results due to chance events. Despite the exceptions for these two children, it is clear that the results of these experiments overall (i.e., as indicated by the pattern of pooled p values) support the hypothesis that the children who had transcoding deficits as identified by the SRT would obtain higher scores in response to the AMI intervention in comparison to the CTL and PMP treatment conditions.

Comparison to Other Motor Learning Approaches

Maas et al. (2014) concluded that the integral stimulation approach is supported by the most number of controlled studies currently used to treat CAS. The AMI approach shares characteristics with the integral stimulation approach in that high-intensity practice was implemented with rapid adaptation of support provided to the child in a dynamic fashion in response to the child's trial-to-trial performance. Care was taken to provide the least amount of stimulation required prior to the child's production attempt and to withhold clinician feedback about the accuracy of production attempts until the end of each five-trial practice block as long as the child could maintain challenge point performance with minimal external feedback. One significant difference was the use of pseudoword stimuli, but this change does not seem to have interfered with the impact of the intervention on children's speech learning because all of the children demonstrated a significant treatment result and transferred learning to real words. Furthermore, the children maintained that learning over short intervals, and some children maintained performance over the longer term to the follow-up assessment interval that took place a month or two after completion of the treatment protocol.

Several approaches to treatment for children with CAS incorporate tactile cues (Dale & Hayden, 2013; Strand & Skinder, 1999). No tactile cues were used in any of the treatment conditions employed here with any of the children. The hypothesis suggested by Terband and Maassen (2010) in which weak forward models arise as a consequence of poor somatosensory feedback might be seen as a reason to provide tactile guidance during speech practice. As reviewed by Maas et al. (2014), however, the studies on this approach (e.g., Dale & Hayden, 2013) have not been conducted with sufficient experimental control to know whether these cues provide a useful substitute for self-produced sensory inputs. Another potential solution is to substitute visual feedback for degraded somatosensory feedback. Single-subject studies have yielded some encouraging results for this technique (Preston et al., 2013, 2017), but there is as yet no evidence that the ultrasound feedback itself is a necessary component of the treatment program or a more effective manipulation than, for example, the AMI approach implemented in this study.

One procedure used in the AMI intervention in this study was the focus on holistic word shapes and the requirement that the child produce the complete target accurately, including segmental content, syllable shape, and prosody. Furthermore, every child had one target that focused specifically on one or two stress patterns (e.g., weak–strong–weak vs. strong–weak–strong word shapes). These procedures are similar to those employed in the ReST program, recently validated in a randomized controlled trial (Murray et al., 2015). In our study, as in Murray et al., good transfer from pseudowords to real words was observed with a 9-point change in PCC reported in Murray et al. and a 12-point change in PCC obtained by the participants in this study. Murray et al. also reported that the ReST approach yielded results that were roughly similar to the Nuffield Dyspraxia Program (P. Williams, 2009), although longer term maintenance of treatment effects may have been better for children who received the ReST program. The Nuffield program that targets real words and phrases differs by taking a "bottom-up" approach to the development of increasing complexity in the acquisition of speech motor control. Overall, the message appears to be that, given sufficient intensity, children with CAS can make good gains from different approaches to speech therapy. However, marginal benefits may be observed if the primary underlying psycholinguistic deficits are targeted in a systematic fashion.

Phonological Planning Group

The PMP prepractice procedures were designed to make explicit the phonological plan for each specific target and to help the child compile the phonological plan independently during speech practice, using visual cues if necessary but without the aid of a prior auditory model. These procedures have been shown to be effective for children with IPD (Crosbie et al., 2005; Dodd & Bradford, 2000; McIntosh & Dodd, 2008; see also Moriarty & Gillon, 2006). The overall results indicated that all the children showed a differential treatment effect but the multiple comparisons were not all in the expected direction. Three children achieved a significantly better result in the PMP condition compared to the AMI condition, and the pooled *p* value confirmed that these children with PMP deficits overall responded more favorably to the PMP condition than to the AMI condition. However, the comparison of PMP condition to the CTL condition yielded unpredictable results across children, and ultimately, the difference between these interventions was not found to be statistically significant. The clear result in comparison to the AMI condition suggests that children who do not present with transcoding errors and other clear indicators of a motor planning disorder are in fact in need of a different treatment approach that is individualized to their specific profile of underlying psycholinguistic deficits. However, any conclusion beyond that is difficult to draw on the basis of the results reported here. There are at least three possibilities. One is that the PMP intervention as implemented here was not effective enough to meet the needs of children with a phonological planning deficit. Perhaps, functional core vocabulary targets and a home practice component are essential components of this approach (Crosbie et al., 2005). A second possibility is that the intervention as implemented here was effective with some of the children (as suggested by the results of the statistical analyses for TASC05, TASC20, and TASC26) but not others because of differences in their underlying psycholinguistic profiles. TASC09 and TASC21 produced very low encoding scores on the SRT and achieved unusually low accuracy at all three syllable lengths, whereas the more classic profile for children with a phonological memory deficit is to produce accurate repetitions of two-syllable items with a marked decline for three- and four-syllable items. It can be seen in Table 2 that TASC21 also obtained a dysarthria score of 2 because his single-syllable repetition rates were very slow. A third possibility is that the CTL intervention as implemented here is equally as effective as the PMP intervention for children with a phonological planning deficit. This is the least likely explanation because the pattern of results shown in Table 5 indicates considerable variability in responding across children rather than a consistently good response. Nonetheless, some practice parameters in the CTL condition might be helpful such as the random practice schedule. Although there were overlapping treatment procedures between the AMI condition and the other two conditions, the student SLPs tended to practice stimuli according to a blocked practice schedule in the AMI condition, and they were encouraged to provide many auditory models; under these practice conditions, the child may have had fewer opportunities to independently assemble the phonological plan for the target words.

Limitations

The limitations of this study include aspects of the research participants, the study design, the clinicians, and the interventions. These will be discussed in turn. The participants have some complex features, including delays in the verbal and nonverbal cognitive domains. Although some of the basic science literature on CAS has described children who are relatively uncomplicated (e.g., Grigos & Case, 2018; Peter & Stoel-Gammon, 2005), it is common for children recruited with CAS to have co-occurring disorders in multiple domains (e.g., Preston et al., 2013; Shriberg et al., 2012). This heterogeneity, despite the small size of this population, is not surprising given diagnostic criteria that focus on impaired speech movements regardless of the presence or absence of comorbid idiopathic or neurogenic disorders (American Speech-Language-Hearing Association, 2007). However, it creates something of a problem for clinical generalization from intervention trials to any given child on a clinical caseload. The pooled *p* values for the seven children who presented with a motor planning deficit in this study were consistent with the hypothesis of a superior treatment response to the AMI intervention relative to the CTL or PMP interventions, but the effect sizes varied widely from child to child. Therefore, further research is necessary to identify the limits to which this finding can be replicated and generalized. This difficulty is even more acute in the case of the PMP intervention, which was successful for only three of the five children who were judged to have phonological planning difficulties. Further to the question of whether children with the same SRT profiles that were described in this report will show similar treatment responses in future studies is the question of whether these diagnostic profiles are stable with age and treatment history. The short-term nature of this study provides no clues as to the long-term stability of these underlying deficits. It seems most likely to us that the children will change with treatment and that the children will require different approaches to intervention as they face new speech and language challenges as they grow older (for further discussion, see Rvachew & Brosseau-Lapré, 2018).

The single subject randomization design that was used in this study provides excellent control for threats to internal validity through randomization: In this case, treatment goals were randomly assigned to treatment conditions and treatment conditions were randomly assigned to treatment sessions using a block randomization schedule. On any given day, probe scores may be impacted by many different variables besides the intervention. In any study, history effects (events in the child's life beyond the treatment program) will be of concern as will be any variables that might cause a time trend in the probe scores (maturation, practice, and fatigue effects). In this study, specific concerns related to the treatment protocol arise, in particular, differences in the skill level of the student SLPs, which changed from day to day, and differences in the difficulty level of the treatment goals. The randomization procedure controls for the effect of these extraneous variables by allowing resampling tests and the pooling of the *p* values—tests that essentially conclude that, in the event that the AMI treatment was not more effective than the other conditions, it would be very unlikely to randomly allocate the AMI treatment condition to the easiest target and the best student SLP each week for 6 consecutive weeks and then again over seven consecutive participants. Therefore, it is parsimonious to conclude that the pattern of results reflects a treatment effect. Nonetheless, the presence of these known extraneous variables introduces some interpretative uncertainty for the phonological planning group. In this case, the pooled *p* value suggests that there is no difference in outcome for the PMP versus CTL condition. Therefore, when viewing the individual experiments, it is possible that the apparent effects in favor of CTL for TASC21 and in favor of PMP for TASC26 may in fact be due to differences in target difficulty for those participants. Unfortunately, it is impossible to control for target difficulty in any other way except through random assignment because, (a) even if all the participants are treated for the same goal (e.g., McAllister Byun, 2017), difficulty for each child will vary and (b) there is no accepted metric for ensuring equivalent difficulty of different targets within or across children except possibly by choosing unstimulable targets, which is not always wise (Rvachew & Nowak, 2001). In fact, our experience with the treatment approaches that we implemented here is that they are not effective when applied to unstimulable targets (e.g., see CTL target for TASC30, an unusual case in which we were forced to choose an unstimulable specific phoneme goal).

Further to the issue of skill level of the student SLPs, it is possible that the use of experienced SLPs to conduct the intake assessments would have led to different diagnostic decisions in some cases. In this study, student SLPs collected the assessment data under the supervision of the second author and student research assistants coded those data under the supervision of the first author, with both authors being experienced SLPs with expertise in motor speech disorders. However, the diagnosis of CAS is a difficult process that requires clinical judgment when observing the child's performance over a variety of speech and nonspeech tasks. In some cases, the student SLPs were unable to obtain the child's cooperation for completion of a task (TASC04 and TASC30); in others, the evidence that the child might have demonstrated some features consistent with subclinical dysarthria did not emerge for many months after the child had completed the treatment protocol (TASC16, TASC34, and TASC21). Despite this limitation, however, the diagnostic decisions made in this study did result in two groups of children that demonstrated a distinct pattern of treatment response, which supports the validity of the SRT as a useful and easily administered tool for differential diagnosis.

Finally, with respect to the interventions, any generalizability of the results reported herein is limited to the treatment protocol as implemented. The same results might not be obtained if a child with a motor planning deficit were to receive the AMI intervention once per week, targeting a single goal, rather than three times per week, targeting three goals. It is possible that the treatment targeting the other two goals plays a role in the success of any one approach. Certainly, there are many indications that intensity alone (in terms of ensuring both many practice trials per session and several treatment sessions per week) is critical for achieving transfer and maintenance for children with motor planning disorders (Edeal & Gildersleeve-Neumann, 2011; Maas et al., 2014; Namasivayam et al., 2015). A randomized controlled parallel-group design would be best suited to assessing differences in outcomes for different goal attack strategies (i.e., variable vs. constant practice).

Summary and Conclusions

Although it is clear that children with inconsistent speech errors benefit from intense speech practice implemented in accordance with the principles of motor learning, their learning can be supported with different prepractice approaches. In this study, an AMI approach to prepractice was implemented that included auditory bombardment, focused stimulation, error detection, and self-monitoring in order to prepare the child to use auditory feedback more effectively during practice. This approach was compared to a PMP approach to prepractice in which children were taught to segment the target pseudowords into their constituent phonemes, pair those phonemes with a visual cue, practice the phonemes singly, and then chain them back together to reconstruct the word. The children were encouraged to use these strategies to assemble their own phonological plans during spontaneous practice of the words during drill-play activities. Seven children who were judged to have motor planning deficits on the basis of an excessive number of transcoding (addition) errors on the SRT obtained the highest probe scores in the AMI treatment condition compared to the PMP condition and a CTL condition that involved speech practice with no prepractice. Five children who were judged to have phonological memory difficulties but no transcoding deficit responded best to the PMP treatment condition overall, although two of these children also responded well to the CTL condition. This study requires replication and extension to address limitations regarding the heterogeneous sample, aspects of the design, and the specific combination of treatment parameters that were employed.

Acknowledgments

This project was supported by the Childhood Apraxia of Speech Association of North America; the Natural Sciences and Engineering Research Council; the Ruth Ratner Miller Foundation; the Centre for Research in Brain, Language and Music; and the Centre de Recherche Interdisciplinaire en Réadaptation du Montréal Métropolitain.

We thank the children and their parents who participated in the research project. We are also grateful for assistance with recruitment provided by the MAB-MacKay Rehabilitation Centre and the Montréal Children Hospital. Additional thanks to staff, students, research assistants, and volunteers: Alexandra Barbeau-Morrison, Madeleine Bellemare, Olivia Bouskill, Emily Brady, Meagan Buttle, Elizabeth Carolan, Roxane Carrière, Tara Commandeur, Cassandra Cowie, Nicholas Cunningham, Raphaëlle Curis, Larissa Der, Marla Folden, Kalyna Franko, Emily Goodwin, Maggie Haughey, Danya Hiob, Rachel Hong, Roxanna Ibrahim, Mary Ann Lacey, Renata Masucci, Émalee Manns, Jillian Mills, Leah Modlin, Jeanie Morry, Elizabeth Murphy, Rebecca Nishimura, Amanda Ovadia, Sara Perillo, Alele Rangel, and Omar Obregozo Zalava.

References

American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech* [Tech. Rep.]. Retrieved from http://www.asha.org/policy, http://www.asha.org/docs/html/TR2007-00278.html#r8

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.

Ballard, K. J., Wambaugh, J. L., Duffy, J. R., Layfield, C., Maas, E., Mauszycki, S., & McNeil, M. R. (2015). Treatment for acquired apraxia of speech: A systematic review of intervention research between 2004 and 2012. *American Journal of Speech-Language Pathology, 24*(2), 316–337. https://doi.org/10.1044/2015_AJSLP-14-0118

Bernhardt, B. H., Stemberger, J. P., & Major, E. (2006). General and nonlinear phonological intervention perspectives for a child with resistent phonological impairment. *Advances in Speech-Language Pathology, 8,* 190–206.

Bradford, A., & Dodd, B. (1996). Do all speech-disordered children have motor deficits? *Clinical Linguistics & Phonetics, 10,* 77–101.

Campbell, T. F. (1999). Functional treatment outcomes in young children with motor speech disorders. In A. Caruso & E. A. Strand (Eds.), *Clinical management of motor speech disorders in children* (pp. 385–395). New York, NY: Thieme.

Crary, M. A. (1993). *Developmental motor speech disorders*. San Diego, CA: Singular.

Crosbie, S., Holm, A., & Dodd, B. (2005). Intervention for children with severe speech disorder: A comparison of two approaches. *International Journal of Language & Communication Disorders, 40*(4), 467–491.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*(2), 170–180.

Dale, P. S., & Hayden, D. (2013). Treating speech subsystems in childhood apraxia of speech with tactual input: The PROMPT approach. *American Journal of Speech-Language Pathology, 22,* 644–661.

Dodd, B. (2014). Differential diagnosis of pediatric speech sound disorder. *Current Developmental Disorders Reports, 1*(3), 189–196.

Dodd, B., & Bradford, A. (2000). A comparison of three therapy methods for children with different types of developmental disorder. *International Journal of Language & Communication Disorders, 35*(2), 189–209.

Dodd, B., Ttofari-Eecen, K., Brommeyer, K., Ng, K., Reilly, S., & Morgan, A. (2017). Delayed and disordered development of articulation and phonology between four and seven years. *Child Language Teaching and Therapy, 34,* 87–99. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/0265659017735958

Dodd, B., Zhu, H., Crosbie, S., Holm, A., & Ozanne, A. (2006). *Diagnostic Evaluation of Articulation and Phonology (DEAP)*. London, United Kingdom: Pearson Education.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–III (PPVT-III)*. 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*(2), 95–110.

Edgington, E. S. (1972). An additive method for combining probability values from independent experiments. *The Journal of Psychology, 80*(2), 351–363.

Edgington, E. S. (1987). Randomized single-subject experiments and statistical tests. *Journal of Counseling Psychology, 34,* 437–442.

Flinker, A., Korzeniewska, A., Shestyuk, A. Y., Franaszczuk, P. J., Dronkers, N. F., Knight, R. T., & Crone, N. E. (2015). Redefining the role of Broca's area in speech. *Proceedings of the National Academy of Sciences of the United States of America, 112*(9), 2871–2875.

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

Forrest, K., Dinnsen, D. A., & Elbert, M. (1997). Impact of substitution patterns on phonological learning by misarticulating children. *Clinical Linguistics & Phonetics, 11,* 63–76.

Forrest, K., Elbert, M., & Dinnsen, D. A. (2000). The effect of substitution patterns on phonological treatment outcomes. *Clinical Linguistics & Phonetics, 14,* 519–531.

Grigos, M. I., & Case, J. (2018). Changes in movement transitions across a practice period in childhood apraxia of speech. *Clinical Linguistics & Phonetics, 32*(7), 661–687.

Hesketh, A., Adams, C., Nightingale, C., & Hall, R. (2000). Phonological awareness therapy and articulatory training approaches for children with phonological disorders: A comparative outcome study. *International Journal of Language & Communication Disorders, 35*(3), 337–354.

Hickok, G., Rogalsky, C., Chen, R., Herskovits, E. H., Townsley, S., & Hillis, A. (2014). Partially overlapping sensorimotor networks underlie speech praxis and verbal short-term memory: Evidence from apraxia of speech following acute stroke. *Frontiers in Human Neuroscience, 8,* 649. Retrieved from https://www.frontiersin.org/articles/10.3389/fnhum.2014.00649/full

Iuzzini, J., & Forrest, K. (2010). Evaluation of a combined treatment approach for childhood apraxia of speech. *Clinical Linguistics & Phonetics, 24,* 335–345.

Iuzzini-Seigel, J., Hogan, T. P., & Green, J. R. (2017). Speech inconsistency in children with childhood apraxia of speech, language impairment, and speech delay: Depends on the stimuli. *Journal of Speech, Language, and Hearing Research, 60*(5), 1194–1210.

Iuzzini-Seigel, J., Hogan, T. P., Guarino, A. J., & Green, J. R. (2015). Reliance on auditory feedback in children with childhood apraxia of speech. *Journal of Communication Disorders, 54,* 32–42.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test–Second Edition (KBIT-2)*. Circle Pines, MN: AGS.

Kaufman, N. R. (1995). *Kaufman Speech Praxis Test for Children*. Detroit, MI: Wayne State University Press.

Law, J., Garrett, Z., & Nye, C. (2003). Speech and language therapy interventions for children with primary speech and language delay or disorder (Cochrane Review). *Cochrane Database of Systematic Reviews,* (3), CD004110. Retrieved from https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD004110/abstract

Liégeois, F. J., Turner, S. J., Mayes, A., Bonthrone, A. F., Boys, A., Smith, L., . . . Morgan, A. T. (2019). Dorsal language stream anomalies in an inherited speech disorder. *Brain, 142*(4), 966–977. https://doi.org/10.1093/brain/awz018

Lohmeier, H. L., & Shriberg, L. D. (2011). *Reference data for the Syllable Repetition Task (SRT)*. Phonology Project, Waisman Center, University of Wisconsin-Madison.

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

McAllister Byun, T. (2017). Efficacy of visual–acoustic biofeedback intervention for residual rhotic errors: A single-subject randomization study. *Journal of Speech, Language, and Hearing Research, 60*(5), 1175–1193.

McIntosh, B., & Dodd, B. (2008). Evaluation of the core vocabulary intervention for treatment of inconsistent phonological disorder: Three treatment case studies. *Child Language Teaching and Therapy, 24,* 307–327.

McReynolds, L. V., & Bennett, S. (1972). Distinctive feature generalization in articulation training. *Journal of Speech and Hearing Disorders, 37,* 462–470.

Moriarty, B. C., & Gillon, G. T. (2006). Phonological awareness intervention for children with childhood apraxia of speech. *International Journal of Language & Communication Disorders, 41*(6), 713–734.

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*(3), 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*(3), 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*(1), 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 & Communication Disorders, 50,* 529–546.

Olswang, L. B., & Bain, B. A. (1985). The natural occurrence of generalization during articulation treatment. *Journal of Communication Disorders, 18,* 109–129.

Ozanne, A. (1995). The search for developmental apraxia. In B. Dodd (Ed.), *The differential diagnoses and treatment of children with speech disorder* (pp. 91–109). San Diego, CA: Singular.

Peter, B., & Stoel-Gammon, C. (2005). Timing errors in two children with suspected childhood apraxia of speech (sCAS) during speech and music-related tasks. *Clinical Linguistics & Phonetics, 19*(2), 67–87.

Powell, J., & McReynolds, L. (1969). A procedure for testing position generalization from articulation training. *Journal of Speech and Hearing Research, 12*(3), 629–645.

Preston, J. L., Brick, N., & Landi, N. (2013). Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *American Journal of Speech-Language Pathology, 22*(4), 627–643.

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.

Rvachew, S., & Brosseau-Lapré, F. (2018). *Developmental phonological disorders: Foundations of clinical practice* (2nd ed.). San Diego, CA: Plural.

Rvachew, S., & Grawburg, M. (2006). Correlates of phonological awareness in preschoolers with speech sound disorders. *Journal of Speech, Language, and Hearing Research, 49,* 74–87.

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*(4), 146–157.

Rvachew, S., & Matthews, T. (2017a). Demonstrating treatment efficacy using the single subject randomization design: A tutorial and demonstration. *Journal of Communication Disorders, 67,* 1–13. https://doi.org/10.1016/j.jcomdis.2017.04.003

Rvachew, S., & Matthews, T. (2017b). Using the syllable repetition task to reveal underlying speech processes in childhood apraxia of speech: A tutorial. *Canadian Journal of Speech-Language Pathology and Audiology, 41*(1), 106–126.

Rvachew, S., & Nowak, M. (2001). The effect of target-selection strategy on phonological learning. *Journal of Speech, Language, and Hearing Research, 44,* 610–623.

Shiller, D. M., & Rochon, M.-L. (2014). Auditory–perceptual learning improves speech motor adaptation in children. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1308–1315.

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America, 125*(2), 1103–1113.

Shriberg, L. D., Aram, D. M., & Kwiatkowski, J. (1997). Developmental apraxia of speech: II. Toward a diagnostic marker. *Journal of Speech, Language, and Hearing Research, 40,* 286–312.

Shriberg, L. D., Lohmeier, H. L., Campbell, T. F., Dollaghan, C. A., Green, J. R., & Moore, C. A. (2009). A nonword repetition task for speakers with misarticulations: The syllable repetition task (SRT). *Journal of Speech, Language, and Hearing Research, 52,* 1189–1212.

Shriberg, L. D., Lohmeier, H. L., Strand, E. A., & Jakielski, K. J. (2012). Encoding, memory, and transcoding deficits in childhood apraxia of speech. *Clinical Linguistics & Phonetics, 26,* 445–482.

Strand, E. A., & Debertine, P. (2000). The efficacy of integral stimulation intervention with developmental apraxia of speech. *Journal of Medical Speech-Language Pathology, 8*(4), 295–300.

Strand, E. A., & Skinder, A. (1999). Treatment of developmental apraxia of speech: Integral stimulation methods. In A. Caruso & E. A. Strand (Eds.), *Clinical management of motor speech disorders in children* (pp. 109–148). 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., van Brenk, F., & van Doornik-van der Zee, A. (2014). Auditory feedback perturbation in children with developmental speech sound disorders. *Journal of Communication Disorders, 51,* 64–77.

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

Velleman, S. L., & Shriberg, L. D. (1999). Metrical analysis of the speech of children with suspected developmental apraxia of speech. *Journal of Speech, Language, and Hearing Research, 42,* 1444–1460.

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, A. L. (2012). Intensity in phonological intervention: Is there a prescribed amount. *International Journal of Speech-Language Pathology, 14*(5), 456–461.

Williams, P. (2009). The nuffield centre dyspraxia programme third edition (NDP3). In C. Bowen (Ed.), *Children's speech sound disorders* (pp. 269–275). Chichester, United Kingdom: Wiley-Blackwell.