Behavioral Interventions to Treat Speech Sound Disorders in Children With Autism

Sridhar Aravamudhan¹ • Smita Awasthi¹

Published online: 20 June 2019

© Association for Behavior Analysis International 2019

Abstract

Children with autism are at a higher risk of being affected by speech disorders and often require remedial intervention. Eikeseth and Nasset (Journal of Applied Behavior Analysis, 36(3), 325–337, 2003) used sufficient-response exemplar training of vocal imitation in conjunction with prompting, chaining, and shaping procedures to successfully teach 2 typically developing children to articulate several Norwegian words with blends. The present study extends and adapts these procedures to children with autism. Participants were TA, an 11-year-old boy, and KS, a 15-year-old girl, both with autism and speech sound disorders. For each participant, 3 sets of 10 words with specific blends in the initial position were targeted for training. Vocal imitation training with within-stimulus prompts was used for both participants. For KS, lip-tongue-teeth position prompts and chaining were added during the training of certain words. A multiple-baseline across-behaviors (word sets with target blends) design demonstrated improvement in the articulation of trained words and generalization of correct articulation to untrained words with both participants. The findings suggest that speech sound disorders in children with autism can be addressed with behavioral interventions.

Keywords Autism · Speech sound disorder · Phonological disorder · Articulation disorder · Sufficient-response exemplar training · Echoic · Vocal imitation training · Chaining · Lip-tongue-teeth prompts

Speech sound disorders, previously known as phonological disorders, affect a significant proportion of the population of children with autism. They fall under the broad category of communication disorders according to the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013). They involve omission (“top” for “stop”), distortion (“tate” for “teeth”), or substitution (“chewy” for “cry”) of phonemes or syllables (Porter, 2016) while uttering words. A diagnosis of speech sound disorder requires ruling out difficulties in speech production related to structural or motor impairments, such as dysarthria seen in cerebral palsy, ethnic and regional variations, or hearing impairments.

Speech disorders have been observed in several populations affected by different conditions, such as hearing impairment, intellectual disabilities, autism, cerebral palsy, and cleft palate (Gibbon, 1999). Studies of vocal-verbal 3- to 9-year-old children

with autism place prevalence estimates for speech delay and speech disorders between 12% and 33% (Cleland, Gibbon, Peppe, O’Hare, & Rutherford, 2010; Rapin, Dunn, Allen, Stevens, & Fein, 2009; Shriberg, Paul, Black, & Van Santen, 2011). In comparison, the prevalence of speech errors at 8 years of age in the general population is only 7.9% (Wren, Roulstone, Miller, Emond, & Peters, 2009). After normalization for age differences, these results indicate a substantially higher risk for concomitant speech errors in children with autism (Shriberg et al., 2011).

Articulation training is a much-neglected area in the field of behavior analysis, and this is evident from the very limited number of peer-reviewed studies available. Hardly any evidence currently exists to inform a practitioner of the protocols that can be successful in articulation training. The interest in this area could also be limited for want of evidence that improved articulation leads to better outcomes for persons with autism and speech sound disorders. It is important to explore if, in this population of children with autism and speech sound disorders, improved articulation can be achieved with behavioral interventions and whether such improvements can lead to improved educational and social outcomes.

✉ Smita Awasthi
smita.awasthi@behaviormomentum.com

¹ Behavior Momentum India, 407, 7th Main, 80 ft. Road, HRBR Layout, Bangalore 560043, India

Interventions that are commonly used by speech-language pathologists to address articulation problems include non-speech oral motor exercises (NSOMEs) and prompts for restructuring muscular and oral targets (PROMPT; Dale & Hayden, 2013). There is little empirical evidence to support either method for improving speech production or the management of speech disorders (Clark, 2005; Lof & Watson, 2008; McCauley, Strand, Lof, Schooling, & Frymark, 2009). Lass and Pannbacker (2008) reviewed 45 articles and reports in peer-reviewed and non-peer-reviewed journals over a 15-year period. They reported that the ones with a strong experimental design actually produced evidence against the use of NSOMEs for modifying speech.

Vocal imitation training (VIT) with shaping has been used to shape the articulation of whole words (Lovaas, Berberich, Perloff, & Schaeffer, 1966) by targeting incremental improvements in the words said by a speaker. An echoic operant occurs when a person, after hearing another, repeats what was heard. It is a behavior under the control of an antecedent auditory stimulus (a sound heard) and generates an auditory stimulus (a sound produced by the speaker), which has point-to-point correspondence and formal similarity with the antecedent stimulus (Michael, 1982). Skinner (1957), while discussing the value of an echoic repertoire, explained that it can be used to bring responses under the control of other stimulus conditions. Echoic training has also been used for the acquisition of mands and tacts (Kodak, Clements, & Ninness, 2009), sentence production by prelingual and deaf children (Golfeto & de Souza, 2015), reading improvement (Neville, 1968), and improvement of the complexity of echoics (Tarbox, Madrid, Aguilar, Jacobo, & Schiff, 2009). With persons who have speech sound disorders, it may be more useful for practitioners to improve articulation of words said under echoic control before mand, tact, or intraverbal training.

Hegde and Pena-Brooks (2007) examined the evidential basis of treatment protocols for phonological disorders from 1970 to 2007 and recommended the use of discrete-trial methods incorporating behavioral techniques such as prompting, manual guidance, fading, positive reinforcement, corrective feedback, and shaping. They further state that these treatments are well established with adequately controlled and replicated experiments.

Eikeseth and Nasset (2003) used a promising treatment package comprising sufficient-response exemplar training (SRET) of vocal imitation, prompting, chaining, and shaping to treat speech sound disorders of two typically developing school-aged Norwegian children. The study included a selection of target words, beginning with blends that the children had difficulty in articulating, such as “sk,” “st,” and “r” for one participant and “bl,” “dr,” and “sl” for the second. Blends are groups of two or three consonants wherein each sound may be heard. Examples of blends include “bl,” “cl,” “fl,” “gl,” “pl,” “fr,” and “tr.” Target sounds were based on articulation assessments. Baseline assessments for each participant suggested

speech errors in all 30 target words selected. In conjunction with the presentation of vocal models, they used exaggerated models of difficult sounds (within-stimulus prompts); repeated presentation of specific omitted or substituted sounds several times before the presentation of the whole word; presentation of component sounds separately and then brought closer together until the whole word was articulated correctly (chaining); and modeling lip, mouth, and teeth positions (extra-stimulus prompts).

Learning the correct articulation of a targeted sound in a single word or exemplar may not result in the generalization of articulation of that sound in a novel word. Stokes and Baer (1977), while discussing a technology of generalization programming, list seven techniques, namely, (a) train and hope, (b) sequential modification, (c) the introduction of natural maintaining contingencies, (d) training sufficient exemplars, (e) training loosely, (f) using discriminable contingencies, and (g) programming common stimuli. When one exemplar is taught, mastery may not have generalized beyond the one example. One way to address this problem would be to train another exemplar and another until generalization effects to nontrained examples are seen at targeted levels. In this method, if the student performs poorly on generalization probes, the instructor trains additional examples and assesses generalization to a new set of untrained examples. Wunderlich, Vollmer, Donaldson, and Phillips (2014) evaluated serial multiple-exemplar training (S-MET) versus concurrent multiple-exemplar training (C-MET) and evaluated generalization to untrained examples. They found concurrent presentations better in terms of requiring fewer sessions and greater generalization to untrained examples. However, Schnell, Vladescu, Kodak, and Nottingham (2018) evaluated S-MET and C-MET with three children with autism and found S-MET to be more effective with two of the participants.

Chaining involves linking sequences of stimuli and responses to get new performances (Cooper, Heron, & Heward, 2007, p. 436). Chaining and task analysis have been used successfully to teach a variety of skills, such as side-of-the-foot soccer passes, computer tasks, and Internet skills (Jerome, Frantino, & Sturmey, 2007). It is possible to conceptualize articulation of words, phrases, and sentences as behavior chains that involve the emission of a series of sounds in a sequence. For students with difficulty in articulating words, it may be possible to train the component sounds in isolation, bring the responses closer together, and then have them emitted at one time without a break. Tarbox et al. (2009) successfully used a modified chaining procedure to teach complex echoics (e.g., “mun” + “day” to teach “Monday”) to two children with autism and one with developmental delay.

Another component of the Eikeseth and Nasset (2003) intervention was prompting. Dyer (2009) describes the use of within-stimulus prompts to troubleshoot incorrectly articulated sounds. He used an exaggerated presentation of a part of

the model for better stimulus salience. For example, to teach “m” in the final position, the therapist model could have a sustained presentation of the “m” sound as in “dimmmme.” Dyer’s study also refers to extra-stimulus prompts such as holding the child’s jaw slightly open, prompting the tongue position using a tongue depressor, or touching the cheeks.

In the Eikeseth and Nasset (2003) study, both participants acquired accurate vocal imitation of all target sounds within a few sessions of intervention. In the study, with training on one or two words in a word set, correct articulation generalized to remaining untrained target words. The authors identified certain limitations with their study, such as limited generality, a lack of comparison with other methods, and a limited examination of basic behavioral processes underlying the effects.

The current study attempts to replicate and extend the Eikeseth and Nasset (2003) study to a different population—namely, children with autism—where the prevalence of speech sound disorders is significantly higher.

Method

Participants

Participants in this study were TA, an 11-year-old boy, and KS, a 15-year-old girl, both with autism. TA’s parent reported that he was diagnosed with autism by a pediatrician in California when he was 2 years old. He underwent intensive behavioral intervention (IBI) based on applied behavior analysis (Baer, Wolf, & Risley, 1968) from age 7 and made gains across several learning domains. A behavioral language assessment (BLA; Sundberg & Partington, 1998), used to assess his language and verbal repertoires, showed low vocal play (score of 2 out of 5), fair vocal imitation (4 out of 5), and strong tacting, requesting, match-to-sample, and listener responding repertoires. His overall BLA score was 47 out of 60. TA could articulate most functional words clearly and had difficulty only with words that started with certain blends.

A pediatric neurologist in Mumbai, India, diagnosed KS with autism when she was 28 months old. She was nonvocal until 13 years of age. She acquired a functional speech repertoire of about 30 short word approximations with IBI. These were intelligible only to people familiar with her. Her overall BLA score was 41 out of a possible 60. She scored 2 out of 5 in vocal play, 2 out of 5 in vocal imitation, and 3 out of 5 in social interactions. A speech-language pathologist used a set of 80 words listed in Table 1, in imitation trials, and reported consonant strengths and weaknesses in the initial, medial, and final positions (Table 2). These confirmed that KS had a profound speech sound disorder.

An important prerequisite for VIT is auditory discrimination. Both participants had normal hearing and auditory discrimination. They could produce a correct response in several

listener responding tasks. They could perform a number of one-step actions under auditory discriminative stimuli (e.g., clap hands, salute, wave goodbye), orient to an adult when their name was called or when told to “look,” select a stimulus from an array based on adult instructions for over a 100 instructions (“Point to the horse,” “Point to the toy,” “Which one do you iron clothes with?”), echo a number of vowel and consonant sounds, and respond to questions requiring auditory discrimination (e.g., “How do you come to school?” “What is your mother’s name?”).

Procedures

Assessment and selection of target words For both the participants, the second author conducted an assessment with a list of words starting with “st,” “sp,” “sm,” “sn,” “cr,” “bl,” and “fl.” Of these, “sp,” “st,” and “sm” were selected as target blends for TA. A set of words were shortlisted from a long list of words based on a behavioral assessment of the words that could be used in mand, tact, or intraverbal training. In addition, his parents were also consulted on the words that could be useful for him in the home and in community settings. Care was taken to avoid long words with more than five to six letters. His 30 word targets based on these blends in the initial position are presented in Table 3. For KS, similarly, “st,” “sp,” and “sc”/“sk” were selected, and the word targets for each blend are presented in Table 4.

Setting and staff The interventions were carried out in a center providing IBI based on applied behavior analysis (Baer et al., 1968) with a focus on operant verbal behavior (Skinner, 1957). The articulation training was delivered in one-on-one teaching sessions by therapists with BA, BSc, or MSc degrees who had at least 1 year of experience in delivering behavioral interventions to children with autism under the supervision of a Board Certified Behavior Analyst. A therapist with more than 2 years’ experience in implementing behavioral interventions served as an independent examiner and conducted probe sessions. For each participant, the same therapist and examiner conducted the training and probe sessions, respectively, throughout the study. The first and second authors, who are Board Certified Behavior Analysts, trained the therapists and the examiners on running the articulation training sessions and probe sessions and conducted treatment integrity (TI) checks.

Stimulus preference assessment A list of preferred items or actions were first identified for each participant using stimulus assessment procedures such as free operant observation (Cooper et al., 2007, p. 297) and multiple stimuli without replacement preference assessments (Halvey & Rehfeldt, 2005). Before each probe or training session, the therapist or examiner presented an array of the five highest ranked items from the preference assessments for a duration of 1 min

Table 1 Syllables/Words Used by the Speech-Language Pathologist for KS's Articulation Assessment

saw	chee	kaa	koo	shy	choo	show	shoe	high	j	g	ho
hay	zoo	ray	low	row	yay	four	five	zee	lay	lie	la
loo	k	y	wow	goeey	foo,	key	hi	c	sow	say	Go
guy	see-saw	tea	toe	two	tie	boo	bau (bow-bow)	bow	bee	baa	bye
mow	may	me	moo	no	neigh	knee	pooh	pea	pay	p	pie
mow	day	do	da (duh-duh)	dee							

(Carroll & Klatt, 2008). The items that the participant did not touch, reach out for, or engage with were removed, and the remaining items were randomly used during probe and training sessions. Both participants always showed clear preferences before a training or probe session.

Target behavior The primary dependent variable in this study was the vocal imitation of target words modeled by the independent examiner during probe sessions. A correct response was defined as an exact imitation (point-to-point correspondence) during probe sessions in all the positions for TA (i.e., initial, medial, and final position). For KS, it was defined as the targeted blend (“st,” “sp” or “sc”/“sk”) being correct and at least one of the other parts of the emitted word being correct. For instance, “spay,” with the omission of “r,” was accepted as a correct response to the modeled “spray” (refer to Table 5 for the list of words and accepted approximations). The reason for this accommodation for KS’s responses is as follows. Whereas TA had problems with the articulation of only selected blends, KS had articulation problems with several consonants as well. In her case, the mastery of correct articulation of the blend part, by itself, may not lead to the correct articulation of the whole word, though it would be a significant improvement from the baseline level. For instance, given the challenges that came up in training the “r” sound in “spray,” it was felt by the experimenters that being able to say “spay” for “spray” is a significant improvement over saying “ay.” The authors believed that teaching KS to correctly articulate “spray” with the “r” would have involved additional training that would have been a distraction in the current study. A secondary dependent variable—namely, trials to mastery—was also measured for each trained word. This was calculated as the number of training trials between the commencement of articulation training on a word and its subsequent declaration as mastered in a probe session by the examiner. On each day of

training, the number of sessions and target trials in each session were recorded.

Probe sessions to measure target behavior The examiner conducted probe sessions in each of the four conditions: baseline, articulation training, maintenance, and follow-up. The probe session procedures were identical in all these conditions. The examiner assessed if (a) the participant’s vocal imitation of the target word under training was to the standard required and (b) the correct vocal imitation also transferred to other untrained words in the same word set. During the articulation training phase, a probe session was conducted immediately following the correct vocal imitation of any target word on more than 90% of the trials in a training session. Because every probe session measured the articulation of all the words in all the word sets, any probe session served as a mastery assessment of the words and word set under training, as a maintenance or follow-up probe for previously mastered word sets, and last as a baseline measurement for target word sets on which training had not commenced.

The examiner sat across from the participant at a table at about a distance of 1 m facing the participant and completed a brief stimulus preference assessment to identify backup reinforcers. These were then delivered in exchange for tokens. Differential reinforcement was not used in probe sessions. Both correct and incorrect responses resulted in the delivery of a token, and the examiner said “nice try” in a neutral tone. The examiner presented one to three previously mastered familiar words before presenting a target word from the list of 30 target words. This was done to ensure a higher rate of correct responding, to allow for discrimination in responding to target words versus familiar words, and to minimize possible emotional reactions caused by the presentation of novel target stimuli. The examiner started with the first word of the word set, and if the participant emitted a vocal response that

Table 2 Speech-Language Pathologist’s Assessment, Participant KS

Consonant	Initial Position	Medial Position	Final Position
Mastered consonants	p, b, m, t, d, n, s, f, w	p, m, t, d, n, k, y, w	t, n
Emerging consonants	h, y	f	p
Weak/absent	c, g, j, k, q, v, x, z	b, c, d, g, h, j, l, m, n, q, r, s, v, x, z	All except t, n, and p

Table 3 Target Blends, Word Sets, and Words, Participant TA

Blend	Word Targets
st	stem, stop, stool, step, stone, stick, style, store, stove, stack
sp	spice, spoon, spill, spot, spin, spider, spit, spade, speak, spy
sm	smart, smear, smile, smog, smell, smooth, small, smoky, smash, smug

matched the examiner's model (in the case of TA) or a predefined approximation (in the case of KS), the examiner said "nice try," recorded the response as correct, and delivered a token. To ensure that differential reinforcement was not used in probe sessions, even incorrect responses were consequted with a token and a "nice try." If the vocal imitation was incorrect, the examiner provided another opportunity with the same word and recorded the response as correct or incorrect at the end of the second trial. Where such a second opportunity was provided, at the end of the trial, whether the response was correct or incorrect, the examiner delivered a token paired with "nice try" in a neutral tone. The examiner then proceeded to the next target word in the list and continued until the participant obtained 12 tokens, at which point a backup reinforcer was delivered. Then, a break of 40 s to 1 min was provided before resumption of the probe with the next target word. The probe session ended when the examiner had successfully completed the trials with all 30 targeted words from the three word sets. If the participant showed any signs of distress or emitted any escape behaviors, the probe session was terminated and resumed after a play break of at least 10 min, during which no demands were placed on the participant.

Interobserver agreement A trained second observer recorded data on target behavior for 83% of TA's and 45% of KS's probe sessions. A therapist acting as a second observer stood within hearing distance from the participant but at least 2 m away from the examiner. For each session, interobserver agreement (IOA) was calculated as the number of agreements divided by the number of agreements plus disagreements, and multiplied by 100. IOA was 100% for the baseline probe session for both participants. For probes during the intervention, maintenance, and follow-up conditions, the mean IOA for TA was 98.7% (range 93%–100%) and for KS was 91% (range 80%–100%).

General procedures The procedures had several steps that were conditional upon successful completion or unsuccessful completion of the previous step. The sequence of events

during the course of the study, the decision points, and the timing of baseline, probes, and training sessions are depicted in the flow diagram in Fig. 1.

Experimental design A concurrent multiple-baseline across-behaviors (articulation of word sets) design was used in this study.

Baseline After the selection of targets and before the intervention, probes were conducted on all 10 words in each of the three word sets for each participant as described in the probe procedure.

Articulation training sessions There were four to six articulation training sessions with an average duration of 3 min each, embedded within 2 hours of an IBI Program for both the participants. The training took place every day, Monday through Friday, except on holidays, vacations, or occasions when the participants were ill. The articulation training sessions were interspersed between training on other skills according to the participant's lesson plan for IBI. Backup reinforcers were made available for 45–60 s. To avoid confounds, none of the target words were trained in any other context or with any other trainer. In the home setting, too, the parent who was a primary caretaker confirmed that no echoic training or any other form of vocalization training was provided at home.

The therapist and the participant sat in chairs across from each other at a small table. In target trials, the therapist presented the specific target word identified for training from a word set (e.g., the therapist said, "Say 'stop'"). The target word trial was conducted after one to four trials on previously mastered words or sounds. These mastered words or sounds had a history of positive reinforcement for the participant and helped ensure better cooperation and compliance. For example, with TA, one to four words such as "daddy," "mummy," or "chair" from a pool of words that he could articulate correctly were presented before the target word "stop." In addition, care was taken to ensure that the target blend did not appear in the previously mastered words or sounds used for nontarget trials to reduce the potential for inadvertent practice. Only one target word was trained at a time. Each training session was approximately 3 min in duration with 8–12 trials of target sounds, interspersed with 24–28 trials of previously acquired sounds or words. Tokens were delivered contingent on correct responses of the target word. If the response to a modeled target word was not correct, an additional presentation of the target word was made. This was to provide increased practice opportunities when error responses occurred. If the response was correct in the second attempt, a token paired with social praise was delivered. If the response was incorrect in the second attempt, the therapist presented a mastered sound that obtained a correct response and then delivered a token. Thus, a slight delay to reinforcement was introduced for incorrect responses. Access to preferred

Table 4 Target Blends, Word Sets, and Words, Participant KS

Blend	Word Targets
st	stop, stim, stone, stay, stew, staple, stuck, steady, stove, stack
sp	spit, spy, spoon, spine, spot, spell, spray, spike, spat, speak
sc	sky, scoop, skip, scar, ski, skew, skin, skate, scan, scare

Table 5 Approximations Accepted for Some Responses, Participant KS

Word Targets	Baseline	Errors	Approximation Accepted
stim	“tm”	Omission of “s,” “I”	stum
stew	“too”	Omission of “s”	stu/stoo
steady	“ethy”	Substituted “oo” for “ew” Omitted “st,” Substituted “th” for “d”	steathy
spoon	“un”	Omitted “sp” substituted “u” for “oo”	spun
spray	“ay”	Omitted “sp” and “r”	spay (“r” omitted)

items or activities was provided once 12 tokens were obtained. When correct articulation of the targeted word occurred in 90% of the trials in a session, a probe session was scheduled to confirm mastery of the word. Articulation of a word was deemed mastered whenever the participant echoed the word correctly during the probe session (i.e., the examiner marked the response as correct while implementing the probe procedure). The examiner also identified if correct articulation generalized to other words in the same word set. If, in the word set under training, during the probe, the examiner scored less than 8 out of 10 words as correctly articulated, the next target word from the same word set was trained. On the other hand, when responding was correct with 8 or more words out of 10 words in the word set in a probe session on the target word set, the word set was declared as mastered and the first word from next untrained word set (if any were remaining) was targeted. This followed the S-MET arrangement. This sequence of events is depicted in Fig. 1.

Within-stimulus prompts, lip-tongue-teeth position prompt, and chaining procedures used in articulation training sessions

Within-stimulus prompting, lip-tongue-teeth position prompts, and chaining procedures were used to improve the participants’ articulation of target words and sounds. Merely providing a vocal model was not enough to train correct articulation of the target words for either participant. For TA, within-stimulus prompting was the only additional component used in the treatment package for all the trained target words. The use of within-stimulus prompting was implemented by exaggerating the part of the target word or sound the participant had difficulty with in the model presented by the therapist. With both participants, in the first words trained, the initial “s” sound had to be accentuated. For example, the trainer said “ssssstop” to model “stop” initially. The exaggeration was gradually faded until the participant could accurately imitate a normal vocal model.

Fig. 1 Flow diagram for training sessions and triggers for probe sessions. A check mark represents a correct response, $\geq$ represents “greater than or equal to,” boxes represent processes, and rhombuses represent decision points

With KS, additional difficulties were observed during the training of certain words. In these instances, extra-stimulus prompts and chaining were used. Once she started articulating the “st” or “sp” part, within-stimulus prompting was further used for the sound “i” in “stim” with a slightly elongated “i” sound. Within-stimulus prompts were also used for “n” in “stone.”

For the final part of the word “stove,” pronounced “stuvh” in India, it became clear that within-stimulus prompting alone was not effective in improving the probability of a correct response. So, an additional, lip-tongue-teeth position prompt was used. The therapist provided a model prompt involving the lower lip folding over the lower row of teeth. The participant then imitated the lip and teeth position and articulated the “ove” part of the word. The therapist faded the prompts in successive sessions. With the word “stuck,” KS was seen substituting the “th” sound for the “ck” sound. The correct pronunciation of the “ck” sound required KS to retract her tongue slightly. For this word, a lip-tongue-teeth prompt was used followed by a chaining procedure. Training on the “ck” sound was done separately. The approval of a physician and her mother’s approval for the procedure was secured before the training. The therapist used a normal-sized plastic Ikea spoon to prompt her to retract the tongue less than a millimeter when the model “k” was presented. The participant cooperated well with the procedure and did not display any signs of anxiety or distress. This prompt was faded in 3 days’ time to the use of the participant’s own finger as a self-prompt when the sound “k” was presented. The self-prompt, too, was faded, and once the sound “k” came under echoic control, a modified chaining procedure was implemented. The therapist first presented the “stu” sound and paused briefly. KS repeated “stu” after her. The therapist then presented “ck” to which KS responded with “ck.” The pause between the two sounds was reduced in successive days until KS responded with “stuck” in one try.

For the word “spoon,” only chaining was used. The parts were first presented separately as “spoo” and “n” and then closer and closer together in time across successive trials until the complete word was articulated as “spun,” the predefined approximation, without any pause in between.

Maintenance Once the mastery criterion was met for any word set (i.e., the participant’s vocal imitation was determined to be correct in at least 8 out of the 10 words in the set), articulation training on that word set ended. Thereafter, the maintenance condition was in effect for 8 weeks. In this condition, to maintain correct articulation, a different therapist provided additional practice opportunities in a different session. For example, if articulation training sessions were run from 9 a.m. to 11 a.m., maintenance training on mastered words was embedded in a 12–2 p.m. IBI session. The second therapist provided one to two opportunities for a correct articulation of such mastered words daily and delivered tokens for correct responding as per the schedule in effect for other behavioral interventions. Such

maintenance sessions were conducted for 8 weeks after the words were acquired for each word set. If any word was articulated incorrectly in maintenance sessions, no remedial action was taken, so as not to interfere with the treatment that was underway at the same time with another word. The probes triggered by mastery of correct articulation of new targets would cover all the previously mastered words and record retention or loss, if any, in maintenance or follow-up conditions. Retraining on any word whose correct articulation was lost in the maintenance or follow-up phase was not attempted during this study.

Follow-up For any word set, the follow-up phase started at the end of the 8-week maintenance period. At this stage, maintenance trials on mastered words were withdrawn. The probe sessions conducted after the maintenance period of 8 weeks served to assess retention or loss of correct articulation of each word.

Postintervention echoic responding and assessments during mand, tact, and intraverbal training Postintervention, based on the participants’ lesson plans, several words, though not all, whose correct articulation was acquired were targeted for mand, tact, and intraverbal training. For example, in a play involving throwing small stones into a water puddle, when participant KS declared motivation for stones by reaching out, she was given an echoic prompt for the word “stone” during mand training. In an assessment exercise, the examiner recorded whether she could articulate “stone” correctly under motivation and echoic prompt. Additional examples include echoic prompts given to TA to tact “stem” or to provide an intraverbal response of “stove” to the antecedent verbal stimulus “What do you cook food on?” The examiner assessed the response only for the correct articulation of the relevant target word. The examiner recorded the response to the echoic prompt as correct or incorrect using the criteria used earlier in probe sessions during the training phase. In these contexts, articulation scoring was done for each of the selected words, a minimum of two times, during a 2-day observation for each participant.

TI. TI in this study required examination of both probe sessions and training sessions. Five trials from video-recorded probe sessions were scored by one of the supervisors as correct or incorrect on four parameters: correct and distinct presentation of target words, the absence of prompts, delivery of tokens, and the recording data immediately on completion of the response by the participant. A score of 20 was the maximum possible in each TI assessment of probe sessions. TI scores were then computed as the number correct divided by 20 and multiplied by 100. TI checks were conducted for 28% of TA’s probe sessions and 23.5% of KS’s probe sessions. TI scores for probes averaged 88% with a range of 72% to 100% for the study.

TI checks were also carried out during articulation training sessions. TA underwent training on four words and KS on nine words. TI checks were done by observing at least one session on the 2nd and 3rd day of training on each word. For each trained word, one of the authors examined the TI on two of the most important implementation components. The first was the use of within-stimulus prompts, lip-tongue-teeth position prompts, and chaining procedures specific to the word. The second was the use of differential reinforcement procedures. The TI score for a session was 0, 1, or 2 depending on whether none of the components was done correctly, one of the two components was done correctly, or both components were done correctly. A total of 40 TI audits were done during the study, in which 32 sessions were scored 2, 6 were scored 1, and 2 were scored 0. In the eight sessions where the score was less than 2, additional training was provided immediately until the therapist could score 2 in role-plays and in the next training session.

Results

Figs. 2 and 3 show the results of the probe sessions conducted during the baseline, training, maintenance, and follow-up sessions for TA and KS, respectively. Table 6 displays, for each trained word, days, sessions, and trials to mastery. Correct articulation in baseline conditions was zero for all the target words for both the participants in baseline conditions.

TA reached a level of 80% or more correct in all the three word sets for “st,” “sp,” and “sm” after 467, 1,230, and 432 trials over 9, 23, and 7 days, respectively. After 467 training trials over 9 days, “stop” was emitted with the immediate generalization of correct pronunciation to the eight remaining untrained words in the set. Correct articulation of the 10th word, “store,” occurred during the last probe in the maintenance phase, with no specific training in the intervening period. Correct articulation of all the words continued in the follow-up phase. After 1,230 trials of training on “spin” over 23 days, correct articulation of “spin” occurred. The correct articulation also transferred to the remaining targeted words in the “sp” word set. Correct articulation of the word “spider” was lost in one maintenance session and one follow-up probe but was correct in the second and final probe in the follow-up condition. Intervention on the third set resulted in the acquisition of correct pronunciation of “smooth” and “small” in 72 trials (1 day) and 360 trials (6 days), respectively, along with generalization of correct articulation to the remaining words in the “sm” word set. Performance was at 10 of 10 words in the maintenance phase, but the correct articulation of one word, “smart,” was lost in the follow-up condition.

As with TA, the improvement from baseline levels occurred in both the “st” and “sp” word sets for KS only after the treatment package was applied. Intervention could not be

carried out in the third word set as she suffered from multiple illnesses and was irregular in attendance for prolonged periods. With the limited time available, other learning targets were prioritized by the clinicians, and the study ended with her mastery of the second word set. The third word set (“sc”/“sk”) was under baseline conditions throughout the study, and performance was at zero correct in all the probe sessions.

KS reached a level of 80% or more correct in the two word sets “st” and “sp” after 2,894 and 3,956 training trials over 54 and 71 days, respectively. In the first word set, “st,” KS required training on six words in succession, with correct articulation generalizing to two more untrained words. With this, her articulation met the 8 out of 10 mastery criterion for the “st” word set. Two additional words, “staple” and “stack,” met criteria in follow-up probes without additional training. Additional 1-year follow-up probes confirmed retention of correct articulation with all 10 “st” words. In the “sp” word set, three words were trained in succession, and articulation generalized to five other untrained words. Correct articulation did not transfer to two other words, “spell” and “spat.” In the maintenance phase, the correctly pronounced words dropped to seven with the word “spit” accounting for the loss. Two 1-year follow-up probes confirmed retention of correct articulation of six words in this word set with loss of one additional word, “spot.”

In Table 5, for KS, specifically for the words where there were additional errors apart from the omission of the initial sound, the pronunciation of words in baseline assessment, the errors of omission and substitution in baseline, and the predefined acceptable approximations achieved are presented.

Postintervention assessments, done a minimum of two times per word, during mand, tact, and intraverbal training showed that both participants could articulate the words correctly when given an echoic prompt by a therapist. Table 7 displays the words that were used in training under different stimulus conditions. TA was trained, and he responded with correct articulation for 4 words during mand training, 11 words during tact training, and 14 words during intraverbal training. KS was similarly successful with 5 words during mand training, 9 during tact training, and 5 during intraverbal training.

Discussion

The intervention package of VIT with added within-stimulus prompts was effective in improving the articulation of targeted words by TA. SRET helped confirm generalization of correct articulation to other targeted words without specific training. In the case of KS, in addition to VIT and within-stimulus prompts, lip-tongue-teeth prompts, chaining, and reinforcement of close approximations were selectively required for training with different words. The improvements in articulation from

Fig. 2 Effects of a treatment packaged comprising SRET and within-stimulus prompts on the articulation of blends by TA during probes.

The panel on the right shows trials to mastery for each trained word. VIT = vocal imitation training; WSP = within-stimulus prompts

baseline conditions were seen only after the intervention was begun in each word set. The findings extend the Eikeseth and Nasset (2003) study to the autism population and suggest that VIT, in conjunction with prompting and chaining procedures, can produce generalized vocal imitative articulation improvements in children with autism, possibly including children whose vocalizations emerged late and those with profound speech sound disorders.

It is important to note that the components of the lip-tongue-teeth position prompts and chaining were not introduced or withdrawn in a systematic fashion as with studies that aimed to do a detailed component analysis. The components were introduced for specific words, based on the difficulties the participant KS had in emitting specific sounds or in emitting the sounds in a word as a single chain. As such, during visual analysis, it would not be correct to infer that the performance continued to improve even after withdrawal of a component, as the component may not have been relevant for the next target word. With only five

instances where the additional components were used, it can only be said that they possibly had an additive effect in improving the intervention's effectiveness for the specific component sounds and words targeted. It is difficult to make a definitive statement, based on data from the current study, that they were necessary components or that they had an additive effect. For example, whether KS could have acquired the "ck" sound in the word "stuck" without the lip-tongue-teeth prompts—with just VIT and within-stimulus prompts—was not tested in this study. Nor is it possible to infer as to what extent chaining was instrumental in reducing the trials required to train articulation of "spoon." The value of these components for each sound could be addressed in future studies using systematic introduction and withdrawal of the components during training and measuring improvements in the participant's articulation.

It may also be useful to identify specific prompting topographies that are successful in speeding up the acquisition of articulation of specific speech sounds or phonemes. The role

Fig. 3 Effects of a treatment package comprising SRET, within-stimulus prompts, prompts for lip-tongue-teeth position, and shaping and chaining on the articulation of blends by KS during probes. The panel on the right displays the number of training sessions for each trained word before

mastery was achieved. Note that the intervention was not started on the third word set (“sc”/“sk”). VIT = vocal imitation training; WSP = within-stimulus prompts; L = lip-tongue-teeth position prompts; C = chaining

Table 6 Days, Sessions, and Trials to Mastery for Trained Words

Participant	Trained Word	Days to Mastery	Sessions to Mastery	Trials to Mastery
TA	stop	9	52	467
	spin	23	125	1230
	smooth	1	6	72
KS	small	6	36	360
	stop	11	60	566
	stim	10	58	543
	steady	6	32	344
	stove	13	73	736
	stuck	9	46	441
	stone	5	30	264
	spine	5	30	312
	spoon	36	201	2044
	spike	30	164	1600

of and techniques for training aspiration and expiration of breath as components could be examined.

With only two participants, with different levels of severity of speech sound disorders, more studies are required with participants with autism and speech sound disorders to draw more definitive inferences and build the evidence base for the effectiveness of the behavioral interventions.

SRET was used in this study. In this instructional arrangement, one exemplar was trained at a time, and when mastery was achieved, probes were conducted for ascertaining generalization to untrained examples. Additional exemplars were trained until mastery of the word set was achieved. Future studies can compare this arrangement with another arrangement where C-MET is used.

Articulation of a subset of trained words for both participants was further examined during mand, tact, and intraverbal

Table 7 Words Correctly Articulated Postintervention by TA and KS During Training Under Other Stimulus Conditions With Echoic Prompts

Participant	Antecedent	Operant	Words Articulated Correctly With Echoic Prompts
TA	Motivating operation for specific item or actions	Mand	stop, stone, spin, stick (4)
TA	Contact with stimuli through sight or touch	Tact	stem, stool, step, stick, store, stove, spoon, spider, spade, smear, smooth (11)
TA	Antecedent verbal stimulus without point-to-point correspondence	Intraverbal	stop, step, stem, stool, store, stick, stove, smile, smell, spoon, spot, spin, smart, smile (14)
KS	Motivating operation for specific item or actions	Mand	stop, stone, stay, stew, spoon (5)
KS	Contact with stimuli through sight or touch	Tact	spoon, stone, stew, staple, stuck, spay (for spray), spike, spine, stove (9)
KS	Antecedent verbal stimulus without point-to-point correspondence with expected response	Intraverbal	stop, stew, steady, stove, spoon (5)

training. The correct articulation of all these words under other stimulus conditions with an echoic prompt during mand, tact, and intraverbal training suggests that improved articulation brought under echoic control can generalize to other stimulus conditions combined with echoic prompts. Despite this possibility and potential, this study did not examine further articulation of the words in the context of natural speech, and this somewhat limits the social validity of the study.

TA could achieve complete point-to-point correspondence in his vocal imitation on all targets. Although KS successfully acquired imitation of the targeted blend part of all the words, she had not acquired complete point-to-point correspondence in five of the mastered words. This suggests that although generalized vocal imitation can be achieved in children with autism and speech sound disorders, complete point-to-point correspondence may require additional training or components. The improvements from baseline in KS's articulation of the words "stim," "stew," "steady," "spoon," and "spray" are presented in Table 5. It is possible that the improved utterances have a higher probability of being reinforced by a verbal community than her utterances of the same word in baseline levels. However, this could not be ascertained, as use of the words in natural speech contexts was not explored in this study. It is possible that, for longer functional words, with more syllables, a complete point-to-point correspondence may only be achieved in several stages. Future studies could look at the use of various components in effecting stepwise incremental improvements in the articulation of words, progressing toward a terminal goal of articulating the whole word correctly, using shaping principles.

When the authors examined the trials to mastery for each trained word and the number of words that required specific training, both were found to be higher for KS. This could be attributed to the greater severity of her speech sound disorder. The words "spoon" and "spike" in the second word set took 2,044 (36 days) and 1,600 trials (30 days), respectively. Additional prompting options and less complex approximations for these two words could have helped reduce the overall instructional time.

The number of days of training and the number of words in a set that needed to be trained before generalization of correct articulation to other words differed across the two subjects. The factors that could govern the pace of acquisition have not been studied. Future studies could examine the prerequisites that account for a faster acquisition of correct articulation, component skills that could be pretrained, and the underlying behavioral processes and expand the range of sound-specific strategies used for prompting.

Eikeseth and Nasset (2003) examined the prerequisites for correct articulation. These are a generalized ability to imitate vocal models, the ability to produce or utter the word, and the ability to discriminate similarities and differences between the word heard and the word said. They could not conclude which of these three factors their intervention addressed. The fact that both participants needed training suggests that a generalized vocal imitation repertoire would be weak in persons affected by speech sound disorders. The ability to produce or utter a word could depend on additional variables such as control over vocal musculature and intake or expulsion of air. Examination of these factors was not taken up in the current study. It seems plausible to say that through a process of shaping, the treatment package addresses discrimination of similarities and differences between the word heard and said and leads to improvements in control over the production of the targeted sounds. This could account for the improved articulation outcomes. For example, with KS, to the therapist's model "spy," the response "spee" was immediately followed by a self-corrected response of "spy." Anecdotal observations of participants engaging in the self-correction of responses call for a more detailed exploration of such discrimination.

Overall, the present study suggests that VIT with an SRET instructional arrangement, and with the addition of within-stimulus prompts, lip-tongue-teeth position prompts, shaping, and chaining, can be used with children with autism to achieve correct or near-correct articulation of words. This study offers additional support to explore prompting strategies for children with autism and speech sound disorders.

Acknowledgements We thank Dr. Maurice Feldman for his valuable suggestions on an earlier draft of the manuscript; Dr. Vani Rupella, Speech-Language Pathologist, PhD, for explaining one of the participant's speech assessments; parents of the participants for their consent to share the findings; and Ms. Madhavi Rao, Ms. Stella, and other therapists for their help with conducting this study. The contents of this paper are solely the responsibility of the authors and do not necessarily represent official views of the organization they are affiliated with.

Funding The study has not received any funding from any authority.

Compliance with Ethical Standards

Conflict of Interest The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1(1), 91–97. <https://doi.org/10.1901/jaba.1968.1-91>.
- Carroll, R. A., & Klatt, K. P. (2008). Using stimulus-stimulus pairing and direct reinforcement to teach vocal verbal behavior to young children with autism. *The Analysis of Verbal Behavior*, 24, 135–146.
- Clark, H. M. (2005). Clinical decision making and oral motor treatments. *The ASHA Leader*, 10(8), 8–9.
- Cleland, J., Gibbon, F. E., Peppe, S. J., O'Hare, A., & Rutherford, M. (2010). Phonetic and phonological errors in children with high functioning autism and Asperger syndrome. *International Journal of Speech-Language Pathology*, 12(1), 69–76.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River: Pearson.
- Dale, P. S., & Hayden, D. A. (2013). Treating speech subsystems in childhood apraxia of speech with tactual input: The PROMPT approach. *American Journal of Speech-Language Pathology*, 22(4), 644–661.
- Dyer, K. (2009). Clinical application of speech intelligibility research: The River Street Autism Program at Coltsville. *The Journal of Speech and Language Pathology—Applied Behavior Analysis*, 4(1), 190–203. <https://doi.org/10.1037/h0100259>.
- Eikeseth, S., & Nasset, R. (2003). Behavioral treatment of children with phonological disorder: The efficacy of vocal imitation and sufficient-response-exemplar training. *Journal of Applied Behavior Analysis*, 36(3), 325–337.
- Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *Journal of Speech, Language, and Hearing Research*, 42(2), 382–397.
- Golfeto, R. M., & de Souza, D. G. (2015). Sentence production after listener and echoic training by prelingual deaf children with cochlear implants. *Journal of Applied Behavior Analysis*, 48(2), 363–375.
- Halvey, C., & Rehfeldt, R. A. (2005). Expanding vocal requesting repertoires via relational responding in adults with severe developmental disabilities. *Journal of Applied Behavior Analysis*, 38(1), 101–105.
- Hegde, M. N., & Pena-Brooks, A. (2007). Introduction to treatment protocols and the CD resource. In *Treatment protocols for articulation and phonological disorders* (p. xii). San Diego: Plural Publishing.
- https://books.google.co.in/books?id=9F10CQAAQBAJ&pg=PR12&lpq=PR5&focus=viewport&dq=hegde+pena-brooks&output=html_text.
- Jerome, J., Frantino, E. P., & Sturmey, P. (2007). The effects of errorless learning and backward chaining on the acquisition of internet skills in adults with developmental disabilities. *Journal of Applied Behavior Analysis*, 40(1), 185–189. <https://doi.org/10.1901/jaba.2007.41-06>.
- Kodak, T., Clements, A., & Ninness, C. (2009). Acquisition of mands and tacts with concurrent echoic training. *Journal of Applied Behavior Analysis*, 42(4), 839–843.
- Lass, N. J., & Pannbacker, M. (2008). The application of evidence-based practice to nonspeech oral motor treatments. *Language, Speech and Hearing Services in Schools*, 39(3), 408. [https://doi.org/10.1044/0161-1461\(2008/038\)](https://doi.org/10.1044/0161-1461(2008/038)).
- Lof, G. L., & Watson, M. M. (2008). A nationwide survey of non-speech oral motor exercise use: Implications for evidence-based practice. *Language, Speech and Hearing Services in Schools*, 39(3), 392–407.
- Lovaas, O. I., Berberich, J. P., Perloff, B. F., & Schaeffer, B. (1966). Acquisition of imitative speech by schizophrenic children. *Science*, 151(3711), 705–707.
- McCauley, R. J., Strand, E., Lof, G. L., Schooling, T., & Frymark, T. (2009). Evidence-based systematic review: Effects of nonspeech oral motor exercises on speech. *American Journal of Speech-Language Pathology*, 18(4), 343. [https://doi.org/10.1044/1058-0360\(2009/09-0006\)](https://doi.org/10.1044/1058-0360(2009/09-0006)).
- Michael, J. (1982). Skinner's elementary verbal relations: Some new categories. *The Analysis of Verbal Behavior*, 1(1), 1–3. <https://doi.org/10.1007/bf03392791>.
- Neville, M. H. (1968). Effects of oral and echoic responses in beginning reading. *Journal of Educational Psychology*, 59(5), 362.
- Porter, D. (2016). *DSM-5 category: Communication disorders*. Retrieved from [http://www.theravive.com/therapedia/Speech-Sound-Disorder-\(Phonological-Disorder\)-DSM-5-315.39-\(F80.0\)](http://www.theravive.com/therapedia/Speech-Sound-Disorder-(Phonological-Disorder)-DSM-5-315.39-(F80.0))
- Rapin, I., Dunn, M. A., Allen, D. A., Stevens, M. C., & Fein, D. (2009). Subtypes of language disorders in school-age children with autism. *Developmental Neuropsychology*, 34(1), 66–84.
- Schnell, L. K., Vladescu, J. C., Kodak, T., & Nottingham, C. L. (2018). Comparing procedures on the acquisition and generalization of tacts for children with autism spectrum disorder. *Journal of Applied Behavior Analysis*, 51(4), 769–783. <https://doi.org/10.1002/jaba.480>.
- Shriberg, L. D., Paul, R., Black, L. M., & Van Santen, J. P. (2011). The hypothesis of apraxia of speech in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 41(4), 405–426.
- Skinner, B. F. (1957). *Verbal Behavior*. Appleton-Century-Crofts, New York.
- Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, 10(2), 349–367. <https://doi.org/10.1901/jaba.1977.10-349>.
- Sundberg, M. L., & Partington, J. W. (1998). *Teaching language to children with autism or other developmental disabilities*. Pleasant Hill: Behavior Analysts.
- Tarbox, J., Madrid, W., Aguilar, B., Jacobo, W., & Schiff, A. (2009). Use of chaining to increase complexity of echoics in children with autism. *Journal of Applied Behavior Analysis*, 42(4), 901–906. <https://doi.org/10.1901/jaba.2009.42-901>.
- Wren, Y., Roulstone, S., Miller, L., Emond, A., & Peters, T. (2009 June). *Prevalence of speech impairment in 8-year-old-children*. Madison: Poster presented at the 30th Annual Symposium on Research in Child Language Disorders.
- Wunderlich, K. L., Vollmer, T. R., Donaldson, J. M., & Phillips, C. L. (2014). Effects of serial and concurrent training on acquisition and generalization. *Journal of Applied Behavior Analysis*, 47(4), 723–737. <https://doi.org/10.1002/jaba.154>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.