Received: 13 January 2019 | Revised: 7 November 2019 | Accepted: 14 November 2019

DOI: 10.1002/bin.1700

RESEARCH ARTICLE

WILEY

An evaluation of the effects of response interruption redirection and matched stimulation on vocal stereotypy

Cara Gibney | Katrina J. Phillips | Angela Arnold-Saritepe | Sarah Ann Taylor

School of Psychology, University of Auckland, Auckland, New Zealand

Correspondence
Katrina Phillips, School of Psychology, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand.
Email: kj.phillips@auckland.ac.nz

Stereotypy is one of the core diagnostic features of autism spectrum disorder and warrants behavioral intervention when it negatively impacts the person's life. The aim of this study was to evaluate the effectiveness of noncontingent matched stimulation (NCR-M) compared with response interruption redirection (RIRD) on reducing levels of vocal stereotypy in the natural environment. Interventions were compared using an alternating treatments design followed by generalization and maintenance phases, which utilized lay people as instructors. Results showed that both RIRD and NCR-M resulted in a reduction in vocal stereotypy for three out of four participants within an educational setting. An increase in appropriate vocalizations was found for two out of four participants for both interventions. These results suggest that both NCR-M and RIRD were effective in reducing levels of vocal stereotypy, yet their effects on appropriate communication remain undetermined. Social validity scores were obtained following the study, which demonstrated that both interventions were met with approval across a variety of measures.

KEYWORDS
autism spectrum disorder, noncontingent matched stimulation, response interruption redirection, vocal stereotypy

1 | INTRODUCTION

Vocal stereotypy is defined as the production of noncontextual, repetitive speech, which occurs either immediately (i.e., echolalia) or sometime after (i.e., delayed echolalia) an auditory cue (Wolery, Kirk, & Gast, 1985). It is generally accepted that stereotypy is maintained by the type of sensory stimulation it produces (i.e., automatic reinforcement; Iwata, Lerman, & Wallace, 1999). However, research has provided examples of stereotypic behaviors, which are sensitive to social contingencies. For example, chronic hand mouthing has been shown to increase when followed by verbal reprimands, which functioned as social positive reinforcement (Goh et al., 1995). Kennedy, Meyer, Knowles, and Shukla (2000) similarly found that social positive and negative reinforcers maintained motor stereotypy for three out of five participants with autism spectrum disorder (ASD), leading the researchers to conclude that the causes behind stereotypy are more complex than previously thought. These findings support the behavioral premise to assess behavior based on function rather than topography. The function is especially important, as it has been proven that treatments based upon functional assessment are more effective (Repp, Felce, & Barton, 1988).

Stereotypy can be problematic as it may reduce learning opportunities, prohibit social inclusion, restrict access to new environments, and can attract negative attention (Rotholz & Luce, 1983; Smith & Van Houten, 1996). Within the literature that is currently available, two interventions appear to demonstrate a reduction in vocal stereotypy: noncontingent stimulation, also known as noncontingent reinforcement (NCR), and response interruption redirection (RIRD).

RIRD was originally adapted for vocal behavior by Ahearn, Clark, MacDonald, and Chung (2007) from a response blocking procedure used to decrease motor stereotypy (Hagopian & Adelinis, 2001). RIRD involves issuing a series of three vocal demands contingent on vocal stereotypy. Following the individual's compliance with the sequential demands, praise and escape from further demands are provided. Ahearn et al. (2007) demonstrated a reduction in vocal stereotypy and an increase in appropriate vocalizations. Replications of RIRD have also demonstrated reductions in vocal stereotypy (e.g., Cassella, Sidener, Sidener, & Progar, 2011; Colon, Ahearn, Clark, & Malasky, 2012; Love, Miguel, Fernand, & Labrie, 2012; Miguel, Clark, Tereshko, & Ahearn, 2009).

RIRD is thought to work on a positive punishment contingency, in which the presentation of the vocal demands decreases the future likelihood of vocal stereotypy. Ethical requirements (e.g., Behavior Analyst Certification Board, 2017) require clinicians to utilize no aversive methods or punishment procedures until all reinforcement options have been exhausted. RIRD does contain a social reinforcement component whereby praise follows vocal responses from the participant, which may help to temper the response blocking component of the intervention. To date, there has been minimal assessment of the treatment acceptability of RIRD with only a few studies reporting social validity measures administered to parents or participants (Cassella et al., 2011; Cividini-Motta, Garcia, Livingston, & MacNaul, 2019).

NCR, as an intervention for vocal stereotypic behavior, employs the presentation of preferred stimuli on a predetermined time-based schedule. The stimuli can be matched (NCR-M) or unmatched (NCR-UM). NCR-M uses items that may equate to the type of sensory stimulation produced by the individual's vocal behavior (Lanovaz & Sladeczek, 2011), such as musical toys, story tapes, or the radio. In contrast, NCR-UM provides any preferred leisure items to the individual irrespective of the type of sensory stimulation produced by the stereotypic behavior (Rapp, 2007). Examples of un-matched items for vocal stereotypy may include fidget toys, edibles, or a picture book. Although research has found NCR-M stimuli to be more effective than NCR-UM in decreasing vocal stereotypy (Lanovaz, Fletcher, & Rapp, 2009; Lanovaz & Sladeczek, 2011; Rapp, 2007), others have demonstrated the utility of both NCR-M and NCR-UM (Ahearn, Clark, DeBar, & Florentino, 2005; Rapp, 2006). One reason for varying effectiveness of NCR-M in relation to NCR-UM may be the manner in which studies have selected "matched" stimuli. In some cases, stimuli have been selected based on the researcher's hypothesis of the underlying sensory feedback (Saylor, Sidener, Reeve, Fetherston, & Progar, 2012) or the

results of preference assessments (Higbee, Chang, & Endicott, 2005; Lanovaz & Argumedes, 2010). These indirect methods may identify potential competing stimuli, but do not involve experimental analysis, which may serve as a more accurate measure. Without accurate identification, stimuli may only be serving as a response competitor, where the target behavior can still be selected over the unmatched stimulus (Green & Freed, 1993). Evaluation should involve measuring the rate of the target behavior in the presence of the proposed stimuli to more accurately identify matched potential (Love et al., 2012; Piazza, Adelinis, Hanley, Goh, & Delia, 2000; Rapp, 2007).

Love et al. (2012) compared the effects for NCR-M and RIRD implemented in isolation as well as implemented in combination. The data suggest that both interventions resulted in reductions of vocal stereotypy, and an even greater suppressive effect was demonstrated in the combined condition. The authors also saw a greater number of appropriate vocalizations presented when RIRD was being implemented than in the NCR-M alone condition. One limitation of the Love et al. (2012) study is the lack of preassessment for matching stimuli in the NCR condition. In addition, the study failed to assess the effectiveness of the two interventions in less-controlled settings when implemented by nonbehavior analysts. Across the RIRD literature, clinical utility has not been consistently demonstrated regarding the ability for general personnel to implement the intervention in natural environments (Martinez & Betz, 2013).

The current study compared the effects of matched stimulation (NCR-M) and RIRD on vocal stereotypy in four children with ASD. The aim was to assess the efficacy of both treatments in decreasing stereotypic vocalizations and increasing appropriate vocalizations. Interventions were conducted in the participants' natural environments and implemented by the researcher, staff, and parents. Finally, this study aimed to examine the social validity of the two interventions.

2 | METHOD

2.1 | Participants and setting

Participants were four adolescents with ASD who received their diagnoses by licensed pediatricians within New Zealand. Each participant was referred for their high rates of inappropriate vocalizations, which affected their inclusion in educational and social settings. Ben was a 16-year-old male who communicated using full sentences and simple intraverbal exchanges, such as returning a greeting or answering a simple question. His mother reported that the persistent inappropriate vocalizations prohibited the family from using air travel, public transport, and going on community outings. This was largely due to the increased volume at which the vocalizations were produced. Andy was a 16-year-old male who also communicated using simple intraverbal exchanges. His stereotypy was identified by teachers as interfering with his attending during classroom routines and subsequently impacting his learning. Kallum was a 14-year-old male who was able to participate in intraverbal exchanges, including holding conversations, describing events, and answering questions. During the course of the study, Kallum was prescribed several nonspecified medications, which targeted hyperactivity. Harriet was a 12-year-old female who communicated using simple mand frames and had a limited tacting repertoire. Harriet infrequently engaged in low-intensity self-injurious behavior in the presence of highly aversive auditory stimuli. For Kallum and Harriet, teachers reported that stereotypy was intrusive to the point of impacting participation in group academics and inclusion in assembly, resulting in removal so as to avoid disturbing other students.

All participants attended special needs satellite classrooms situated in mainstream elementary schools. Assessments were conducted in empty rooms within the participants' school, and intervention phases were conducted within classrooms. All rooms were approximately 73 m^2. Each participant's classroom varied in layout, but all contained lesson materials, one teacher, two assistant teachers, and between 12 to 15 peers.

2.2 | Response definition and measurement

Data were collected on instances of vocal stereotypy and appropriate vocalizations. Vocal stereotypy was defined as any instance of repetitive and invariant vocalization. For Ben and Kallum, examples of vocal stereotypy included immediate and delayed echolalia, noncontextual singing, and muttering (e.g., voiceless, indistinguishable vocalizations made during an audible inhale or exhale). For Andy and Harriet, vocal stereotypy included immediate echolalia, delayed echolalia, noncontextual singing, and squealing (e.g., high pitched, lengthy squeals). For Harriet, an additional topography of noncontextual laughing was included within the definition (e.g., spontaneous laughter where an antecedent in the current environment was unknown). Singing was regarded as noncontextual when it occurred during class time, when engaged in a social interaction not related to music, and during any school-based activity that required silent attendance. Appropriate vocalizations were defined as any instance of manding, tacting, or intraverbal speech directed at a listener that appeared to serve a social function.

The duration of vocal stereotypy and frequency of appropriate vocalizations were recorded continuously on a handheld iPAQ® computer using Obsdata software. Onset of vocal stereotypy was recorded as soon as the observer had audibly identified it as stereotypic, and offset was recorded once the behavior was absent for 2 s. During the RIRD intervention, the session clock was stopped and no recording of vocal stereotypy or appropriate vocalizations occurred during the implementation of the RIRD procedure. Data were only collected on stereotypic vocalizations, which were free to occur for 10 min total each session to ensure consistent recording of freely occurring vocalizations. The session clock ensured control, as the data were not affected by the varying time spent implementing the interruption procedure.

2.3 | Interobserver agreement and procedural integrity

Interobserver agreement (IOA) data were collected across 42% of assessment, intervention, and follow-up phases. IOA data were collected for both stereotypic and appropriate vocalizations from videotaped footage obtained from a Flip Mino video camera. Exact agreement was used with IOA for stereotypic vocalizations and was calculated by dividing the shorter duration by the longer duration and multiplying by 100. IOA for the frequency of appropriate vocalizations was calculated using total agreement in which the smaller cumulative score was divided by the larger score then multiplied by 100. Across all participants, overall IOA for vocal stereotypy for the study was 97% (range, 67–100%) and overall IOA for appropriate vocalizations was 88% (range, 50–100%).

Procedural integrity was measured by a secondary observer for RIRD (33% of all sessions) and NCR-M (20% of all sessions). Procedural integrity during experimental conditions was calculated by dividing the number of correct responses by the number of opportunities and multiplying by 100. For RIRD, procedural integrity averaged 85% (range, 40–100%) and NCR-M averaged 92% (range, 80–100%). Procedural integrity was also measured for teachers or parents conducting the intervention during the follow-up phase. For RIRD, teacher and parent integrity averaged 54% (range, 0–80%) and NCR-M averaged 60% (range, 10–80%).

2.4 | Procedure: Preassessments

2.4.1 | Functional analysis

A standard functional analysis was conducted with each participant to rule out any socially mediated function. The functional analysis consisted of four randomly ordered conditions: alone, attention, play, and demand. The conditions followed the method outlined by Iwata, Dorsey, Slifer, Bauman, and Richmond (1994), and each condition lasted for 5 min with sessions running until a functional trend could be observed. The alone condition provided no scheduled

attention or materials and had no programmed consequences. In the attention condition, the participant was informed that the researcher was busy with work so they were required to "wait." For the duration of the condition, the researcher attended to the participant's vocal stereotypy on an FR1 schedule, providing both eye contact and no more than 2 s of verbal interaction. Leisure items were not provided during this time. During the demand condition, gross motor demands were delivered to the participant on an FT5s schedule. Contingent on vocal stereotypy, the participant was freed from the demand for a period of 15 s. If a target response occurred within the final 5 s of the 15-s period, then another 5 s of escape would be granted until 5 s free of vocal stereotypy occurred. During the play condition, the researcher provided attention on an FT15s schedule. A wide range of preferred leisure items were freely available to the participant for the entirety of the session. There were no programmed consequences for vocal stereotypy. Functional analysis results showed that vocal stereotypy persisted regardless of the social consequences. For Kallum, data were elevated in all conditions with the exception of the control condition, where regular attention and leisure items were available. These results suggested a behavioral suppression effect of the enriched environment. Overall, it was concluded that all four participants' vocal stereotypy was maintained by automatic reinforcement (see Figure 1).

2.4.2 | Preference assessment

Both parents and teachers completed a Reinforcer Assessment for Individuals with Severe Disabilities (Fisher, Piazza, Bowman, & Amari, 1996) for each participant, which guided the selection of hypothetically matched items in the form of audio clips without visual accompaniment. Stimuli selected from the Reinforcer Assessment for Individuals with Severe Disabilities included sound effects, music, scripted dialogue extracts, and theme songs from television and movies. The stimuli for each participant can be found in Table 1. These items were presented to participants in the form of a visual choice board, during 10-min free-operant preference assessment sessions. Participants indicated their preference by pointing to or exchanging a picture from the choice board. The selected audio file was played until the participant made a different selection or requested cessation. The duration of engagement with each audio clip was recorded using a stopwatch application on an iPhone 4. A participant was determined to be engaging with the auditory track when they permitted it to play without requesting a different track or attempting to stop the audio. Items with the greatest cumulative duration of engagement over 5 min were determined to be highly preferred, and items with minimal selection and under 5 min engagement were classed as low preferred (see Table 1). A further hierarchy within the high or low preferred categories was not developed.

2.4.3 | Matched stimuli assessment

This procedure was conducted to assess if level of preference was an effective indicator of matched potential. The matched stimuli assessment utilized a reversal design, where auditory access and control conditions were compared to determine which stimuli produced a suppressive effect. During each phase, a control condition was rotated with the target condition: low preference, high preference, and samples of stereotypic recording previously obtained from the participant. Only one category of auditory stimulus was available during a specified condition. A final verification phase involved the stimulus that showed greatest differentiation in levels of vocal stereotypy between audio stimuli and control conditions. Each phase consisted of three 5-min control conditions alternated with three 5-min auditory access conditions (i.e., low preferred, high preferred, or stereotypic recording). Both control and auditory access conditions were conducted in an empty room with a single table and chair, with minimal social interaction. All stimuli used within the assessment were based on preference assessment results (see Table 1). During the control condition, participants had no access to audio stimuli, leisure items, or social interaction. The participant was provided with a chair and desk, but was permitted to move about the room.

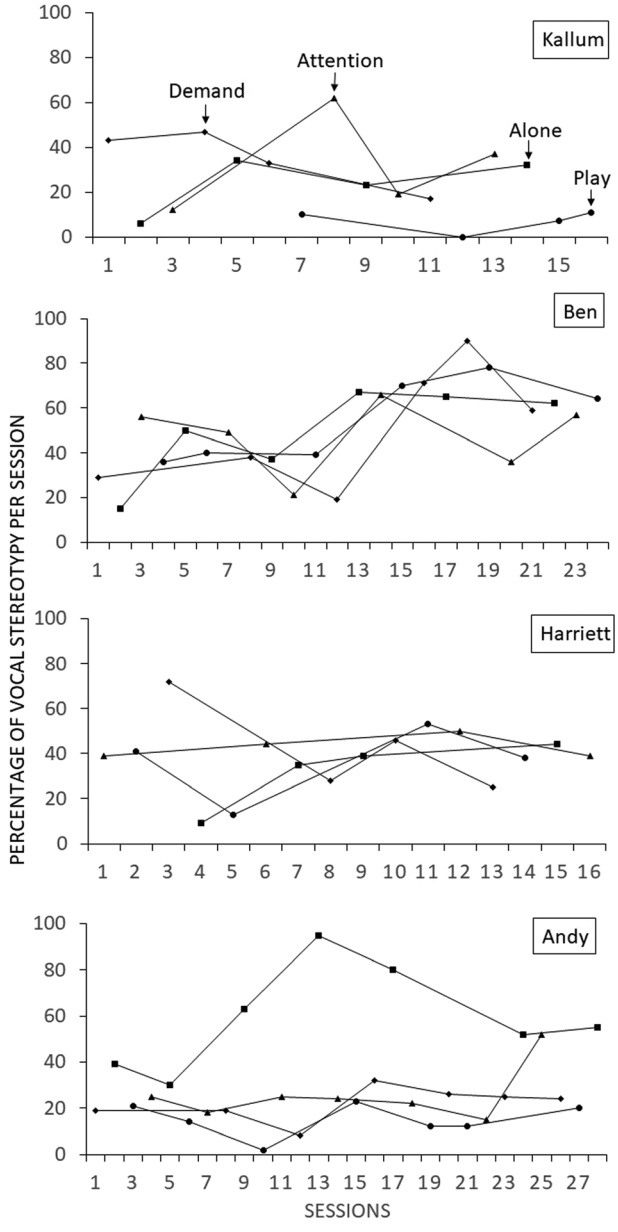

FIGURE 1 Percentage of vocal stereotypy per session for participants during all conditions of the functional analysis

TABLE 1 Degree of preference as indicated by the Reinforcer Assessment for Individuals with Severe Disabilities

Participant	Low preference	High preference
Andy	Sound effects, popular music, Ice Age©	Spongebob Squarepants©, Shrek©
Ben	Musicals, popular music, Spongebob Squarepants©	Smurfs©
Kallum	Aladdin©, Avengers©, Smurfs©	Lion King©, Ice Age©, Shrek©
Harriet	Spongebob Squarepants©, Shrek©, Ice Age©	Barney the Dinosaur©

The auditory access conditions involved files being played consistently for the duration of the condition, using Sony® headphones and a Sony® MP3 player. Audio items were randomly rotated during the 5-min conditions with a maximum play of 1 min per audio file at a time. The final condition involved playback of samples containing participants' own previously recorded vocal stereotypy. These files were randomly collected prior to assessment during opportunities where vocal stereotypy could freely occur.

Results for the matched stimulation assessment are presented in Figure 2. For Andy and Ben, access to highly preferred auditory stimuli reduced vocal stereotypy by 90% compared with baseline. For Andy, access to vocal

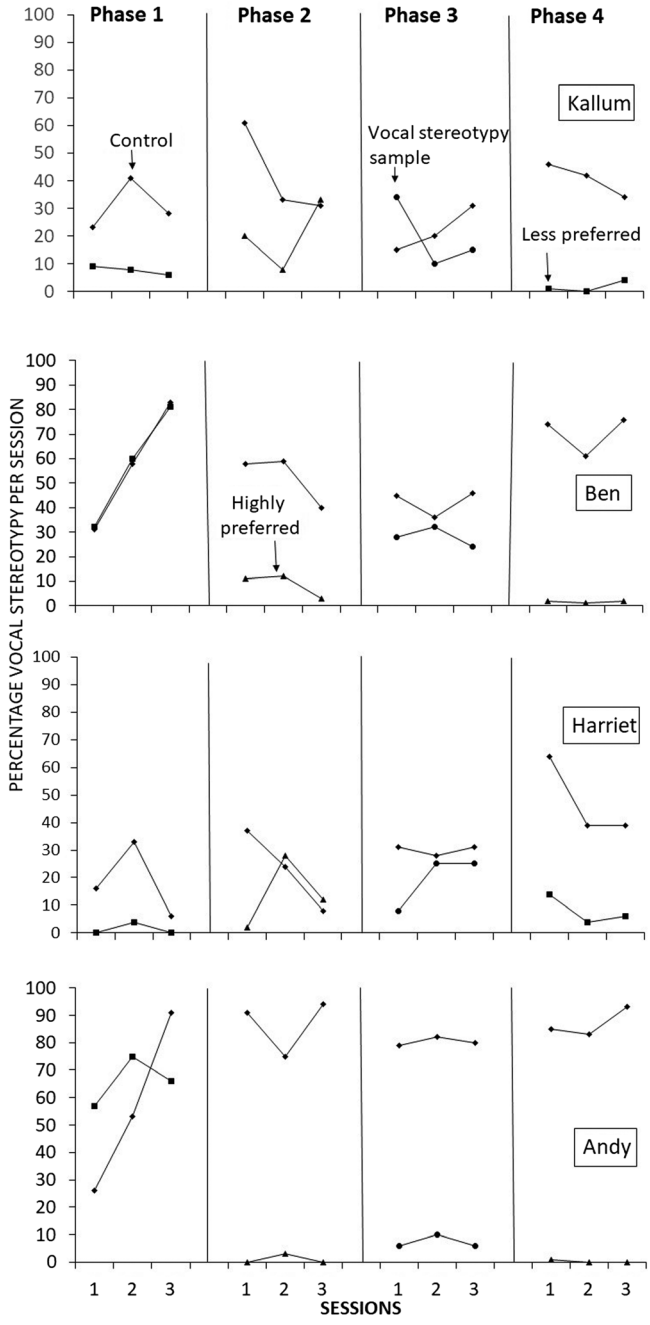

FIGURE 2 Percentage of vocal stereotypy per session for participants during matched stimuli analysis

stereotypy playback also demonstrated a reductive effect (90% decrease) and less preferred stimuli showed little effect (16%). Ben's vocal stereotypy showed a smaller reduction during stereotypy playback (33%), and a minor increase in vocal stereotypy was seen during access to less preferred stimuli (0.5%). For Kallum and Harriet, access to less preferred auditory stimuli demonstrated the greatest reduction in stereotypy compared with control trials, by 85% and 86%, respectively on average. For Kallum, access to highly preferred stimuli showed a moderate reduction in stereotypy at 51%, and access to vocal stereotypy playback had little effect (10% decrease). For Harriet, both highly preferred stimuli and vocal stereotypy playback showed similar reductions at 39% and 35%, respectively.

2.4.4 | RIRD vocal probes

We conducted vocal probes to identify the most effective vocal demands to include in the RIRD procedure. Vocal probe examples for all participants are presented in Table 2. For Ben, Andy, and Kallum, probes included tacting colors, shapes, objects, personal information statements, and reading words. Examples include "What color is this?", "What did you draw?", "Tell me your name," and pure tacts where the object was simply presented to the participant. Each question was asked three times per probe, and probes were repeated two to three times until the experimenter had created a list of mastered items for the purposes of RIRD. Mastery criterion required for list inclusion was 100% accuracy across three consecutive probes. Because Harriet demonstrated less vocal ability, her probes included color and label tacts as well as echoic phonemes, which scored 100% accuracy across all presentations included in the RIRD procedure. Examples include "What color?", "Say 'Pencil'", "Say 'mmm'".

2.5 | Procedure: Intervention

2.5.1 | Experimental design

An alternating treatments design with an initial baseline was used to compare RIRD and NCR-M. Following this, a maintenance and generalization phase with the best treatment was implemented, as well as a second generalization phase in the home environment where possible. Sessions were conducted for 1 hr, where 10-min randomized intervention conditions were run back-to-back for the duration of the sessions. Random rotation was determined by flipping a coin for each selection, with heads representing NCR-M and tails representing RIRD. A standard school day lasted for 6 hr per day, and sessions were rotated across in-class hours for a minimum of 10 sessions for each intervention, with 20 sessions in total.

TABLE 2 RIRD vocal probe examples

Participant	RIRD vocal probe examples
Andy	Color, shape and object label tacts, personal information statements, what intraverbal
Ben	Color, shape and object label tacts, personal information statements, reading words, what and where intraverbal
Kallum	Color, shape and label tacts, personal information statements, reading words, WH rotated intraverbal
Harriet	Color, object label tacts, and echoic phonemes

Abbreviation: RIRD, response interruption redirection.

Baseline

During baseline conditions, the participant followed the classroom routine under the teacher's instruction. Data were collected during 10-min observations, with a minimum of three sessions conducted per participant. The researcher shadowed the participant, and no specific consequences were programmed for vocal stereotypy. During baseline, the teacher delivered natural consequences, which sometimes resulted in the teacher attempting to gain the participant's attention by calling their name or reminding them to lower their voice. This natural consequence was also present throughout intervention conditions.

Matched stimulation (NCR-M)

During NCR-M sessions, the participant followed the classroom routine as outlined in baseline. The participant had noncontingent access to headphones and the MP3 player that continuously played matched audio stimuli as identified within the preassessment. Stimuli were considered matched based on the results of the matched stimuli assessment (see Table 1).

During the sessions, participants were required to keep one headphone ear bud in at a time so they were available to listen and attend to teacher instruction. In addition, participants were not permitted to select audio tracks or manipulate the MP3 player and any attempt to do so was immediately redirected by the researcher. In the event the participant attempted to remove the headphones and stop listening, the researcher would instruct the participant to "keep listening" and re-present the headphones where necessary. If the participant refused to use the headphones, the audio was stopped and the session was terminated. Data from terminated sessions were discarded. It was expected that participants would continue to complete classroom activities. Appropriate vocal behavior was responded with a short succinct contextually appropriate response from researcher and teacher, followed by redirection back to the task.

Response interruption redirection

During RIRD, participants also followed the described classroom routine while being shadowed by the researcher for the session duration noted for NCR-M. Upon every instance of vocal stereotypy, the researcher interrupted the participant by calling their name and gaining their attention and eye contact. Once sufficient attending behavior had been obtained, the researcher issued three consecutive vocal demands. The participant was required to answer each of these demands in a row, without producing vocal stereotypy between the demands. If the participant engaged in vocal stereotypy between the demands, the researcher gave neutral feedback such as "Try again," and reissued the three demands. Once three correct vocal responses occurred, the researcher would deliver praise and allow the participant to return to the activity to which they were initially attending. For vocal demands issued per participant, see Table 2. During times when silence from participants was required such as circle time or show and tell, the researcher would discretely deliver the treatment using a whispered tone. During RIRD, participants were expected to continue classroom activities, and appropriate vocalizations were responded to in the same manner as during NCR-M.

Follow-up

Teacher and parent training sessions occurred 2 weeks after the end of the intervention. If both interventions had been demonstrated as equally effective, the teacher or parent selected their preferred option. The sessions were conducted by the researcher during two 30-min generalization probes with a ratio of one per week on two consecutive weeks. Probes were conducted across people with class teachers in the classroom and across environments with parents in the home setting. Training was composed of a step-by-step procedural handout and associated role play and rehearsal with the researcher. Feedback was provided on a moment-by-moment basis to ensure optimal learning. Mastery of the procedure was required to proceed, involving 80–100% adherence during two role play scenarios. During the next training session, the researcher observed the teachers implementing the treatment during usual class hours and parental implementation at home after school. Specific verbal feedback on the teacher or parents' implementation was provided immediately following the observation.

2.5.2 | Social validity

Upon completion of the study, a social validity questionnaire was distributed to teachers and parents (see Appendix A). The questionnaire contained seven statements, which required an indication of satisfaction through the use of a Likert scale (where 1 was *strongly disagree* and 5 was *strongly agree*). Questions related to social satisfaction of outcomes, feasibility of the implementation, and researcher behavior and conduct for both interventions. Three of the four teachers filled in social validity questionnaires at the end of the study. As shown in Table 3, all recorded that they highly agreed with the treatment procedure, implementation, and outcomes for the participants (average score 4). Further, all recorded that they were highly satisfied with the researcher's conduct with themselves, their participant and general behavior in the classroom. Two of the four parents also filled in questionnaires. Both marked that they highly agreed with the feasibility of the intervention as well as the outcomes, providing scores of 4 and 5. Scores provided were not separated according to treatment.

3 | RESULTS

Figure 3 provides the percentage of intervals with vocal stereotypy occurring during baseline, the alternation between RIRD and NCR-M, and follow-up. The average baseline level of target behavior was 32% (range, 15–57%) for Andy, 38% (range, 23–51%) for Ben, 30% (range, 23–40%) for Kallum, and 16% (range, 8–23%) for Harriet. RIRD was effective in reducing vocal stereotypy for Andy, Ben, and Kallum. The vocalizations for these participants reduced from baseline to an average of 3% (range, 0–10%) for Andy, 12% (range, 1–37%) for Ben, 11% (range, 4–19%) for Kallum, and an average of 1%, 11%, and 7% over the last three sessions for each participant, respectively. For Harriet, the decrease could not be considered clinically significant with a reduction to an average of 15% (range, 1–22%) and 16% over the last three sessions. NCR-M was effective in reducing target behavior for Andy and Ben. Vocal stereotypy was reduced from baseline to an average of 1% (range, 0–3%) for Andy and 6% (range, 0–17%) for Ben with an average 2% and 6% over the last three sessions for each participant, respectively. For Kallum, a lesser reduction to 15% on average (range, 2–38%) was recorded, and Harriet demonstrated the smallest effect on target behavior with a clinically insignificant result of 14% (range, 3–28%). Average vocalizations recorded for the last three sessions were 16% for Kallum and 15% for Harriet.

The generalization and maintenance phases of this study occurred at school and home settings, with staff and family members implementing the RIRD protocol (see Figure 3). Average levels of reduction in vocal stereotypy decreased within the school placement for Andy with an average of 1% (range, 0–2%) yet increased to 8% (range, 2–13%) within the home environment. Ben's average level of vocal stereotypy maintained within the school setting at 13% (range, 9–19%) and home at 12% (range, 6–21%), with consistent variability. Kallum demonstrated a decrease from previous conditions at 1% (range, 0–2%) average stereotypic behavior, with a low steady trend observed. The

TABLE 3 Social validity scores for response interruption redirection and noncontingent matched stimulation

"I was satisfied with … "	Teacher 1	Teacher 2	Teacher 3	Parent 1	Parent 2
Initial teacher interview	4	5	4	N/A	N/A
Behavior assessment procedures	4	5	4	4	5
Intervention/teaching methods in classroom	4	5	4	5	5
Researchers behavior in classroom toward participants and teachers	5	5	5	N/A	N/A
Information provided during and after the study	4	5	4	4	5
Intervention results of the participants	5	4	4	5	5
Summary of research findings	4	4	4	5	4

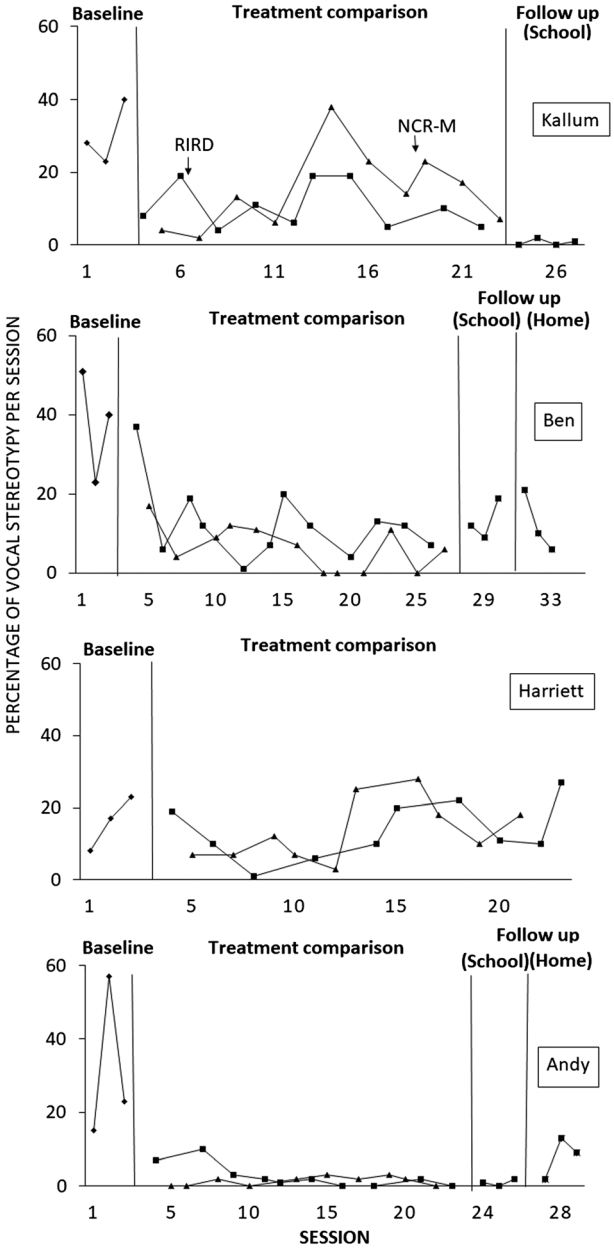

FIGURE 3 Percentage of session containing verbal stereotypy for each participant during baseline, treatment comparison (RIRD and NCR-M), and follow-up. NCR-M, noncontingent matched stimulation; RIRD, response interruption redirection

only follow-up condition for Kallum was the school setting. No follow-up sessions were conducted for Harriet due to the inefficacy of both treatments.

Appropriate vocalizations are presented in Figure 4. Average rates per hour during baseline were 0.3 (range, 0–1) for Andy, 0 (range, 0–0) for Ben, 12 (range, 6–21) for Kallum, and 4 (range, 0–7) for Harriet. During RIRD, appropriate vocalizations increased from baseline to an average of 1 for Andy (range, 0–3), 1 for Ben (range, 0–8), and 5 for Harriet (range, 0–15). An increasing trend was noted during the final seven sessions for Andy, and high variability in responding was demonstrated by Harriet. A decrease in appropriate vocalizations was recorded from baseline during both interventions for Kallum, with an average of 6 (range, 0–24) during RIRD and 7 (range, 0–15) during NCR-M. Responding was highly variable for Kallum within both conditions, with a significant increase in appropriate vocalizations in the final session of RIRD. A greater number of appropriate

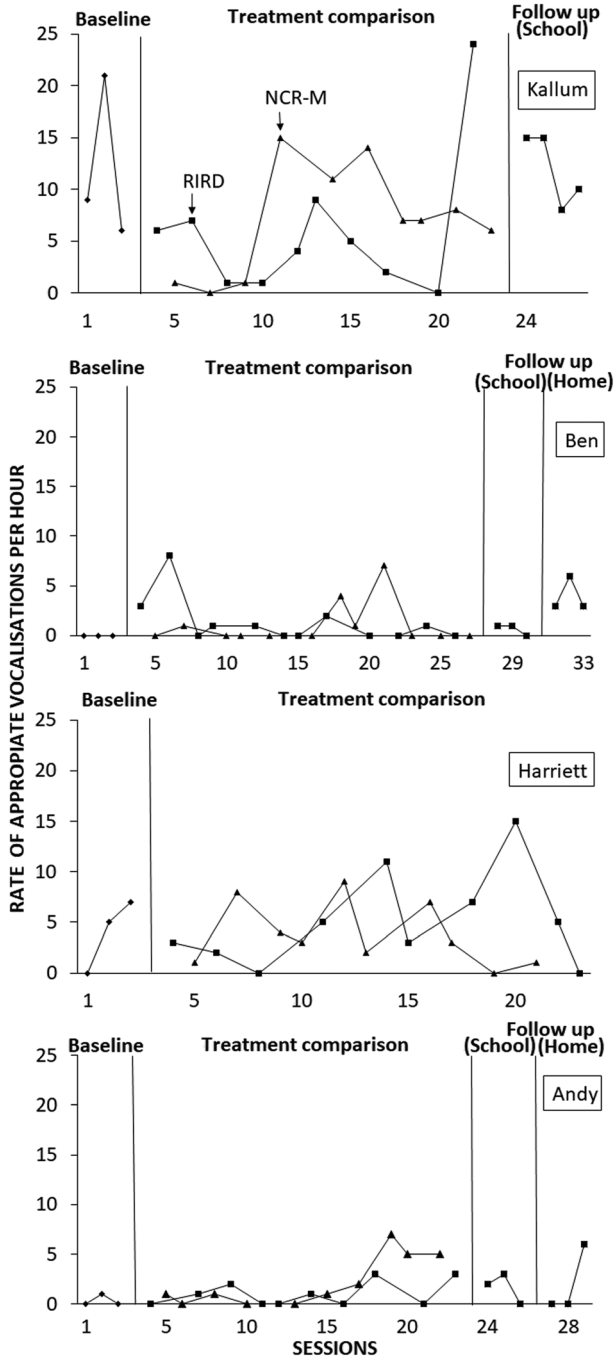

FIGURE 4 Rate of appropriate vocalizations per session for each participant during baseline, treatment comparison (RIRD and NCR-M), and follow-up. NCR-M, noncontingent matched stimulation; RIRD, response interruption redirection

vocalizations from baseline were only noted for Andy with an average of 2 (range, 0–7) and 1 for Ben (range, 0–7) with no increase from baseline for Harriet with 4 average vocalizations during NCR-M (range, 0–9). A steady increase in rates during the final seven conditions was again noted during NCR-M for Andy with a similar pattern seen during RIRD.

In Figure 4, appropriate vocalizations occurring during generalization and maintenance phases of RIRD were recorded for Andy, Ben, and Kallum. For Andy, rates remained consistent across both school at 2 (range, 0–3) and home (range, 0–6), which is a minor increase from previous conditions. For Ben, rates remained low and stable during

school follow-up at an average of 1 (range, 0–1) and increased in the home setting to an average of 4 (range, 3–6). Vocalizations returned to baseline rates for Kallum at 12 (range, 8–15), and no follow-up data were collected for Harriet.

4 | DISCUSSION

Overall, both interventions were effective in decreasing vocal stereotypic behavior for three out of four participants. In addition, only minor increases in untrained appropriate vocalizations were noted for two out of four participants. These findings support the results of Love et al. (2012) in demonstrating the efficacy of RIRD in reducing unwanted vocalizations; however, these findings did not produce the same effects on appropriate vocalizations. RIRD was considered the preferred treatment by teachers and parents alike due to perceived feasibility of implementation and therefore was selected for generalization and maintenance.

The results of the study indicated that RIRD was clinically effective in reducing vocal stereotypy. These results support the previous literature, which replicates the findings of Ahearn and colleagues (Ahrens, Lerman, Kodak, Worsdell, & Keegan, 2011; Cassella et al., 2011; Colon et al., 2012; Love et al., 2012; Miguel et al., 2009). Despite a significant reduction being demonstrated for Ben, Andy, and Kallum, RIRD was not successful for Harriet. An initial decrease in behavior level was observed; however, Harriet's level of vocal stereotypy returned to baseline over the course of the study. A potential confound regarding the effectiveness of RIRD may be due to the previous learning repertoires of the participations. For example, RIRD was likely ineffective in reducing Harriet's vocal stereotypy due to her lack of ready responding and compliance with RIRD demands. It is therefore recommended that screening be conducted for ready responding or instructional control prior to commencing RIRD for both research and clinical applications.

NCR-M was also found to be effective for reducing vocal stereotypy. These findings support the results of multiple studies, which demonstrate the utility of NCR-M (Lanovaz et al., 2009; Lanovaz & Sladeczek, 2011; Rapp, 2007). For Andy and Ben, the reductive effect was considerable and immediate, whereas Kallum and Harriet showed initial reductions followed by variable levels of responding. Harriet's lack of engagement with NCR-M was also noted, as 30% of sessions were terminated due to refusal to engage with auditory stimuli. Refusal was noted as walking away from the audio and removing headphones. The two participants whose behavior most greatly responded to NCR-M had a higher degree of preference for the audio stimuli during the preintervention assessments. These findings suggest that the use of the preintervention assessments may help determine if an intervention is likely to be effective. Additional research into the role of preference as an indicator of effectiveness of NCR-M is recommended to further contribute to the literature base.

The finding that three of the participants demonstrated a decrease in stereotypic vocalizations in the NCR-M intervention suggests that the behavior might be reduced in the absence of punishment-based interventions, which is useful given the ethical code. However, the current study did not make a comparison between NCR-M and NCR-UM, nor between the highly preferred and most effective for the participants whose matched stimulus analysis found these to be different. As such, it is unclear if it was the access to a preferred matched item that was important or the access to an alternate preferred item that provided similar sensory stimulation. This is also an area for future research.

An increase in appropriate vocalizations compared with baseline was only noted for two of the three participants during the follow-up phase when RIRD was in place. Although no research suggests that NCR-M should result in increased vocalization, the current findings do not support the effect found by Ahearn et al. (2007), Ahrens et al. (2011), and Miguel et al. (2009) despite replicating the protocol. The effects of implementing the intervention within the natural environment on appropriate vocalizations may have impacted the results; however, this was not controlled for nor directly assessed. Whereas previous research does demonstrate an increase in appropriate vocalizations during RIRD, a recent study supports the current findings where no significant increase was observed (Cividini-

Motta et al., 2019). These results support the importance of programming for language acquisition when designing an intervention to decrease stereotypic behavior in clinical practice.

Generalization and maintenance sessions were completed with three out of four participants, with only two out of four completing the generalization phase to the home environment. Environments were programmed based on parent willingness to receive training and allow sessions within the home environment. For Andy, school-based follow-up rendered greater suppression of stereotypy than the home environment. A likely factor in this discrepancy could be due to greater procedural integrity regarding the teaching staff. For Ben, maintenance of stereotypic level was observed consistently across home and school. Kallum was the only participant to receive follow-up solely at school. High fidelity to intervention protocol was observed by the teaching staff and may account for a further reduction in stereotypy during follow-up. Generalization and maintenance were not programmed for Harriet due to inefficacy of treatments.

Although both interventions were shown to be effective in suppressing the level of stereotypic vocalizations for some of the participants, there was no assessment of stereotypy rate after a session was completed. Other researchers have investigated and demonstrated the importance of assessing post-session effects of treatment on target behavior (Lanovaz & Argumedes, 2010; Rapp, 2007). If rates are likely to decrease after sessions, then the value of the treatment may possibly be increased. Lanovaz and Argumedes (2010) found that post-session effects of access to matched stimuli demonstrated a suppression effect on stereotypy. This suppression effect may go some way to assisting clinicians in successfully thinning schedules of NCR-M access. The result may increase feasibility within the natural environment where stimuli may not be as discrete.

A strength of this study relates to the implementation of intervention in the natural environment, as well as the utilization of parents and teachers as interventionists during the follow-up phase. The majority of research supporting RIRD and NCR-M has been either located in a clinic setting (Ahearn et al., 2007; Ahrens et al., 2011; Rapp, 2007) or in a restrictive environment with clinicians conducting the procedure (Ahrens et al., 2011; Love et al., 2012). A focus on generality as a core construct of ABA (Baer, Wolf, & Risley, 1968) is frequently considered within applied settings and should also be a priority within the research. A limitation of conducting research using persons with minimal training is the impact of procedural integrity. As noted within this study, procedural integrity scores had great variability for RIRD during the follow-up phase. The inclusion of parent or teacher preference when selecting the follow-up treatment was not significant in impacting fidelity. Therefore, a more rigorous training protocol and continued use of Behavior Skills Training would be recommended for future researchers to assist in greater efficacy when generalizing treatments.

Social validity data were collected after the cessation of the study and indicated a high degree of satisfaction with both interventions across a variety of measures (see Table 3). Collection of social validity data can be considered a strength of this study, as parent preference of treatments has been limited within the research base relating to RIRD. A limitation of the survey is related to measures taken on efficacy, conduct, and procedures on combined treatments, which likely impacts specific feedback on NCR-M compared with RIRD.

Although RIRD was noted to effectively reduce vocal stereotypy in this study, it is important to note that data were not collected and included during the intervals where the session clock was stopped. The inability to present and compare this data may demonstrate a possible limitation of the study. It has been suggested that exclusion of such data may overestimate the efficacy of RIRD. This was the finding for some participants whose stereotypy occurred at baseline levels when data were reported and included from implementation intervals (Wunderlich & Vollmer, 2015). However, evidence for vocal stereotypy suppression when interval data are included was demonstrated by Cividini-Motta et al. (2019). This will likely be an area for further investigation and comparison to ensure accurate data reporting within the future research of RIRD.

In conclusion, the results of this study support the effectiveness and feasibility of RIRD and NCR-M when applied to automatically maintained vocal stereotypy in the natural environment. Despite a reductive effect being demonstrated on target behavior, this study failed to identify an increase in untrained appropriate vocalizations for all participants. Future research into vocal stereotypy should have the intention of not only reducing unwanted

vocalizations, but also actively training and increasing engagement in appropriate vocal communication. This way, individuals with disabilities may have access to corrective teaching alongside reductive strategies.

ETHICAL ASSURANCE

All procedures performed in this study were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

ACKNOWLEDGMENTS

We acknowledge the participants, their family, and school staff for their involvement in this study.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Katrina J. Phillips https://orcid.org/0000-0002-3670-2798

Sarah Ann Taylor https://orcid.org/0000-0002-3236-1287

REFERENCES

Ahearn, W. H., Clark, K. M., DeBar, R., & Florentino, C. (2005). On the role of preference and response competition. Journal of Applied Behavior Analysis, 38, 247–250. https://doi.org/10.1901/jaba.2005.36-04

Ahearn, W. H., Clark, K. M., MacDonald, R. P. F., & Chung, B. I. (2007). Assessing and treating vocal stereotypy in children with autism. Journal of Applied Behavior Analysis, 40, 263–275. https://doi.org/10.1901/jaba.2007.30-06

Ahrens, E. N., Lerman, D. C., Kodak, T., Worsdell, A. S., & Keegan, C. (2011). Further evaluation of response interruption and redirection as treatment for stereotypy. Journal of Applied Behavior Analysis, 44, 95–108. https://doi.org/10.1901/jaba.2011.44-95

Baer, D., Wolf, M., & Risley, T. (1968). Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1, 91–97. https://doi.org/10.1901/jaba.1968.1-91

Behavior Analyst Certification Board (2017). BACB professional and ethical compliance code for behavior analysts. https://www.bacb.com/wp-content/uploads/2017/09/170706-compliance-code-english.pdf

Cassella, M. D., Sidener, T. M., Sidener, D. W., & Progar, P. R. (2011). Response interruption and redirection for vocal stereotypy in children with autism: A systematic replication. Journal of Applied Behavior Analysis, 44, 169–173. https://doi.org/10.1901/jaba.2011.44-169

Cividini-Motta, C., Garcia, A. R., Livingston, C., & MacNaul, H. L. (2019). The effect of response interruption and redirection with and without a differential reinforcement of alternative behavior component on stereotypy and appropriate responses. Behavioral Interventions, 34, 3–18. https://doi.org/10.1002/bin.1654

Colon, D. L., Ahearn, W. H., Clark, K. M., & Malasky, J. (2012). The effects of verbal operant training and response interruption and redirection on appropriate and inappropriate vocalizations. Journal of Applied Behavior Analysis, 45, 107–120. https://doi.org/10.1901/jaba.2012.45-107

Fisher, W. W., Piazza, C. C., Bowman, L. G., & Amari, A. (1996). Integrating parent report with a systematic choice assessment. American Journal on Mental Retardation, 101, 15–25.

Goh, H., Iwata, B. A., Shore, B. A., DeLeon, I. G., Lerman, D. C., Ulrich, S. M., & Smith, R. G. (1995). An analysis of the reinforcing properties of hand mouthing. Journal of Applied Behavior Analysis, 282, 269–283. https://doi.org/10.1901/jaba.1995.28-269

Green, L., & Freed, D. E. (1993). The substitutability of reinforcers. Journal of the Experimental Analysis of Behavior, 60, 141-158.

Hagopian, L. P., & Adelinis, J. D. (2001). Response blocking with and without redirection for the treatment of pica. Journal of Applied Behavior Analysis, 34, 527–530. https://doi.org/10.1901/jaba.2001.34-527

Higbee, T. S., Chang, S., & Endicott (2005). Noncontingent access to preferred sensory stimuli as a treatment for automatically reinforced stereotypy. Behavioral Interventions, 20, 177–184. https://doi.org/10.1002/bin.190

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richmond, G. S. (1994). Toward a functional analysis of self-injury. Journal of Applied Behavior Analysis, 27, 197–209. https://doi.org/10.1901/jaba.1994.27-197

Iwata, B. A., Lerman, D. C., & Wallace, M. D. (1999). Side effects of extinction: Prevalence of bursting and aggression during the treatment of self-injurious behavior. Journal of Applied Behavior Analysis, 32, 1–8. https://doi.org/10.1901/jaba.1999.32-1

Kennedy, C. H., Meyer, K. A., Knowles, T., & Shukla, S. (2000). Analyzing the multiple functions of stereotypical behavior for participants with autism: Implications for assessment and treatment. Journal of Applied Behavior Analysis, 33, 559–571. https://doi.org/10.1901/jaba.2000.33-559

Lanovaz, M. J., & Argumedes, M. (2010). Immediate and subsequent effects of differential reinforcement of other behavior and noncontingent matched stimulation on stereotypy. Behavioral Interventions, 25, 229–238. https://doi.org/10.1002/bin.308

Lanovaz, M. J., Fletcher, S. E., & Rapp, J. T. (2009). Identifying stimuli that alter immediate and subsequent levels of vocal stereotypy: A further analysis of functionally matched stimulation. Behavior Modification, 33, 682–704. https://doi.org/10.1177/0145445509344972

Lanovaz, M. J., & Sladeczek, I. E. (2011). Vocal stereotypy in individuals with autism spectrum disorders: A review of behavioral interventions. Behavior Modification, 36, 146–164. https://doi.org/10.1177/0145445511427192

Love, J. J., Miguel, C. F., Fernand, J. K., & Labrie, J. K. (2012). The effects of matched stimulation and response interruption and redirection on vocal stereotypy. Journal of Applied Behavior Analysis, 45, 549–564. https://doi.org/10.1901/jaba.2012.45-549

Martinez, C. K., & Betz, A. M. (2013). Response interruption and redirection: Current research trends and clinical application. Journal of Applied Behavior Analysis, 46, 549–554. https://doi.org/10.1002/jaba.38

Miguel, C. F., Clark, K., Tereshko, L., & Ahearn, W. H. (2009). The effects of response interruption redirection and sertraline on vocal stereotypy. Journal of Applied Behavior Analysis, 42, 883–888. https://doi.org/10.1901/jaba.2009.42-883

Piazza, C. C., Adelinis, J. D., Hanley, G. P., Goh, H. L., & Delia, M. D. (2000). An evaluation of the effects of matched stimuli on behaviors maintained by automatic reinforcement. Journal of Applied Behavior Analysis, 33, 13–27. https://doi.org/10.1901/jaba.2000.33-13

Rapp, J. T. (2006). Toward an empirical method for identifying matched stimulation of automatically reinforced behavior: A preliminary investigation. Journal of Applied Behavior Analysis, 39, 137–140. https://doi.org/10.1901/jaba.2006.37-05

Rapp, J. T. (2007). Further evaluation of methods to identify matched stimulation. Journal of Applied Behavior Analysis, 40, 73–88. https://doi.org/10.1901/jaba.2007.142-05

Repp, A. C., Felce, D., & Barton, L. E. (1988). Basing the treatment of stereotypic and self-injurious behaviors on hypotheses of their causes. Journal of Applied Behavior Analysis, 219, 281–289. https://doi.org/10.1901/jaba.1988.21-281

Rotholz, D. Z., & Luce, S. C. (1983). Alternative reinforcement strategies for the reduction of self-stimulatory behavior in autistic youth. Education and Treatment of Children, 6, 363–377.

Saylor, S., Sidener, T. M., Reeve, S. A., Fetherston, A., & Progar, P. R. (2012). Effects of three types of noncontingent auditory stimulation on vocal stereotypy in children with autism. Journal of Applied Behavior Analysis, 45, 185–190. https://doi.org/10.1901/jaba.2012.45-185

Smith, E. A., & Van Houten, R. (1996). A comparison of the characteristics of self- stimulatory behaviors in normal children and child with developmental delays. Research in Developmental Disabilities, 17, 254–268.

Wolery, M., Kirk, K., & Gast, D. L. (1985). Stereotypic behavior as a reinforcer: Effects and side effects. Journal of Autism and Developmental Disorders, 15, 149–161. https://doi.org/10.1007/BF01531601

Wunderlich, K. L., & Vollmer, T. R. (2015). Data analysis of response interruption redirection as a treatment for vocal stereotypy. Journal of Applied Behavior Analysis, 48, 749–764. https://doi.org/10.1002/jaba.227

How to cite this article: Gibney C, Phillips KJ, Arnold-Saritepe A, Taylor SA. An evaluation of the effects of response interruption redirection and matched stimulation on vocal stereotypy. *Behavioral Interventions*. 2020;35:114–130. https://doi.org/10.1002/bin.1700

APPENDIX A

Teacher Social Validity Questionnaire

Thank you for taking the time to complete this questionnaire. Once completed please return to the researchers using the pre-paid, self-addressed postage envelope. Please note that this questionnaire is voluntary, anonymous, and all information provided will only be made available to the supervising and participants researchers of this study. Please answer as honestly as you can.

"I was satisfied with …"	1 *Strongly disagree*	2 *Disagree*	3 *Neither agree or disagree*	4 *Agree*	5 *Strongly agree*
Initial teacher interview					
Behaviour assessment procedures					
Intervention/teaching methods used in sessions					
Researcher behavior in the classroom toward participants and teachers					
Information provided before and during the study					
Intervention results of the participants					
Summary of research findings					