Journal of
Applied Behavior Analysis

JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2018, **51,** 899–914 NUMBER 4 (FALL)

THE EFFECTS OF NONCONTINGENT MUSIC AND RESPONSE INTERRUPTION AND REDIRECTION ON VOCAL STEREOTYPY

ASHLEY R. GIBBS, CHRISTOPHER A. TULLIS, RAVEN THOMAS AND BRITTANY ELKINS

GEORGIA STATE UNIVERSITY

Vocal stereotypy is a commonly occurring challenging behavior in children with autism spectrum disorder (ASD) that is frequently maintained by automatic reinforcement and often interferes with skill acquisition. Matched stimulation (MS), and response interruption and redirection (RIRD) are two interventions that have been demonstrated to be effective in reducing the occurrence of vocal stereotypy with participants with ASD. The current study sought to determine if the combination of MS (noncontingent music) and RIRD was more effective at reducing vocal stereotypy than RIRD alone and if the parents of children with ASD found the combination of MS and RIRD more socially valid than RIRD alone. The results suggested that the combined intervention resulted in greater suppression of vocal stereotypy and increased occurrences of on-task behavior in both participants. Additionally, RIRD required fewer implementations and had a shorter duration when combined with MS. Results suggest that the combination of MS and RIRD may be an effective intervention outside of highly controlled settings.

Key words: autism spectrum disorder, response interruption and redirection, vocal stereotypy

Vocal stereotypy is a response often observed in people with autism spectrum disorder (ASD) and other developmental disabilities that is frequently targeted for behavior change within applied settings (American Psychiatric Association, 2013; Durand & Carr, 1987; Koegel & Covert, 1972). Vocal stereotypy may interfere with acquisition of adaptive behavior and can be socially stigmatizing (Liu-Gitz & Banda, 2010). Reducing the occurrence of vocal stereotypy allows for more opportunities to learn and expand upon communication, adaptive, and social skills vital for community inclusion (Wells, Collier, & Sheehey, 2016). In practice, vocal stereotypy presents a unique challenge for interventionists as it is frequently maintained by automatic reinforcement (Martinez & Betz, 2013), and it cannot be physically redirected by another person (Lanovaz & Sladeczek, 2012).

An antecedent intervention, matched stimulation (MS; Piazza, Adelinis, Hanley, Goh, & Delia, 2000) and a consequence intervention, response interruption and redirection (RIRD; Ahearn, Clark, MacDonald, & Chung, 2007), are two procedures that have been effective in reducing automatically maintained vocal stereotypy. MS manipulates motivating operations (MOs; Laraway, Snycerski, Michael, & Poling, 2003) by providing individuals with noncontingent access to stimuli that are hypothesized to match the consequences of the target behavior (Piazza et al., 2000). RIRD, presumably a punishment-based procedure, interrupts stereotypy and redirects individuals to emit three verbal responses correctly and in the absence of further vocal stereotypy (Ahearn et al., 2007).

To date, only one study has examined the effectiveness of combining MS and RIRD (i.e., MS + RIRD) to create a treatment package that manipulates both the antecedent conditions and consequences for engaging in vocal

Address correspondence to: Christopher A. Tullis, Department of Educational Psychology, Special Education and Communication Disorders, Georgia State University, P.O. Box 3965, Atlanta, GA 30302; ctullis2@gsu.edu

doi: 10.1002/jaba.485

stereotypy. Love, Miguel, Fernand, and LaBrie (2012) used a reversal design to examine the effect of RIRD and MS individually and a combination intervention of MS + RIRD on the percentage of time engaged in vocal stereotypy, frequency of appropriate vocalizations, frequency of RIRD implementation, and session length for two participants diagnosed with ASD. The authors used duration recording to measure the percentage of session engaged in vocal stereotypy and frequency recording to measure appropriate vocalizations and RIRD implementation. Data collection was interrupted, meaning that data were not collected during the RIRD procedure, excluding this information from the reported results. The results of the study demonstrated that MS + RIRD produced lower levels of vocal stereotypy in a shorter session time with fewer RIRD implementations than either intervention alone for both participants, though one participant had increased variability in appropriate vocalizations within the combined intervention condition as compared to during RIRD alone.

The results of Love et al. (2012) should be viewed in light of four considerations. First, the authors used moderately to highly preferred toys that produced sound as the source of matched stimulation, which could lead to a reduction in on-task behavior during demands requiring behavior other than toy manipulation (Lanovaz & Sladeczek, 2012). Additionally, the toys did not necessarily provide continued access to MS during RIRD implementation as they were removed contingent on vocal stereotypy, though they were not silenced if they produced sound at the time RIRD was implemented, limiting the investigation of whether the presence of MS during RIRD implementation influences its frequency or duration. Second, Love et al. evaluated intervention effects in the context of play activities only; the generality of the effects of MS + RIRD in other contexts (e.g., demands) should be investigated. Third, Love et al. and previous studies on RIRD used interrupted data collection. That is, the time in RIRD was removed, and data on the occurrence of vocal stereotypy was either not collected or excluded during RIRD implementation (e.g., Dickman, Bright, Montgomery, & Miguel, 2012; Liu-Gitz & Banda, 2010), which could have overestimated the effects of RIRD on vocal stereotypy (Wunderlich & Vollmer, 2015). Finally, the positive results reported by Love et al. were limited to a clinical setting and staff implementers, precluding evaluation of whether the effects of the intervention would have generality in the participant's natural environment with caregiver implementation.

Thus, the current study sought to replicate and extend the procedures of Love et al. (2012) in several ways. We provided noncontingent (NC) music via headphones, rather than musical toys, as MS. We evaluated both RIRD and MS + RIRD during task demands, rather than during play, to determine which was more effective at reducing the level of vocal stereotypy and increasing on-task behavior. We used uninterrupted data-collection procedures during RIRD and MS + RIRD to determine the overall effects of each. We also evaluated the generality of treatment effects when implemented by parents and in different contexts and conducted social validity assessments of the treatments with parents.

METHOD

Participants

Participants were two children who had been diagnosed with ASD by an independent physician. Both participants received occupational therapy and/or ABA services at a private, outpatient clinic between 2 and 5 days per week. Both participants communicated vocally and had an established repertoire of vocal imitation. Both participants remained seated and engaged in an adult-directed task for at least 10 min without aggression (e.g., hitting, kicking,

scratching, biting), self-injury (e.g., hitting, pinching, or biting self), or property disruption (e.g., swiping materials off of the table), and engaged in vocal stereotypy that was at least partially maintained by automatic reinforcement as determined by a brief functional analysis. In addition, both participants tolerated wearing over-the-ear headphones that played music between 40-45 dB, as measured by a sound-level meter, for at least 10 min without prompts during 80% of opportunities. The participants' treatment therapists referred them for inclusion in the study because their vocal stereotypy occurred often enough to interfere with their engagement in academic and social activities.

Elizabeth was a 4-year-old girl who communicated vocally. Based on the Vineland Adaptive Behavior Scales, Second Edition (VABS-II; Sparrow, Cicchetti, & Balla, 2005), Elizabeth's adaptive behavior composite was low and categorized as a mild deficit. The VABS-II categorized her receptive communication skills as moderately low, with an approximate age equivalency of 2.5 years, and her expressive skills as low, with an approximate age equivalency of 2 years. She could request highly preferred items using the mand frame, "I want —," tact a variety of everyday stimuli, complete simple intraverbal fill-ins, and engage in echoic responses. She received occupational therapy twice per week and speech therapy once per week at the time of the study, and she had never previously received behavioral intervention. Elizabeth took anti-epileptic medication for the duration of the study.

Matthew was a 7-year-old boy who communicated vocally. Matthew's adaptive behavior composite on the VABS-II (Sparrow et al., 2005) was low, categorized as a mild deficit. His receptive communication skills were moderately low, with an approximate age equivalency of 4 years, while his expressive communication skills were categorized as low, with an approximate age equivalency of 2.5 years. Matthew requested items using full sentences (e.g., "I want the black iPad, please," or, "May I have the book?"). He could also tact a variety of stimuli, complete intraverbal fill-ins, answer personal information and WH-questions, and engage in echoic responses. He received occupational and speech therapy once per week and intensive ABA intervention five times per week. Matthew received between 8 and 10 hr of intensive intervention services for over 4 years at the time of the study, and he had previously been exposed to RIRD without a maintained positive outcome. Matthew did not take any medication for the duration of the study. For both participants, the results of a brief functional analysis (Northup et al., 1991) indicated that vocal stereotypy was maintained at least in part by automatic reinforcement.

Setting and Materials

Experimental sessions were conducted in one of three treatment rooms (each approximately 4.6 m by 3.4 m) in a clinic equipped with a table and chairs, closed cabinets, and bookcases or shelves of nonessential materials that were occluded by a cloth cover. During experimental sessions, only the participant and the experimenter were present in the treatment room. Generality probes for Elizabeth occurred in the clinic with her mother implementing the intervention and included the same tasks used during experimental sessions. Generality probes for Matthew occurred in his home with his mother implementing the intervention, first while completing the same tasks used during experimental sessions, and later during an activity in the home that was reported to be associated with a high rate of vocal stereotypy.

Experimental materials included a sound-level meter, Beats™ wireless over-the-ear headphones, and an Apple iPod touch second generation®. Additional experimental materials related to the adult-delivered tasks included: 9 to 24 piece jigsaw puzzles, 12 shape sorter

boxes, sequencing cards, number cards, beads and string, pegs and a pegboard, and an AlphaSmart word processing keyboard (Matthew only). Data collection materials included an interval timer, a stopwatch, paper data sheets, a pencil, and a small, portable video camera.

Response Definitions and Measurement

There were four primary dependent variables: percentage of intervals with vocal stereotypy, percentage of on-task, the frequency of RIRD implementation, and the length of RIRD implementation. Each experimental session was videotaped for data collection purposes.

Vocal stereotypy was defined as any instance of contextually inappropriate vocalization lasting at least 3 s. This included contextually inappropriate singing, laughing, babbling, or saying words or phrases unrelated to the present context. It also included repetitive sounds, rhythmic breathing patterns (e.g., opening the mouth making a "huh" sound repetitively; clenching teeth, retracting lips, and repetitively breathing in audibly), blowing of air, squeals, lip popping, repetitive sounds with a closed mouth (i.e., mouth closed and lips vibrating together). Occurrences of whining or crying were not included. Sounds made in the context of task completion, such as sighing or groaning, as well as appropriate vocalizations, such as manding for help, were not counted unless the same vocalization was repeated within 3 s. Vocal stereotypy was measured using 10-s partial-interval recording in all pre-experimental and experimental conditions. Observers used interval timers to break each session into 10-s intervals (60 intervals) while watching the video footage of the session. If vocal stereotypy occurred at any point during the 10-s interval, the observer marked an "X" on a data sheet, including while RIRD was being implemented. The percentage of intervals during which vocal stereotypy occurred was calculated by dividing the number of intervals with vocal stereotypy by the total number of intervals and multiplying by 100.

On-task behavior during instructional contexts was defined as the participant sitting in his or her seat and engaging with the task materials. Engagement included the participant looking at the activity, and manipulating materials with his or her hands to complete the task. During the baking condition of Matthew's generality probes in his home, the definition of on-task behavior changed and was recorded if Matthew (a) remained within 0.3 m of the kitchen work surface while he or his mother engaged with the task materials, or (b) left the work area to obtain a required material when directed by his mother. During both experimental and generality sessions, participants were considered to be on task if he or she manded for help while attempting to manipulate task materials. Across all sessions, on-task behavior was measured from the video footage of the session using duration-per-occurrence recording. Observers activated a stopwatch when the participant exhibited on-task behavior and stopped the clock following 3 s without on-task behavior. The duration was recorded on a data sheet, and the stopwatch was reset. The duration of each occurrence of on-task was recorded in this manner for the entirety of the session video, and the percentage of on-task behavior during the session was calculated by dividing the total number of seconds on task by the total number of seconds in the session, multiplied by 100. We removed time spent implementing RIRD from the calculation of participant's on-task behavior, as compliance to RIRD demands was not included in the calculation.

The duration of RIRD implementation also was measured from the video footage of the session using duration-per-occurrence recording. Observers activated a stopwatch when the experimenter initiated the RIRD procedure by

interrupting the participant, and stopped the clock when the experimenter concluded RIRD by directing the participant to the previous task. Total duration of RIRD implementation (in minutes) within each session was calculated by summing all RIRD implementations and dividing by 60. The frequency of RIRD implementation and the frequency of task completion were collected and recorded on a data sheet during the session. Sessions lasted for a total of 10 min.

Interobserver Agreement and Procedural Fidelity

A second, trained observer scored data for a minimum of 33% of sessions in all conditions for each participant from video recordings. Interobserver agreement on occurrence of vocal stereotypy was assessed using the point-by-point method (Gast & Ledford, 2014), with an agreement scored if both observers recorded the occurrence or nonoccurrence of vocal stereotypy for an interval. Percentage of agreement was calculated by dividing agreements by the total number of intervals, multiplied by 100. Interobserver agreement for both on-task behavior and length of RIRD implementation was assessed using total duration. Each observer's durations recorded for occurrences of on-task behavior were summed, and the smaller duration was divided by the larger duration and multiplied by 100. For Elizabeth, mean agreement was 92.7% for vocal stereotypy (range, 85%-100%), 91.9% for on-task behavior (range, 83.3%-98.6%), and 96.9% for length of RIRD implementation (range, 84.7%-99.5%). For Matthew, mean agreement was 96% for vocal stereotypy (range, 83.3%-100%), 94.6% for on-task behavior (range, 80.3%-99.5%), and 98.7% for length of RIRD implementation (range, 96.9%-99.7%).

Procedural fidelity data were collected for RIRD implementation and the use of MS. Procedural fidelity was measured from video recordings for a minimum of 33% of the sessions in each condition for each participant. Observers used a checklist outlining the correct use of MS; correct implementation of MS required accurate performance of all steps on the checklist. Event recording for each part of the RIRD procedure was used to evaluate RIRD fidelity. The percentage of correct implementations of the entire RIRD procedure, and of each individual component, was calculated and analyzed. Across both procedures and participants, procedural fidelity averaged 97.3% (range, 87.5% to 100%).

Pre-experimental Procedures

Brief functional analysis. We conducted a brief functional analysis of each participant's vocal stereotypy to confirm that vocal stereotypy was at least partially maintained by automatic reinforcement. The brief functional analysis consisted of a single exposure to 5-min test and control conditions (Northup et al., 1991). The test conditions included attention, escape, and ignore conditions, and play was used as a control condition.

During the attention condition, the experimenter and participant sat together at a table with familiar toys present. The experimenter said s/he was busy and could not play with the participant, after which the experimenter turned away from the participant and appeared to complete paperwork. Contingent on vocal stereotypy, the experimenter turned to the participant, touched him or her on the shoulder, and stated, "Don't do that please" or, "That's too loud."

During the escape condition, the experimenter presented demands to engage in four nonvocal activities that were not in the participant's skill repertoire, which were similar to those presented to the participant at school or in therapy. The experimenter instructed the participant to begin the activity and used least-to-most prompting following incorrect or

incomplete responses. Contingent on vocal stereotypy, the experimenter stated, "You don't have to do it," and removed the task for 30 s.

During the ignore condition, the participant and the experimenter were both present in the treatment room, which was absent of toys or activities. The experimenter withheld all social interaction (e.g., body orientation, eye contact, conversation, physical touch), and no programmed consequences were delivered contingent on vocal stereotypy.

During the toy play condition, the participant and experimenter were both present in the treatment room, and the participant had free access to toys and preferred leisure items. The experimenter provided frequent verbal attention did not present task demands. No programmed consequences were provided following vocal stereotypy.

Preference assessment. Each participant's parents filled out a modified Reinforcer Assessment for Individuals with Severe Disabilities (RAIS-D; Fisher, Piazza, Bowman, & Amari, 1996) to determine the types of auditory stimuli that s/he potentially preferred (e.g., specific songs, music with higher pitches, faster beats). From the RAIS-D, six potential auditory stimuli consisting of songs without words were selected for each participant and evaluated in a multiple stimulus without replacement (MSWO; DeLeon & Iwata, 1996) preference assessment. Prior to the MSWO, each auditory stimulus was correlated with a specific picture card; these cards were the stimuli selected within the MSWO assessment. The experimenter conducted the MSWO five times and scored the results using the weighted scoring procedures described by Ciccone, Graff, and Ahearn (2005). Each selection rank was assigned a score (e.g., rank 1 = 6, rank 2 = 5). If a selection did not occur, the stimulus or remaining stimuli were assigned a score of 0. At the end of the assessment, all scores (including zero scores) were summed to yield a total score. From the ranked list of six auditory stimuli, each participant's most preferred (highest weighted score), moderately preferred (median weighted score), and least preferred (lowest weighted score) stimulus was selected for assessment within the competing stimulus assessment.

Competing stimulus assessment. To identify the stimulus that most competed with vocal stereotypy, each participant had a single exposure to a no-interaction baseline and three individual test conditions consisting of his or her most preferred, moderately preferred, and least preferred auditory stimuli as identified in the MSWO. Each condition was 5 min. No demands were placed during any condition, and the experimenter did not interact with the participant outside of minimizing any potential safety hazards (e.g., climbing on furniture). There were no programmed consequences for engagement in vocal stereotypy or other challenging behavior (e.g., throwing objects). The percentage of intervals with vocal stereotypy was calculated by dividing the number of intervals during which vocal stereotypy occurred by the total number of intervals, multiplied by 100. The condition with the lowest percentage of vocal stereotypy was selected as the music used during subsequent experimental sessions.

RIRD and task probes. To identify RIRD commands and tasks within the participant's skill repertoire, the experimenter conducted probes of RIRD commands and task demands. These probes were similar to those in Love et al. (2012). After consulting with each participant's current therapists, the experimenter conducted two probes: (a) Commands requiring a vocal response, and (b) preacademic and academic tasks for each participant. RIRD commands included sound imitation (e.g., "Say 'eeeee'"; "Say "ah'"), word imitation (e.g., "Say 'shoe'"; "Say 'dog'"), fill-ins (e.g., "A cow says _____"; "You smell with your _____"), and answering questions (e.g., "How old are you?"; "What color is this item?"). Task demands included visual performance tasks (e.g., jigsaw

puzzles, shape sorters, picture sequences), fine motor tasks (e.g., stringing beads, placing pegs in a peg board), and writing tasks (e.g., typing sentences from a model).

RIRD commands were placed onto a list of commands for use during sessions contingent on the participants' emission of the target response for a minimum of 80% of trials within each probe. Prior to each experimental session, the experimenter randomly selected 10 RIRD commands from each participant's list to use during the RIRD procedure.

Task demands were chosen for inclusion in session activities if the participant could complete the task demand within 10 min with no more than model or gestural prompts from the experimenter during both probes. The experimenter chose five activities that could be completed within 10 min.

Experimental Design

The effects of RIRD only and MS + RIRD on the percentage of intervals with vocal stereotypy and the occurrence of on-task behavior were evaluated using an ABAB reversal design (Gast & Ledford, 2014). For each participant, MS was added to the RIRD procedure after data in the RIRD-only condition reflected a countertherapeutic trend (i.e., stereotypy increased and time on task decreased) for three consecutive sessions.

Procedure

All sessions began when the experimenter delivered the first task directive (e.g., string the beads), and lasted 10 min. If the participant stopped engaging in the task for at least 3 s but did not engage in vocal stereotypy, the experimenter prompted the participant back on task using a verbal, gestural, or physical prompt. Each participant had the same five activities available per session, and the activities were randomly ordered within each session. When participant finished an activity before the 10 min had elapsed, the experimenter presented the next activity on the list. When possible, certain materials were rotated to prevent practice effects (e.g., varying the specific nine-piece jigsaw puzzle used in each session). No more than four sessions occurred per day, and a minimum of 30 min elapsed between experimental sessions in a given day.

RIRD only. The experimenter directed the participant to the table and gave the designated S^D for the session task. Contingent on vocal stereotypy, the experimenter immediately interrupted the task and initiated RIRD using the same procedure as in Love et al. (2012). Three correct, consecutive responses without vocal stereotypy were required to end RIRD, at which time the participant was redirected back to the task with either a verbal or gestural prompt. General praise (e.g., "Good job") was delivered following a correct response within 5 s of a command during RIRD. If 5 s elapsed without a correct response, or the participant continued to engage in vocal stereotypy, the experimenter stated the correct answer and reissued the command (e.g., "A cat says [5 s elapses] meow. A cat says ___"). If the participant again did not emit the correct response or continued to engage in vocal stereotypy, the experimenter issued the next command on the list. If the participant emitted an appropriate vocalization or mand during the RIRD procedure, the experimenter responded either by praising the appropriate vocalization (e.g., "Nice using your words") or acknowledging the mand (e.g., "When we are done you can have ____") before continuing to issue commands. Items corresponding to participant mands were not made available during experimental sessions, but access to these items was provided outside of sessions.

MS + RIRD. The procedures for issuing task demands and implementing RIRD used during this condition were similar to those described in the RIRD-only condition; however, prior to the start of sessions, the experimenter used the

sound-level meter to confirm that the volume of the music coming through the headphones was within the prescribed limit of 40-45 dB. This decibel level was chosen to ensure that the participant could hear the RIRD commands, as typical conversation measures between 55-67 dB depending on the level of ambient noise in the environment (Pearsons, Bennett, & Fidell, 1977). The experimenter told the participant, "Today, you get to listen to music while you work," and placed the headphones on the participant before presenting the session task. The music that competed most effectively with each participant's vocal stereotypy was played on a loop throughout the entirety of the session, and headphones were not removed during the implementation of RIRD. Participants were free to remove the headphones during session, as no consequences were programmed contingent on headphone removal.

Generality probes. The experimenter conducted training to implement RIRD with the parents of each participant prior to the initiation of generality probes. Behavioral skills training consisted of instructions, modeling, rehearsal, and feedback (Ward-Horner & Sturmey, 2012). When parents could implement RIRD with at least 90% fidelity, generality probes were conducted. Data were collected on the same four dependent variables as during experimental sessions. Generality probes were shortened to a duration of 5 min, as both participants' parents indicated that this is the typical duration of their child's task engagement at home. The procedure remained the same as in the MS + RIRD condition.

For Elizabeth, the therapist collected data on the effectiveness of the intervention in the clinic when her mother implemented the same tasks as were completed during experimental sessions. For Matthew, the therapist collected data in his home when his mother delivered the intervention (a) while he completed the same instructional tasks conducting during experimental sessions in the clinic, and (b) during baking. This activity was chosen because Matthew's mother reported that it was something she did with Matthew frequently during which he engaged in vocal stereotypy with high frequency.

Social Validity

After completion of the intervention evaluation in the clinic, parents of each participant viewed a video recording of their child during RIRD only and a second recording during MS + RIRD. To determine what videos would be shown, participants were first randomly assigned to either the initial or secondary reversal to ensure that the videos came from directly adjacent treatment conditions, after which the video that most closely reflected the average levels of vocal stereotypy and time on task for that condition was selected. After viewing the recording, the parents completed a questionnaire adapted from the Treatment Evaluation Inventory – Short Form (TEI-SF; Kelly, Heffer, Gresham, & Elliot, 1989) measuring treatment acceptability for each condition. The questionnaire consisted of a set of nine statements that parents responded to using a Likert-type scale (1 = strongly disagree; 5 = strongly agree). From the completed surveys, the experimenter calculated the mean acceptability of both the RIRD-only and MS + RIRD conditions. When calculating mean acceptability, the score given by each participant's parents to item six was adjusted to address the reverse scoring of the item (i.e., if the parents rated this item a 1, strongly disagree, this was reversed to a 5 for calculation of mean acceptability).

RESULTS

The results of both participants' brief functional analyses indicated that their vocal stereotypy appeared to be maintained by automatic reinforcement (Elizabeth) or multiply controlled (Matthew). Elizabeth's and Matthew's vocal stereotypy was highest in ignore and

control conditions; however, Matthew also engaged in elevated levels of vocal stereotypy during the attention condition. Results of the competing stimulus assessment showed that Elizabeth's most preferred stimulus competed most effectively with her vocal stereotypy. By contrast, Matthew's moderately preferred stimulus was found to compete most effectively with his vocal stereotypy. (Data available upon request.)

Figure 1 depicts the percentage of intervals with vocal stereotypy and the percentage of on-task behavior for both participants. For Elizabeth, both vocal stereotypy and on-task behavior occurred at moderate levels during the initial RIRD-only condition (*M* = 36.4%; range, 21.7%-58.3%; *M* = 38.2%; range, 24.8%-66.7%, respectively). During MS + RIRD, vocal stereotypy immediately decreased to low levels (*M* = 11.3%; range, 1.7%-16.7%) and on-task behavior rapidly increased (*M* = 64.5%; range, 43.5%-77.1%). Elizabeth's vocal stereotypy immediately increased and exceeded initial levels during the return to the RIRD-only condition (*M* = 49%; range, 36.7%-66.7%), whereas her on-task behavior quickly decreased below initial RIRD-only levels (*M* = 30.5%; range, 4.6%-50.6%). When MS + RIRD was reintroduced, Elizabeth's vocal stereotypy returned to low levels (*M* = 13.1%; range, 6.7%-16.7%) and her on-task behavior increased above the level observed in the initial NC music + RIRD phase (*M* = 67%; range, 57.6%-72.4%). During the generality probe conducted by Elizabeth's mother in the clinic, Elizabeth's vocal stereotypy remained low (16.7%) and her on-task behavior remained elevated (64.8%).

For Matthew, vocal stereotypy occurred at high levels (*M* = 51.7%; range, 45%-58.3%), while on-task behavior occurred at moderate levels (*M* = 34.3%; range, 32.2% to 38.3%) during the initial RIRD-only condition. After introducing MS + RIRD, Matthew's vocal

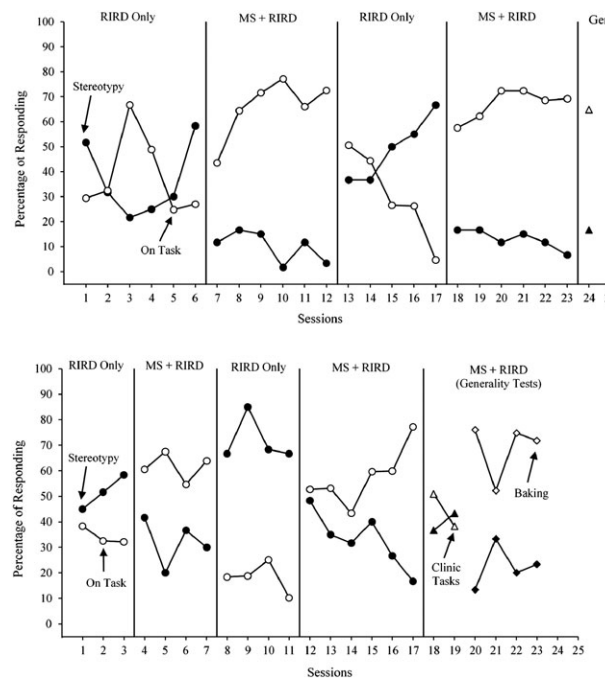

Figure 1. Results for Elizabeth (top panel) and Matthew (bottom panel). Closed points denote the percentage of intervals with vocal stereotypy, and open points denote the percentage of intervals with on-task behavior.

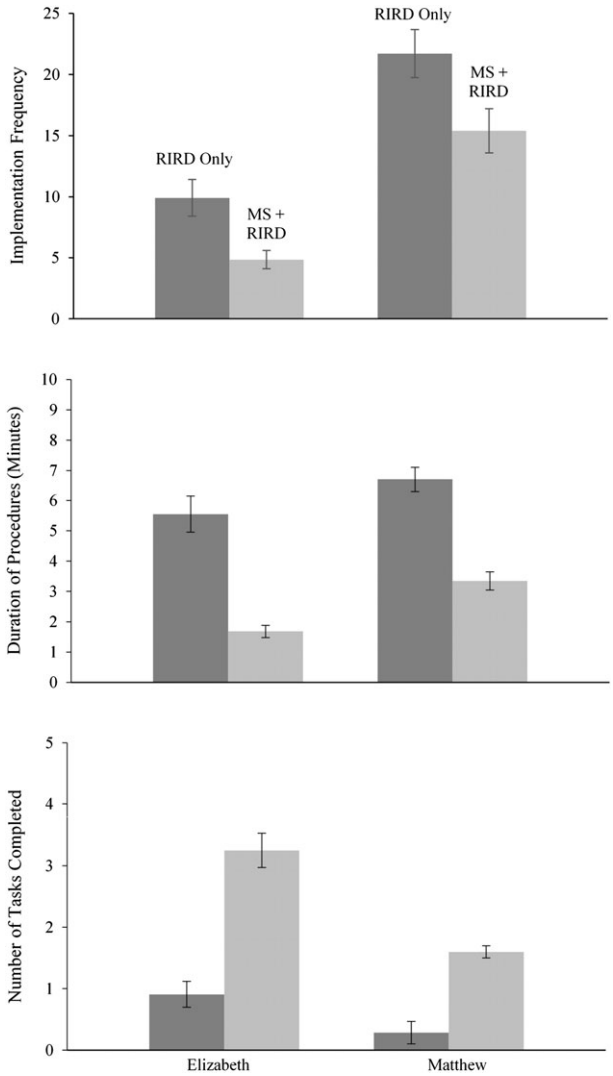

Figure 2. Averages for frequency of RIRD implementation (top), duration of RIRD procedures (middle), and tasks completed per session (bottom).

stereotypy decreased and on-task behavior increased (*M* = 32.1%; range, 20% to 41.7%; *M* = 61.6%; range, 54.6% to 67.5%, respectively). Once RIRD only was reintroduced, Matthew's vocal stereotypy quickly increased and exceeded levels in the initial RIRD-only condition (*M* = 71.7%; range, 66.7% to 85%) while on-task behavior immediately decreased and occurred well below levels in the initial RIRD-only condition (*M* = 18.1%; range, 10.2% to 25.1%). After MS + RIRD was reintroduced, Matthew's vocal stereotypy decreased (*M* = 33.1%; range, 16.7% to 48.3%) and his percentage of on-task behavior increased (*M* = 57.7%; range, 43.4% to 77.2%), with stereotypy on a decreasing trend and on-task behavior on an increasing trend at the conclusion of the intervention phase. During the instructional context generality probes conducted by Matthew's mother in his home, he

Table 1
Results from the Treatment Evaluation Inventory-Short Form (TEI-SF) for Elizabeth and Matthew

Question	Elizabeth		Matthew	
	RIRD Only	MS + RIRD	RIRD Only	MS + RIRD
1. I find this intervention to be an acceptable way of dealing with my child's vocal stereotypy	3	5	4	5
2. I would be willing to use this intervention at home to address my child's vocal stereotypy	2.5	4.5	3	5
3. I believe that it would be acceptable to use this intervention without my child's consent	4	4.5	5	5
4. I like the procedures used in this intervention	2.5	5	5	5
5. I believe this intervention is effective for my child	2	5	5	5
6. I believe that my child experiences discomfort during this intervention	4	1	1	1
7. I believe this intervention is likely to result in permanent improvement in the amount my child engages in vocal stereotypy	2.5	4.5	3	5
8. I believe that it would be acceptable to use this intervention with children who cannot choose interventions for themselves	3	5	4	5
9. Overall, I have a positive reaction to this intervention	2	5	4	5

Note. Items were rated on a Likert-type scale of 1-5 with 1 indicating strongly disagree, 3 indicating neutral, and 5 indicating strongly agree. Item 6 was reverse scored.

demonstrated a slightly higher percentage of vocal stereotypy (M = 40%) and a lower percentage of on-task behavior (M = 44.6%) as compared to sessions conducted by the experimenter in the clinic. However, when his mother implemented MS + RIRD during the natural context generality probes, Matthew had lower levels of vocal stereotypy (M = 22.5%) and higher levels of on-task behavior (M = 68.7%)

Figure 2 displays the average frequency of RIRD (top), duration of RIRD implementation (middle), and the average number of completed tasks in each condition (bottom) for both participants. RIRD was implemented more frequently during the RIRD-only condition for both Elizabeth (M = 9.9) and Matthew (M = 21.7) as compared to implementations during the MS + RIRD condition (M = 4.8 for Elizabeth; M = 15.4 for Matthew). More time was spent implementing RIRD during the RIRD-only conditions for both Elizabeth (M = 5.5 min) and Matthew (M = 6.7 min) as compared to the MS + RIRD conditions (M = 1.7 min for Elizabeth; M = 3.3 min for Matthew). Fewer tasks were completed during the RIRD-only conditions by both Elizabeth (M = 0.9 tasks) and Matthew (M = 0.3 tasks) as compared to the MS + RIRD conditions (M = 3.3 tasks for Elizabeth; M = 1.6 tasks for Matthew).

Table 1 depicts the item-by-item results of the TEI-SF (Kelly et al., 1989). Across both participants, MS + RIRD was rated more favorably than RIRD only. For RIRD only, Elizabeth's parents scored a mean acceptability of 2.6, while Matthew's parents found the intervention more acceptable and scored a mean acceptability of 4.2. By contrast, for MS + RIRD, Elizabeth's parents scored a mean acceptability of 4.8, and Matthew's parents scored a mean acceptability of 5. Both parents indicated that they would like to use the MS + RIRD intervention at home, and Elizabeth's parents expressed interest in pursuing the implementation of the intervention within her classroom.

DISCUSSION

The addition of MS to RIRD decreased vocal stereotypy and increased on-task behavior for

both participants. Also, the combined intervention decreased both the frequency and the duration of RIRD implementations for each participant, potentially making RIRD less cumbersome to implement with fidelity (Carroll & Kodak, 2014). O'Reilly et al. (2008) argued for similar modifications in instances in which implementation of a solely consequence-based intervention may interrupt ongoing activities. For example, implementation of RIRD during instructional activities similar to those implemented in the current study may result in loss of instructional time, even though vocal stereotypy is reduced. When these instances are present, the addition of an antecedent strategy (e.g., NC music) may be preferable as it may create an abolishing operation to engage in stereotypic behavior, which reduces the probability that the consequence-based intervention will need to be implemented frequently. Thus, the combination of NC music as an antecedent intervention paired with RIRD as a consequence-based intervention not only reduced vocal stereotypy, but also increased both participants' availability for instruction in a relatively nonobtrusive fashion.

The current findings are partially consistent with the results of Love et al. (2012), which demonstrated a reduction in vocal stereotypy when RIRD was combined with MS from musical toys. The reduction in vocal stereotypy following the addition of auditory stimulation supports the hypothesis that noncontingent access to auditory stimuli functions as an abolishing operation for engaging in vocal stereotypy. In contrast to the results of Love et al. (2012), however, RIRD only was ineffective in decreasing vocal stereotypy to clinically significant levels. Both participants in the current study had higher levels of vocal stereotypy during RIRD only. The results also further extend the findings by Rapp (2006, 2007) that stereotypy is responsive to the manipulation of MOs given that stereotypy could still occur while noncontingent music was present.

The current investigation failed to replicate the results of previous studies on RIRD (e.g., Ahearn et al., 2007; Liu-Gitz & Banda, 2010; Wells et al., 2016), as both participants demonstrated accelerating trends in vocal stereotypy during one (Elizabeth) or both (Matthew) of the RIRD-only conditions. These findings are potentially the result of uninterrupted measurement procedures that included data on the occurrence of vocal stereotypy while RIRD was implemented (Carroll & Kodak, 2014; Wunderlich & Vollmer, 2015). It is clear that RIRD has been an effective intervention for reducing vocal stereotypy for some participants (Ahearn et al., 2007; Liu-Gitz & Banda, 2010; Wells et al., 2016). However, in the current investigation, the onset of RIRD may have evoked additional vocal stereotypy for both Elizabeth and Matthew. These data support the hypothesis that automatically reinforced responses may increase when socially mediated punishment (e.g., contingent demands) is implemented (Wunderlich & Vollmer, 2015), which may not be detected if data collection procedures are interrupted.

However, caution must be observed when interpreting these results as a limitation of the current study is the absence of a condition evaluating the effects of MS alone, which deviated from the procedures described in Love et al. (2012). Noncontingent music has successfully been used to reduce immediate engagement in vocal stereotypy (Lanovaz, Sladeczek, & Rapp, 2011; Saylor, Sidener, Reeve, Featherston, & Progar, 2012), though we are not aware of any evaluations of its effect on on-task behavior. Given that RIRD has appeared to function as a punishment procedure for some individuals (Ahrens, Lerman, Kodak, Worsdell, & Keegan, 2011; Shawler & Miguel, 2015) and that MS + RIRD was found to be more socially valid to the current participants' parents, future research should compare the effects of MS alone to both RIRD alone and MS + RIRD on vocal stereotypy and on-task

behavior. These evaluations can help identify whether MS alone may have enough of an abative effect on stereotypy to increase the probability of on-task behavior, absent the implementation of RIRD.

Few published studies demonstrate generality of the effects of RIRD outside of a controlled environment or with relevant implementers (Cassella, Sidener, Sidener, & Progar, 2011; Frewing, Tanner, Bonner, Baxter, & Pastrana, 2015). Although Frewing et al. (2015) demonstrated generality of treatment effects to additional environments, there were no reported attempts to extend the effects of RIRD to additional implementers. By contrast, Cassella et al. (2011) attempted to demonstrate the generality of the reductive effect of RIRD on vocal stereotypy across multiple settings and therapists; vocal stereotypy remained at baseline levels, however. In the current study, generality probes were completed with the participants' parents conducting sessions in both the clinic (Elizabeth) and at home across two contexts (Matthew). As previously reported, the results of the generality probes were positive; both participants had reduced stereotypy and increased time on-task when MS + RIRD was implemented by parents (Elizabeth) outside of the clinical setting and within a natural, rather than instructional, context (Matthew). However, a limitation of the current study is the absence of baseline measures of vocal stereotypy and on-task behavior with caregivers in the natural environment prior to treatment. As a result of this limitation, the conclusions that can be drawn regarding the efficacy of the intervention within these untrained contexts are somewhat weakened. Further research may be necessary to more precisely understand the generality of effects of MS + RIRD on behavior.

Similar to the inclusion of data on the generalization of treatment effects, few published studies evaluating the effects of either MS (e.g., Saylor et al., 2012) or RIRD (e.g., Cassella et al., 2011; Giles, St. Peter, Pence, & Gibson, 2012; Shawler & Miguel, 2015) on vocal stereotypy have included data on the social validity of their intervention. In instances where these data were collected and reported, the results are mixed. Giles et al. (2012) reported high social validity for RIRD to reduce motor stereotypy with participants themselves using a concurrent chain procedure to determine preference for response blocking or motor RIRD as a behavior reduction procedure for motor stereotypy. These strong social validity results are similar to that found in the current investigation, as well as others (e.g., Love et al., 2012; Saylor et al., 2012). In comparison, other evaluations of the social validity of RIRD have produced weaker or mixed results. For instance, Cassella et al. (2011) reported mixed results, with the caregivers of one participant finding RIRD to be highly acceptable, while the caregivers of the second participant reported lower treatment acceptability. In their comparison of the effects of motor RIRD as compared to vocal RIRD, Shawler and Miguel (2015) reported that both procedures received social validity scores only slightly above a 'neutral' designation by participants' parents. Future research should include social validity data on the acceptability of both MS and RIRD with key stakeholders in participants' daily lives. Additional research into the variables that may impact social acceptability outcomes may also be warranted.

Two additional limitations deserve consideration and may serve as areas of further inquiry. First, we did not measure appropriate vocalizations during experimental sessions, although previous investigations found that appropriate vocalizations increased during RIRD (e.g., Ahearn et al., 2007; Love et al., 2012). Although MS + RIRD resulted in increases in on-task behavior within the current investigation, it is unknown whether the intervention would result in a comparable increase in appropriate vocalizations. To extend the current findings, future researchers may measure

appropriate vocalizations to evaluate whether MS + RIRD results in a concurrent increase in appropriate vocalizations and on-task behavior.

Last, the specific aspect of the music that suppressed stereotypic behavior is not known. Parents completed a RAIS-D from which hypothesized preferred musical qualities were identified (e.g., music with high beats per minute, music with higher pitches), and a preference assessment was conducted with songs that had these qualities. As preference has been determined to play a role in response competition to reduce automatically maintained behavior (Ahearn, Clark, DeBar, & Florentino, 2005), and highly preferred music has demonstrated a greater abative effect on vocal stereotypy than lesser preferred music (Lanovaz, Rapp, & Ferguson, 2012), the identification of musical selections containing the specific preferred quality may enhance the effects of the intervention. Given the wide array of musical selections that could potentially contain a specific quality (e.g., beats per minute, pitch, dynamic), the use of a preference assessment designed to identify categories of preference, as described by Ciccone, Graff, and Ahearn (2015), may be indicated. Future research may extend the current results by examining the properties of music that suppress stereotypy.

Our data suggest that the use of an intervention package that includes antecedent- and consequence-based strategies can reduce the occurrence of vocal stereotypy and increase the occurrence of on-task behavior. Further, the results suggest that the addition of MS can reduce both the frequency and duration of RIRD implementation, potentially facilitating the use of RIRD outside of controlled settings. Additional research on MS + RIRD is necessary to evaluate the effectiveness of these procedures with individuals for whom RIRD only does not produce a reduction in vocal stereotypy. Researchers also should further evaluate the utility of MS + RIRD in natural environments, and determine its efficacy when implemented by parents, teachers, and other practitioners.

REFERENCES

Ahearn, W. H., Clark, K. M., DeBar, R., & Florentino, C. (2005) On the role of preference in 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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Arlington, VA: American Psychiatric Pub.

Carroll, R. A., & Kodak, T. (2014). An evaluation of interrupted and uninterrupted measurement of vocal stereotypy on perceived treatment outcomes. *Journal of Applied Behavior Analysis, 47*, 264-276. https://doi.org/10.1002/jaba.118

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

Ciccone, F. J., Graff, R. B., & Ahearn, W. H. (2005). An alternate scoring method for the multiple stimulus without replacement preference assessment. *Behavioral Interventions, 20*, 121-127.

Ciccone, F. J., Graff, R. B., & Ahearn, W. H. (2015). Increasing the efficacy of paired-stimulus preference assessments by identifying categories of preference. *Journal of Applied Behavior Analysis, 48*, 221-226. https://doi.org/10.1002/jaba.190

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*, 519-532.

Dickman, S. E., Bright, C. N., Montgomery, D. H., & Miguel, C. F. (2012). The effects of response interruption and redirection (RIRD) and differential reinforcement on vocal stereotypy and appropriate vocalizations. *Behavioral Interventions, 27*, 185-192. https://doi.org/10.1002/bin.1348

Durand, V. M., & Carr, E. G. (1987). Social influences on "self-stimulatory" behavior: Analysis and

treatment application. *Journal of Applied Behavior Analysis, 20*, 119-132. https://doi.org/10.1901/jaba.1987.20-119

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

Frewing, T. M., Tanner, A., Bonner, A. C., Baxter, S., & Pastrana, S. J. (2015). The immediate and subsequent effects of response interruption and redirection on vocal stereotypy, motor stereotypy, and heart rate. *European Journal of Behavior Analysis, 16*, 351-361.

Gast, D. L., & Ledford, J. R. (2014). *Single subject research methodology in behavioral sciences*. Routledge: New York.

Giles, A. F., St. Peter, C. C., Pence, S. T., & Gibson, A. B. (2012). Preference for blocking or response redirection during stereotypy treatment. *Research in Developmental Disabilities, 33*, 1691-1700. https://doi.org/10.1016/j.ridd.2012.05.008

Kelly, M. L., Heffer, R. W., Gresham, F. M., & Elliot, S. N. (1989). Development of a modified treatment evaluation inventory. *Journal of Psychopathology and Behavioral Assessment, 11*, 235-247.

Koegel, R. L., & Covert, A. (1972). The relationship of self-stimulation to learning in autistic children. *Journal of Applied Behavior Analysis, 5*, 381-387. https://doi.org/10.1901/jaba.1972.5-381.

Lanovaz, M. J., Rapp, J. T., & Ferguson, S. (2012). The utility of assessing musical preference before implementation of noncontingent music to reduce vocal stereotypy. *Journal of Applied Behavior Analysis, 45*, 845-851. https://doi.org/10.1901/jaba.2012.45-845

Lanovaz, M. J., & Sladeczek, I. E. (2012). Vocal stereotypy in individuals with autism spectrum disorders: A review of behavioral interventions. *Behavior Modification, 36*, 146-164. https://doi.org/10.1177/0145445511427192

Lanovaz, M. J., Sladeczek, I. E., & Rapp, J. T. (2011). Effects of music on vocal stereotypy in children with autism. *Journal of Applied Behavior Analysis, 44*, 647-651. https://doi.org/10.1901/jaba.2011.44-647

Laraway, S., Snycerski, S., Michael, J., & Poling, A. (2003). Motivating operations and terms to describe them: Some further refinements. *Journal of Applied Behavior Analysis*, *36*, 407-414. https://doi.org/10.1902/jaba.2003.36-407

Liu-Gitz, L., & Banda, D. R. (2010). A replication of the RIRD strategy to decrease vocal stereotypy in a student with autism. *Behavioral Interventions, 25*, 77-87. https://doi.org/10.1002/bin

Love, J. L., 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

Northup, J., Wacker, D., Sasso, G., Steege, M., Cigrand, K., Cook, J., & DeRaad, A. (1991). A brief functional analysis of aggressive and alternative behavior in an outclinic setting. *Journal of Applied Behavior Analysis, 24*, 509-522. https://doi.org/10.1901/jaba.1991.24-509

O'Reilly, M. F., Sigafoos, J., Lancioni, G., Rispoli, M., Lang, R., Chan, J., ..., Langthorne, P. (2008). Manipulating the behavior-altering effect of the motivating operation: Examination of the influence on challenging behavior during leisure activities. *Research in Developmental Disabilities, 29*, 333-340.

Pearsons, K., Bennett, R., & Fidell, S. (1977). *Speech levels in various noise environments*. Washington, DC: Office of Health and Ecological Effects, Office of Research and Development, U.S. Environmental Protection Agency.

Piazza, C. C., Adelinis, J. D., Hanley, G. P., Goh, H., & 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.

Rapp, J. T. (2006). Toward an empirical method for identifying matched stimulation for automatically reinforced behavior: A preliminary investigation. *Journal of Applied Behavior Analysis, 39*, 137-140.

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

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

Shawler, L. A., & Miguel, C. F. (2015). The effects of motor and vocal response interruption and redirection on vocal stereotypy and appropriate vocalizations. *Behavioral Interventions, 30*, 112-134. https://doi.org/10.1002/bin

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland adaptive behavior scales: (Vineland II), survey interview form/caregiver rating form*. Livonia, MN: Pearson Assessments

Ward-Horner, J., & Sturmey, P. (2012). Component analysis of behavior skills training in functional analysis. *Behavioral Interventions, 27*, 75-92. https://doi.org/10.1002/bin.1339

Wells, J., Collier, L., & Sheehey, P. H. (2016). Decreasing vocal stereotypy of a youth with autism in a classroom setting. *Child & Family Behavior Therapy, 38*, 164-174. https://doi.org/10.1080/07317107.2016.1172887

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

Received June 19, 2017
Final acceptance November 5, 2017
Action Editor, Tiffany Kodak