Discriminative Control of Vocal Stereotypy Using an Empirically Identified Punishment Procedure

By
© 2021
Hayley D. Hypse
B.G.S., University of Kansas, 2013

Submitted to the graduate degree program in The Department of Applied Behavioral Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

Chair: David P. Jarmolowicz, Ph.D.

Florence D. DiGennaro Reed, Ph.D., BCBA-D

Allison Tetreault, Ph.D., BCBA-D

Date Defended: 5 May 2021

The thesis committee for Hayley D. Hypse certifies that this is the approved version of the following thesis:

Discriminative Control of Vocal Stereotypy Using an Empirically Identified Punishment Procedure

Chair: David P. Jarmolowicz, Ph.D.

Date Approved: 6 June 2021

Abstract

The purpose of Study 1 was to evaluate whether two reinforcement procedures could reduce vocal stereotypy (VS) as previous literature has stated. We examined the VS of three boys diagnosed with an Autism Spectrum Disorder (ASD). Results of functional analyses indicated that each participant's VS was maintained by nonsocial consequences. Secondly, we examined noncontingent reinforcement (NCR) with a matched and unmatched stimulus. We then layered differential reinforcement of other behavior (DRO) onto of the implementation of NCR. Results indicated that both procedures were successful at the reduction of VS. Noncontingent reinforcement with a matched stimulus had the most success at the reduction of VS, but was deemed inappropriate outside of the research setting. Noncontingent reinforcement with an unmatched stimulus plus DRO were not as robust in the reduction of VS. Therefore, the purpose of study 2, was to evaluate whether an empirically identified punishment procedure could reduce VS and if those reductions could be maintained via inhibitory stimulus control. A punisher selection interview with their clinicians was then implemented to identify socially acceptable punishers. After an effective punishment procedure was identified, discrimination training was then conducted to bring the responses under stimulus control. Results showed that VS decreased to low levels in the presence of the inhibitory stimulus for all three participants. However, several modifications were required throughout the treatment evaluation and stimulus control was only established for one participant.

Keywords: vocal stereotypy, reinforcement, stimulus control, punishment, autism, and discrimination training.

Table of Contents

Discriminative Control of Vocal Stereotypy Using an Empirically Identified Punishment

Stereotypy is one of the core diagnostic features of Autism Spectrum Disorder (ASD) (American Psychiatric Association, 2013). The *Diagnostic Statistical Manual of Mental Disorders 5th Edition* (DSM-5) defines stereotypy as restricted, repetitive patterns of behavior, such as motor movements or speech, which may be maintained by nonsocial consequences (American Psychiatric Association, 2013). There are at least two forms of stereotypy, motor stereotypy (e.g., hand flapping, body rocking) and vocal stereotypy (VS; e.g., repetitive use of language, indistinguishable sounds). The DSM-5 states that stereotypical behaviors can cause clinically significant social or occupational impairment, or impairment in other important areas of functioning (e.g., communication). In their systematic review, Chebli and colleagues (2016) found that 88% of participants with developmental disabilities displayed stereotypical behaviors.

There is a vast literature on the treatment of motor stereotypy (e.g., Brusa & Richman, 2008; DeRosa et al., 2019; Doughty et al., 2007; Giles et al., 2012; Potter et al., 2013; Tiger et al., 2016; Verriden & Roscoe, 2018). While there are relatively fewer studies on the treatment of VS, research is emerging (e.g., Ahearn et al., 2007; Piazza et al., 2000; Rapp et al., 2009; Rapp et al., 2017; Watkins et al., 2011). Unfortunately, there are still gaps in the literature on the effectiveness of treatment for VS. For instance, most studies have been conducted with school-aged (5-12 years-old) children diagnosed with ASD (e.g., Ahearn et al., 2007; Ahrens et al., 2011; Love et al., 2012; O'Connor et al., 2011; Rapp et al., 2017; Rapp et al., 2009; Watkins et al., 2011). Interventions need to be conducted with younger children as well as adults diagnosed with ASD and other intellectual disabilities to further examine the efficacy of the treatment methods. Furthermore, most studies have been conducted in highly controlled environments. This leaves the effectiveness of treatment in a less-controlled environment (e.g., classroom,

home, habilitative programs) unknown. Finally, there is a need for more research directly comparing different interventions. Outcomes of such studies would equip clinicians with information to allow them to select and deliver the most effective treatments for their clients (Van Houten et al. 1988).

Rapp and Vollmer (2005) indicated that stereotypy is generally maintained by automatic reinforcement and consists of repetitive body movements or vocalizations. Thus, VS consists of repetitive vocalizations that tend to exist in the absence of social consequences. These repetitive vocalizations can include grunting (e.g., Ahearn et al., 2003), laughing (e.g., Gibney et al., 2019), humming (e.g., Taylor et al., 2005), and repeating words previously heard (e.g., Mancina et al., 2000). Vocal stereotypy is problematic when it interferes with learning and when individuals are ostracized by their peers due to their behavior (Cunningham & Schreibman, 2008; Koegel & Covert, 1972; Matson et al., 1997).

Vocal stereotypy is common in individuals with ASD (Mayes & Calhoun, 2011) and is often associated with negative outcomes, such as a delay in acquiring skills (e.g., Dunlap et al., 1983; Matson & Nebel-Schwalm, 2007), stigmatization (e.g., DiGennaro Reed et al., 2012), and impaired social interactions (e.g., Wolery et al., 1985). For some individuals, VS can impair participation in their daily activities (Koegel & Covert, 1972; Matson et al., 1997).

Stereotypic behaviors are displayed by typically developing children throughout their early years (Foster, 1998). Such behaviors, however, decrease in frequency over time (MacDonald et al., 2007). By contrast, stereotypy tends to persist in children with ASD. For example, MacDonald et al. (2007) compared stereotypical behavior (both vocal and motor) in children with autism and typically developing children. They observed the behaviors of children aged 2–4 years while playing and while undertaking academic tasks (e.g., motor imitation,

receptive language). On average, stereotypy increased with age in children with autism, but stayed the same or decreased slightly with age for typically developing children. As a result, when comparing the same age across both populations, the 4-year-olds with autism exhibited a significantly higher mean level of stereotypical behavior than the typically developing 4-year-old children (3.1% for typically developing children versus 38.8% for children with ASD).

Stereotypical behaviors - including VS - can severely restrict young people with ASD's peer/adult interactions and access to general education settings and the community (Cunningham & Schreibman, 2008; Dunlap et al., 1983; Jones et al., 1990; Wolery et al., 1985). Social restrictions resulting from VS may eliminate opportunities for individuals with ASD to live independent lives. More than half of young adults with ASD remain unemployed and unenrolled in higher education (Autism Speaks, 2021), leaving them reliant on caregivers. This outcome limits their opportunity to develop adaptive and social skills (Matson et al., 1997; Matson et al., 2006).

Individuals with ASD who engage in VS tend to demonstrate a delay in social development, learning, and independent functioning (Matson et al., 2006; Koegel & Covert, 1972), which can result in their being ostracized (Jones et al., 1990). Jones et al. (1990), for instance, assessed how individuals with developmental disabilities who displayed stereotyped behaviors were viewed by a large sample of young people. Two hundred and five 14–15-year-old participants were divided into four groups and viewed a recording of an actor engaging in routine household tasks (e.g., buttering toast, pouring water from a kettle into a cup). For two groups, the actor was seen engaging in typical behaviors, while for the other two groups, she

performed stereotypical behaviors.[1] Of the groups that observed the typical behaviors, one was informed that she was a university student, while for the other was told that she had a disability[2]. The same information was given to the other two groups with the exception that they observed the actor engaging in stereotypical behaviors. All participants filled out a questionnaire after watching the recording. Significantly more negative responses were attributed to the actor when she presented stereotypies, regardless of her disability status. Thus, participants perceived an individual displaying stereotypical behavior more negatively than an individual who did not. The disability label did not influence the nature of the responses attributed to the actor; rather, her behavior did. Thus, there are negative social consequences of stereotypy, potentially limiting individuals' social interactions within the community and/or restricting their opportunities for a fully integrated lifestyle. Jones et al. (1990) demonstrated that stereotypic behaviors can lead to individuals with intellectual disabilities being ostracized and, thus, impeding social opportunities with peers, learning opportunities within the community, and potential independent functioning.

Lanovaz and Sladeczek (2012) reviewed the research on behavioral interventions for reducing VS in individuals with ASD. They found that many behavioral approaches successfully treated VS with either antecedent-based interventions and/or consequence-based interventions. Antecedent-based interventions are designed to reduce the target behavior by manipulating the environmental circumstances that precede the target behavior (Wong et al., 2015). Consequence-based interventions decrease the future frequency of the target behavior by

[1] Typical behaviors are being referred to as non-restricted and/or repetitive behaviors. Behaviors that typically developing individuals engage in (e.g., not engaging in hand flapping, body rocking).

[2] Participants were informed that the actor was mentally handicapped.

manipulating environmental events that occur immediately after the target behavior is exhibited (Thomeer et al., 2017).

Antecedent-Based Interventions

Noncontingent Reinforcement

A common reinforcement-based intervention used to address automatically reinforced VS is the manipulation of motivating operations through noncontingent reinforcement (NCR) (e.g., Enloe & Rapp, 2014; Lanovaz et al., 2013; Lanovaz et al., 2011; Rapp et al., 2017; Saylor et al., 2012). Noncontingent reinforcement either involves time-based or continuous delivery of reinforcers (Vollmer, 1994). Lanovaz et al. (2011) studied the effects of manipulating the volume of music on the engagement in VS for two children aged 5- and 6-years-old, who were diagnosed with ASD. They used a reversal design that was combined with a three-component multiple-schedule and a multielement design in order to test effects of the volume of the music on VS. The authors found that noncontingent access to music (for 5 min) decreased VS and that the volume of the music did not impact VS. However, when continuous access to music was removed, it produced insignificant effects on VS.

In another study employing music as a treatment method for the reduction of VS, Saylor and colleagues (2012) compared three types of auditory stimulation – music, recordings of participants own VS, and white noise – for two adolescents diagnosed with ASD. The authors used a reversal design with an embedded alternating treatments design. Results demonstrated that white noise was the least effective and had similar percentage of time spent engaging in VS as in the baseline phase. Voice recording of the participant's VS demonstrated significantly lower levels of VS in comparison to the baseline phase. However, for both participants, the music caused VS to drop to zero levels of responding.

Matched/Unmatched Stimuli. An alternative form of NCR is the implementation of matched or unmatched stimulation. Matched stimulation can be defined as stimulation that shares similar properties to the stimulation produced by the stereotypy. Unmatched stimulation can be defined as stimulation that does not share any properties with the stimulation produced by the stereotypy (Piazza et al., 2000).

Rapp (2007) examined the effects of preferred stimulation (e.g., musical keyboard, music from CD player, blocks, figurines) on VS in two 9-year-old boys with ASD. Both participants' VS was assessed and found to be maintained through automatic reinforcement. Rapp conducted a free-operant preference assessment (Roane et al., 1998) to identify preferred objects that were either matched or unmatched to the stimulation generated by VS. Based on the results, matched stimulation was selected for one participant (Brian) and unmatched stimulation was selected for the other participant (Nevin). A reversal design was used to evaluate the effects of music and matched and unmatched toys separately and in combination. Vocal stereotypy reduced when Brian manipulated toys that generated auditory stimulation. However, when given the same toys that did not generate auditory stimulation, VS occurred frequently. Results for Nevin revealed that the letter board and blocks were not as effective as auditory stimulation from the CD player for decreasing his VS. Nevin's VS was low during music, moderate during music and toys, and high during no-interaction and toys conditions. These results suggest that the unmatched toys exerted little or no effect on Nevin's VS. Therefore, results demonstrated that auditory stimulation decreased VS for both participants and that Nevin's VS increased when toys were added to music (i.e., matched stimulation).

Ahearn and colleagues (2005) also examined the effects of matched and unmatched stimulation on motor stereotypy and VS for two boys with ASD (ages 11- and 13-years old). For

both participants, motor stereotypy and VS was assessed; data suggested that their stereotypy was maintained by automatic reinforcement. The researchers implemented a multielement design to compare baseline levels of stereotypy to a condition in which the participant had continuous access to a high preferred item that was assumed to match (large therapy ball for Tim; a videotape for Cris) or not match (blocks for Tim; books for Cris) the stimulation produced by their stereotypy. Items were assessed alone and compared across conditions. Stereotypies decreased for both participants in the matched and unmatched stimulation conditions. However, unmatched stimulatory items were associated with the lowest levels of stereotypy for both participants. These results suggest that the effects of unmatched and matched stimulation, in comparison to previously described studies on VS, were idiosyncratic.

Stimulus Control

Another antecedent-based intervention used to decrease VS is the utilization of stimulus control. Stimulus control is an intervention that uses a discriminative stimulus (e.g., a red bracelet) to signal that reinforcement is available contingent on a particular response (Malott, 2007) and can use a discriminative stimulus for punishment (e.g., a black bracelet) to signal that punishment will be delivered contingent on a particular response. Once the targeted behavior comes under consequence control, this control can be maintained by presenting antecedent stimuli that place the target behavior under stimulus control. However, to place the target behavior under stimulus control discrimination training must occur (see discrimination training below).

Consequence-Based Interventions

Differential Reinforcement of Other Behavior

A common consequence-based procedure to decrease VS is differential reinforcement of other behavior (DRO; e.g., Mancina et al., 2000; Taylor et al., 2005). Taylor et al. (2005) compared the effects of fixed-time reinforcement (NCR; every 60 s) and a DRO schedule of reinforcement (1-min interval) on the occurrence of VS, using an ABCBC reversal design. During fixed-time reinforcement, 30 s of access to preferred auditory toys was delivered every 60 s. During the DRO, 30 s of access to preferred auditory toys was delivered contingent on the absence of VS for 60 s. If VS occurred during the DRO, the participant was told 'No, that's not quiet, I have to reset your time.' The timer was then reset for 60 s. Results revealed that fixed-time access to auditory stimuli did not reduce VS. However, the DRO decreased VS to low levels. The researchers were able to successfully increase the DRO interval to 10 min. This procedure yielded a promising reduction of VS.

Similarly, Mancina and colleagues (2000) taught a 12-year-old girl diagnosed with ASD to monitor her own stereotypic behaviors (e.g., humming, tongue clicking, echolalic words/phrases) through the implementation of a 5-s and 10-s DRO (Study 1). A multiple baseline design across tasks was used to teach the participant to use self-management procedures. The participant was taught to correctly identify her VS through modeling and then was taught self-reinforcement (e.g., popcorn, cereal, stickers) in the absence of VS. The participant was required to refrain from VS originally for 5 s with an increase to 10 s. The participant delivered self-reinforcement contingent on the absence of VS for the specified duration. The authors successfully reduced VS through the implementation of the DRO. Nonetheless, Mancina and colleagues extended their findings and trained the participant's teacher to implement the treatment in Study 2. The participant's teacher was trained on the implementation of the self-management procedure (procedures were identical to Study 1).

However, there was limited independence for the participant in Study 2. The participant frequently required gestural prompts from the teacher. Consequently, the teacher was not able to fade out support or proximity from the participant. Hence, the implementation of the procedure could be too laborious for teachers or paraprofessionals to implement with integrity throughout the student's school day.

Response Interruption and Redirection

Multiple studies have used response interruption and redirection (RIRD) to decrease VS (e.g., Ahearn et al., 2007; Ahrens et al., 2011; Cassella et al., 2011; Liu-Gitz & Banda, 2010; Pastrana et al., 2013; Shawler & Miguel, 2015; and Wunderlich & Vollmer, 2015). Response interruption and redirection is a consequence-based procedure used to treat stereotypic behaviors (i.e., motor and/or vocal stereotypy) by interrupting the participant's stereotypy. Typically, these interruptions consist of the researchers delivering multiple instructions (e.g., motor or vocal response) contingent on the occurrence of stereotypy (Ahearn et al., 2007).

Ahearn and colleagues (2007) established the methodology for RIRD. They implemented the RIRD procedure (vocal demands) within an ABAB reversal design to evaluate whether RIRD could successfully reduce VS. The procedure was evaluated with four children with ASD whom engaged in VS maintained by nonsocial consequences. Specifically, during RIRD, vocal demands were contingent on VS. These demands continued until participants completed three consecutive vocal demands without engaging in VS. All participants demonstrated lower levels of VS in the RIRD phase in comparison to the baseline phase.

Ahrens et al. (2011) replicated and extended the above study by using both vocal and motor demands to treat VS. In the first study, the authors utilized a combined reversal and multielement design to examine the effects of the two specific RIRD techniques (motor demands

and vocal demands). During study 1, vocal RIRD and motor RIRD were used within the same phase for VS. For example, if the participant engaged in VS the therapist would ask social questions that required a vocal response during vocal RIRD. During motor RIRD, motor imitation behaviors were implemented contingent on VS. In study 2, the authors implemented a combined reversal and multielement design to compare the effects of vocal and motor RIRD on both vocal and motor stereotypy for two the participants. The same procedures from Study 1 were utilized with the exception of prompts being delivered between 2-3 seconds instead of 5 seconds. The authors found that RIRD was effective at the reduction of VS regardless of the technical difference or the topography of the stereotypy (i.e., motor or vocal stereotypy).

Ahearn et al. (2007) and Ahrens et al. (2011) both successfully used RIRD to reduce VS. Both studies, however, used discontinuous measurement (i.e., data collection for VS was interrupted/paused when RIRD was being implemented). This procedural choice has been varied across studies attempting to replicate RIRD. For example, Wunderlich and Vollmer (2015) repeated the same procedures as Ahearn et al. (2007) while comparing continuous and discontinuous data collection. Continuous data collection consisted of collecting data on VS outside and during the implementation of RIRD (i.e., percentage of each session in which the subject engaged in VS excluding RIRD intervention time). Discontinuous data collection consisted of collecting data on VS outside of the implementation of RIRD (i.e., percentage of each session in which the subject engaged in VS during the entirety of the session). There were seven participants, six diagnosed with ASD and one diagnosed with Trisomy 9 and intellectual disabilities. Participants were between 4 to 20 years of age. Response interruption and redirection was evaluated using a reversal design for three participants. Vocal RIRD and motor RIRD sessions were alternated during the treatment condition. To evaluate the data collection

and analysis of procedures used in RIRD, data collection was graphed for both continuous and discontinuous data. The researchers found that RIRD appeared to be effective for reducing VS when data collection was not continuous (i.e., VS was excluded from data collection during the implementation of RIRD). However, when data collection was continuous, RIRD did not have the same reductive effect. For two participants, a reduction in VS was obtained during the entirety of the session (i.e., continuous data), although less of a reduction in comparison to the nonintervention time (i.e., discontinuous data). These findings pose a question on the validity of treatment pertaining to RIRD on the reduction of VS.

Punishment

Punishment may be necessary when VS is severe, extinction cannot be applied, and other reinforcement-based interventions have been unsuccessful. Reducing VS has been difficult, historically, because it often persists independent of social contingencies. Implementing punishment procedures can effectively reduce VS. Nevertheless, special considerations should be made before implementing. The Behavior Analyst Certification Board's (BACB) Professional and Ethical Compliance Code for Behavior Analysts states that it is the ethical obligation of a behavior analyst to implement reinforcement procedures whenever possible, and, if punishment procedures are believed essential, they should be supplemented by differential reinforcement procedures (BACB, 2014).

Many studies have successfully used punishment procedures to reduce VS. For example, Anderson and Le (2011) compared the effects of four different treatments (i.e., matched stimulation, response cost, DRO, and overcorrection) on VS. The researchers conducted a series of reversals to compare the effects of the treatments on the reduction of VS. The effects of continuous access to music (i.e., matched stimulation), response cost (removal of either kids'

music or DVD), and DRO were evaluated before the overcorrection procedure was implemented with the DVD response cost. The participant was a 7-year-old boy diagnosed with ASD whose VS was maintained by nonsocial consequences. They found that matched stimulation successfully reduced VS. However, VS returned to high levels when treatment was removed. Response cost reduced levels of VS below 5% of intervals; however, these levels were not consistent across sessions (i.e., range during NCR 3-93%; mean level of VS of around 50%). Differential reinforcement of other behaviors did not robustly reduce VS (i.e., range during DRO 2-97%), with a mean level of VS of around 48% of intervals persisting throughout treatment. Nevertheless, when positive practice overcorrection was implemented (e.g., physically guiding the participant's hand to make a "shush" sign 100 times; implementation took between 30 and 40 s), VS decreased to an average of 3% intervals (implementation of the procedure was not subtracted from session duration) with a range from 0% to 27% of intervals. These data suggest that matched stimulation was not an effective procedure long term due to the increase of VS when the procedure was removed and that response cost and DRO were only mildly effective at reducing VS. Overcorrection successfully reduced VS to low levels; however, overcorrection is laborious to implement, especially for teachers or parents within the community.

Falcomata et al. (2004) examined the effects of NCR with and without response cost on VS. They compared baseline levels of responding to those seen in two treatment conditions, NCR and NCR plus response cost, using a reversal (ABCACBC) design. The participant was an 18-year-old man diagnosed with ASD who engage in VS maintained by nonsocial reinforcement. In the early phases of the study, the participant was given continuous access to the preferred stimulus (i.e., a radio) and there were no consequences for VS. Noncontingent reinforcement decreased VS below baseline levels; yet, the reduction was not clinically significant. Adding

response cost (i.e., removal of the radio for 5 s) to the NCR procedure, however, yielded an immediate reduction of VS to near zero levels.

Discrimination Training

Discrimination training has been a promising treatment method for establishing stimulus control for VS. Discrimination training consists of the implementation of a multiple schedule in which a reinforcement schedule is in effect in one component but not in the other component. Each component is associated with a distinctive stimulus and contingencies differ across schedules (e.g., reinforcement, no consequences). Discrimination training offers a more ethical approach for the reduction of VS, in that it does not completely restrict access to VS. This is of value given the potentially limited number of obtainable reinforcers, and autonomy to access those reinforcers, for individuals with ASD.

Haley and colleagues (2010) examined the use of stimulus control to reduce VS for an 8-year-old boy diagnosed with ASD. The authors utilized an alternating treatments design that included baseline and reversal phases. Two visual stimuli were used to signal when reinforcement was available (during presentation of the green card) and when reinforcement was not available (during presentation of the red card). If the participant engaged in VS during the visual cue that noted no VS, the paraprofessional would pick up the red card and hold it approximately 6 inches in front of the participants face. There were no programmed consequences for VS during the visual cue that noted VS was appropriate (green card). Their results demonstrated a reduction in VS when the red card was implemented during intervention phases in comparison to baseline phases. During the implementation of the green card, VS continued to remain at or around baseline levels. When the intervention was removed, there was an immediate increase in the level of VS. Results of the study demonstrated stimulus control by

the reduction of VS in the presence of the red card and unchanged responding in the presence of the green card.

Rapp et al. (2009) attempted to establish inhibitory stimulus control of VS by implementing discrimination training by using multiple schedules (i.e., different contingencies across schedules). For Study 2 (Study 1 evaluated the function of VS), a combination of a non-concurrent multiple baseline design across participants with a multielement and a reversal design were conducted to evaluate the effects of verbal reprimands on VS for two participants. Participants were twin 8-year-old boys diagnosed with ASD whose VS was maintained by automatic reinforcement. Similar to Haley et al. (2010), two visuals–red and green card–were implemented to signal either reinforcement for VS (green card) or punishment for VS (red card). In the presence of the green card there were no programmed consequences for VS. However, in the presence of the red card, the therapist would hold up the card and state 'red time, no movie' contingent on VS (verbal reprimand). The therapist labeled the participant's VS as 'movie time'. Results from Study 2 showed a decrease in VS for both participants in the presence of the red card. Vocal stereotypy continued at moderate levels in the presence of the green card. However, for one participant, VS was high and undifferentiated in the presence of the green card and absence of any card. Consequently, the researchers evaluated the addition of another punishment procedure, response cost, in another participant. Vocal stereotypy was evaluated through the effects of verbal reprimands and then contingent removal of preferred stimulation using a combined multielement and ABCDD'B' design. The participant was a 5-year-old boy who was diagnosed with ASD and whose VS was automatically maintained. For that participant a preferred toy was removed for 10 s contingent on each instance of VS, in addition to the delivery of the verbal reprimand in the presence of the red card. There were no programmed

consequences for VS in the presence of the green card. That participant's VS decreased and eventually was suppressed in just the presence of the red card. These data suggests that VS had come under inhibitory stimulus control. However, Rapp et al. (2009) only succeeded in putting VS under inhibitory stimulus control with one participant. The findings also found that verbal reprimands alone were not effective for inhibiting VS, suggesting that verbal reprimands were not an effective form of punishment for these three participants.

In summary, a range of procedures have been used to decrease the emission of VS in individuals with ASD. As discussed previously, these findings need to be replicated amongst other populations of individuals with intellectual disabilities. Additionally, these findings need to be replicated to further demonstrate the effectiveness of treatment for VS, and to extend findings outside of the research setting (i.e., natural environment). Thus, the purpose of Study 1 was to evaluate the effects of two reinforcement procedures (NCR and DRO) on VS in a more naturalistic environment (e.g., classroom setting).

Study 1 Method: Treatment Evaluation

Procedures used in this study were approved by the University of Kansas' Institutional Review Board. The study identification number is 00143927.

Automatically maintained behaviors can be extremely difficult to reduce through reinforcement procedures alone due to (a) difficulties with being able to identify the specific source of reinforcement or (b) identifying reinforcers that cannot be manipulated (e.g., VS; Vollmer, 1994). Although there are conditions under which punishment may be necessary to decrease automatically maintained behaviors (Lerman & Vorndran, 2002), behavior analysts should attempt to implement reinforcement-based procedures in the behavior-change program (Bailey & Burch, 2016).

Therefore, the purpose of this study was to assess two reinforcement-based procedures and their effects on VS with the hopes that a punishment procedure could be avoided. A common procedure used for the reduction of VS is NCR (e.g., Lanovaz et al., 2011), which is a feasible procedure that can be conducted by most people (e.g., caregivers, teachers). Another reinforcement-based procedure is DRO – it has effectively reduced VS (e.g., Taylor et al., 2005). This procedure has been believed advantageous due to the simplicity of implementing it and it does not interfere with ongoing activities when on a lean schedule. This study evaluated the effects of NCR and DRO on VS.

Participants and Setting

Participants were recruited from an applied behavior analysis (ABA) clinic that provides treatment for behavioral excesses and deficits associated with autism. Some of the research team members were involved with the participants' treatment (i.e., one therapist worked directly with two participants outside of this study, see Table 1). The participants were three boys with ASD

who engaged in VS maintained by sensory stimulation. None of the participants had a history of treatment for VS. John was an 8-year-old boy who communicated verbally and had a repertoire of following direct instructions. His clinician referred him to the study for treatment of VS, which was an objective on his treatment plan that posed educational concerns. Ethan was a 6-year-old boy who communicated verbally and had a repertoire of following direct instructions. His clinician requested participation in the study due to the same reasons as John. Nathan was a 6-year-old boy, who was Ethan's identical twin. Nathan communicated verbally and had a repertoire of following direct instructions as well. His clinician requested participation in the study due to an increase in his VS during therapy sessions, which interfered with his learning opportunities.

Sessions were conducted 3-5 days per week (2-6 sessions per day) in an assessment room at the treatment center (approximately 14 ft x 14 ft). The room was furnished with two small tables, one larger table, four chairs, two stacked tubs, and two locked cabinets. Two one-way mirrors ran along adjacent walls, with a central microphone allowing for discrete observation and data collection.

Data Collection and Measurement

Vocal Stereotypy

The primary dependent variable was the percentage of 10-s intervals in which the participants engaged in VS. Vocal stereotypy was scored using 10-s partial interval recording. Data were summarized as the percentage of 10-s intervals in which the participant engaged in VS. Total session duration was 10 min.

John's VS was defined as any vocal response that could not be identified as a word or phrase (e.g., repetitive sounds such as squeals, humming, mumbling, clearing his throat).

Ethan's VS was defined as any vocal response that was not appropriate to the context (e.g., talking about SpongeBob while drawing the planets), could not be identified as a word or phrase, or repeating a word/phrase three or more times within 10 s. Topographies of Ethan's VS included repetitive grunts, snorts, screeches, crash sounds, and non-functional/non-contextual statements. Nathan's VS was defined as the same as Ethan's. For all participants, if VS occurred during play and was not related to the play activity, then VS was scored. However, if the behavior was related to the play activity (e.g., talking about the planets while drawing them), then VS was not scored. The percentage of total intervals was calculated by dividing the total number of intervals containing VS by the total number of intervals in the session or component, then multiplying by 100.

Appropriate Engagement

Appropriate engagement was assessed using 10-s whole-interval recording. Data were summarized as the percentage of 10-s intervals with appropriate engagement. For all participants, appropriate engagement with a leisure item was defined as physical manipulation of an item or activity (e.g., manipulating play-doh) with one or two hands. Appropriate engagement with an auditory item was defined as remaining at (or, at least) a 2 ft distance of the table on which it played (Rapp, 2007).

Emotional Responding

Emotional responding was defined as whining, crying, screaming, aggressive behavior, self-injury, and/or attempts to escape from or physically resist the procedure (i.e., pushing the therapist away during the initiation of a procedure, turning their body away, pulling away from the procedure, dropping to the floor). The frequency of emotional responding was recorded and

summarized as responses per minute (i.e., number of instances divided by the total duration of the component; Verriden & Roscoe, 2018).

Interobserver Agreement

A second observer independently collected data during a minimum of 30% of the sessions in each condition for all participants. Interobserver agreement (IOA) for VS and appropriate engagement was calculated on an interval-by-interval method. To calculate IOA each session was divided into 10-s intervals and the number of intervals in which the two observers agreed was divided by the total number of intervals and that number was multiplied by 100. An agreement was scored if either (a) both observers recorded that a response occurred within a specified interval or (b) neither observer recorded a response within an interval.

For emotional responding, meant count per interval IOA was calculated. The smaller count of emotional response was divided by the larger count of emotional response, then multiplied by 100. The percent agreement for all intervals was then averaged to determine the percent agreement for the entire session. The mean IOA for John's VS was 95% (range, 83% to 100%), appropriate engagement was 95% (range, 83% to 100%), and emotional response was 97% (range, 83% to 100%). The mean IOA for Ethan's VS was 95% (range, 82% to 100%), appropriate engagement was 98% (range, 80% to 100%), and emotional response was 98% (range, 80% to 100%). The mean IOA for Nathan's VS was 92% (range, 72% to 100%), appropriate engagement was 94% (range, 90% to 100%), and emotional response was 90% (range, 50% to 100%; IOA was low due to issues with the video and the observer not detecting all instances of emotional responding).

Pre-assessments

A battery of assessments were conducted before the treatment was implemented. First, a multiple stimulus without replacement (MSWO) preference assessment was conducted to identify preferred items (DeLeon & Iwata, 1996; Appendix A) to be used in the analog functional analysis (FA; Iwata et al., 1982/1994). Next an analog FA was conducted to identify the function of VS for each participant (Appendix B). Next, participants completed a competing-items assessment. This assessment identified stimuli (i.e., auditory and non-auditory) that were successful in lowering their levels of vocal stereotypical behaviors, to be used in the NCR phase of the treatment evaluation (Appendix C). Next, a paired-stimulus preference assessment was conducted to identify preferred edible items (e.g., ring pops) to be consumed (by licking) in the DRO phase of the treatment evaluation (Appendix D).

Treatment Evaluation

Two treatments (i.e., NCR and DRO) were evaluated to determine the effects on VS through the implementation of a reversal design (see Table 1 for order of conditions). The sequence of phases were modified for each participant contingent on the pattern of VS across exposure to each phase. The order of conditions for each participant can be found in Table 2.

No Interaction Baseline

Baseline consisted of the last three to five sessions of the FA (Appendix B). During those sessions, participants were left alone in the room with no materials aside from the furniture already in the room and there were no programmed consequences for VS.

NCR with Matched and Unmatched Stimulation

During this condition, participants had access to auditory and non-auditory items identified by the competing items preference assessment (Piazza et al., 1996; Appendix C). Specifically, participants were given continuous access to one item that was suspected to match

the sensory modality of their VS, along with one item not suspected to match the sensory modality of their VS. Vocal stereotypy did not result in programmed consequences.

NCR with Matched Stimulation

During this condition, participants had access to an auditory item identified by the competing items preference assessment (Piazza et al., 1996; Appendix C). Specifically, participants were given continuous access to one item that was suspected to match the sensory modality of their VS. Vocal stereotypy did not result in programmed consequences.

NCR with Unmatched Stimulation

During this condition, participants had access to a leisure item identified by the competing items preference assessment (Piazza et al., 1996; Appendix C). Specifically, participants were given continuous access to one leisure item that was not suspected to match the sensory output produced by the participant's VS. Vocal stereotypy did not result in programmed consequences.

NCR with Unmatched Stimulation plus DRO 10-s

At the start of each session, participants were given a choice between their two highest preferred edible items from the initial paired-stimulus preference assessment (Fisher et al., 1992; Appendix D). Their chosen edible item was provided following the absence of VS for 10 s. If the participant engaged in VS, the 10-s timer was restarted. The edible item was then delivered (e.g., a single lick of the item) contingent on the absence of VS for 10 s. Participants had access to a leisure item identified by the competing items preference assessment (Piazza et al., 1996; Appendix C). Specifically, participants were given continuous access to one leisure item that was not suspected to match the sensory output produced by the participant's VS. The therapist

presented gestural prompts and/or verbal prompts for item engagement every 10 s if the participant had not engaged with the item during the previous 10 s.

Study 1 Results and Discussion

Results for John are shown in the top panel of Figure 1. During the baseline phase, John engaged in high levels of VS. In the NCR conditions, with the matched stimulus (i.e., Ryan's Review), levels of VS decreased to between 0% and 7% of the intervals, with high levels of appropriate engagement (i.e., 95% to 100% of the intervals) and zero instances of emotional responding. However, when the matched stimulus was presented with the unmatched stimulus (i.e., dinosaur toys), VS increased while appropriate engagement remained at high levels (i.e., 73% to 100% of intervals) and there were zero instances of emotional responding. The research team then implemented a 10-s DRO while the unmatched stimulus was freely available, providing the participant with an edible item in the absence of VS (i.e., a ring pop). This procedure resulted in highly variable patterns of VS. Appropriate engagement initially remained high, however, the last data point dropped to moderate levels (i.e., 30% to 100% of intervals) with zero instances of emotional responding.

Ethan's results are shown in the middle panel of Figure 1. During baseline, there were moderate to high levels of VS (i.e., 47% to 90% of intervals). Once NCR using a matched stimulus (i.e., Kidz Bop) and the unmatched stimulus (i.e., the coloring book) were implemented, low levels of VS occurred (i.e., at 5% of intervals), relative to baseline. Appropriate engagement remained high (i.e., 88% to 90% of intervals) with zero instances of emotional response. Noncontingent reinforcement with unmatched stimulus only was implemented. There were low levels of VS (i.e., 12% to 25% of intervals), with high levels of appropriate engagement (i.e., 97% to 100% of intervals), and there were zero instances of emotional responding. Noncontingent reinforcement of matched stimulus and unmatched stimulus were reintroduced and VS continued at low levels (i.e., 8% to 35% of intervals), appropriate engagement dropped

initially, but then returned to high levels (i.e., 30% to 100% of intervals), emotional responding remained at zero instances. Noncontingent reinforcement of unmatched stimulus was reintroduced and VS remained at low levels across sessions with the exception of the last session increasing (i.e., 13% to 53% of intervals), appropriate engagement remained at high levels (i.e., 73% to 100% of intervals), and emotional responding remained at zero instances. Finally, NCR using unmatched plus DRO was implemented and we see VS at moderate levels across sessions with an increase at the last session (i.e., 25% to 67% of intervals), appropriate responding remained at high levels (i.e., 73% to 100% of intervals), while emotional responding remained at zero instances.

Nathan's results are shown in the bottom panel of Figure 1. During baseline, Nathan engaged in moderate to high levels of VS (i.e., 53% to 70% of intervals), with high levels of appropriate engagement (i.e., all 100% of intervals), and there were zero instances of emotional responding. In NCR using a matched stimulus (i.e., listening to the digestive system) and the unmatched stimulus (i.e., play-doh), moderate to high levels of VS were seen (i.e., 22% to 77% of intervals), appropriate engagement remained at high levels (i.e., all at 100% of intervals), and there were zero instances of emotional responding. Therefore, we removed the unmatched stimulus and only assessed the matched stimulus, triggering a decrease in VS to low levels (i.e., 42% to 0% of intervals), appropriate engagement remained at high levels (i.e., all at 100% of intervals), and there were zero instances of emotional responding. We reinstated the unmatched stimulus with the matched stimulus and VS decreased to low levels of responding (i.e., 3% to 7% of intervals), appropriate engagement remained at high levels (i.e., all at 100% of intervals), and there were zero instances of emotional responding. The unmatched stimulus was then removed, and the matched stimulus was assessed alone. We noted higher levels of VS (i.e., 35% to 88% of

intervals) relative to the baseline, appropriate engagement remained at high levels (i.e., 88% to 100% of intervals), and there were zero instances of emotional responding. The unmatched stimulus was studied once again with the matched stimulus and, again, we noted low levels of VS (i.e., 2% and 8% of intervals), appropriate engagement remained at high levels (i.e., 100% of intervals), and there were zero instances of emotional responding. The matched stimulus was removed again, and the unmatched stimulus was assessed alone once more. We noted higher levels of VS (i.e., 78% and 87% of intervals) relative to the baseline, appropriate engagement remained at high levels (i.e., all at 100% of intervals), and there were zero instances of emotional responding. Thus, the DRO was layered onto the NCR. Relative to the other two participants' data, we again saw a highly variable pattern of VS (i.e., 30% to 85% of intervals), appropriate engagement remained at high levels (i.e., 98% to 100% of intervals), and zero instances of emotional responding. This pattern of responding resulted in the removal of the DRO. The matched stimulus was then removed, and the unmatched stimulus was assessed alone. We noted higher levels of VS (i.e., 65% to 83% of intervals) relative to the baseline, appropriate engagement remained high (i.e., 88% to 93% of intervals), and zero instances of emotional responding.

Across all participants, we successfully reduced VS through the implementation of two reinforcement-based procedures. Implementation of the different treatments produced reductions in comparison to the baseline phase. Specifically, NCR using a matched stimulus most robustly decreased VS across participants. All participants' VS decreased to zero or near zero levels of responding in the NCR using a matched stimulus. However, we wanted to identify a procedure that could be replicated in a less controlled environment (e.g., a classroom setting during instructional opportunities). Therefore, from a practical standpoint, continuous access to an

auditory item could not be recommended as a treatment for VS in a classroom environment during instructional times. Continuous access to auditory stimulation could interfere with the student attending and/or listening to the teacher's instructions throughout the learning day. The purpose of reducing VS was to allow more learning and social opportunities for the participants during their daily activities (e.g., school). However, if we gave clinical recommendations for our participant's teachers and paraprofessionals to implement NCR using a match stimulus in the classroom setting, we would be doing a disservice for our participants. As behavior analysts, we are required to provide the most effective treatment (Van Houten et al. 1988) and are responsible for operating in a manner that is best for our participants (Bailey & Burch, 2016).

We originally evaluated both matched and unmatched procedures to replicate previous literature (Ahearn et al., 2005; Rapp, 2007) and when we did, our findings suggested that the matched stimulus was by far more effective in comparison to the unmatched stimulus. These data required the research team to further investigate a procedure that could be replicated in a classroom setting during instructional opportunities, but would also be effective in the reduction of VS. Nonetheless, NCR using matched stimulation demonstrated that it would be an effective procedure in the reduction of VS during non-classroom time and would be a good starting spot for a comprehensive behavior plan. Regardless, an item that was also associated with lower levels of VS in the competing-items preference assessment (Piazza et al., 1996), but more closely mimicked play, was also evaluated in the NCR phase. The reason behind this decision was that if we could reduce VS through NCR using an unmatched item, then we could potentially select items that could be discretely manipulated or manipulated in a way that would not be distracting towards others in the classroom (e.g., fidget spinner). For all participants, we successfully

reduced VS with NCR using an unmatched item; however, reductions were not clinically significant. Consequently, we layered the DRO over NCR using an unmatched item.

Noncontingent reinforcement using an unmatched stimulus with DRO (NCR:UM/DRO), had a unique effect on all participant's VS in comparison to Verriden and Roscoe (2018). Previous studies (e.g., Mancina et al., 2000; Taylor et al., 2005) found that DRO robustly reduced VS. However, when NCR:UM/DRO was evaluated, we saw variable responding with an increasing trend for all participants. All participants' VS, at some point within the NCR:UM/DRO condition increased to baseline levels. These data suggest that NCR:UM/DRO did not reduce VS to clinically acceptable levels. Consequently, the DRO was removed from the treatment package.

The procedures from Study 1 were either not acceptable to implement in a natural setting or did not substantially reduce VS for all participants. Therefore, in Study 2 the research team evaluated a procedure that could reduce VS and be feasible for implementation in the classroom setting during instructional opportunities.

Study 2 Method: Discrimination Training

Prior to initiation of the punisher assessment (described below), the procedures were approved by the University of Kansas' Institutional Review Board. The study identification number is 00143927. In addition, each of the punishment procedures was discussed in detail with and approved by the child's clinical team.

Inhibitory Stimulus Control

As discussed previously, one approach to maintaining low levels of VS is placing it under stimulus control. Stimulus control is an antecedent-based strategy that uses the presence of discriminative stimuli to signal that reinforcement is available and the presence of a stimulus delta to signal that reinforcement is not available (Cooper et al., 2007). Stimulus control is developed through discrimination training. Discrimination training consists of the implementation of a multiple schedule in which a reinforcement schedule is in effect in one component but not in the other component. Each component is associated with a distinctive stimulus and contingencies differ across schedules (e.g., reinforcement, no consequences).

Inhibitory stimulus control is a type of stimulus control. However, inhibitory stimulus control utilizes two antecedent stimuli to signal periods of reinforcement for engaging in the target behavior (e.g., VS) and periods of punishment for engaging in the target behavior. Punishment suppresses rates of behavior in the presence of the inhibitory stimulus, relative to in the presence of the non-punishment stimulus. Inhibitory stimulus control, once accomplished, can reduce the frequency of punishers delivered, thereby offering an alternative to other, more intensive and restrictive interventions for challenging behavior. In this way, inhibitory stimulus control can be more desirable and socially valid.

A promising example of inhibitory stimulus control was a study conducted by Doughty and colleagues (2007). Participants were three adults with intellectual disabilities who exhibited motor stereotypy not maintained by social consequences. Doughty et al. brought the behavior of three individuals with intellectual development disorder (IDD) under the control of multiple schedules. One schedule was denoted by the *presence* of a bracelet, and the other schedule was denoted by the *absence* of the bracelet. In the presence of the bracelet, there were no programmed consequences for stereotypy. In the absence of the bracelet (i.e., participant not wearing a bracelet), stereotypy resulted in the implementation of the punisher, hands down (e.g., guiding the participants hands down, preventing stereotypy). Results demonstrated that the antecedent stimulus (i.e., the absence of bracelet), when correlated with punishment, served as an effective stimulus that signaled punishment for all participants; as intended, the stereotypy of all participants decreased when the bracelet was absent.

Doughty et al. brought motor stereotypy under stimulus control. However, subsequent studies (e.g., Rapp et al., 2009; Tiger et al., 2016) did not establish robust nor consistent stimulus control. Their challenges with establishing inhibitory stimulus control may have been due to the use of ineffective punishers. A formal assessment of potential punishers may have identified an effective punisher for implementation. Therefore, using a punisher assessment to systematically identify potential punishment procedures (e.g., response cost, time-out, verbal reprimands) may improve treatment outcomes (Fisher et al., 1994; Verriden & Roscoe, 2018).

Punisher Assessment

A punisher assessment is an experimental analysis of different punishers. Fisher and colleagues (1994) originally created an empirical method for identifying procedures that may serve as punishers for individuals with IDD who engaged in pica. The authors identified

potential punishers through a two-step process, stimulus avoidance assessment and a punisher assessment. The stimulus avoidance assessment was built off of the stimulus preference assessment developed by Pace et al. (1985). Fisher et al. (1994) believed that if a preference assessment can potentially identify reinforcers, then an avoidance assessment may identify potential punishers as well. They selected nine procedures from the literature and exposed each procedure, individually to the participant. Each potential punisher was implemented 10 times in a single session noncontingently. The authors selected potential punishers based off of off negative vocalizations and avoidance movements, believing that these behaviors were a sure indicator of a punishing effect. From the stimulus avoidance assessment, they selected punishers based off of low, medium, or high emotional responding and then included them in the punisher assessment. They utilized an ABA multielement design to evaluate the efficacy of the procedures for pica. During the punisher assessment, each punisher was delivered contingent on pica. The procedure that produced the highest reduction in pica during the punisher assessment was then selected as the punishment procedure for treatment. Fisher et al. (1994) was a considerable contribution to the literature in that it created a new method for identifying potentially effective punishment procedures. However, Fisher and colleagues identified potential punishers based on the emotional responding they produced. This approach may pose some ethical concerns about implementing procedures known to produce high levels of emotional responding despite their effectiveness.

Verriden and Roscoe (2018) extended the punisher identification research for use in reducing stereotypical behaviors. They conducted similar methods as Fisher et al. (1994) in regards to implementing a punisher assessment and collecting data on emotional responding. However, Verriden and Roscoe made ethically appropriate changes to ensure that participant's

dignity was protected in the process of identifying potential punishers. This was accomplished by eliminating the stimulus avoidance assessment, adding preferred items from a competing-items assessment, collecting data on appropriate engagement, and selecting procedures that were associated with low levels of responding, but also were effective at the reduction of stereotypical behaviors. First, they evaluated whether two reinforcement procedures could reduce stereotypy to clinically significant levels of responding. They found that NCR was not effective as a stand-alone treatment. Next, they added differential reinforcement of alternative behavior (DRA) to NCR, but did not produce desired reductions in stereotypy. Subsequently, they utilized the data that they had collected as a baseline against which to evaluate the effects of several different punishment procedures on automatically maintained behaviors (i.e., mouthing, motor stereotypy, and hair manipulation). All punishment procedures assessed were selected in advance by the participants' clinicians. During the assessment, the researchers randomly selected one of the clinician-suggested punishment procedures before each session and then implemented that procedure contingent on the participants' automatically maintained behaviors. Throughout the study, the rate of a participant's automatically maintained behavior, their emotional responding, and the duration of appropriate play with items selected from competing-items preference assessment (Piazza et al., 1996) were considered when evaluating the efficacy of each punisher. The punishment procedure that most effectively reduced stereotypical behaviors, had the highest percentage of appropriate engagement, and had the lowest rate of emotional responding was then selected for use as a treatment. The researchers found that when they then applied the most effective punishment procedure for each participant, stereotypical behaviors rapidly decreased to near-zero levels for all participants. A compelling approach to reducing VS is the combination of Verriden and Roscoe's (2018) procedure with inhibitory stimulus control, which may

humanely suppress VS for instructional purposes. Verriden and Roscoe (outlined an approach that could be replicated and utilized to select a punishment procedure that enables inhibitory stimulus control to be established over VS. Therefore, the purpose of Study 2 was to investigate if an empirically identified punishment procedure could be used to reduce VS through inhibitory stimulus control.

Participants and Setting

All three participants from Study 1 (i.e., John, Ethan, and Nathan) participated in Study 2. Therapists conducted all sessions during Study 2 in the same environment as in Study 1. Materials used during the punishment assessment included a chair for time-out, a bucket that was used for placing an item following a response cost, and another bucket that was filled with ripped up paper for tidiness training. Additional materials that were used during the discrimination training included two Snap-On black bracelets and a blackboard that was utilized for the punishment component.

Data Collection and Measurement

Vocal Stereotypy

The primary dependent variable was the percentage of 10-s intervals in which the participants engaged in VS. The definitions for each participant were identical to the definitions in Study 1. Vocal stereotypy was scored using 10-s partial interval recording. Data were summarized as the percentage of 10-s intervals during which the participant engaged in VS. Each session lasted 10 min.

Appropriate Engagement

As in Study 1, appropriate engagement was defined as physical manipulation of an item or activity (e.g., manipulating play-doh) with one or two hands. Appropriate engagement with an auditory item was defined as remaining at (or, at least) a 2 ft distance of the table on which it

played (Rapp, 2007). Appropriate engagement was assessed using 10-s whole-interval recording Data were summarized as the percentage of 10-s intervals with appropriate engagement.

Emotional Responding

As in Study 1, emotional responding was defined as whining, crying, screaming, aggressive behavior, self-injury, and/or attempts to escape from or physically resist the procedure (i.e., pushing the therapist away during the initiation of a procedure, turning their body away, pulling away from the procedure, dropping to the floor). The frequency of emotional responding was recorded and summarized as responses per minute (i.e., number of instances divided by the total duration of the component; Verriden & Roscoe, 2018).

Intervals Until Vocal Stereotypy

Intervals until VS was defined as the percentage of 10 s intervals that had elapsed prior to the first instance of VS. This measure was used to determine if discrimination training had established stimulus control. Specifically, stimulus control was said to have been established if the percentage of the component that elapsed before the first interval scored with a VS was less than 30% in the no-punishment scenario (i.e., in the absence of the bracelet) and greater than 70% in the punishment scenario (i.e., with the bracelet on), for six consecutive sessions.

Interobserver Agreement

A second observer independently collected data during a minimum of 30% of the sessions for each condition for all participants. Interobserver agreement for VS and appropriate engagement was calculated on an interval-by-interval basis by dividing the number of agreements plus disagreements and multiplying by 100.

For emotional responding, meant count per interval IOA was calculated. The smaller count of emotional responding was divided by the larger count of emotional responding, then

multiplied by 100. The percent agreement for all intervals was then averaged to determine the percent agreement for the entire session. The mean IOA for John's VS was 95% (range, 83% to 100%), appropriate engagement was 95% (range, 83% to 100%), and emotional response was 97% (range, 83% to 100%). The mean IOA for Ethan's VS was 95% (range, 82% to 100%), appropriate engagement was 98% (range, 80% to 100%), and emotional response was 98% (range, 80% to 100%). The mean IOA for Nathan's VS was 92% (range, 72% to 100%), appropriate engagement was 94% (range, 90% to 100%), and emotional response was 90% (range, 50% to 100%; IOA was low due to issues with the video and the observer not detecting all instances of emotional responding).

Punishment Procedure Assessment

Five to eight potential punishment procedures, selected from the punishment procedure selection interview (Appendix E), were evaluated simultaneously with NCR:UM/DRO procedures. Punishment procedures were selected for participants considering three variables: reduction of VS, the highest percentage of appropriate engagement, and the lowest rate of emotional responding.

Like the procedures employed by Verriden and Roscoe (2018), a punishment procedure assessment was implemented to test effects of punishment above and beyond those of our treatment (i.e., NCR plus DRO). Punishment procedures were evaluated using a multi-element design with an initial baseline and reversals back to NCR plus DRO. The punishment procedure that demonstrated the lowest percentage of VS, the highest percentage of appropriate engagement, and the lowest rate of emotional responding, was selected as the most successful. The order of the punishment procedure assessment was different for each participant, the conditions were as follows.

In-Seat Timeout

This condition included the in-seat timeout procedure contingent on VS, along with NCR and DRO components. Contingent on VS, the therapist instructed the participant to go to time-out (seat facing the wall in the corner of the room) using least-to-most prompts (i.e., verbal prompts, model prompts, and physical prompts). Time outs were 30 s in duration and leisure items from the competing-items preference assessment (Piazza et al., 1996) were not accessible during time out. Noncontingent reinforcement and DRO were implemented as was previously described, with the exception that they were not implemented during the punishment procedure (Toole et al., 2004).

Verbal Reprimands

This condition included a verbal reprimands procedure (Verriden & Roscoe, 2018) contingent on VS, along with NCR and DRO components. The therapist delivered one reprimand (e.g., "stop that") contingent on the occurrence of VS.

Response Cost

This condition included a response cost procedure (Rapp et al., 2009) contingent on VS, along with NCR and DRO components. The therapist would immediately remove the leisure item(s) for 30 s contingent on the occurrence of VS.

Facial Screen

This condition included a facial screen procedure (Toole et al., 2004) contingent on VS, along with NCR and DRO components. The therapist would stand behind the participant and place their hand over the participant's eyes for 30 s (therapist's hand was not pressed against participants face), blocking them with their other arm if the participant tried to remove the therapist's hand.

Tidiness Training

This condition included a tidiness training procedure (Toole et al., 2004) contingent on VS, along with NCR and DRO components. The therapist would dump crumpled paper onto the floor of the room and instruct the participant to put the paper in the trash can. If the participant did not begin the task after 5 s, the therapist used the minimum amount of physical prompting necessary to guide the participant to complete the task. This procedure lasted approximately 30 s.

Overcorrection-Positive Practice

This condition included an overcorrection-positive practice procedure (Anderson & Le, 2011) contingent on VS, along with NCR and DRO components. The therapist would verbally prompt—by reinstructing and gesturing to their mouth—a more appropriate response (e.g., "I like to color") for 30 s.

Contingent Motor Demands

This condition included a contingent motor demands procedure (Ahearn et al., 2007) contingent on VS, along with NCR and DRO components. The therapist would deliver motor demands (e.g., clap hands, stomp feet) contingent on the occurrence of VS until the participant emitted three consecutive motor responses in the absence of VS. Least-to-most prompting was employed if the participant did not comply with the demands.

Hands Down

This condition included a hands down procedure (Toole et al., 2004) contingent on VS, along with NCR and DRO components. The therapist would gently hold the participant's hands to their lap for 30 s contingent on the occurrence of VS.

NCR with Punisher Assessment

This condition included NCR with unmatched leisure items paired with the punisher assessment. Each punisher (described above) was evaluated as the previous NCR:UM/DRO with punisher assessment condition, with the exception of the removal of the DRO from the treatment package.

Discrimination Training

Procedures similar to those used by Doughty et al. (2007) were applied to establish inhibitory stimulus control, with the exception that programmed consequences were implemented during the bracelet component. Initially, a two-component multiple schedule was implemented with an embedded alternating treatments design (i.e., each 10 min session was either the participant wearing a bracelet or not wearing a bracelet). Each session comprised two 5-min components separated by the time it took to change the stimuli (approximately 1–2 s) during the multiple schedule. However, an alternating treatments design was conducted for John (started on session 65) and Ethan (sessions 53-102). Each session during the alternating treatments design comprised one 10-min component. The order of the two components was randomized with the following restrictions: the same order did not occur more than three consecutive times and each sequence would occur six times in every 12-session block. Two to ten sessions were conducted each day, 3–5 days per week.

Due to specific circumstances (i.e., a stay-at-home order), not all participants were able to participate in all of the phases of the training.

Baseline

During baseline, participants had continuous access to a leisure item identified by the competing items preference assessment (Piazza et al., 1996; Appendix C) and the presentation of the stimulus (i.e., the black bracelet) during portions of the session. A two-component multiple

schedule was employed. One component involved the participant wearing the stimulus (i.e., the black bracelet), and the other involved the absence of the stimulus. There were no programmed consequences for VS in either component. Components alternated after 5 min within the session (e.g., the bracelet was present for the first 5 min, then absent for the last 5 min of the session). The order of the components was randomized with the exception that the same order did not occur more than three consecutive times.

Multiple Schedule

During this phase, a two-component multiple schedule was employed. One component involved the participant wearing the stimulus (i.e., the black bracelet; punishment component), and the other involved the absence of the stimulus (no punishment component). Components alternated after 5 min within the session (e.g., the bracelet was present for the first 5 min, then absent for the last 5 min of the session). The order of the components was randomized with the exception that the same order did not occur more than three consecutive times. During the component with the stimulus present, contingent on VS the selected punishment procedure was delivered. When the black bracelet was absent, there were no programmed consequences for VS. The participant still received continuous access to the leisure item identified by the competing items preference assessment (Piazza et al., 1996; Appendix C) outside of the implementation of the punishment procedure.

Extra Stimuli

During this phase, a two-component multiple schedule was employed. One component involved the participant wearing multiple stimuli (i.e., two black bracelets; punishment component) with the presentation of a black board sitting across from participants, and the other involved the absence of the stimuli (no punishment component). Components alternated after 5

min within the session (e.g., bracelets and black board were present for the first 5 min, then absent for the last 5 min of the session). The order of the components was randomized with the exception that the same order did not occur more than three consecutive times. During the component with the stimuli present, contingent on VS the selected punishment procedure was delivered. When the extra stimuli were absent, there were no programmed consequences for VS. The participant still received continuous access to the leisure item identified by the competing items preference assessment (Piazza et al., 1996; Appendix C) outside of the implementation of the punishment procedure.

Instructions

During this phase, a two-component multiple schedule was employed. One component involved the participant wearing multiple stimuli (i.e., two black bracelets; punishment component) with the presentation of a black board sitting across from participants, and the other involved the absence of the stimuli (no punishment component). Instructions were added to enhance stimulus control. Vocal instructions were associated with the presence and absence of the stimuli. That is, at the onset of the component that included the extra stimuli, participants were told, "You have to have a cool voice while wearing your bracelets or *punishment procedure selected*." At the onset of component where the extra stimuli were absent, participants were told, "You can make silly sounds and say what you want." The onset was defined as the time when the participant walked into the assessment room. Components alternated after 5 min within the session (e.g., bracelets and black board were present for the first 5 min, then absent for the last 5 min of the session). The order of the components was randomized with the exception that the same order did not occur more than three consecutive times. During the component with the stimuli present, contingent on VS the selected punishment procedure was delivered. When the

extra stimuli were absent, there were no programmed consequences for VS. The participant still received continuous access to the leisure item identified by the competing items preference assessment (Piazza et al., 1996; Appendix C) outside of the implementation of the punishment procedure.

Extended Exposure

The multiple schedule was then changed to an alternating treatments design. This included an increase in the length of exposure (i.e., 10 min) to either the absence or the presence of the stimuli. Treatments involved either the participant with the presentation of the extra stimuli (punishment) or the absence of the extra stimuli (no punishment). The order of the two treatments (i.e., punishment and no punishment) were randomized with the exception that the same order did not occur more than three consecutive times. During the punishment treatment, extra stimuli were present and contingent on VS the selected punishment procedure was delivered. During the no punishment treatment, extra stimuli were absent and there were no programmed consequences for VS. The participant still received continuous access to the leisure item identified by the competing items preference assessment (Piazza et al., 1996; Appendix C) outside of the implementation of the punishment procedure.

Randomized Punishment Procedures

For one participant (John), we took the three most effective punishment procedures from the punishment procedure assessment and randomized the order of their implementation per session. Punishment procedures were randomized with replacement, with the exception that a procedure did not occur more than three consecutive sessions. The participant still received continuous access to the leisure item.

Social Validity

Following the completion of the treatment analysis, we administered a close-ended questionnaire (adapted from that used by Potter et al., 2013) with the participant's lead clinician. The questionnaire consisted of four questions (see Table 3) on the acceptability of the procedures, the outcomes, and the goals, as well as the feasibility of the treatment implementation.[3] The lead clinician was asked, for each punisher, (a) if the procedure had been used with the participant, (b) if it would be effective in decreasing the target behavior, (c) if the participant would dislike it, (d) if they were willing to include it in the participant's behavioral support plan if found to be effective, (e) and if they thought the procedure could be implemented by direct care staff with integrity.

Study 2 Results and Discussion: Discrimination Training

Punishment Procedure Assessment

The results for John are shown in Figure 2, the top panel shows VS, the middle panel shows appropriate engagement, and the bottom panel shows emotional responding. During baseline, John engaged in moderate to high levels of VS (i.e., 5% to 88% of intervals), high levels of appropriate engagement (i.e., 30% to 100% of intervals) with a drop in the last data point, and zero instances of emotional responding. In the initial punishment procedure assessment, the leisure item was freely available along with an edible item and the selected punishment procedure was delivered contingent on VS. All punishment procedures (except contingent demands) were effective at reducing VS to below baseline levels. All punishment procedures (except overcorrection) were associated with high levels of appropriate engagement.

[3] Due to specific circumstances, only one lead clinician was able to complete the questionnaire.

For emotional responding, there was an increase across all punishers, relative to baseline (except verbal reprimands; i.e., from 0 to 3.2 per minute).

When weighing reductions in VS, percent of intervals with appropriate engagement, and rate of emotional responding, several procedures appeared effective. Response cost yielded an 89% decrease in intervals with VS (i.e., 0% to 42% of intervals), 87% of intervals with appropriate engagement (i.e., 80% to 100% of intervals), and 0.23 emotional responses per minute (i.e., 0 to 1.3 responses per minute). Time out yielded an 85% decrease in intervals with VS (i.e., from 3% to 27% of intervals), 86% of intervals with appropriate engagement (i.e., 74% to 100% of intervals), and 0.18 emotional responses per minute (i.e., 0 to 0.9 responses per minute). Tidiness training yielded an 85% decrease in intervals with VS (i.e., 0% to 38% of intervals), 93% of intervals with appropriate engagement (i.e., 67% to 100% of intervals), and 0.3 emotional responses per minute (i.e., 0 to 3.2 responses per minute). Overcorrection reduced VS (85% decrease) but was associated with high rates of emotional responding (i.e., average of 1.26 responses per minute; 0.3 to 2.3 responses per minute) and low levels of appropriate engagement (i.e., 38% to 92% of intervals). As a result, overcorrection was not selected.

Upon reversing to NCR with DRO, levels of VS increased (peaking at 72% of intervals) while appropriate engagement remained high (i.e., 97% to 100% of intervals) and there were no emotional responding. Differential reinforcement of other behavior did not appear to reduce VS; thus, it was removed from the treatment package. The punishment procedure assessment was then repeated (against an NCR plus DRO baseline) for the three most effective procedures (i.e., response cost, tidiness training, and time-out). All of the procedures reduced VS (i.e., 0% to 42% of intervals) relative to the NCR baseline data. However, it was response cost that generated the lowest levels of VS (i.e., an average for the last three data points of 8%), the

highest levels of appropriate engagement (i.e., an average for the last three data points, of 93%), and the lowest levels of emotional response (i.e., zero instances across that last five data points). A final reversal to NCR was conducted in which the punishment procedures were removed and only continuous access to the leisure item was available. The level of response was the same as for the auditory and leisure items in previous conditions (i.e., 25% to 27% of intervals), with high levels of appropriate engagement (i.e., 80% and 100%), and zero instances of emotional response. As a result, response cost was selected for discrimination training.

Ethan's results are shown in Figure 3, the top panel shows VS, the middle panel shows appropriate engagement, and the bottom panel shows emotional responding. During baseline (NCR plus DRO), Ethan engaged in moderate to high levels of VS (i.e., 25% to 67% of intervals), high levels of appropriate engagement (i.e., 73% to 100% of intervals), and zero instances of emotional responding. In the initial punishment procedure assessment, the leisure item was freely available along with an edible item and the selected punishment procedure was delivered contingent on VS. All punishment procedures were effective at reducing VS to below baseline levels. All punishment procedures (except overcorrection) were associated with high levels of appropriate engagement. For emotional responding, there was an increase across all punishers, relative to baseline (except verbal reprimands; i.e., 0 to 6 per minute).

When weighing reductions in VS, percent of intervals with appropriate engagement, and rate of emotional responding, several procedures appeared effective. Contingent demands yielded a 68% decrease in intervals with VS (i.e., 2% to 27% of intervals), 86% of intervals with appropriate engagement (i.e., 25% to 100% of intervals), and 0.04 emotional responses per minute (i.e., 0 to 0.3 responses per minute). Overcorrection yielded a 91% decrease in intervals with VS (i.e., 2% to 20% of intervals), 71% of intervals with appropriate engagement (i.e., 18%

to 100% of intervals), and 0.44 emotional responses per minute (i.e., 0 to 1.9 responses per minute). Time out yielded an 88% decrease in intervals with VS (i.e., 0% to 15% of intervals), 85% of intervals with appropriate engagement (i.e., 47% to 100% of intervals), and 0.17 emotional responses per minute (i.e., 0 to 0.5 responses per minute). However, the last three sessions increased to an average of 10% of intervals, yielding an 85% decrease in intervals with VS. Tidiness training yielded a 79% decrease in intervals with VS (i.e., 0% to 48% of intervals), 89% of intervals with appropriate engagement (i.e., from 58% to 100% of intervals), and 0.4 emotional responses per minute (i.e., 0 to 1.5 responses per minute).

Upon reversing to NCR with DRO, levels of VS increased (peaking at 77% of intervals) while appropriate engagement remained high (i.e., 95% to 100% of intervals) and there were no emotional responding. Differential reinforcement of other behavior did not appear to reduce VS; thus, it was removed from the treatment package. Noncontingent reinforcement for unmatched stimulus was implemented and initially VS was at moderate levels, however levels decreased across sessions to low levels (i.e., 10% to 50% of intervals), appropriate engagement levels were high (i.e., 95% to 100% of intervals), and there no emotional responding. Noncontingent reinforcement with matched and unmatched stimuli were conducted and VS immediately dropped to low levels (i.e., 2% to 8% of intervals), moderate to high levels of appropriate responding (i.e., 40% to 100% of intervals), and no emotional responding. Noncontingent reinforcement for unmatched stimulus was reinstated and, again, we see low to moderate levels of VS (i.e., 10% to 72% of intervals), moderate to high levels of appropriate engagement (i.e., 58% to 100% of intervals), and no emotional responding. The punishment procedure assessment was then repeated (against an NCR baseline) for the four most effective procedures (i.e., contingent demands, overcorrection, tidiness training, and time-out). All of the procedures

reduced VS (i.e., 0% to 15% of intervals) relative to the NCR plus DRO baseline data. However, it was tidiness training that generated the lowest levels of VS (i.e., an average for the last three data points of 4%), the highest levels of appropriate engagement (i.e., an average for the last three data points, of 97%), and the lowest levels of emotional response (i.e., an average of 0.5 responses per minute across that last three data points). A final reversal to NCR was conducted in which the punishment procedures were removed and only continuous access to the leisure item was available. Vocal stereotypy was at low levels with a slight increase on the last session (i.e., 10% to 27% of intervals), appropriate engagement remained high (i.e., 70% to 100% of intervals), and no emotional responding. As a result, tidiness training was selected for discrimination training.

The results for Nathan are shown in Figure 4, the top panel shows VS, the middle panel shows appropriate engagement, and the bottom panel shows emotional responding. During baseline, Nathan engaged in moderate to high levels of VS (i.e., 65% to 83% of intervals), high levels of appropriate engagement (i.e., 88% to 93% of intervals), and zero instances of emotional responding. Based on the information that was obtained from the first two participants (John and Ethan), the punishment procedure assessment with NCR using the leisure item was conducted sooner. All punishment procedures were effective at reducing VS to below baseline levels. All punishment procedures were associated with high levels of appropriate engagement. For emotional responding, across all punishment procedures, were either zero or low levels of responding (i.e., 0.0 to 0.4 responses per minute).

When evaluating reductions in VS, percent of intervals with appropriate engagement, and rate of emotional responding, all procedures appeared effective. Hands down yielded a 99% decrease in intervals with VS (i.e., 0% to 2% of intervals), 100% of intervals with appropriate

engagement, and 0.0 emotional responses per minute. Verbal reprimand yielded an 89% decrease in intervals with VS (i.e., 3% to 13% of intervals), 97% of intervals with appropriate engagement (i.e., 94% to 100% of intervals), and 0.0 emotional responses per minute. Response cost yielded a 93% decrease in intervals with VS (i.e., 3% to 7% of intervals), 91% of intervals with appropriate engagement (i.e., 80% to 96% of intervals), and 0.0 emotional responses per minute. Facial screen yielded a 99% decrease in intervals with VS (i.e., 0% to 2% of intervals), 99% of intervals with appropriate engagement (i.e., 97% to 100% of intervals), and 0.0 emotional responses per minute. Tidiness training yielded a 95% decrease in intervals with VS (i.e., 2% to 5% of intervals), 98% of intervals with appropriate engagement (i.e., 96% to 100% of intervals), and 0.0 emotional responses per minute. Contingent demands yielded a 62% decrease in intervals with VS (i.e., 5% to 32% of intervals), 98% of intervals with appropriate engagement (i.e., 96% to 100% of intervals), 0.0 emotional responses per minute. Overcorrection yielded a 92% decrease in intervals with VS (i.e., 0% to 12% of intervals), 98% of intervals with appropriate engagement (i.e., 95% to 100% of intervals), and 0.0 emotional responses per minute. Time out yielded a 93% decrease in intervals with VS (i.e., 2% to 5% of intervals), 92% of intervals with appropriate engagement (i.e., 87% to 100% of intervals), 0.0 emotional responses per minute.

A final reversal to NCR was conducted in which the punishment procedures were removed and only continuous access to the leisure item was available. The level of VS was low with an increase the last two sessions (i.e., 8% to 28% of intervals), appropriate engagement remained high (i.e., 97% to 100% of intervals), and no emotional responding. As a result, hands down procedure was selected for discrimination training.

Discrimination Training

John's results for VS are shown in the top panel of Figure 5. Undifferentiated levels of VS were observed across both components (i.e., 7% to 23% of intervals). These data do not show that the bracelets exhibited any inhibitory control over John's VS prior to their explicit training. In discrimination training (i.e., during multiple schedules), undifferentiated patterns of response across both components (i.e., 0% to 77% of intervals) were noted. When extra stimuli were added, low undifferentiated patterns of response (i.e., 0% to 33% of intervals) were observed, along with variable latencies (i.e., 0% to 93% of component; see Figure 6) across both components.

Further on, in the alternating treatment condition, an increase in VS with variable patterns across both components occurred (i.e., 5% to 53% of intervals). Due to an increase in the variable patterns of response across both components the punishment procedure was randomized for each session (Charlop et al., 1988; Toole et al., 2004). The three most effective procedures from the punishment procedure assessment were used. This included response cost, tidiness training, and time-out. The research team continued to see low to moderate undifferentiated levels of response (i.e., 2% to 60% of intervals). Stimulus control, however, was never developed over John's behavior before the project was cut short due to the COVID-19 stay at home order.

Ethan's data are shown in the middle panel of Figure 5. In the baseline phase, moderate levels of VS took place (i.e., 33% to 60% of intervals), with an increasing trend across both components. These, data suggest that the bracelets did not have any inhibitory control over Ethan's VS prior to discrimination training.

During discrimination training (i.e., Multiple schedules), VS decreased in both components, relative to the baseline (i.e., 33% – 60% intervals in baseline to 0% - 50% of

intervals). However, responding was undifferentiated between both components. In the extra stimuli condition, we continued to observe undifferentiated patterns of VS (i.e., 0% to 23% of intervals).

During the alternating treatment condition, we observed low levels of VS in both components. However, in the last 12 sessions, we saw differentiation between the two components, meeting Doughty et al.'s stimulus control criterion (i.e., bracelet component, 0% to 10%; no bracelet component, 0% to 17%). Nevertheless, we reversed back to the multiple schedule condition and almost immediately witnessed a loss in differentiation (i.e., 0% to 13% of intervals).

Due to a lack of stability in the discrimination, instructions were added at the beginning of each component (Tiger et al., 2016). With the addition of instructions, however, VS decreased to zero levels of responding across both components. Therefore, a less effective unmatched stimulus (i.e., sand; see the second panel of Figure 8) was used in the NCR procedure. Differentiation was immediately obtained across both components, meeting stimulus control criterion within the last six sessions (i.e., short latencies for the absence of the bracelet and long latencies for the presence of the bracelet). To replicate Doughty et al., instructions were discontinued. Following this condition, discrimination across both components was observed, thus meeting the stimulus control criterion.

With the removal of the extra stimuli, discrimination continued to meet the stimulus control criterion. To further assess the efficacy of the bracelets, the leisure item was removed (i.e., playing with sand) and the coloring book was reintroduced. This again, proved to be effective with regards to the participant's ability to discriminate between the two components while meeting stimulus control criterion.

The results for Nathan's discrimination training are shown in the bottom panel of Figure 5. Throughout the baseline phase, there was a variable pattern for the no bracelet component (i.e., 3% to 70% of intervals), while low to moderate levels of response for the bracelet component (i.e., 13% to 30% of intervals) were observed. These data do not provide any indication that the bracelets had an inhibitory control prior to discrimination training.

During discrimination training (i.e., multiple schedules) with the extra stimuli and instructions, we noted differentiation immediately across both components (i.e., bracelet component, 0% of all intervals and no bracelet component, 2% to 57% of intervals). Nathan met stimulus control criterion. for the last five sessions and we had the intention of removing the instructions and extra stimuli once stimulus control criterion was met for the last six sessions. Unfortunately, the COVID-19 stay at home order prevented the completion of this last phase.

In summary, Study 2 was successful at identifying an empirical punishment procedure for each participant, all potential punishers were successful at the reduction of VS in comparison to baseline, and one participant's VS was placed under inhibitory stimulus control.

The punisher assessment was an effective method that reliably identified punishment procedures for each participant. All procedures reduced VS below baseline levels across participants. The punisher assessment also identified procedures that were just as effective in the reduction of VS, but also had low rates of emotional responding, then procedures that demonstrated higher rates of emotional responding. This deviates from previous studies utilizing procedures that had higher rates of emotional responding (Fisher et al., 1994).

Establishing inhibitory stimulus control through discrimination training posed many obstacles across all participants. In Study 2, we were only successful with placing VS under inhibitory stimulus control for Ethan before the global pandemic outbreak. Nathan's data

demonstrated promising effects, suggesting that inhibitory stimulus control could have been achieved.

General Discussion

The purpose of Study 1 was to assess two reinforcement-based procedures and their effects on VS with the hopes that a punishment procedure could be avoided. Results showed that NCR using matched stimulation was an effective procedure for reducing VS. However, continuous access to an auditory item was not deemed appropriate for implementation in the classroom, especially during instructional opportunities. Therefore, NCR using unmatched stimulation was selected. Once assessed, we discovered that the two treatments (i.e., NCR with unmatched and DRO) were not successful in reducing automatically maintained VS to low levels. Thus, the purpose of Study 2 was to investigate if an empirically identified punishment procedure could be used to reduce VS through inhibitory stimulus control. We used a punisher assessment to identify effective punishers for each participant and attempted to bring VS under inhibitory stimulus control. Results revealed that the punisher assessment was effective at identifying multiple punishers for all participants. Through empirically identifying a punishment procedure in the punisher assessment, we successfully brought one participant's VS under inhibitory stimulus control through discrimination training.

The current findings are consistent with previous studies showing that NCR using matched stimulation is an effective treatment for reducing VS. However, apart from NCR using matched stimulation, our findings are consistent with prior demonstrations that other reinforcement-based treatments (e.g., DRO) can be insufficient to reduce automatically maintained behaviors (e.g., VS), suggesting that punishment may sometimes be needed (Lerman & Vorndran, 2002). Notably, Verriden and Roscoe (2018) were unable to reduce stereotypy to low levels using two reinforcement-based procedures (i.e., NCR and DRA) – necessitating the use of empirically identified punishers. Like Verriden and Roscoe, we were unable to reduce VS

to low levels with NCR using unmatched stimulation and DRO – yet we were able to reduce VS to low levels using an empirically identified punisher.

Combining the DRO and NCR procedures decreased VS, but not consistently. The DRO initially decreased VS for all participants. However, across sessions, VS was variable, and slowly increasing, suggesting that the effectiveness of the DRO faded across sessions. These findings were consistent with those of Verriden and Roscoe (2018), whom used DRA with NCR. In Verriden and Roscoe's study, they observed similar patterns that we did during the implementation of the two reinforcement-based procedures. The treatment package (i.e., NCR plus DRA) utilized in their study did originally reduce stereotypy below baseline levels for all participants. However, reductions were not considered clinically significant nor did the reductions maintain across sessions. We saw the same pattern of responding as their participants. The treatment package (i.e., NCR plus DRO) was effective at reducing VS below baseline levels, but was not successful at maintaining those levels across sessions.

The current study systematically replicated Verriden and Roscoe's (2018) punisher assessment for use with VS. We evaluated the effects of numerous punishers delivered contingent on VS using a multielement design. This approach of utilizing a punisher assessment is likely an improvement over previous studies (e.g., Anderson & Le, 2011; Rapp et al., 2009) wherein punishment procedures were arbitrarily selected. Moreover, the punisher assessment allowed an opportunity for us to identify multiple punishment procedures that could be effective for treatment.

Although VS was brought under inhibitory stimulus control for one participant (Ethan) and debatably another (Nathan), the process required considerable personalization. Although, we were able to inhibit Ethan's VS in the presence of the bracelet, the process was challenging.

Not all participants responded to the procedures in the same way. As a result, substantial changes were made for each participant. John did not demonstrate inhibitory stimulus control during the multiple schedule; therefore, extra stimuli were added to increase discrimination. John's VS, however, remained undifferentiated – prompting the use of an alternating treatment design to allow his behavior more time to come in contact with the contingencies. Despite this step, John's VS remained undifferentiated. Therefore, we moved to randomly selecting one of the three most effective punishment procedures from his punishment procedure assessment for use in each session (Charlop et al., 1988). John's VS remained undifferentiated with an increasing trend. The procedures necessary for Ethan, however, differed from John's. It took a large number of changes, but we did establish discriminative effects with Ethan, while we were not able to do so with John. For Ethan, there were low undifferentiated levels across components even with the addition of extra stimuli and a change to an alternating treatment design. Consequently, we added instructions to at the beginning of each component (Tiger et al., 2016), but immediately saw low levels of VS across both components. We then switched the unmatched stimuli and finally saw differentiation across components. Due to these findings with Ethan, Nathan was immediately exposed to all changes (e.g., extra stimuli, instructions) to help reduce the number of sessions and time spent. Nathan's VS was immediately differentiated across both components, contingent on the implementation of the treatment package.

A limitation of this study was that an initially promising intervention (NCR with the matched stimulus) was not practical for classroom use. During the competing-items assessment, all participants' VS was lowest when the matched stimuli (i.e., auditory items) were provided. However, the matched stimuli were not realistic nor feasible for a classroom setting. Specifically, the research team felt that the matched stimuli (auditory) would be distracting to

other learners in the classroom, would compete with the child listening to the teacher, and would hinder learning – making them inappropriate for the classroom. Several researchers have investigated the effects of unmatched and matched stimulation on VS and have found that the effects of unmatched and matched stimuli for VS are idiosyncratic (e.g., Ahearn et al., 2005; Lanovaz & Argumedes, 2009), which is why we initially assessed both. However, our results demonstrated that the matched stimulus was more effective at competing with VS. Due to the lack of feasibility for the matched stimuli, it would have been better to have assessed the unmatched stimuli only. The matched stimuli did, however, generate clinically significant reductions in VS for all participants. This allowed us, the researchers, to realize our capabilities with regards to reducing VS to low levels. We can advise that NCR using matched stimulation is a respectable procedure for a comprehensive behavior plan for the reduction of VS.

Another limitation of this study was the inability to complete the study from start to finish for all participants due to the coronavirus pandemic. Ethan was the only participant who was able to complete the study, while the other two participants were note. Nathan had promising data, suggesting that inhibitory stimulus control could have been achieved. Meanwhile, John was unable to receive the full treatment package (i.e., extra stimuli and instructions) as Ethan, therefore, we were unable to say what his data would have been.

A final limitation of this study was the number of sessions that were needed to establish inhibitory stimulus control. For Ethan, it took over 140 sessions to attain stimulus control. The number of modifications that were made to achieve stimulus control were extensive. As a result, we were unable to complete the treatment package for all participants before social distancing came into effect. Future research should study the effects of immediate exposure of participants

to the treatment package to determine whether there is a more rapid differentiation across components, as a result.

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*(2), 247–250. https://doi.org/10.1901/jaba.2005.36-04

Ahearn, W. H., Clark, K. M., Gardenier, N. C., Chung, B. I., & Dube, W. V. (2003). Persistence of stereotypic behavior: Examining the effects of external reinforcers. *Journal of Applied Behavior Analysis, 36*(4), 439–448. https://doi.org/10.1901/jaba.2003.36-439

Ahearn, W. H., Clark, K. M., MacDonald, R. P., & Chung, B. I. (2007). Assessing and treating vocal stereotypy in children with autism. *Journal of Applied Behavior Analysis, 40*(2), 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*(1), 95–108. https://doi.org/10.1901/jaba.2011.44-95

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association Publishing. https://doi.org/10.1176/appi.books.9780890425596

Anderson, J., & Le, D. D. (2011). Abatement of intractable vocal stereotypy using an overcorrection procedure. *Behavioral Interventions, 26*(2), 134–146. https://doi.org/10.1002/bin.326

Autism Speaks. 2021. *Autism Statistics and Facts | Autism Speaks*. [online] Available at: <https://www.autismspeaks.org/autism-statistics> [Accessed 9 January, 2021]

Bailey, J. S., & Burch, M. (2016). In *Ethics for behavior analysts*. Taylor & Francis Ltd.

Behavior Analyst Certification Board. (2014). Professional and Ethical Compliance Code for Behavior Analysts. Retrieved from https://www.bacb.com/wpcontent/uploads/2017/09/170706-compliance-code-english.pd

Bodfish, J. W., Symons, F. J., Parker, D. E., & Lewis, M. H. (2000). Varieties of repetitive behavior in autism: Comparisons to mental retardation. *Journal of Autism and Developmental Disorders, 30*(3), 237–243. https://doi.org/10.1023/a:1005596502855

Brusa, E., & Richman, D. (2008). Developing stimulus control for occurrences of stereotypy exhibited by a child with autism. *International Journal of Behavioral Consultation and Therapy, 4*(3), 264–269. https://doi.org/10.1037/h0100855

Carr, J. E., Dozier, C. L., Patel, M. R., Adams, A. N., & Martin, N. (2002). Treatment of automatically reinforced object mouthing with noncontingent reinforcement and response blocking: Experimental analysis and social validation. *Research in Developmental Disabilities, 23*(1), 37–44. https://doi.org/10.1016/s0891-4222(01)00090-7

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*(1), 169–173. https://doi.org/10.1901/jaba.2011.44-169

Charlop, M. H., Burgio, L. D., Iwata, B. A., & Ivancic, M. T. (1988). Stimulus variation as a means of enhancing punishment effects. *Journal of Applied Behavior Analysis, 21*(1), 89–95. https://doi.org/10.1901/jaba.1988.21-89

Charlop, M. H., Kurtz, P. F., & Casey, F. G. (1990). Using aberrant behaviors as reinforcers for autistic children. *Journal of Applied Behavior Analysis, 23*(2), 163–181. https://doi.org/10.1901/jaba.1990.23-163

Chebli, S., Martin, V., & Lanovaz, M. (2016). Prevalence of stereotypy in individuals with developmental disabilities: a systematic review. *Journal of Autism and Developmental Disorders, 3*(2), 107-118. doi:10.1007/s40489-106-0069-x

Cook, J. L., Rapp, J. T., Gomes, L. A., Frazer, T. J., & Lindblad, T. L. (2014). Effects of verbal reprimands on targeted and untargeted stereotypy. *Behavioral Interventions, 29*(2), 106–124. https://doi.org/10.1002/bin.1378

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Pearson Education.

Cunningham, A. B., & Schreibman, L. (2008). Stereotypy in autism: The importance of function. *Research in Autism Spectrum Disorders, 2*(3), 469–479. https://doi.org/10.1016/j.rasd.2007.09.006

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*(4), 519–533. https://doi.org/10.1901/jaba.1996.29-519

DeRosa, N. M., Novak, M. D., Morley, A. J., & Roane, H. S. (2019). Comparing response blocking and response interruption/redirection on levels of motor stereotypy: Effects of data analysis procedures. *Journal of Applied Behavior Analysis, 52*(4), 1021–1033. https://doi.org/10.1002/jaba.644

DiGennaro Reed, F. D., Hirst, J. M., & Hyman, S. R. (2012). Assessment and treatment of stereotypic behavior in children with autism and other developmental disabilities: A thirty-year review. *Research in Autism Spectrum Disorders, 6*(1), 422–430. https://doi.org/10.1016/j.rasd.2011.07.003

Doughty, S. S., Anderson, C. M., Doughty, A. H., Williams, D. C., & Saunders, K. J. (2007). Discriminative control of punished stereotyped behavior in humans. *Journal of the Experimental Analysis of Behavior, 87*(3), 325–336. https://doi.org/10.1901/jeab.2007.39-05

Dunlap, G., Dyer, K., & Koegel, R. L. (1983). Autistic self-stimulation and intertrial interval duration. *American Journal of Mental Deficiency, 88*(2), 194–202.

Durand, V. M., & Carr, E. G. (1987). Social influences on "self-stimulatory" behavior: Analysis and treatment application. *Journal of Applied Behavior Analysis, 20*(2), 119–132. https://doi.org/10.1901/jaba.1987.20-119

Enloe, K., & Rapp, J. (2014). Effects of noncontingent social interaction on immediate and subsequent engagement in vocal and motor stereotypy in children with autism. Behavior Modification, 38(3), 374-391. doi:10.1177/0145445513514081

Falcomata, T. S., Roane, H. S., Hovanetz, A. N., Kettering, T. L., & Keeney, K. M. (2004). An evaluation of response cost in the treatment of inappropriate vocalizations maintained by automatic reinforcement. *Journal of Applied Behavior Analysis, 37*(1), 83–87. https://doi.org/10.1901/jaba.2004.37-83

Fisher, W. W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis, 25*(2), 491–498. https://doi.org/10.1901/jaba.1992.25-491

Fisher, W. W., Piazza, C. C., Bowman, L. G., Kurtz, P. F., Sherer, M. R., & Lachman, S. R. (1994). A preliminary evaluation of empirically derived consequences for the treatment

of pica. *Journal of Applied Behavior Analysis, 27*(3), 447–457. https://doi.org/10.1901/jaba.1994.27-447

Foster, L. G. (1998). Nervous habits and stereotyped behaviors in preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*(7), 711–717. https://doi.org/10.1097/00004583-199807000-00010

Friman, P. C., & Poling, A. (1995). Making life easier with effort: Basic findings and applied research on response effort. *Journal of Applied Behavior Analysis, 28*(4), 583-590. https://doi: 10.1901/jaba.1995.28-583

Gibney, C., Phillips, K. J., Arnold-Saritepe, A., & Taylor, S. A. (2019). An evaluation of the effects of response interruption redirection and matched stimulation on vocal stereotypy. *Behavioral Interventions*, *35*(1), 114–130. https://doi.org/10.1002/bin.1700

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*(6), 1691–1700. https://doi.org/10.1016/j.ridd.2012.05.008

Gould, K. M., Harper, J. M., Gillich, E., Luiselli, J. K. (2018). Intervention, stimulus control, and generalization effects of response interruption and redirection on motor stereotypy. *Behavioral Interventions,* 1-9. https://doi.org/10.1002/bin.1652

Hammond, J. L., Iwata, B. A., Rooker, G. W., Fritz, J. N., & Bloom, S. E. (2013). Effects of fixed verses random condition sequencing during multielement functional analyses. *Journal of Applied Behavior Analysis*, *46*(1), 22–30. https://doi.org/10.1002/jaba.7

Hanley, G. P., Iwata, B. A., Thompson, R. H., & Lindberg, J. S. (2000). A component analysis of "stereotypy as reinforcement" for alternative behavior. *Journal of Applied Behavior Analysis*, *33*(3), 285–297. https://doi.org/10.1901/jaba.2000.33-285

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*(2), 197–209. (Reprinted from *Analysis and Intervention in Developmental Disabilities, 2*(1), 3–20, 1982). https://doi.org/10.1901/jaba.1994.27-197

Jones, R. S., Wint, D., & Ellis, N. C. (1990). The social effects of stereotyped behaviour. *Journal of Mental Deficiency Research, 34*(3), 261–268. https://doi.org/10.1111/j.1365-2788.1990.tb01537.x

Kennedy, C. H. (1994). Automatic reinforcement: Oxymoron or hypothetical construct? *Journal of Behavioral Education, 4*(4), 387–395. https://doi.org/10.1007/BF01539540

Koegel, R. L., & Covert, A. (1972). The relationship of self-stimulation to learning in autistic children. *Journal of Applied Behavior Analysis, 5*(4), 381–387. https://doi.org/10.1901/jaba.1972.5-381

Lanovaz, M. J., Rapp, J. T., & Ferguson, S. (2013). Assessment and treatment of vocal stereotypy associated with television: a pilot study. Journal of Applied Behavior Analysis, 46(2), 544-548. doi:10.1002/jaba.35

Lanovaz, M. J., & Sladeczek, I. E. (2012). Vocal stereotypy in individuals with autism spectrum disorders: A review of behavioral interventions. *Behavior Modification, 36*(2), 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*(3), 647–651. https://doi.org/10.1901/jaba.2011.44-647

Lerman, D. C., & Vorndran, C. M. (2002). On the status of knowledge for using punishment: Implications for treating behavior disorders. *Journal of Applied Behavior Analysis, 35*(4), 431–464. https://doi.org/10.1901/jaba.2002.35-431

Lovaas, I., Newsom, C., & Hickman, C. (1987). Self-stimulatory behavior and perceptual reinforcement. *Journal of Applied Behavior Analysis, 20*(1), 45–68. https://doi.org/10.1901/jaba.1987.20-45

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. doi: 10.1901/jaba.2012.45-549

Liu-Gitz, L., & Banda, D. R. (2009). A replication of the RIRD strategy to decrease vocal stereotypy in a student with autism. Behavioral Interventions, 25(1), 77-87. doi:10.1002/bin.297

MacDonald, R., Green, G., Mansfield, R., Geckeler, A., Gardenier, N., Anderson, J., Holcomb, W., & Sanchez, J. (2007). Stereotypy in young children with autism and typically developing children. *Research in Developmental Disabilities, 28*(3), 266–277. https://doi.org/10.1016/j.ridd.2006.01.004

Malott, R. (2007). *Principles of Behaviour*. New Jersey: Pearson Prentice Hall.

Mancina, C., Tankersley, M., Kamps, D., Kravits, T., & Parrett, J. (2000). Brief report: Reduction of inappropriate vocalizations for a child with autism using a self-management treatment program. *Journal of Autism and Developmental Disorders, 30*(6), 599–606. https://doi.org/10.1023/A:1005695512163

Matson, J. L., Kiely, S. L., & Bamburg, J. W. (1997). The effect of stereotypies on adaptive skills as assessed with the DASH-II and vineland adaptive behavior scales. *Research in*

Developmental Disabilities, 18(6), 471–476. https://doi.org/10.1016/s0891-4222(97)00023-1

Matson, J. L., Minshawi, N. F., Gonzalez, M. L., & Mayville, S. B. (2006). The relationship of comorbid problem behaviors to social skills in persons with profound mental retardation. *Behavior Modification, 30*(4), 496–506. https://doi.org/10.1177/0145445505283415

Mayes, S. D., & Calhoun, S. L. (2011). Impact of IQ, age, SES, gender, and race on autistic symptoms. *Research in Autism Spectrum Disorders, 5*(2), 749–757. https://doi.org/10.1016/j.rasd.2010.09.002

McDowell, J. J. (1988). Matching theory in natural human environments. *The Behavior Analyst, 11*(2), 95–109. https://doi.org/10.1007/BF03392462

McKearney, J. W. (1977). Asking questions about behavior. *Perspective in Biology and Medicine, 21*(1), 109–119. https://doi.org/10.1353/pbm.1977.0015

Miltenberger, R. G. (1997). *Behavior modifications: Principles and procedures.* Pacific Grove, CA: Brooks/Cole.

O'Connor, A. S., Prieto, J., Hoffman, B., DeQuinzio, J. A., & Taylor, B. A. (2011). A stimulus control procedure to decrease motor and vocal stereotypy. *Behavioral Interventions, 26*(3), 231–242. https://doi.org/10.1002/bin.335

Pace, G. M., Ivancic, M. T., Edwards, G. L., Iwata, B., & Page, T. J. (1985). Assessment of stimulus preference and reinforcer value with profoundly retarded individuals. *Journal of Applied Behavior Analysis, 18*(3), 249–255. https://doi.org/10.1901/jaba.1985.18-249

Pastrana, S., Rapp, J., & Frewing, T. (2013). Immediate and subsequent effects of response interruption and redirection on targeted and untargeted forms of stereotypy. Behavior Modification, 37(4), 591-610. doi:10.1177/0145445513485751

Peters, L. C., & Thompson, R. H. (2013). Some indirect effects of positive practice overcorrection. *Journal Applied Behavior Analysis, 46*(3), 613–625. https://doi.org/10.1002/jaba.63

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*(1), 13–27. https://doi.org/10.1901/jaba.2000.33-13

Piazza, C. C., Fisher, W. W., Hanley, G. P., Hilker, K., & Derby, K. M. (1996). A preliminary procedure for predicting the positive and negative effects of reinforcement-based procedures. *Journal of Applied Behavior Analysis, 29*(2), 137–152. https://doi.org/10.1901/jaba.1996.29-137

Potter, J. N., Hanley, G. P., Augustine, M., Clay, C. J., & Phelps, M. C. (2013). Treating stereotypy in adolescents diagnosed with autism by refining the tactics of "using stereotypy as reinforcement". *Journal of Applied Behavior Analysis, 46*(2), 407–423. https://doi.org/10.1002/jaba.52

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*(1), 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*(1), 73–88. https://doi.org/10.1901/jaba.2007.142-05

Rapp, J. T., Cook, J. L., McHugh, C., & Mann, K. R. (2017). Decreasing stereotypy using NCR and DRO with functionally matched stimulation: Effects on targeted and non-targeted

stereotypy. *Behavior Modification, 41*(1), 45–83. https://doi.org/10.1177/0145445516652370

Rapp, J. T., Patel, M. R., Ghezzi, P. M., O'Flaherty, C. H., & Titterington, C. J. (2009). Establishing stimulus control of vocal stereotypy displayed by young children with autism. *Behavioral Interventions, 24*(2), 85–105. https://doi.org/10.1002/bin.276

Rapp, J. T., & Vollmer, T. R. (2005). Stereotypy I: A review of behavioral assessment and treatment. *Research in Developmental Disabilities, 26*(6), 527–547. https://doi.org/10.1016/j.ridd.2004.11.005

Rincover, A. (1978). Sensory extinction: A procedure for eliminating self-stimulatory behavior in developmentally disabled children. *Journal of Abnormal Child Psychology, 6*, 299–310. https://doi.org/10.1007/BF00924733

Rincover, A., Cook, R., Peoples, A., & Packard, D. (1979). Sensory extinction and sensory reinforcement principles for programming multiple adaptive behavior change. *Journal of Applied Behavior Analysis, 12*(2), 221–233. https://doi.org/10.1901/jaba.1979.12-221

Roane, H. S., Vollmer, T. R., Ringdahl, J. E., & Marcus, B. A. (1998). Evaluation of a brief stimulus preference assessment. *Journal of Applied Behavior Analysis, 31*(4), 605–620. https://doi.org/10.1901/jaba.1998.31-605

Roantree, C. F., & Kennedy, C. H. (2006). A paradoxical effect of precession attention on stereotypy: Antecedent attention as an establishing, not an abolishing, operation. *Journal of Applied Behavior Analysis, 39*(3), 381–384. https://doi.org/10.1901/jaba.2006.97-05

Rogers, S. J., & Ozonoff, S. (2005). Annotation: What do we know about sensory dysfunction in autism? A critical review of the empirical evidence. *Journal of Child Psychology and Psychiatry, 46*(12), 1255–1268. https://doi.org/10.1111/j.1469-7610.2005.01431.x

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. doi:10.1901/jaba.2012.45-185

Shawler, L. A., Dianda, M., & Miguel, C. F. (2019). A comparison of response interruption and redirection and competing items on vocal stereotypy and appropriate vocalizations. *Journal of Applied Behavior Analysis, 53*(1), 355-365. https://doi.org/10.1002/jaba.596

Shawler, L., & Miguel, C. (2015). The effects of motor and vocal response interruption and redirection on vocal stereotypy and appropriate vocalizations. Behavioral Interventions, 30(2), 112-134. doi:10.1002/bin.1407

Sidman, M. (1960). Appendix. In *Tactics of scientific research* (pp. 405-406). New York: Basic Books, Pub.

Spencer, V. G., & Alkhanji, R. (2018). Response interruption and redirection (RIRD) as a behavioral intervention for vocal stereotypy: A systematic review. *Education and Training in Autism and Developmental Disabilities, 53*(1), 33–43. doi:10.2307/26420425

Taylor, B. A., Hoch, H., & Weissman, M. (2005). The analysis and treatment of vocal stereotypy in a child with autism. *Behavioral Interventions, 20*(4), 239–253. https://doi.org/10.1002/bin.200

Thomeer, M. L., McDonald, C. A., Rodgers, J. D., & Lopata, C. (2017). Highfunctioning autism spectrum disorder: A framework for evidence-based practices. School Mental Health. doi:10.1007/s12310-017-9236-1

Tiger, J. H., Wierzba, B. C., Fisher, W. W., & Benitez, B. B. (2017). Developing and demonstrating inhibitory stimulus control over repetitive behavior. *Behavioral Interventions, 32*(2), 160–174. https://doi.org/10.1002/bin.1472

Toole, L. M., DeLeon, I. G., Kahng, S., Ruffin, G. E., Pletcher, C. A., & Bowman, L. G. (2004). Re-evaluation of constant versus varied punishers using empirically derived consequences. *Research in Developmental Disabilities, 25*(6), 577–586. https://doi.org/10.1016/j.ridd.2004.03.005

Van Houten, R., Axelrod, S., Bailey, J. S., Favell, J. E., Foxx, R. M., Iwata, B. A., & Lovaas, O. I. (1988). The right to effective behavioral treatment. *Journal of Behavior Analysis, 21*(4), 381–384. https://doi.org/10.1901/jaba.1988.21-381

Verriden, A. L., & Roscoe, E. M. (2019). An evaluation of a punisher assessment for decreasing automatically reinforced problem behavior. *Journal of Applied Behavior Analysis, 52*(1), 205–226. https://doi.org/10.1002/jaba.509

Vollmer, T. R. (1994). The concept of automatic reinforcement: Implications for behavioral research in developmental disabilities. *Research in Developmental Disabilities, 15,* 187-207. doi: 10.1016/0891-4222(94)90011-6

Vollmer, T. R., Marcus, B. A., Ringdahl, J. E., & Roane, H. S. (1995). Progressing from brief assessments to extended experimental analyses in the evaluation of aberrant behavior. *Journal of Applied Behavior Analysis, 28*(4), 561-576. https://doi.org/10.1901/jaba.1995.28-561

Watkins, N., Paananen, L., Rudrud, E., & Rapp J. T. (2011). Treating vocal stereotypy with environmental enrichment and response cost. *Journal of Clinical Case Studies, 10*(6), 440-448.

Winett, R. A., Moore, J. F., & Anderson, E. S. (1991). Extending the concept of social validity: Behavior analysis for disease prevention and health promotion. *Journal of Applied Behavioral Analysis, 24*(2), 215–230. https://doi.org/10.1901/jaba.1991.24-215

Wolery, M., Gast, D. L., & Ledford, J. R. (2014). Comparison Designs. In D.L. Gast, B.P. Lloyd, & J.R. Ledford (Eds) *Single case research methodology: Applications in special education and behavioral sciences.* Routledge, New York.

Wolery, M., Kirk, K., & Gast, D. L. (1985). Stereotypic behavior as a reinforcer: Effects and side effects. *Journal of Autism and Developmental Disorders, 15*(2), 149–161. https://doi.org/10.1007/BF01531601

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*(2), 203–214. https://doi.org/10.1901/jaba.1978.11-203

Wong, C., Odum, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., … & Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. Journal of Autism and Developmental Disabilities, 45, 1951-1966. doi:10.1007/s10803- 014-2351-z

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*(4), 749–764. https://doi.org/10.1002/jaba.227

Tables

Table 1
The Order of the Conditions in the Treatment Evaluation

Conditions		
John	Ethan	Nathan
Baseline	Baseline	Baseline
NCR:M	NCR:M & UM	NCR:M & UM
NCR:M & UM	NCR:UM	NCR:M
NCR:M	NCR:M & UM	NCR:M & UM
NCR:M& UM	NCR:UM	NCR:UM
NCR:UM & DRO	NCR:UM & DRO	NCR:M & UM
Punisher Assessment & NCR:UM& DRO	Punisher Assessment & NCR:UM& DRO	NCR:UM
NCR:UM & DRO	NCR:UM & DRO	NCR:UM & DRO
Punisher Assessment & NCR:UM	NCR:M & UM	NCR:UM
NCR:UM	NCR:UM	Punisher Assessment & NCR:UM
	Punisher Assessment & NCR:UM	NCR:UM
	NCR:UM	

Table 2

Social Validity Questions and Lead Clinician Ratings

Questions
1. Do you think the treatment involving noncontingent access to toys and *punishment procedure* contingent on vocal stereotypy were acceptable?
2. Do you think the behavior change was acceptable or sufficient?
3. Do you feel that the goals of this treatment were acceptable, appropriate, and important to the individual client?
4. Do you think that the client's teachers can effectively implement this procedure with integrity?

Items will be scored on a Likert-type scale (1) totally unacceptable (2) unacceptable (3) somewhat unacceptable (4) neutral (5) slight acceptable (6) acceptable (7) perfectly acceptable

Table 3
MSWO Results for FA

Order Selected	Items		
	John	Ethan	Nathan
1	Medieval Knights	City Blocks	Coloring
2	Dinosaur Figurines	Number Board	Reading Books
3	Legos	Play-doh	Slime
4	Hulk Smash	Puzzles	Legos
5	Spiderman	Reading Books	Lightning McQueen
6	Coloring	Jenga Blocks	Fidget Spinners
7	Fidget Spinners	Race tracks	Sand
8	Race Cars	Board game	Puzzles
9	Play-doh	Bubbles	Dinosaur Figurines
10	Reading Books	Race Cars	Bubbles

Table 4

Participant's Average Exposure to Each Punishment Procedure During the Punisher Assessment

	Average Percent of Exposure		
Punishment Procedure	John	Ethan	Nathan
In-Seat Timeout	24	19	16
Verbal Reprimands	13	20	7
Response Cost	16	39	16
Facial Screen	23	N/A	2
Tidiness Training	23	24	11
Overcorrection-Positive Practice	44	23	18
Contingent Motor Demands	49	18	34
Hands Down	43	N/A	2

Figures

Figure 1

Results from Initial Treatment Evaluation

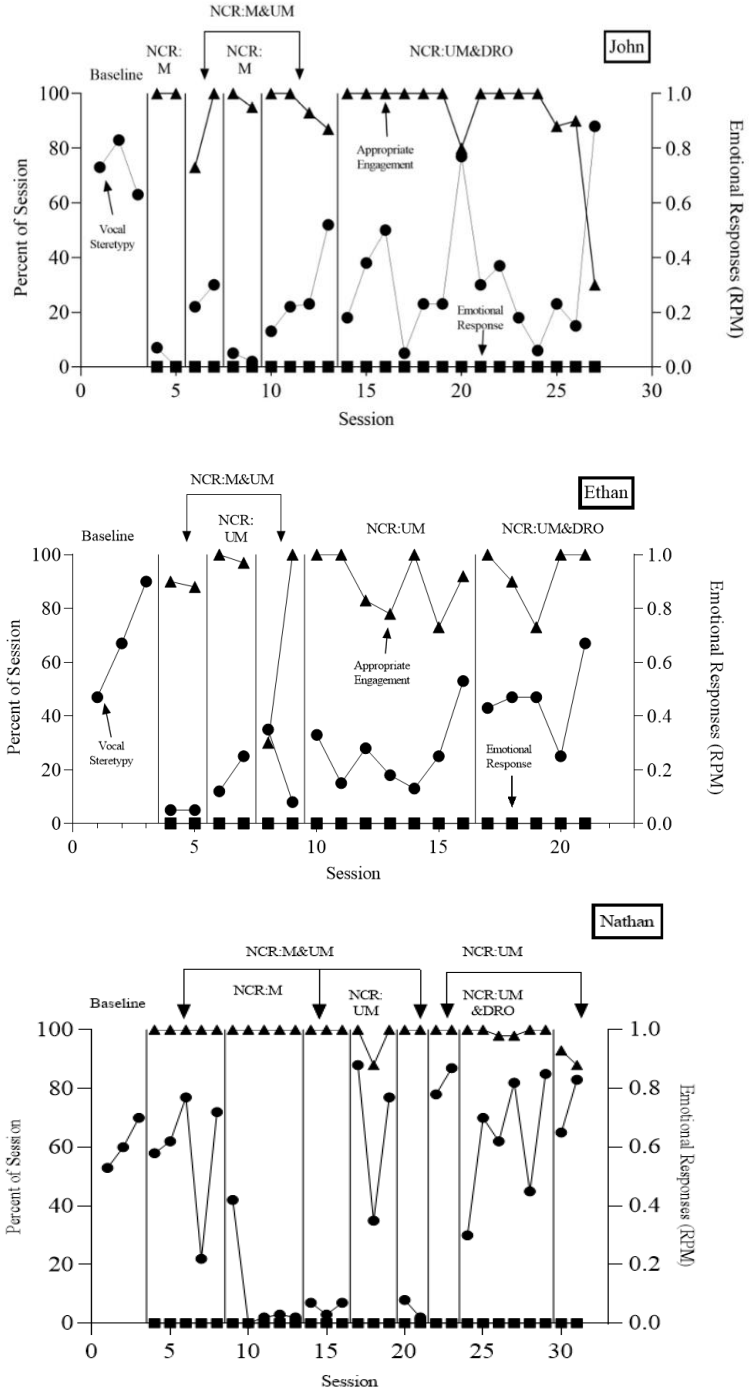


Figure 2

Results from John's Punisher Assessment

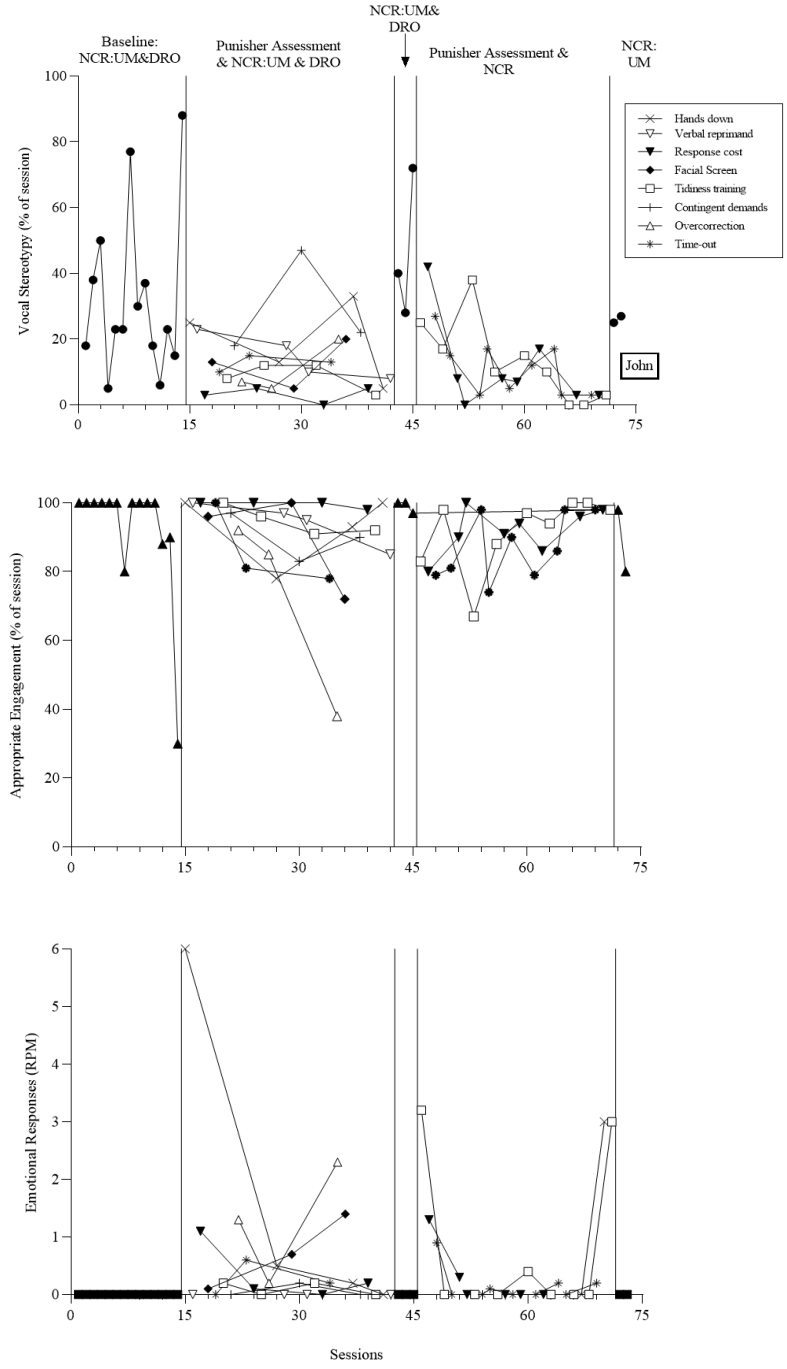


Figure 3

Results from Ethan's Punisher Assessment

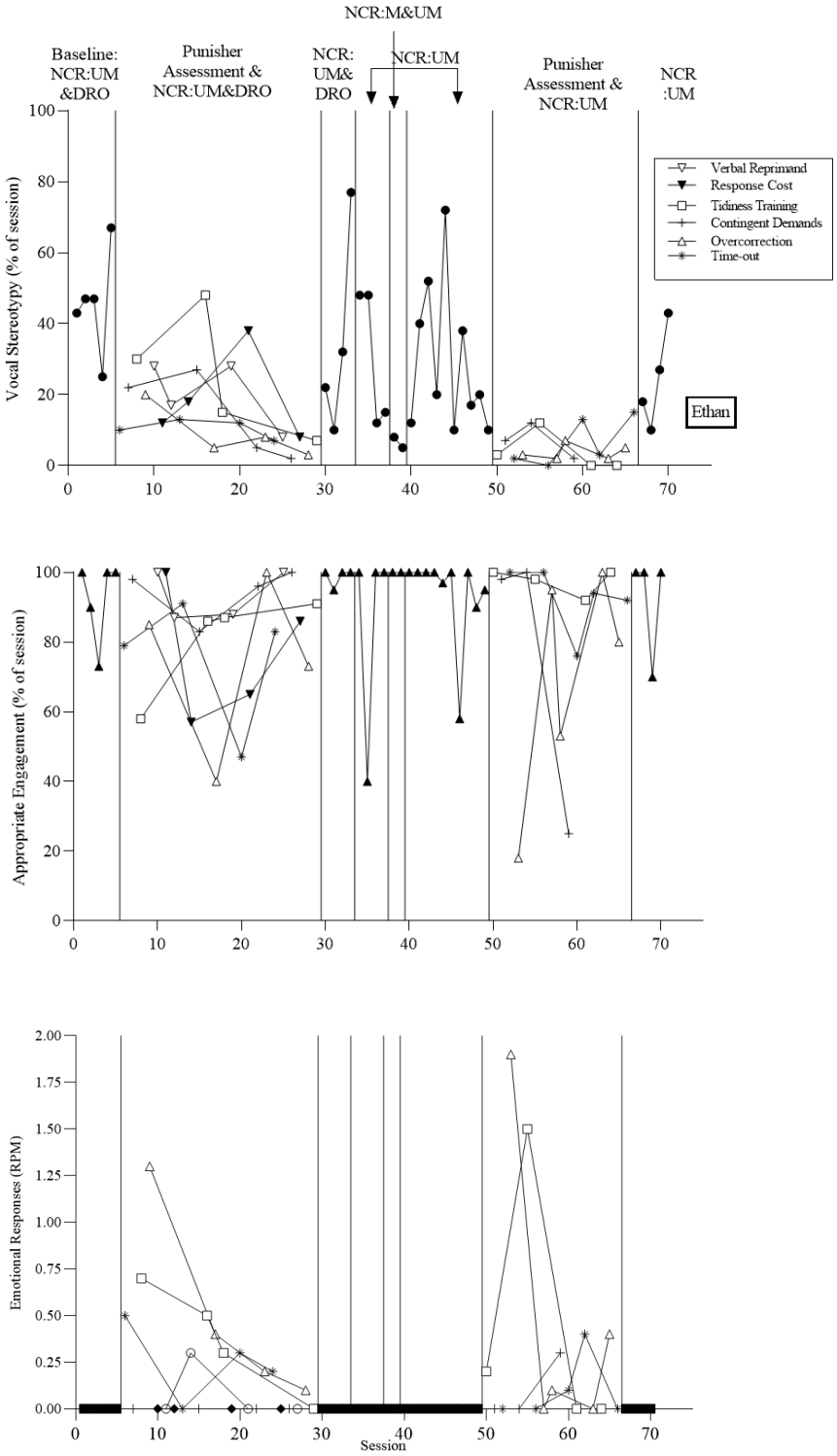


Figure 4

Results from Nathan's Punisher Assessment

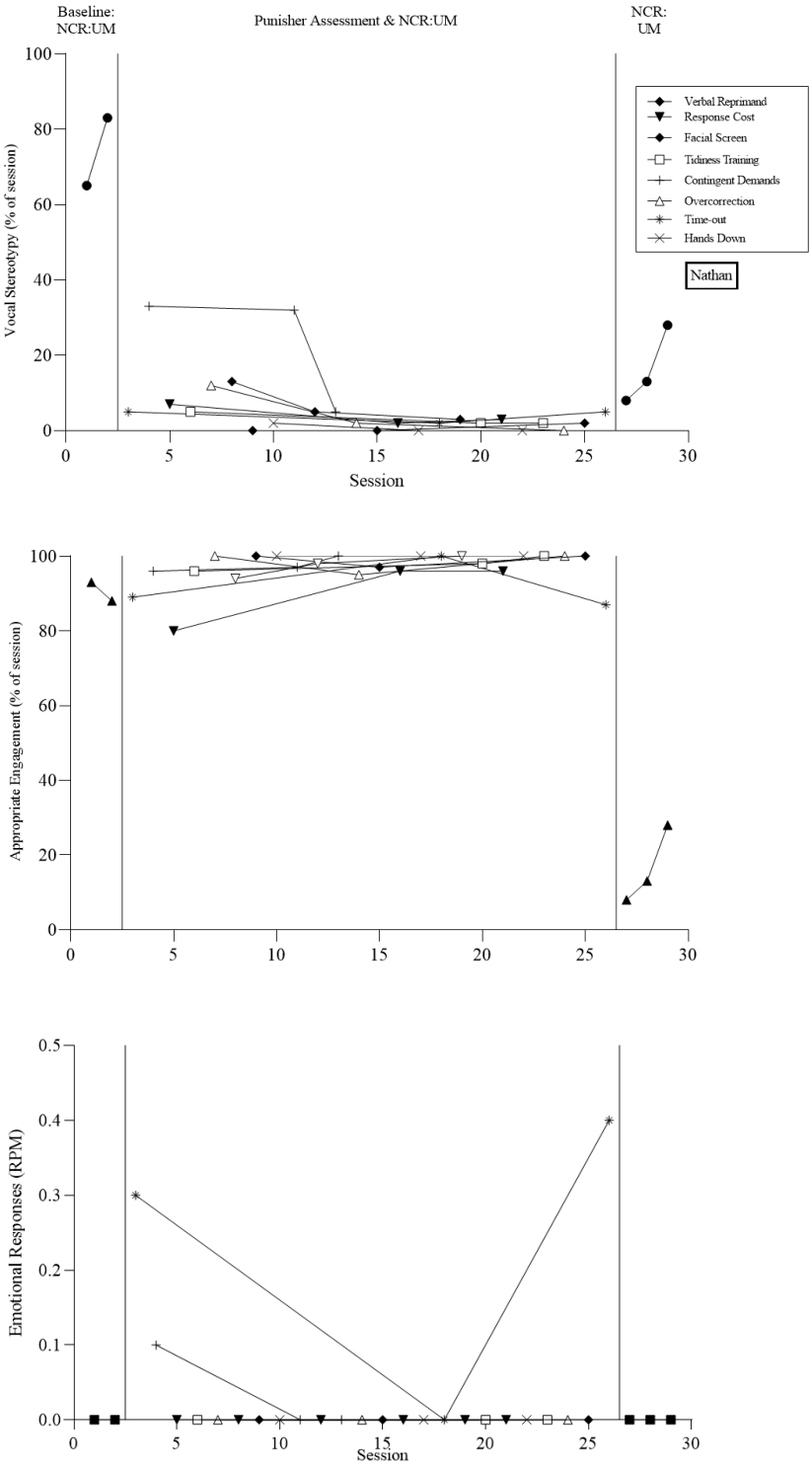


Figure 5

Results from All Participants' Discrimination Training

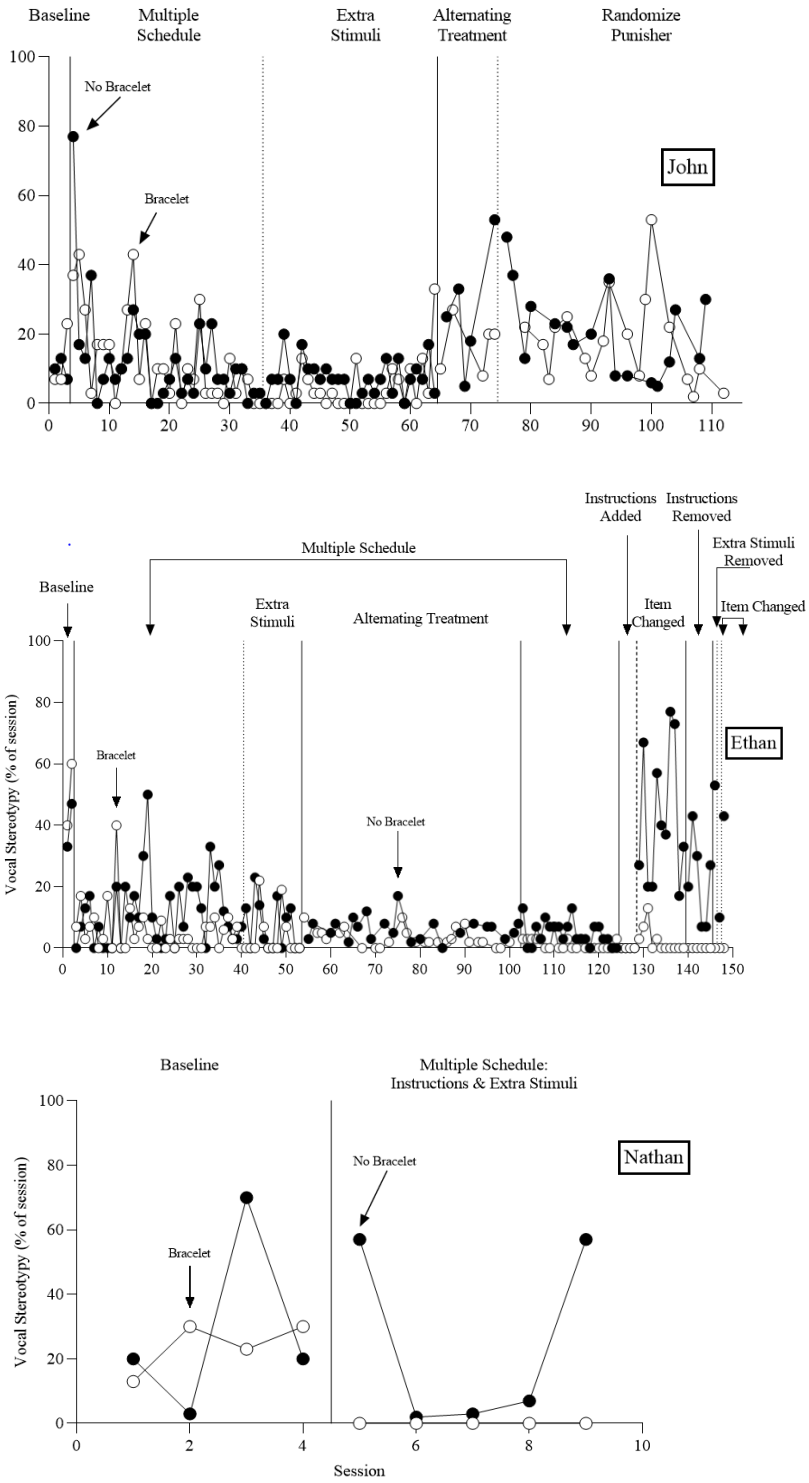
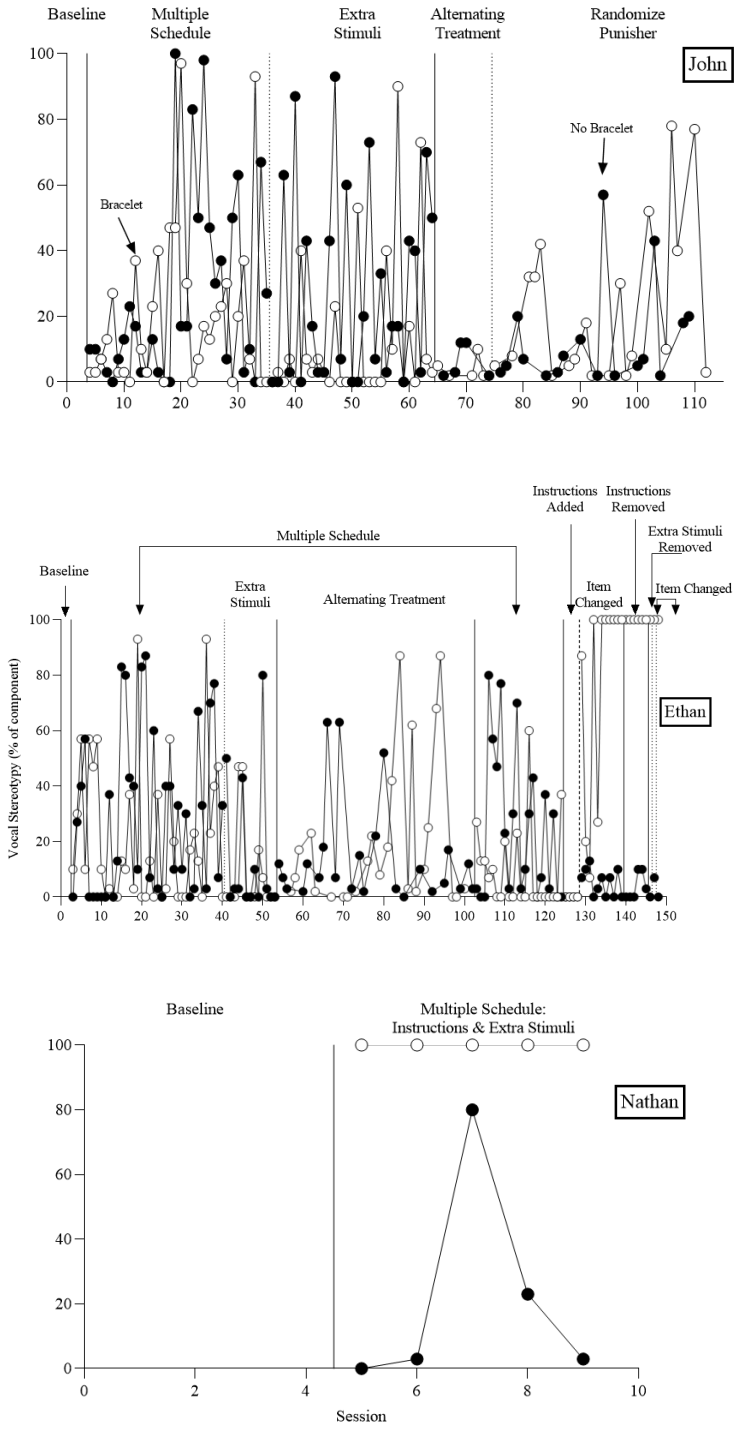


Figure 6

Results from All Participants' Percentage of the Component Elapsed prior to First Interval with Vocal Stereotypy



Appendices

Appendix A

Multiple-Stimulus Without Replacement (MSWO) Preference Assessment

Participants were exposed to only two assessment sessions in this study. The first, an MSWO preference assessment, was conducted for each participant based on the procedures outlined by DeLeon and Iwata (1996) in order to identify leisure items that would then be used during the second assessment session, the functional analysis.

Prior to the beginning of the first session, participants were given 30-s access to each of the leisure items included in the study (10 in total). Leisure items assessed were preferred items recommended by the participants' lead clinician. Each session then began with all items sequenced randomly in a straight line on the floor. The participant was seated on the floor across from the stimulus array, while the therapist was positioned opposite the participant. The therapist instructed the participant to select one item. After a selection was made, the participant was given 30 seconds to engage with the item, then the item was removed from the immediate area. Prior to the next trial, the sequencing of the remaining items was rotated by taking the item at the left end of the line and moving it to the right end, then shifting the other items so that they were again equally spaced on the floor. The second trial then followed immediately. This procedure continued until all items were selected or until a participant made no selection within 30 s of the beginning of a trial. A selection response was recorded when the participant made physical contact with one of the items; this response was then recorded on data sheets that were customized for the MSWO (i.e., recording the first to last selection).

Results from the MSWO preference assessment are shown in Table 4. John's most preferred items were medieval knights and dinosaurs. His moderate preferred items were Legos

and superheroes (e.g., Hulk, Spiderman). For Ethan, his highest preferred items were City Blocks and the Number Board. His moderate preferred items were play-doh and a puzzle. Nathan's highest preferred items were coloring and reading books. His moderate preferred items were slime and Legos.

Appendix B

Functional Analysis (FA)

Functional analyses, based on the procedures by Iwata et al. (1982, 1994), were then conducted in the second assessment session. Five conditions were alternated in a predetermined sequence (i.e., alone, ignore, attention, play, demand). A multielement design was employed to maximize participants' motivation for the reinforcers available in each condition (Hammond et al., 2013). Sessions lasted 5 minutes in duration and were conducted as follows:

- **Alone.** In the alone condition, the participant was alone in the room with only the furniture already in the room. Stereotypy resulted in no programmed consequences.
- **Ignore.** In this condition, the participant and therapist were both in the room with only each other and the furniture. VS resulted in no programmed consequences.
- **Attention.** During the attention condition, the participant had access to moderately preferred items (see Table 5) selected in the MSWO assessment. The therapist sat near the participant and pretended to be busy by looking at different pieces of paper. Contingent on the occurrence of stereotypy, the therapist immediately delivered brief vocal and physical attention for 3–5 s (e.g., telling the participant to not make that sound, while touching their shoulder).
- **Demand.** During the demand condition, one task that was within the participant's skill set was continuously presented using three-step guided compliance. This consisted of the therapist giving the initial instruction (i.e., making a demand); if the participant did not respond within 5 s, the instruction was repeated again. If the participant still refused to comply with the instruction, the therapist would model the correct response. If this was not effective in initiating compliance, the therapist would gently guide the participant to comply

with the original instruction. Contingent on the occurrence of VS, the work task or instruction was immediately removed, while the therapist slightly turned their body away from the participant for 30 s.

- **Play.** During the play condition, the participant had access to highly preferred items (see Table 5) selected in the MSWO and received brief bouts of attention (i.e., vocal and physical) every 30 s. Attention was given in the form of a comment on the participant's play along with some form of physical touch (e.g., a high five). There were no programmed consequences for the occurrence of VS, even if it occurred during the implementation of attention.

The results of functional analyses are shown in Figure 1. In John's FA (top panel), extra assessment in the alone condition (Vollmer et al., 1995) was conducted because his response in the alone condition was undifferentiated initially in the assessment from their response in the play condition.

In the play condition, John exhibited initially high levels of VS, though across sessions VS decreased to relatively low levels (i.e., from 13% to 60% of intervals). John presented elevated levels of VS during the alone condition (i.e., from 37% to 83% of intervals) relative to the play condition. John meanwhile demonstrated relatively low levels of VS during the demand condition (i.e., from 7% to 17% of intervals) and moderate to low levels during the attention condition (i.e., from 20% to 30% of intervals), relative to the play condition. These data suggest that VS was maintained by nonsocial sources of reinforcement.

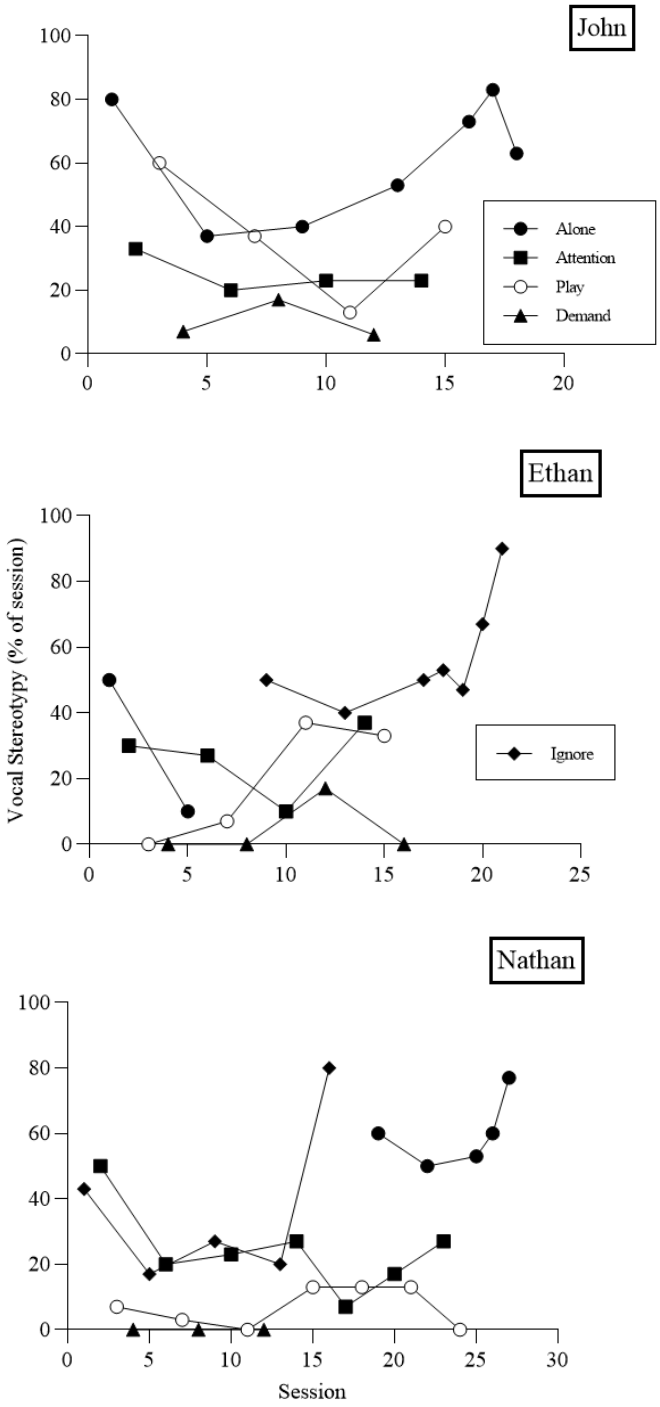
In Ethan's FA (middle panel), he presented moderate to low levels of VS during the play condition (i.e., from 0% to 37% of intervals). In the alone condition, Ethan exhibited an initial moderate high response (i.e., 50%) of VS, but then decreased to low levels (i.e., 10%). Due to

unsafe circumstances, an ignore condition was implemented in place of the alone condition. Ethan exhibited relatively low levels of VS during the demand condition when compared to the play condition (i.e., from 0% to 17% of intervals). In the play and attention conditions there were moderately low levels of VS relative to all other conditions (i.e., from 17% to 37% of intervals). As VS persisted in the play condition, extra assessment in the ignore condition was conducted to determine whether the behavior persisted under repeated exposure to no social consequences (Vollmer et al., 1995). In the repeats of the ignore condition, Ethan demonstrated relatively high levels of VS (i.e., from 40% to 90% of intervals), in comparison to the play condition, and the behavior persisted. These data suggest that VS was maintained by nonsocial sources of reinforcement.

In Nathan's FA (bottom panel), he demonstrated low levels of VS during the play condition (i.e., from 0% to 13% of intervals) and did not engage in any VS during the demand condition. Nathan presented moderately low levels of VS in the attention condition (i.e., from 7% to 50% of intervals) in comparison to the play condition. The ignore condition elicited moderate levels of responding initially, which then increased to high levels (i.e., from 17% to 80% of intervals) in comparison to the play condition. Vocal stereotypy persisted at relatively high levels in the ignore condition (range, 50%to 77%), therefore suggesting that VS was maintained by nonsocial sources of reinforcement.

Figure 7

Results from All Participants' Functional Analyses



Appendix C

Competing-Items Preference Assessment

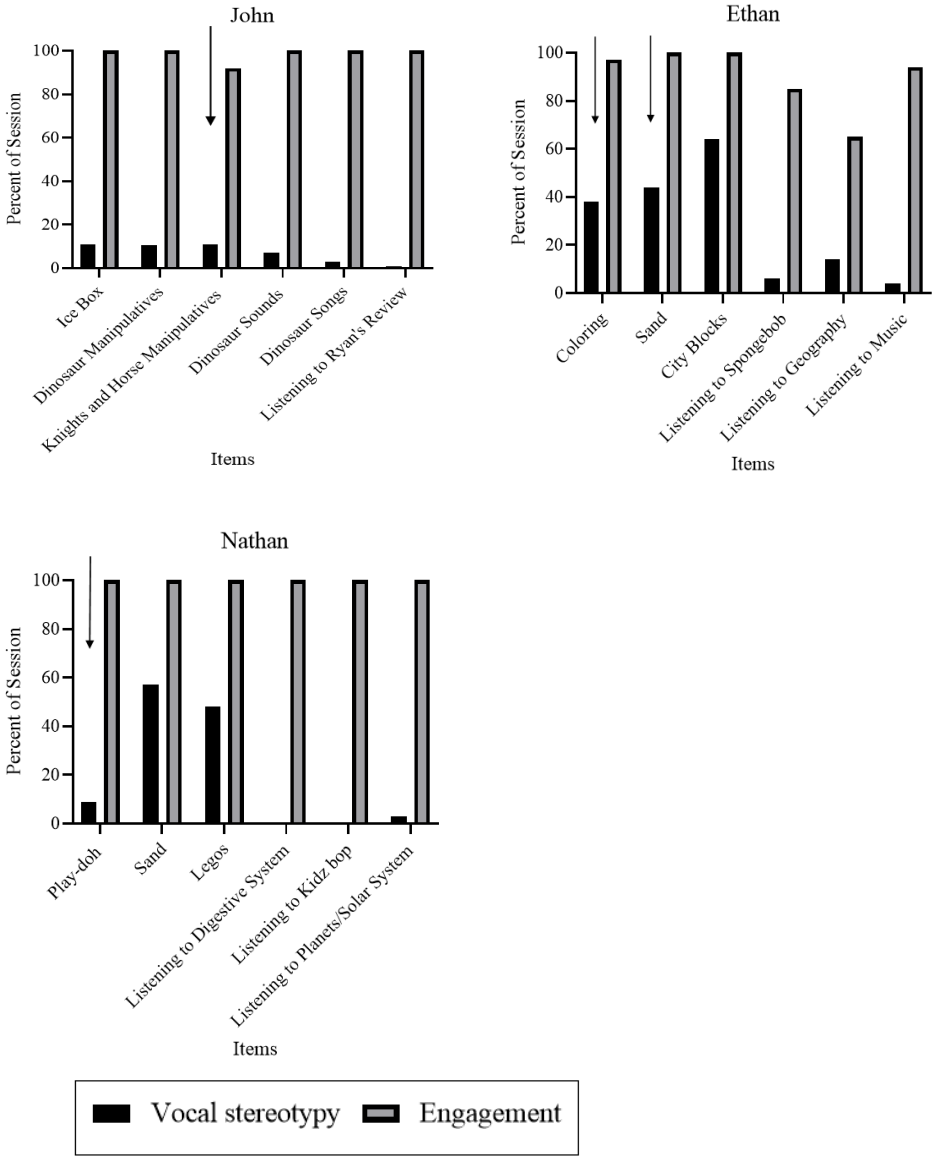
A competing-items preference assessment (Piazza et al., 1996) was conducted to identify leisure items for use during the treatment. Participants' clinicians nominated six items to be included in the corresponding participant's competing-items preference assessment. Three of the items were thought to be matched to the suspected sensory stimulation produced by a participant's VS (i.e., auditory stimulation), while the other three were not (Piazza et al., 2000). The items were presented independently for 3 minutes, two times each, in a random order without replacement. Observers recorded the duration of VS and appropriate engagement exhibited in response by the participant; data were summarized as the percent of the session. Data were recorded at the immediate onset and offset. The leisure item that was associated with the highest percent of engagement and the lowest percent of VS was included in the treatment analysis. However, because an item that produced auditory stimulation was associated with the lowest levels of VS for each participant, a manipulable leisure item (e.g., toy figurine) that was also associated with low levels of VS was selected for the treatment analysis instead as, from a practical standpoint, continuous access to auditory stimulation could not be recommended as a treatment for the behavior in a classroom environment. We had hoped both auditory and leisure items would work equally as they did in Verriden and Roscoe (2018). Therefore, an item that was also associated with low levels of VS relative to those of the auditory item, but that more closely mimicked play and suited educational conditions, was included instead (see Figure 2).

The data from this assessment are shown in Figure 2. All three participants during the competing-items preference assessment engaged in high levels of appropriate engagement for all six items (i.e., from 65% to 100% of session). Across participants, VS was relatively low for the

three items that produced auditory stimulation (i.e., from 0% to 64% of session). John demonstrated moderately low levels of VS for the unmatched items relative to the matched items (i.e., from 10.5% to 11%). Ethan demonstrated higher levels of VS for the unmatched items in comparison to the matched items (i.e., from 38% to 64% of session). Nathan presented higher levels of VS for two of the unmatched items (i.e., sand and Legos; i.e., 48% and 57% of session) with low levels for the other unmatched item (i.e., play-doh; i.e., 9% of session), in comparison to the matched items. All three items that produced auditory stimulation demonstrated relatively low levels of VS (i.e., from 0% to 3% of session).

Figure 8

Results from All Participants' Competing Items Assessment



Arrows denote the items included in treatment.

Appendix D

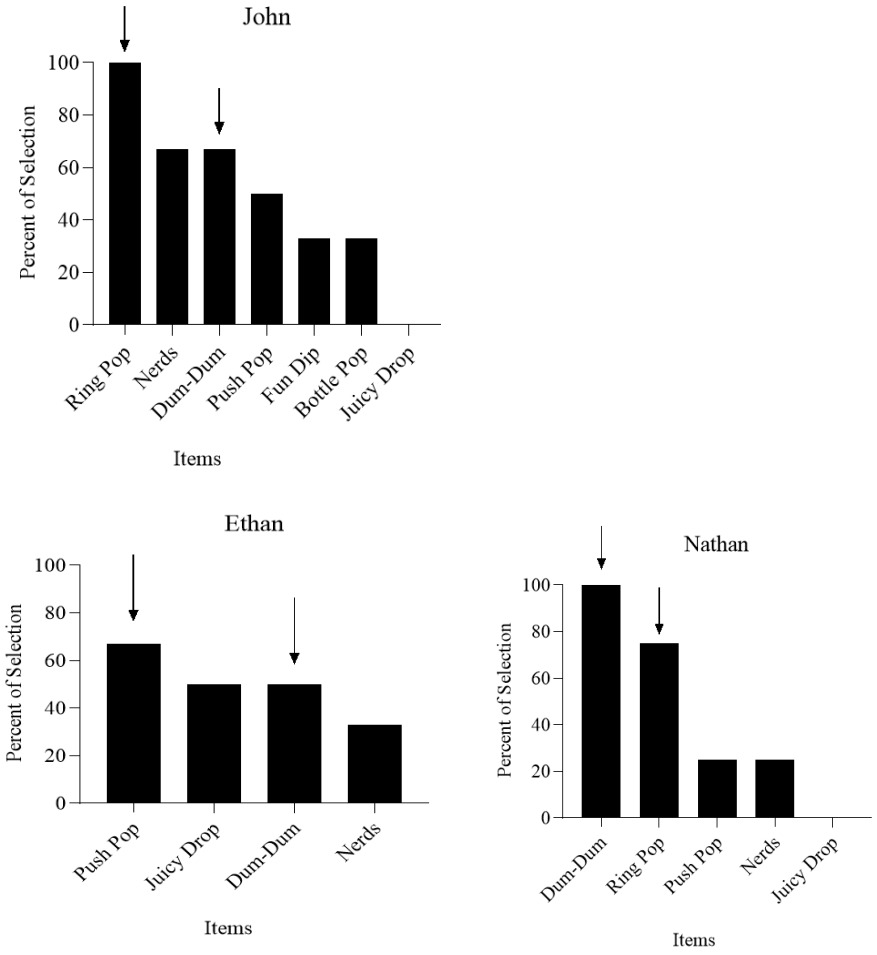
Paired-Stimulus Preference Assessment

A paired-stimulus preference assessment (PSPA; Fisher et al., 1992) was conducted to identify edible items for use during the treatment evaluation. The participants sampled each edible item (i.e., they were allowed to eat it) before their evaluation in the paired-stimulus preference assessment. Once participants tasted each edible item, the items were then presented to the participants in pairs. The therapist would present one edible first and state what it was; the item was then removed. The participant was then presented with the other edible item and told what it was; the item was then removed. After this, the therapist presented both edible items and instructed the participant to pick one. Pairing based on four to seven edible items were assessed in this way in a random order without replacement. All edible items were compared with one another. Data were collected on the selections and summarized as the percent of trials that each item was selected. The edible item associated with the highest percent of selection for each participant was included in their treatment analysis.

Data from the PSPA is shown in Figure 9. John selected a ring pop, dum-dum, and nerds most often, while Ethan preferred the push pop, dum-dum, and juicy drops, and Nathan selected dum-dum and ring pop most often. Before each session scheduled for the delivery of edible items, the participants were presented with a paired choice of the two most preferred to determine which edible would be used for that session.

Figure 9

Results from All Participants' Paired-Stimulus Assessment



Arrows denote the edibles included in treatment.

Appendix E

Punishment Procedure Selection Interview

An interview was conducted with each participant's lead clinician to identify socially acceptable procedures for inclusion in the punisher assessment. Each lead clinician was a Board Certified Behavior Analyst (BCBA) charged with managing the clinical treatment for one of the participants. During the interview, the experimenter described eight procedures with corresponding definitions (see Table 6) that have been frequently reported in the literature to decrease automatically maintained problem behavior (e.g., Ahrens et al., 2011; Carr et al., 2002; Cook et al., 2014; Doughty et al., 2007; Falcomata et al., 2004; Peters & Thompson, 2013). For each of the listed procedures, the experimenter asked if the procedure had been used with the participant, whether the clinician thought that the procedure would be effective in decreasing the VS, whether he or she thought the participant would dislike the procedure, whether he or she would be willing to include the procedure in the participant's behavioral program if it was found to be effective, and whether he or she thought the procedure could be implemented by the participant's direct care staff with integrity. The interview also included an open-ended portion in which the clinicians could list idiosyncratic procedures that they thought should be evaluated.

The interviews with the clinicians responsible for John and Nathan identified eight potential procedures; time-out, hands down, verbal reprimands, response cost, facial screen, tidiness training, contingent motor demands, and overcorrection (see Table 6 for definitions). The interview with the clinician responsible for Ethan identified six potential procedures; time-out, response cost, tidiness training, overcorrection, contingent motor demands, and verbal reprimands. These procedures were then included in the punishment procedure assessment.

Interview Questions (Verriden & Roscoe, 2018)

Date: Participant: Respondent:

1. Some children enjoy small toys (e.g., action figures, toy cars), sensory objects (e.g., lights, mirrors), and/or independent activities (e.g., iPad games, puzzles, coloring books). What are some items/activities that you think *participant* likes and can engage with appropriately?
2.

3. Describe *participant*'s play skills:

4. What are some other things that *participant* prefers?

5. What are some things that *participant* dislikes?

Interventions	**Has *participant* come into contact with the intervention in the natural environment? 1=Never 2=Sometimes 3=Always**	**Do you believe this procedure would be effective or has it been effective? 1=Yes, effective 2= No, ineffective 3=Not applicable**	**Is this procedure something that *participant* dislikes?**	**Would you be willing include this procedure in *participant*'s behavior program?**	**Is this procedure something that teachers could implement with integrity?**
Hands down	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
Contingent demands: vocal response	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
Contingent demands: motor response	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
Response cost	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
In seat timeout	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
Reprimands	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
Overcorrection-positive practice (practicing appropriate things to say during play)	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
Facial screen	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
Tidiness training	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
Are there any additional interventions that have been used and are not listed above? Are there any additional things that the participant dislikes or avoids that could be used contingently?					

	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No