2017

An evaluation of a punisher assessment for decreasing automatically reinforced problem behavior

Amanda L. Verriden
Western New England University

An Evaluation of a Punisher Assessment for Decreasing Automatically Reinforced Problem Behavior

Amanda L. Verriden

Western New England University

The New England Center for Children

A Dissertation Submitted to the Department of Psychology

and the College of Arts and Sciences at

Western New England University in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

Supervised by:

Eileen M. Roscoe, Ph.D., BCBA-D, LABA

Abstract

We extended research on the identification and evaluation of potential punishers for decreasing automatically reinforced problem behavior in 4 individuals with autism spectrum disorder. A punisher selection interview was conducted with lead clinicians to identify socially acceptable punishers. During the treatment evaluation, treatment phases were introduced sequentially and included: noncontingent reinforcement (NCR), NCR and differential reinforcement of alternative behavior (DRA), and NCR and DRA with punishment. During the NCR and DRA with punishment phase, 4 to 5 potential punishers were evaluated using a multielement design. Dependent measures included the target problem behavior, appropriate item engagement, and emotional responding. For all participants, NCR and DRA was not effective and punishment was necessary. However, the most effective punisher identified in the context of NCR and DRA differed across participants.

An Evaluation of a Punisher Assessment for Decreasing Automatically Reinforced Problem Behavior

Punishment is defined as an environmental change contingent on behavior that results in a decrease in responding over time (Michael, 2004). When a stimulus is presented contingent on a response and the future likelihood of that response decreases, this is defined as positive punishment. When a stimulus is removed contingent on a response and the future likelihood of that responses decreases, this is defined as negative punishment. Various punishment procedures (e.g., hands down, overcorrection, response cost) have been evaluated in clinical settings for decades to decrease challenging behavior with great success (Baker, Rapp, & Caroll, 2010; Fisher et al., 1994; Foxx & Azrin, 1973; Singh, Watson, & Winton, 1986; Thompson, Iwata, Conners, & Roscoe, 1999); however, there is controversy associated with the use of punishment because of a movement towards function-based and least intrusive interventions (Lerman & Toole, 2011; Thompson et al., 1999).

Punishment may be necessary when problem behavior is severe, extinction cannot be implemented with integrity, and other reinforcement-based interventions have been attempted and found to be unsuccessful. Another situation in which punishment may be necessary is when the problem behavior persists independent of social contingencies and is likely maintained by automatic reinforcement or the response product (e.g., tactile, auditory, visual stimulation) it produces (Vaughan & Michael, 1982; Vollmer, 1994). Automatically reinforced problem behavior can include a wide range of topographies, including but not limited to self-injury, motor stereotypy, vocal stereotypy, rumination, and ritualistic or repetitive behavior. It is important to identify effective treatments because these responses may pose medical risks (Piazza et al.,

1998), can interfere with skill acquisition (Matson & Nebel-Schwalm, 2007), inhibit social interactions, and be socially stigmatizing (Cunningham & Schreibman, 2008; DiGennaro Reed, Hirst, & Hyman, 2012). However, treatment for automatically reinforced problem behavior can be challenging because the reinforcer cannot be readily presented and withheld.

Reinforcement-based interventions that are commonly used to address automatically reinforced problem behavior include manipulations of motivating operations (e.g., noncontingent reinforcement; NCR) and differential reinforcement (e.g., delivery of reinforcement contingent on an appropriate behavior or contingent on the absence of the automatically maintained target behavior). Noncontingent reinforcement involves time-based or response-independent delivery of reinforcers (Vollmer, 1994) and has frequently been used to treat automatically reinforced problem behavior by allowing continuous access to stimuli that produce tactile, auditory, or visual stimulation (Rapp & Vollmer, 2004). Differential reinforcement involves the delivery of reinforcement contingent on an alternative response (DRA) or contingent on the absence of the problem behavior for a specified interval of time (DRO; Cooper, Heron, & Heward, 2007).

Although NCR alone has been found effective for treating automatically reinforced problem behavior (Buchanan & Fisher, 2002; Piazza, Adelinis, Hanley, Goh, & Delia, 2000; Roscoe, Iwata, & Zhou, 2013), a number of studies have shown that the addition of a punishment procedure (e.g., response cost) was necessary to achieve clinically significant reductions (Barlett, Rapp, Krueger, & Henrickson, 2011; DeRosa, Roane, Bishop, & Silkowski, 2016; Falcomata, Roane, Hovanetz, Kettering, & Keeney, 2004; Peters & Thompson, 2013; Watkins & Rapp, 2014). Falcomata et al. (2004) and Bartlett et al. (2011) found that NCR alone did not reduce inappropriate vocalizations and spitting, respectively, to clinically significant levels and that the addition of response cost resulted in immediate decreases in problem behavior to near-zero

levels. Peters and Thompson (2013) and Watkins and Rapp (2014) found that NCR alone did not produce reductions in automatically reinforced stereotypy whereas NCR combined with overcorrection (Peters & Thompson, 2013) or response cost (Watkins & Rapp, 2014) was effective in reducing stereotypy for all participants.

Differential reinforcement using an arbitrary reinforcer (e.g., an edible or leisure item) has also been found effective for treating automatically reinforced problem behavior (Ringdahl et al., 2002; Taylor, Hoch, & Weissman, 2005; Toussaint & Tiger, 2012) however, researchers have demonstrated the need for additional punishment components before clinically significant outcomes are observed (Anderson & Le, 2011; Baker et al., 2010; Mitteer, Romani, Greer, & Fisher, 2015). Anderson and Le (2011) found that DRO alone (i.e., delivery of preferred edible and video contingent on the absence of stereotypy for a specified duration) was ineffective in decreasing stereotypy. When implementing overcorrection, stereotypy decreased to clinically significant levels. After DRO was ineffective in reducing a child's rumination, Baker et al. (2010) implemented visual screen, and rumination reduced to low levels.

Although differential reinforcement for automatically reinforced problem behavior typically includes an arbitrary reinforcer, DRA using the maintaining reinforcer (e.g., by withholding and delivering access to the target problem behavior has motor stereotypy; Charlop, Kurtz, & Greenberg Casey, 1990; Hanley, Iwata, Thompson, & Lindberg, 2000; Potter, Hanley, Augustine, Clay, & Phelps, 2013) has also been evaluated and found to be effective. A consideration when using DRA with the functional reinforcer for automatically reinforced problem behavior is that it requires restriction of the target response, usually in the form of blocking. As such, this procedure requires the inclusion of a potential punisher to prevent the occurrence of problem behavior during non-reinforcement periods. Although DRA using the

maintaining reinforcer has been found effective for decreasing stereotypy and increasing appropriate behavior, it may not be appropriate when response blocking in ineffective or cannot be implemented with integrity or when the target behavior is too dangerous (e.g., severe self-injury) to be permitted as a reinforcer, even for short durations.

Given that a punishment component is often necessary for treating automatically reinforced problem behavior (Anderson & Le, 2011; Bartlett et al., 2011; Falcomata et al., 2004; Peters & Thompson, 2013; Watkins & Rapp, 2014), it is important to further evaluate a systematic method for identifying potential punishers. When punishment is used for treating automatically reinforced problem behavior, punishers are often selected arbitrarily (i.e., no systematic assessment is used to identify the procedure used). In some cases, the topography of the target behavior has informed the punishing stimulus selected (e.g., mouthwash for rumination; Baker et al., 2010) whereas in other cases the research findings demonstrating the efficacy of the punishing stimulus may have resulted in selection of the procedure (Ahrens, Lerman, Kodak, Worsdell, & Keegan, 2011; Bartlett et al., 2011; Peters & Thompson, 2013; Watkins & Rapp, 2014). One notable exception is a systematic assessment for identifying punishers reported by Fisher et al. (1994).

Fisher et al. (1994) developed an empirical method for identifying potential punishers. The authors evaluated a stimulus avoidance assessment for identifying potential punishers for inclusion in a treatment for pica. Nine potential punishers, including: baskethold time-out, tidiness training, chair time-out, water mist, facial screen, contingent demands, contingent exercise, hands down, and quiet hands, that had been shown to be effective in previous literature, were included. During this assessment, each punisher was randomly selected and singly presented on a time-based schedule for 30 s, 10 times each, and avoidance responses (e.g.,

negative vocalizations, avoidance movements, and attempts to escape from the procedure) were measured. From this assessment, a low, medium or high avoidance stimulus was identified and included in a punisher assessment. During this assessment, each stimulus was delivered contingent on problem behavior (e.g., disruption, inappropriate touching, or pica). The greatest reductions in problem behavior were observed with the high-avoidance stimulus for all participants, suggesting that the stimulus avoidance assessment was a valid predictor of punisher efficacy. After the stimulus avoidance and punisher assessments, the most effective punisher was included in a treatment package that included punishment of pica combined with differential reinforcement of appropriate eating. Treatment was found effective in decreasing pica and increasing appropriate food consumption for all three participants.

Fisher et al. (1994) was a substantial contribution to the literature because it outlined a preliminary technology for identifying punishing stimuli for treating severe problem behavior. It has guided much research on the use of punishment in subsequent publications (Lerman & Vorndran, 2002) and has been cited by other researchers who have reported using it for identifying punishers (DeRosa et al., 2016; Mitteer et al., 2015; Toole et al., 2004). Although there may be utility in the Fisher et al. method, there are some concerns that limit its utility in clinical practice. Most notably, some of the punishers may not be acceptable for use given current ethical guidelines, which mandate the most effective and least intrusive procedures (Behavior Analysis Certification Board, 2014). In addition, many of the included punishers {e.g., water mist (spraying water on an individual's face) and facial screen (placing one's hand over an individual's eyes)} are not accepted practice for use in most educational settings.

A second concern with the Fisher et al. (1994) method is the use of a stimulus avoidance assessment. This assessment was developed to parallel research on reinforcer identification and

involved singly presenting stimuli like the single-stimulus preference assessment by Pace, Ivancic, Edwards, Iwata, and Page (1985). Although the stimulus avoidance assessment successfully identified punishers, it is unclear whether presenting potential punishers on a fixed-time schedule while measuring indirect effects (i.e., emotional responding) was necessary. By definition, a punisher involves contingent delivery of the stimulus on a target behavior and subsequent decreases in that behavior. Therefore, presenting potentially punishing stimuli contingent on the target problem behavior rather than response independently may be more effective and more socially acceptable. In addition, it may be beneficial to identify a punisher that is less likely to elicit emotional responding that may interfere with educational programming. That is, if several stimuli are identified as punishers, then the stimulus that elicits the least emotional responding would be the optimal choice.

There has not been a systematic evaluation of a punisher assessment since the Fisher et al. (1994) publication and researchers continue to report using the stimulus avoidance assessment to inform punisher selection for treatments (DeRosa et al., 2016; Mitteer et al., 2015). An exception was a study by Thompson et al. (1999) who conducted a brief punisher assessment with multiple stimuli using an AB design. The A phase involved no consequence delivery, and the B phase involved delivery of the punisher contingent on the target problem behavior. Because this assessment was not the focus of their study, details concerning the included punishers and results were not provided. Given best practice considerations that have emerged in the last 20 years regarding the type of punishers that are deemed socially acceptable and some of the limitations of the stimulus avoidance assessment, an updated technology for systematically identifying potential punishers seems warranted. Similar to idiosyncratic differences in the reinforcing efficacy of stimuli, there may be individual differences in the punishing efficacy of

stimuli (e.g., hands down may function as a punisher for one individual and as a reinforcer for another).

Therefore, the purpose of the present study was to develop and evaluate an updated technology for identifying potential punishers. First, potential punishers were identified through a systematic interview with lead clinicians to determine if they thought the procedure would be effective and to ensure that they found them to be socially acceptable. Additionally, stimuli included were deemed appropriate for use in educational settings. Next, reinforcement-based interventions (NCR and DRA) were evaluated to ensure that reinforcement alone was not effective in reducing the target behavior. If NCR and DRA alone were unsuccessful, then a punisher assessment was added. We extended previous research by conducting an interview to identify potential punishers to include in the punisher assessment, conducting a punisher assessment in the context of a reinforcement-based intervention, and using multiple dependent variables, such as the target problem behavior and appropriate item engagement, to determine the most effective punisher. Emotional responding was also measured to suggest a potential negative side effect rather than an indicator of punishing efficacy.

Method

Participants and Setting

Participants were of four individuals who attended a residential school for children diagnosed with autism spectrum disorder (ASD). Rory was a 6-year-old boy diagnosed with ASD who communicated using some sign language and gestures. He engaged in problem behavior in the form of object and hand mouthing that interfered with the presentation of academic stimuli and was a health and safety concern. Eric was an 11-year-old boy diagnosed with ASD and mental retardation, who communicated with vocal approximations and a communication book. Jacob was an 8-year-old boy diagnosed with ASD, who communicated

using vocal approximations and phrases. Both Eric and Jacob engaged in problem behavior in the form of motor stereotypy that interfered with skill acquisition and was socially stigmatizing. Sally was a 14-year-old girl diagnosed with ASD who communicated using some sign language and gestures. She engaged in problem behavior in the form of hair manipulation that was social stigmatizing and often resulted in significant knots in her hair.

All participants' functional analysis sessions were conducted in a small experimental room. Eric and Sally's treatment analysis sessions were conducted in the same research room, while Rory and Jacob's treatment analysis sessions were conducted in their private cubby areas located within their respective classrooms. Materials included stimuli relevant to the condition, a table, chairs, and a video camera for recording sessions. During all sessions conducted with Rory, there was a set of non-leisure items baited in the room so that an equal opportunity for mouthing could occur. Because Jacob most frequently engaged in motor stereotypy that involved twirling, tapping, and flapping teaching materials (e.g., pencils, tokens), these items were present in all of his sessions.

Response Measurement

Target problem behavior. Rory's mouthing was defined as any instance in which he placed his hand, an object, or any non-food-related item into his mouth, or put his mouth on a larger item (wall, table, etc.). Eric's motor stereotypy was defined as non-contextual, non-functional motor movements including swift motions of head, arms, and/or legs, head rocking, body rocking/swaying, hand clapping, slapping objects, and limb swinging. Jacob's motor stereotypy was defined as nonfunctional repetitive motor movements including arm flapping, hand flapping, clapping, finger waving/manipulations, twirling objects, moving legs in a kicking motion, and repeated jumping up and down. Sally's hair manipulation was defined as twirling

hair around one or multiple fingers or placing her hair past the plane of her lips. Observers recorded mouthing, motor stereotypy, and hair manipulation using duration in seconds. Data were subsequently summarized as percentage of session by dividing the total number of seconds of problem behavior by the total session duration and multiplying by 100%.

Appropriate engagement. For all participants, appropriate engagement was defined as manipulation of an item or activity as designed (with one or both hands) for at least 2 s. For example, engagement with the See ‘n Say® toy included pulling the lever, turning the pages, turning the arrow, etc. Duration recording was used and data were summarized as percentage of session by dividing the total number of seconds of appropriate engagement by the total session duration and multiplying by 100%. If the participant simultaneously engaged in the target problem behavior and appropriate engagement, only the target problem behavior was scored.

Emotional responses. For all participants, emotional responding was defined as engaging in whining, crying, screaming, aggression, self-injury, attempts to escape from the procedure, or physical resistance. Emotional responding was recorded using frequency, and summarized as responses per min.

Interobserver Agreement

Mean duration per interval interobserver agreement (IOA) was calculated for the target problem behavior and appropriate engagement, and mean count per interval IOA was calculated for emotional responding. A second observer independently collected data during at least 30% of sessions in each condition, across all participants. Interobserver agreement was calculated by dividing each session into 10-s intervals. For problem behavior and appropriate engagement, within each interval the smaller number of seconds of problem behavior or appropriate engagement was divided by the larger number of seconds of problem behavior or appropriate

engagement and multiplied by 100%. For emotional responses, the smaller count of emotional responses was divided by the larger count of emotional responses and multiplied by 100%. The percent agreement for all intervals was averaged to determine the percent agreement for the entire session. Mean IOA for Rory was 87.7% (range, 81.3-93.5%) for mouthing, 89.5% (range, 83.5-92.5%) for appropriate engagement, and 100% for emotional responding. Mean IOA for Eric was 91.3% (range, 85.5-96.3%) for motor stereotypy, 87.5% (range, 84.6-92.5%) for appropriate engagement, and 100% for emotional responding. Mean IOA for Jacob was 85.3% (range, 82.5-89.3%) for motor stereotypy, 89.8% (range, 85.5-95.2%) for appropriate engagement, and 98.5% (range, 95.6-100%) for emotional responding. Mean IOA for Sally was 85.5% (range, 83.5-87.5%) for hair manipulation, 87.8% (range, 85.5-90.3%) for appropriate engagement, and 97.5% (range, 95-100%) for emotional responding.

Pre-Assessments

Functional analysis. A functional analysis was conducted to confirm that the target problem behavior was maintained by automatic reinforcement. A functional analysis based on the procedures by Roscoe et al. (2013) was conducted, which included three conditions (i.e., attention, demand, and alone) that were alternated in a 2:1 ratio of alone to attention or demand. Sessions were 10 min in duration, and a multielement design was used. In the alone condition, the participant was alone in the room with no materials (with the exception of Rory and Jacob, whose rooms were baited with non-leisure items), and there were no programmed consequences. During the attention condition, the therapist sat at a table and acted busy. Contingent on the occurrence of the target problem behavior, the therapist immediately delivered brief vocal and physical attention. During the demand condition, five tasks were continuously presented in a

random order using three-step prompting (vocal, then model, then physical guidance). The target problem behavior resulted in a 30-s break (the demand materials were removed).

Functional analysis results are depicted in Figure 1. Results from Rory's functional analysis showed elevated levels of mouthing across all conditions, with higher levels of mouthing during the alone condition relative to the demand and attention conditions. Eric's levels of motor stereotypy were consistently high during the alone condition, with low levels during the attention and demand conditions. Jacob's levels of motor stereotypy were consistently high across all conditions, with higher levels during the alone condition. Sally's levels of hair manipulation were elevated across all conditions, with higher levels during the alone and demand conditions relative to the attention condition. In summary, functional analysis patterns for all participants suggested maintenance by automatic reinforcement.

Preference assessments. A competing-items preference assessment (Piazza, Fisher, Hanley, Hilker, & Derby, 1996) was conducted to identify leisure items for use during the treatment evaluation. Six items were singly presented for 3 min, two times each. Observers recorded the duration of target problem behavior and item engagement, and data were summarized as the average duration of problem behavior and item engagement for each item. The three leisure items that were associated with the highest percentage of engagement and the lowest percentage of problem behavior were included in the treatment analysis (see Table 2).

A paired-stimulus preference assessment (Fisher et al., 1992) was conducted to identify edible items for use during the treatment evaluation. Seven edible items were presented in pairs for a total of 42 stimulus-pair presentations. Data were collected on selection, and summarized as percentage of trials. The two edibles associated with the highest percentage of selection were included in the treatment analysis (see Table 2).

Punisher selection interview. An interview was conducted with each participant's lead clinician to identify socially acceptable procedures for inclusion in the punisher assessment (see Appendix for interview form). Each participant's lead clinician was a master's level BCBA who oversaw his or her educational and clinical programming. During this interview, the experimenter listed seven procedures and their corresponding definitions (see Table 1) that have been frequently reported in the literature for decreasing automatically reinforced problem behavior (Ahrens et al., 2011; Carr, Dozier, Patel, Adams, & Martin, 2002; Cook, Rapp, Gomes, Frazer, & Lindblad, 2014; Doughty, Anderson, Doughty, Williams, & Saunders, 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 target problem behavior, 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 behavior program, and whether he or she thought the procedure could be implemented by direct care staff with integrity. The interview also included an open-ended portion in which clinicians could list idiosyncratic procedures that they thought should be evaluated. Based on what was predicted to be effective and what each clinician was willing to include in the participant's behavior program, we consulted with the lead clinician in selecting 4-5 potential punishers for inclusion in the punisher assessment. Table 3 shows the procedures selected for inclusion in the punisher assessment with each participant.

Treatment Evaluation

A progressive sequence of phases was introduced including no-interaction baseline, NCR, NCR and DRA, and NCR and DRA with punishment. During the NCR and DRA with punishment phase, a multielement design was used to evaluate the effects of the various

punishers. Following this phase, the effects of NCR and DRA with punishment was evaluated using a reversal design. That is, NCR and DRA was conducted a second time followed by NCR and DRA with punishment using the most effective punisher identified in the multielement design phase. Sessions were 5 min in duration and conducted 3 to 5 days per week. The session clock was paused when punishment was implemented and then restarted when punishment ended. Therefore, session time included 5 min of non-intervention time, but total running time often exceeded 5 min when punishment was in effect. The target problem behavior and appropriate engagement were scored using session time as the denominator, whereas emotional responding was recorded using running time (session time plus punishment time) as the denominator.

No interaction baseline. During baseline, the participant and therapist were in the room, and all target responses resulted in no programmed consequences.

NCR. During this condition, the participants were given continuous access to the three leisure items identified from the competing items preference assessment, and all target responses resulted in no programmed consequences.

NCR and DRA. During this condition, leisure items were continuously available. The therapist presented vocal and gestural prompts for item engagement every 15 s if the participant was not already engaging with an item. Prior to each session, the participant was instructed to choose between two highly preferred edible items associated with the greatest percentage of selection from the paired-stimulus preference assessment. The selected edible was then delivered contingent on appropriate item engagement. If the participant emitted item engagement continuously, then the therapist delivered an edible immediately after consumption of the previous edible.

NCR and DRA with punisher assessment. This phase included a punisher assessment in addition to NCR and DRA components. The punisher assessment involved alternating 4-5 punishers in a random order. NCR and DRA was conducted as described above, and contingent on the occurrence of the target problem behavior, the punisher associated with the respective condition, was delivered. See Table 1 for definitions of each punishment procedure. For example, during the hands down condition, in addition to delivery of a preferred edible contingent on appropriate item engagement, a 30-s hands down procedure was implemented contingent on occurrences of the target problem behavior.

NCR and DRA with single punisher. Following a reversal to NCR and DRA alone, we met with the lead clinician and reviewed the results of the punisher assessment phase with them. Based on the results for problem behavior, appropriate engagement, emotional responding, and the lead clinician's treatment selection, contingent demands, response blocking, hands down, and response blocking for used for Rory, Eric, Jacob, and Sally, respectively. We then re-implemented NCR and DRA with punishment, using the selected punisher (see Table 3) to replicate the effects from the previous punishment phases.

Social Validity

Following completion of the treatment analysis, we reviewed outcomes and session videos with the lead clinician and subsequently administered a closed-ended questionnaire (adapted from that used by Potter, Hanley, Augustine, Clay, & Phelps, 2013). The questionnaire was comprised of four questions (see Table 4) that asked about acceptability of the procedures, the outcomes, and the goals, as well as the feasibility of treatment implementation.

Results

Rory. Results from Rory's treatment evaluation are depicted in Figure 2. During baseline, mouthing occurred at moderate levels. During NCR, mouthing continued to occur at moderate levels with a slight increasing trend, appropriate engagement occurred at moderate levels with a slight decreasing trend, and no emotional responding occurred. When NCR and DRA was initiated, mouthing decreased somewhat but not to clinically significant levels and was on an increasing trend towards the end of the phase, appropriate engagement increased to moderate levels, and no emotional responding occurred. During the punisher assessment phase, mouthing decreased to near zero levels across all conditions, suggesting that all procedures functioned as punishers. Appropriate item engagement increased to high levels across conditions. Emotional responding occurred during only the hands down and overcorrection conditions. After reviewing outcomes from the punisher assessment with Rory's lead clinician, contingent demands was selected for use in the final treatment phase because it was associated with no emotional responding. Although response cost was also associated with no emotional responding, Rory's lead clinician selected contingent demands because it was associated with lower levels of mouthing and higher levels of appropriate item engagement. During the return to NCR and DRA phase, mouthing returned to moderate levels, appropriate item engagement increased to moderate levels, and emotional responding did not occur. During the NCR and DRA with contingent demands phase, mouthing decreased to near zero levels, appropriate item engagement increased to high levels, and emotional responding did not occur. During a 2-month follow-up probe, levels of mouthing remained low, appropriate item engagement remained high, and emotional responding did not occur. After training a teacher in Rory's classroom to implement the intervention, she subsequently conducted sessions. Mouthing remained at near

zero levels, appropriate item engagement remained high, and emotional responding did not occur during these sessions.

Eric. Results from Eric's treatment evaluation are depicted in Figure 3. During baseline motor stereotypy occurred at low-to-moderate levels. During NCR, motor stereotypy increased to moderate-to-high levels, appropriate engagement occurred at low-to-moderate levels, and no emotional responding occurred. When NCR and DRA was initiated, motor stereotypy decreased to baseline levels, appropriate item engagement decreased to low levels, and no emotional responding occurred. During the punisher assessment phase, motor stereotypy decreased to low levels in all conditions, except for verbal reprimands. Appropriate item engagement increased to high levels in all conditions except for verbal reprimands and contingent demands, and emotional responding occurred during only one hands down session. After reviewing outcomes from the punisher assessment with Eric's lead clinician, response blocking was selected because it was associated with low levels of motor stereotypy, high levels of appropriate item engagement, and no emotional responding. Although hands down and overcorrection were also associated with low levels of motor stereotypy and high levels of appropriate item engagement, response blocking was selected because it is less intrusive and because the lead clinician thought it would be less resource intensive. During the return to NCR and DRA, motor stereotypy increased to low-to-moderate and variable levels, appropriate engagement occurred at moderate-to-high levels, and no emotional responding occurred. During the NCR and DRA with response blocking phase, motor stereotypy decreased to low levels, appropriate item engagement increased to high levels, and emotional responding did not occur. During the return to the NCR and DRA phase with one of Eric's teachers, motor stereotypy increased to moderate levels and appropriate item engagement decreased to moderate levels. Following training, when the teacher implemented

NCR and DRA with response blocking, motor stereotypy decreased to low levels, appropriate item engagement increased to high levels, and emotional responding did not occur.

Jacob. Results from Jacob's treatment evaluation are depicted in Figure 4. Motor stereotypy occurred at high levels during baseline. During NCR, motor stereotypy increased slightly, and no appropriate engagement or emotional responding occurred. During NCR and DRA, there was a modest decrease in motor stereotypy, little appropriate item engagement and no emotional responding. During the punisher assessment phase, motor stereotypy decreased to low levels and appropriate item engagement increased to moderate levels in all conditions. Emotional responding occurred during all conditions with the exception of response blocking, but overall levels were low. After reviewing outcomes from the punisher assessment with Jacob's lead clinician, hands down was selected because it was associated with the lowest levels of motor stereotypy, moderate levels of appropriate item engagement, and low levels of emotional responding. Although contingent demands was associated with comparably low levels of motor stereotypy and high levels of appropriate item engagement, hands down was selected because the lead clinician thought it would be the most practical in Jacob's classroom setting. During the second NCR and DRA phase, motor stereotypy increased to moderate levels, appropriate item engagement decreased to low levels, and no emotional responding occurred. During the subsequent NCR and DRA with hands down phase, motor stereotypy decreased to near zero levels, appropriate engagement increased to high levels, and some emotional responding occurred. During the return to NCR and DRA baseline with one of Jacob's teachers, motor stereotypy increased to moderate levels, appropriate item engagement decreased to low-to-moderate levels, and no emotional responding occurred. Following training, when the teacher implemented NCR and DRA with hands down, motor stereotypy decreased to near zero levels

and appropriate engagement returned to high levels. Although emotional responding occurred, it was low and sporadic. When it occurred, it was often in the form of Jacob attempting to pull his arms out of the hands down procedure.

Sally. Results from Sally's treatment analysis are depicted in Figure 5. Hair manipulation occurred at high levels during baseline. During NCR and NCR and DRA phases, hair manipulation remained high, appropriate item engagement occurred at low levels, and no emotional responding occurred. During the punisher assessment phase, hair manipulation decreased to low levels and appropriate engagement increased to high levels in all conditions. Emotional responding occurred during the response blocking, reprimands, and overcorrection conditions. After reviewing outcomes from the punisher assessment with Sally's lead clinician, response blocking was selected because it resulted in the lowest levels of hair manipulation and highest levels of appropriate engagement compared to the other procedures. During the reversal to NCR and DRA, hair manipulation increased to moderate levels, appropriate item engagement decreased to moderate levels, and emotional responding did not occur. During the NCR and DRA with response blocking phase, hair manipulation decreased to near-zero levels, appropriate item engagement increased to levels near 100% of the session, and emotional responding remained at zero. During a NCR and DRA baseline with one of Sally's teachers, hair manipulation increased to moderate levels, appropriate item engagement decreased to moderate levels, and emotional responding did not occur. Following training, when the teacher implemented NCR and DRA with response blocking, hair manipulation returned to near-zero levels, appropriate engagement increased to high levels, and emotional responding did not occur.

Figure 6 shows the rate of punisher delivery for the most effective punisher during the initial punishment phases and the final punishment phase (with teacher) for all participants. The

rate of punisher delivery decreased as treatment progressed for all participants. For Rory and Eric, the decrease was to near zero rates during the final punishment phase, whereas for Jacob and Sally, the rate of punisher delivery decreased and then remained stable at about 1-2 implementations per minute.

Results from the social validity questionnaire are depicted in Table 4. In general, lead clinicians reported that the goals of the intervention were appropriate, the methodology of the intervention was acceptable, the change in problem behavior was clinically sufficient, and that teachers could implement the final intervention with integrity. One exception was Eric's lead clinician who reported *somewhat unacceptable* when asked if the participant's teachers could effectively implement the treatment with integrity.

Discussion

Reinforcement-based interventions alone were ineffective in reducing automatically reinforced problem behavior for Eric, Jacob and Sally, or reducing it to clinically significant levels for Rory. During the punisher assessment phase, when punishers were delivered contingent on the target problem behavior, clinically significant decreases in problem behavior occurred for almost all punishers across participants. The specific punishers that were most effective in reducing problem behavior to near zero levels varied across participants. Results from Eric's functional analysis showed near zero rates of problem behavior during the attention condition, suggesting that verbal reprimands would sufficiently reduce responding. Despite this pattern of responding during the functional analysis, verbal reprimands was the least effective procedure for reducing Eric's motor stereotypy. It was also found that appropriate item engagement was higher during the phases that included punishment compared to the phases that were reinforcement-only for all participants, suggesting that the use of punishment may produce

increases in other forms of appropriate behavior. In the present study, we found low rates of emotional responding across all participants and all procedures. The emotional responses that were typically observed included resisting or attempting to escape from the procedures.

Although NCR and differential reinforcement alone have been found effective for reducing automatically reinforced problem behavior (Buchanan & Fisher, 2002; Ringdahl et al., 2002, Taylor, Hoch, & Weissman, 2005), several studies have demonstrated the need for punishment to reduce problem behavior to acceptable levels. Our results are similar to those found by previous researchers, who first implemented NCR (Bartlett et al., 2011; Falcomata et al., 2004; Peters & Thompson, 2013; Toole et al., 2004; Watkins & Rapp, 2014), DRO (Anderson & Le, 2011), or DRA (Mitteer et al., 2014) without success, and required the addition of punishment to achieve clinically significant reductions in problem behavior.

We added to the existing literature on assessing the effects of potential punishers by including multiple dependent measures. In addition to measuring automatically reinforced problem behavior, we also measured appropriate item engagement and found higher levels of appropriate item engagement when punishment was included in treatment, compared to reinforcement-only. These results are similar to those found by other researchers (Anderson & Le, 2011; Mitteer et al., 2015). For example, Anderson and Le (2011) used a positive practice overcorrection procedure to decrease vocal stereotypy and found high levels of appropriate play when overcorrection was combined with DRA. Overcorrection was then removed from the intervention and appropriate play decreased to near zero levels, suggesting that the punishment component was necessary to sustain high levels of appropriate play. Our results differ from those found by Thompson et al. (1999), who found that for three of four participants, item engagement was lower during reinforcement phases that included punishment compared to reinforcement-

only phases. One possible explanation for this outcome is that for these three participants, reinforcement included NCR (continuous access to leisure items) without a DRA component.

We also measured emotional responding, as authors have asserted that punishment can be associated with negative side effects in the form of increased rates of other problem behavior, for example aggression (Hagopian & Toole, 2009; Kazdin, 2001). We found low rates of emotional responding across participants and procedures. Interestingly, the punishment procedures that were most effective in reducing problem behavior were not always associated with the highest levels of emotional responding. Because authors have cautioned against the use of punishment procedures because of emotional responding as a side effect (LaVigna & Donnellan, 1986), future research seems warranted to determine if this is a reliable side effect. It is possible that we did not observe high rates of emotional responding because reinforcement was simultaneously available for an appropriate alternative response (i.e., item engagement). Although it may be a limitation that we did not evaluate the isolated effects of punishment, we chose to have overlaying reinforcement contingencies as this is a best practice consideration outlined by the professional and ethical compliance code for Behavior Analysts (Behavior Analysis Certification Board, 2014).

We extended previous research on punisher identification for automatically reinforced problem behavior by first conducing a punisher selection interview with lead clinicians to identify socially acceptable procedures to include in the punisher assessment. During this interview, the experimenter listed procedures frequently reported in the literature for decreasing automatically reinforced problem behavior and asked a series of questions regarding each procedure. We asked if the procedure had been used with the participant, whether the clinician thought that the procedure would be effective in decreasing the target problem behavior, 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 behavior program, and whether he or she thought the procedure could be implemented by direct care staff with integrity. Clinicians reported that they were willing to evaluate most procedures, with the exception of response cost and in-seat timeout. When response cost was not selected, clinicians often reported that they did not think it would be effective due to the participants' limited leisure repertoire. That is, because participants rarely engaged with leisure items, clinicians did not think losing access to them would function as a punisher. Clinicians also seemed cautious of including an in-seat timeout procedure, as this procedure would seem more appropriate for problem behavior maintained by access to attention. Lastly, clinicians did not offer any idiosyncratic procedures for evaluation during the open-ended portion of the interview.

We also extended previous research on punisher identification by conducting a punisher assessment that is embedded within the treatment context. More specifically, we evaluated the effects of various procedures, when delivered contingent on problem behavior, using a multielement design. This contrasts with previous research in which punishers were selected for treatment arbitrarily. For example, some researchers have reported selecting punishers that have shown been shown to be effective in the literature (Watkins & Rapp, 2014), or selecting procedures that seem to match the topography of problem behavior (e.g., mouthwash for rumination; Baker et al. 2010). An empirically-based assessment, the stimulus avoidance assessment, has also been used to inform punishers for treatment of automatically-reinforced problem behavior (DeRosa et al., 2016; Mitteer et al., 2015). However, a concern regarding this assessment is that it involves multiple presentations of potential punishers using a fixed-time schedule and measures indirect effects (i.e., emotional responding). The punisher assessment

outlined in the present study is a more direct approach, and may be more clinically acceptable to use over the stimulus avoidance assessment because it presents punishers contingent on problem behavior, rather that response-independently.

Another concern with the stimulus avoidance assessment is that it requires selecting the punishment procedure that elicits the most emotional responding. The results from the present study show that the punishment procedures that were most effective in reducing problem behavior were not always associated with the highest levels of emotional responding. These findings contradict those of Fisher et al. (1994) who found that procedures associated with low levels of emotional responding were ineffective. Because our results suggest that procedures associated with little to no emotional responding were equally or more effective, we suggest conducting a punisher assessment that involves delivering punishers contingent on the target problem behavior, and selecting those associated with low levels of problem behavior and emotional responding. Emotional responding in the form of crying, whining, dropping to the ground, self-injury, aggression, and so forth can further interfere with skill acquisition and be socially stigmatizing; therefore it is optimal to use an intervention that does not evoke these responses.

Given our results, when considering the inclusion of punishment in treatment, we recommend evaluating the direct effects on multiple dependent variables. The punisher analysis did not consistently reveal a superior procedure for decreasing problem behavior for participants, therefore we considered the effects on appropriate item engagement and emotional responding when selecting a final punisher. In addition to measuring at the effects on the target problem behavior, appropriate engagement, and emotional responding, we also consulted with the participant's lead clinician when selecting the final punisher. Receiving stakeholder approval in

the selection of the final punisher is an important consideration that may impact implementation integrity, as it is often not enough for behavioral procedures to be effective; they must also be accepted by those implementing them (Wolf, 1978).

Given that the social acceptability of punishment in treatment may be low due to the movement towards function-based interventions and positive behavioral support, it is important to include an assessment of social validity when implementing punishment procedures. Following the completion of the treatment analysis, we reviewed the data and session videos with the participant's lead clinician and asked them to complete a social validity questionnaire regarding the goals, effects, and acceptability of the treatment package. The social validity questionnaire showed that lead clinicians generally believed the treatment goals were appropriate, the methodology was acceptable, the effects on the dependent measures were clinically sufficient, and that teachers could implement the final intervention with integrity. One exception to this occurred with the final question answered by Eric's lead clinician. The clinician commented that, given the high rates of Eric's motor stereotypy in the natural environment, it would be difficult for teachers to implement response blocking with integrity. Future research would benefit from evaluating how to best extend these procedures to various times throughout the participants' day, while maintaining effectiveness and integrity.

A common concern with the use of punishment is the intensiveness of the intervention and the resources necessary to ensure adequate integrity. We collected data on the rate of punisher implementations and observed a pronounced decrease in rate from the earlier punishment phases to the final punishment phase for all participants. For Rory and Eric, punisher implementations decreased to near zero. For Rory and Erin, however, punisher implementations decreased to one punisher per min.

Given the persistence of problem behavior with Jacob and Sally, an alternative interpretation is that decreases were due to interruption of bouts of stereotypy (i.e., the duration of time between the start and ending point of a stereotypy episode) rather than the process of punishment. To examine this possibility further, we reviewed within session data for the frequency of stereotypy bouts across reinforcement-only and reinforcement with punishment phases. If decreases were due to interruption, we should observe a similar frequency of bouts across these phases. Although bouts of stereotypy decreased during punishment compared to reinforcement-only phases for all participants, the greatest reductions occurred for Rory and Eric, who emitted only 1 or 2 bouts of stereotypy during the final punishment condition. This almost elimination of stereotypy for Rory and Eric suggests a punishment effect, whereas the persistent low level of stereotypy observed with Jacob and Sally may suggest that decreases were the result of redirecting bouts of stereotypy. Therefore, future research on the process associated with decreases during punishment is warranted and may shed further light on the mechanism responsible for behavior change.

Although we administered a social validity questionnaire to the participants' lead clinicians, a limitation of this study is that we did not assess social validity by asking the participants themselves. Previous researchers have assessed client preference for treatment procedures using a concurrent chains arrangement (Hanley, Piazza, Fisher, & Maglieri, 2005; Potter et al., 2013) and have shown that clients may prefer a treatment context that includes a punishment component compared with a reinforcement-only or no intervention context. Future research evaluating the use of multiple punishment procedures should assess client preference for each punishment context. Although negative emotional responding from the client may provide

some evidence regarding perceived aversiveness, a more formal assessment using procedures similar to Hanley et al. (2005) would provide more convincing social validity data.

The present study offers an updated technology for identifying and evaluating the effects of socially acceptable punishers. Each punisher was implemented in the context of reinforcement-based interventions and multiple dependent measures (problem behavior, appropriate item engagement, and emotional responding) were included. Future research would benefit from extending the procedures in the present study to other topographies of automatically reinforced problem behavior, such as self-injury. The procedures from the present study should also be evaluated with socially reinforced problem behavior, particularly when differential reinforcement-based interventions alone and in combination with extinction have been attempted and failed (Hagopian, Fisher, Thibault Sullivan, Acquisto, & LeBlanc, 1998). Furthermore, punishment may be necessary when extinction cannot be implemented with integrity or when the target behavior is too severe to permit its use (e.g., self-injury, aggression, pica, or elopement; Hagopian et al., 1998).

References

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

Anderson, J., & Le, D. D. (2011). Abatement of intractable vocal stereotypy using an overcorrection procedure. *Journal of Applied Behavior Analysis, 26,* 134-146. doi: 10.1002/bin.326.

Baker, L. M., Rapp, J. T., & Carroll, R. A. (2010). Treating operant vomiting with visual screening. *Clinical Case Studies, 9,* 218-224. doi: 10.1177/1534650110372253.

Bartlett, S. M., Rapp, J. T., Krueger, T. K., & Henrickson, M. L. (2011). The use of response cost to treat spitting by a child with autism. *Behavioral Interventions, 26,* 76-83. doi: 10.1002/bin.322.

Behavior Analyst Certification Board. (BACB). (2014). Professional and Ethical Compliance Code. Retrieved from https://bacb.com/wp-content/uploads/2016/03/160321-compliance-code-english.pdf.

Buchanan, J. A., & Fisher, J. E. (2002). Functional assessment and noncontingent reinforcement in the treatment of disruptive vocalization in elderly dementia patients. *Journal of Applied Behavior Analysis, 35,* 99-103. doi: 10.1901/jaba.2002.35-99.

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,* 37-44. Doi: 10.1016/s0891-4222(01)00090-7.

Charlop, M. H., Kurtz, P. F., & Greenberg Casey, F. (1990). Using aberrant behaviors as reinforcers for autistic children. *Journal of Applied Behavior Analysis, 23,* 161-181. doi: 10.1901/jaba.1990.23-163.

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,* 106-124. doi: 10.1002/bin.1378.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis,* 2nd ed. Upper Saddle River, N.J.: Pearson Prentice Hall.

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

DeRosa, N. M., Roane, H. S., Bishop, J. R., & Silkowski, E. L. (2016). The combined effects of noncontingent reinforcement punishment on the reduction of rumination. *The Journal of Applied Behavior Analysis, 49,* 1-6. doi: 10.1002/jaba.304.

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,* 422-430. doi: 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 Experimental Analysis of Behavior, 87,* 325-336. doi.10.1901/jeab.2007.39-05.

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,* 83-87. doi: 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,* 491-498. doi: 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,* 447-457. doi: 10.1901/jaba.1994.27-447.

Foxx, R. M., & Azrin, N. H. (1973). The elimination of autistic self-stimulatory behavior by overcorrection. *Journal of Applied Behavior Analysis, 6,* 1-14. doi: 10.1901/jaba.1973.6-1.

Hagopian, L. P., Fisher, W. W., Thibault Sullivan, M., Acquisto, J., & LeBlanc, L. A. (1998). Effectiveness of functional communication training with and without extinction and punishment: A summary of 21 inpatient cases. *Journal of Applied Behavior Analysis, 31,* 211-235. doi: 10.1901/jaba.1998.31-211.

Hagopian, L. P., & Toole, L. M. (2009). Effects of response blocking and competing stimuli on stereotypic behavior. *Behavioral Interventions, 24,* 117-125. doi: 10.1002/bin.278.

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,* 285-297. doi: 10.1901/jaba.2000.33-285.

Hanley, G. P., Piazza, C. C., Fisher, W. W., & Maglieri, K. A. (2005). On the effectiveness of and preference for punishment and extinction components of function-based interventions. *Journal of Applied Behavior Analysis, 38,* 51-65. doi: 10.1901/jaba.2005.6-04.

Kazdin, A. E. (2001). *Behavior modification in applied settings* (6th ed.). Belmont, CA: Wadsworth/Thompson Learning.

LaVigna, G. W., & Donnellan, A. M. (1986). *Alternatives to punishment: Solving behavior problems with non-aversive strategies.* New York: Irvington.

Lerman, D. C., & Toole, L. M. (2011). Developing function-based punishment procedures for problem behavior. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), Handbook of applied behavior analysis (pp. 348-369). New York, NY: Guilford Press.

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,* 431-464. doi: 10.1901/jaba.2002.35-431.

Matson, J. L., & Nebel-Schwalm, M. (2007). Assessing challenging behaviors in children with autism spectrum disorders: A review. *Research in Developmental Disabilities, 28,* 567-579. doi:10.1016/j.ridd.2006.08.001.

Michael, J. L. (2004). *Concepts and principles of behavior analysis* (rev. ed.). Kalamazoo, MI: Society for the Advancement of Behavior Analysis.

Mitteer, D. R., Romani, P. W., Greer, B. D., & Fisher, W. W. (2015). Assessment and treatment of pica and destruction of holiday decorations. *Journal of Applied Behavior Analysis, 48,* 912-917. doi: 10.1002/jaba.255.

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

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

Piazza, C. C., Adelinis, J. D., Hanley, G. P., Goh, H. & Delia, M. D. (2000). An evaluation of the effects matched stimuli on behaviors maintained by automatic reinforcement. *Journal of Applied Behavior Analysis, 33,* 13-27. doi: 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,* 137-152. doi: 10.1901/jaba.1996.28-137.

Piazza, C. C., Fisher, W. W., Hanley, G. P., LeBlanc, L. A., Worsdell, A. S., Lindauer, S E., & Keeney, K. M. (1998). Treatment of pica through multiple analysis of its reinforcing functions. *Journal of Applied Behavior Analysis, 31,* 165-189. doi: 10.1901/jaba.1998.31-165.

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 tactic of "using stereotypy as reinforcement." *Journal of Applied Behavior Analysis, 46,* 407-423. doi: 10.1002/jaba.52.

Rapp, J. T., & Vollmer, T. R. (2004). Stereotypy I: A review of behavioral assessment and treatment. *Research in Developmental Disabilities, 26,* 527-547. doi: 10.1016/j.ridd.2004.11.005.

Ringdahl, J. E., Andelman, M. S., Kitsukawa, K., Winborn, L. C., Barretto, A., & Wacker, D. P. (2002). Evaluation and treatment of covert stereotypy. *Behavioral Interventions, 17,* 43-49. doi: 10.1002/bin.105.

Roscoe, E. M., Iwata, B. A., & Zhou, L. (2013). Assessment and treatment of chronic hand mouthing. *Journal of Applied Behavior Analysis, 46,* 181-198. doi: 10.1002/jaba.14

Singh, N. N., Watson, J. E., & Winton, A. S. W. (1986). Treating self-injury: Water mist spray versus facial screening or forced arm exercise. *Journal of Applied Behavior Analysis, 19,* 403-410. doi: 10.1901/jaba.1986.19-403.

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

Thompson, R. H., Iwata, B. A., Conners, J., & Roscoe, E. M. (1999). Effects of reinforcement for alternative behavior during punishment of self-injury. *Journal of Applied Behavior Analysis, 32,* 317-328. doi: 10.1901/jaba.1999.32-317.

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,* 577-586. doi: 10.1016/j.ridd.2004.03.005.

Toussaint, K. A., & Tiger, J. H. (2012). Reducing covert self-injurious behavior maintained by automatic reinforcement through a variable momentary DRO procedure. *Journal of Applied Behavior Analysis, 45,* 179-184. doi: 10.1901/jaba.2012.45-179.

Vaughan, M. E., & Michael, J. L. (1982). Automatic reinforcement: An important but ignored concept. *Behaviorism, 10,* 217-227.

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., & LeBlanc, L. (1994). Treatment of self-injury and hand mouthing following inconclusive functional analyses. *Journal of Applied Behavior Analysis, 27,* 331-344. doi: 10.1901/jaba.1994.27-331.

Watkins, N., & Rapp, J. T. (2014). Environmental enrichment and response cost: Immediate and subsequent effects on stereotypy. *Journal of Applied Behavior Analysis, 47,* 186-191. doi: 10.1002/jaba.97.

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,* 203-214.

Table 1

Operational Definitions for Punishment Procedures

Hands down	The therapist gently holds the participants' hands to their lap or on the table for 30 s
Response blocking	The therapist briefly and gently blocks the target problem behavior from continuing to occur (1-2 s)
Contingent demands	The therapist delivers demands until the participant emits 3 demands with the absence of the target problem behavior. Least to most physical guidance will be used if the participant does not comply with the demand
Response Cost	The therapist will remove leisure items for 30 s
In-seat timeout	The therapist instructs the participant to go to time-out (seat in the corner of the room) using least to most guidance. Time-out duration is 30 s
Reprimands	The therapist delivers one reprimand (e.g., stop that)
Overcorrection-positive practice	The therapist uses manual guidance to prompt a more appropriate response (e.g., item engagement) for 30 s

Table 2

High Preference Items Selected from the Preference Assessments

Participant	PS preference assessment items	CI preference assessment items
Rory	1. Chips 2. Chocolate	1. See n' Say ® 2. Book 3. Toy piano
Eric	1. Granola Bar 2. Cookie	1. Toy guitar 2. Play-doh 3. Legos
Jacob	1. Chocolate 2. Marshmallows	1. iPad, 2. Toy cars 3. Trains
Sally	1. Mixed veggies 2. Grapes	1. Kindle 2. Bead toy 3. Slinky

Table 3

Punishers Evaluated in Punisher Assessment Phase and Final Punisher Selected

Participant	Punisher Assessment	Final Punisher
Rory	1. Response cost 2. Hands down 3. Overcorrection 4. Contingent demands	Contingent demands
Eric	1. Reprimands 2. Hands down 3. Response blocking 4. Contingent demands 5. Overcorrection	Response blocking
Jacob	1. Reprimands 2. Contingent demands 3. Overcorrection 4. Hands down 5. Response blocking	Hands down
Sally	1. Hands down 2. Reprimands 3. Contingent demands 4. Response blocking 5. Overcorrection	Response blocking

Table 4

Social Validity Questions and Lead Clinician Ratings

Questions	Responses			
	Eric	Rory	Jacob	Sally
1. Do you think the treatment involving noncontingent access to toys, differential reinforcement for appropriate play, and X (e.g., hands down) contingent on problem behavior was acceptable?	6	7	7	7
2. Do you think the behavior change was acceptable or sufficient?	7	7	6	7
3. Do you feel that the goals of this treatment were acceptable, appropriate, and important to the individual client?	6	7	7	7
4. Do you think that the client's teachers can effectively implement this procedure with integrity?	3	6	6	5

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

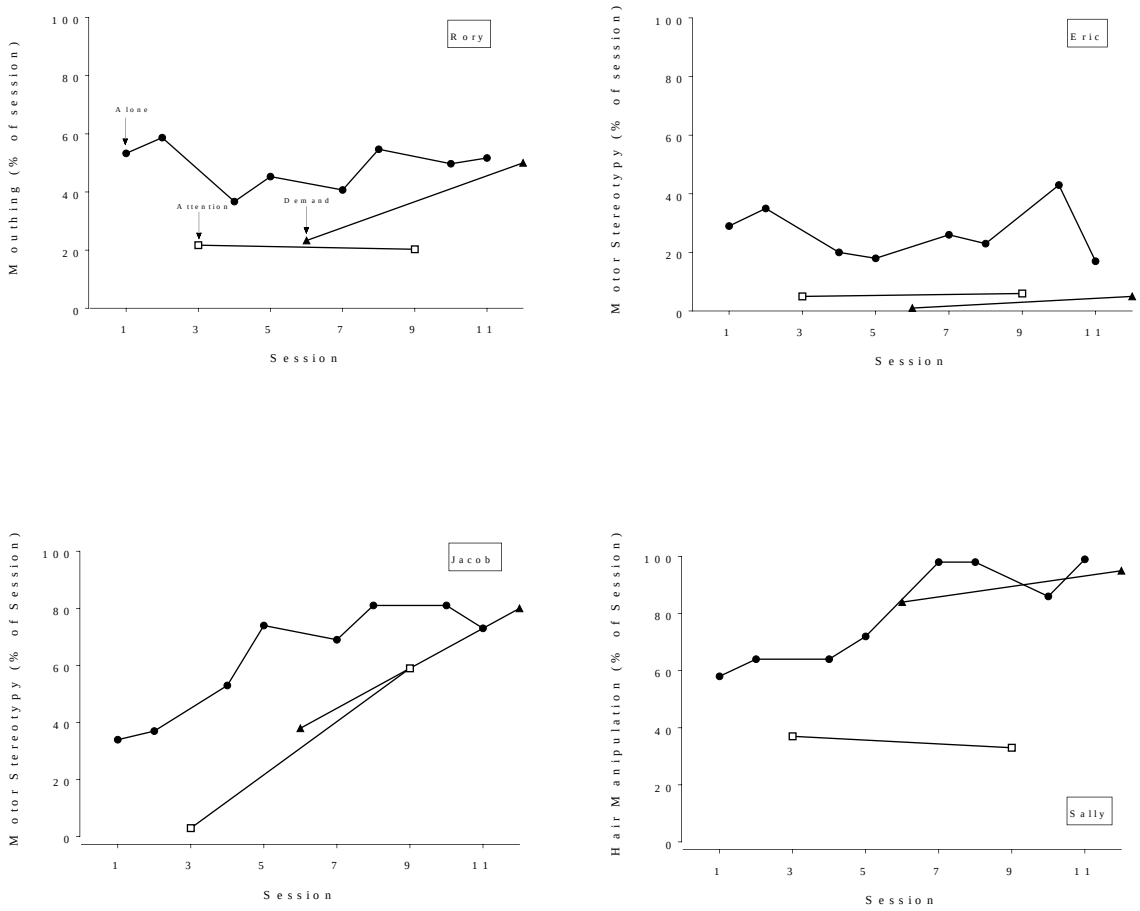

Figure 1. Results from all participants' functional analyses.

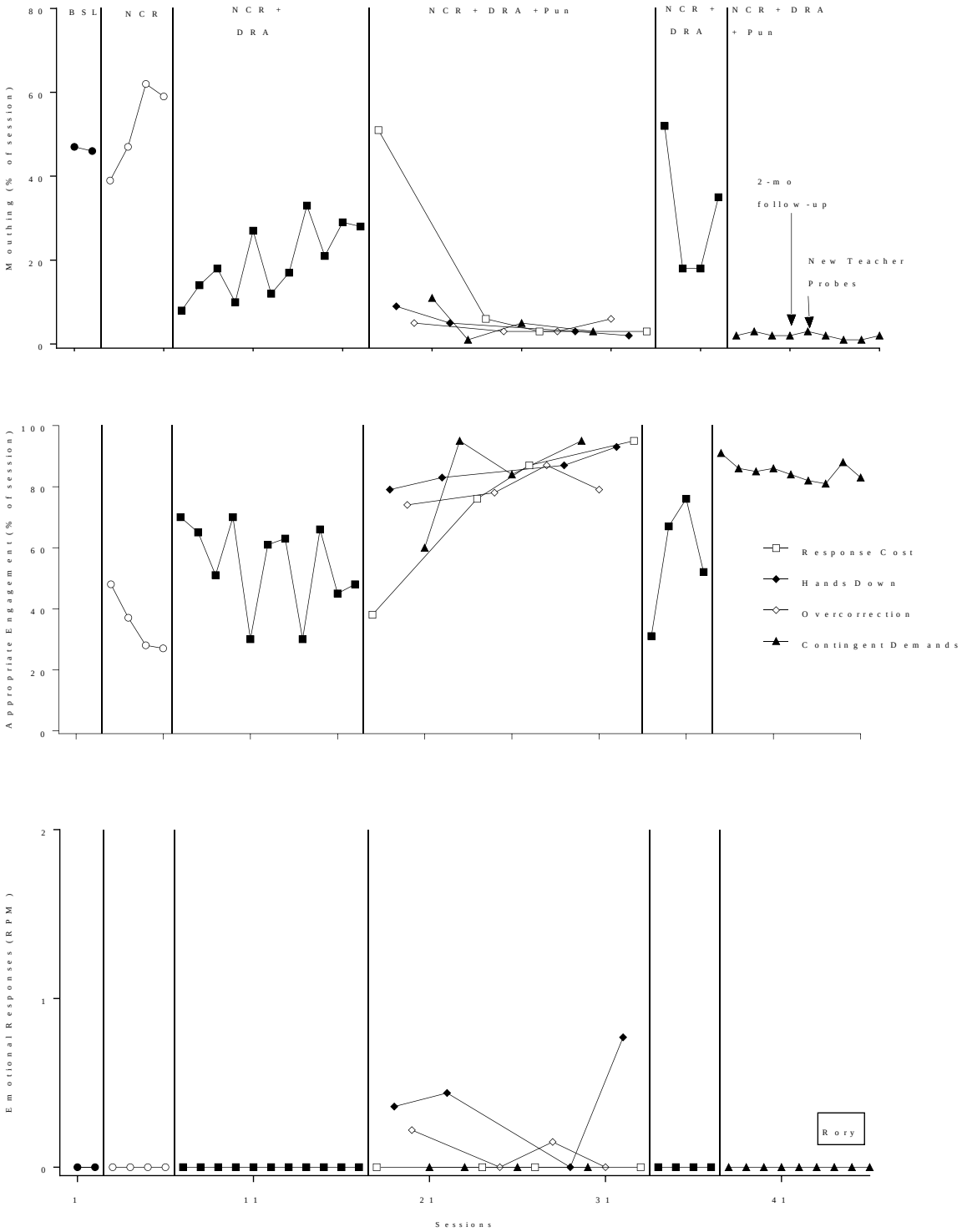

Figure 2. Results from Rory's treatment evaluation.

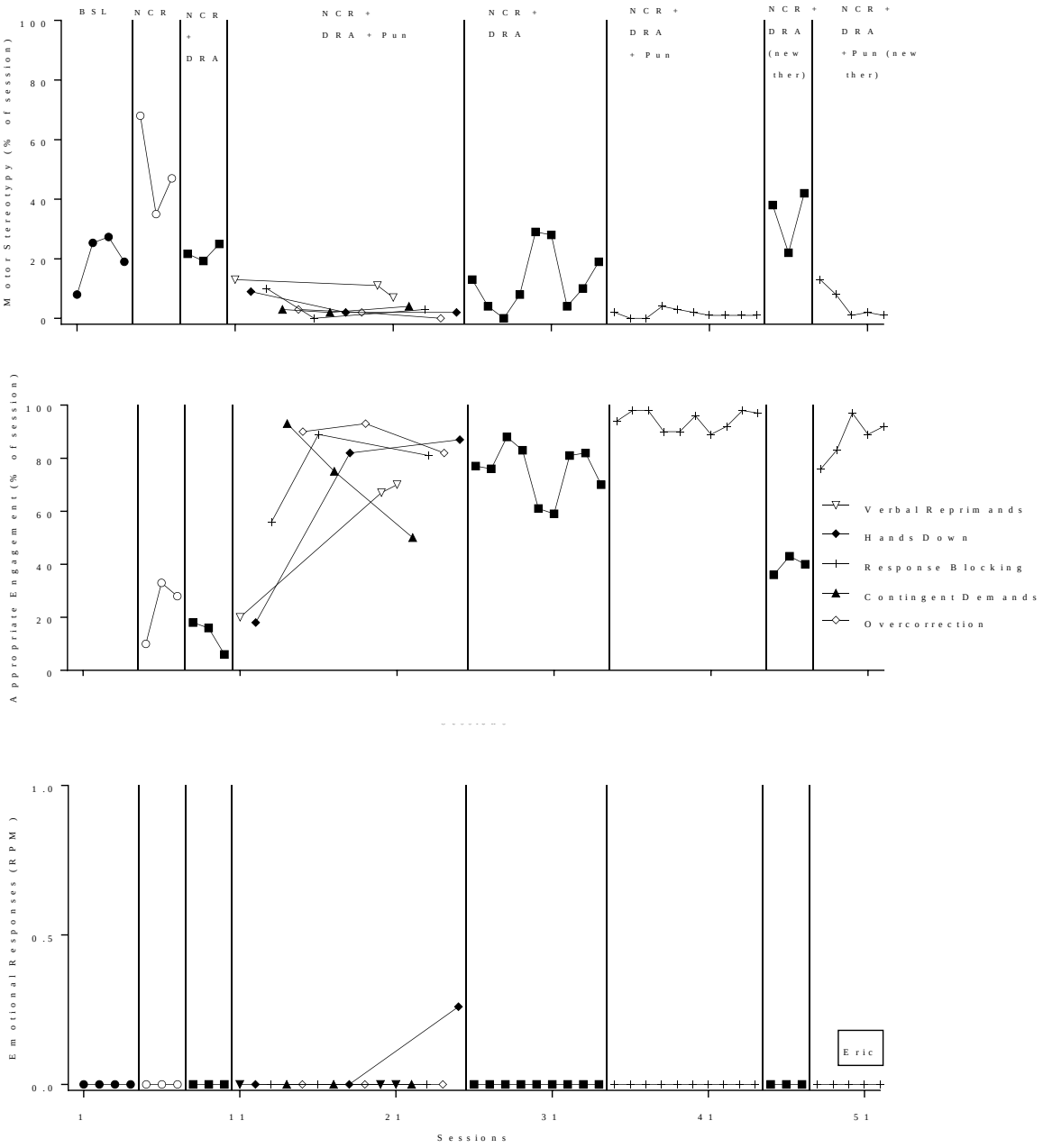

Figure 3. Results from Eric's treatment evaluation.

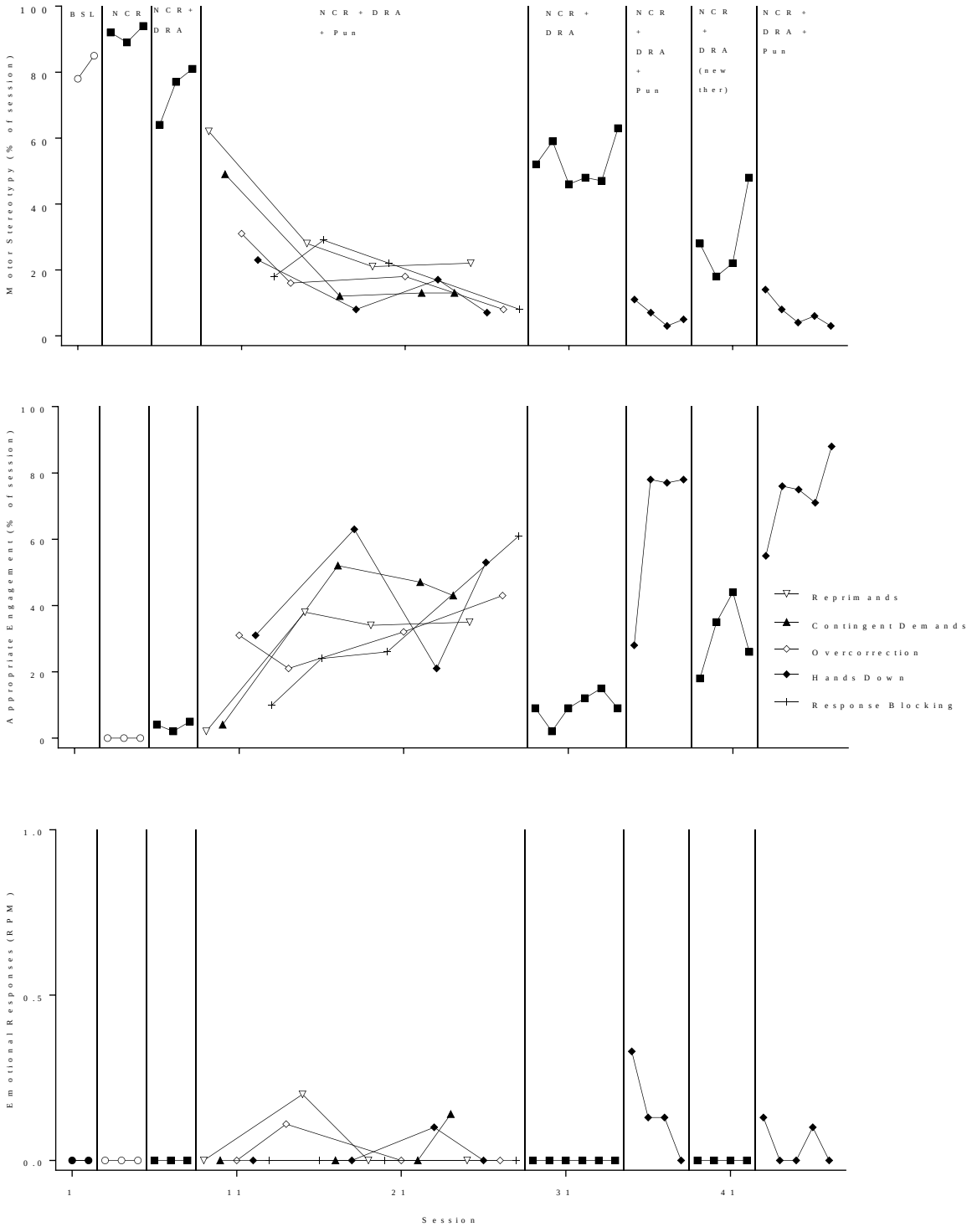

Figure 4. Results from Jacob's treatment evaluation.

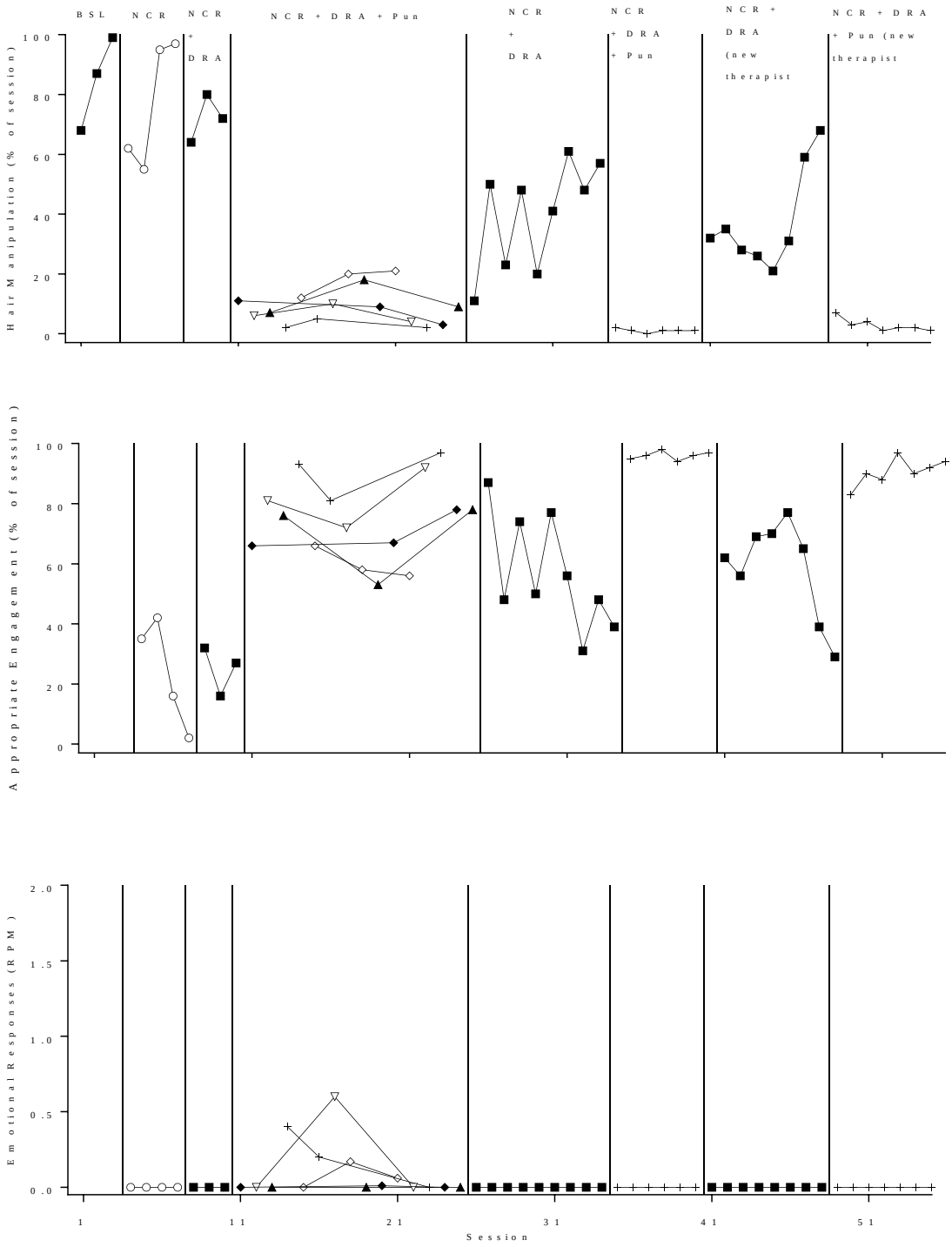

Figure 5. Results from Sally's treatment analysis.

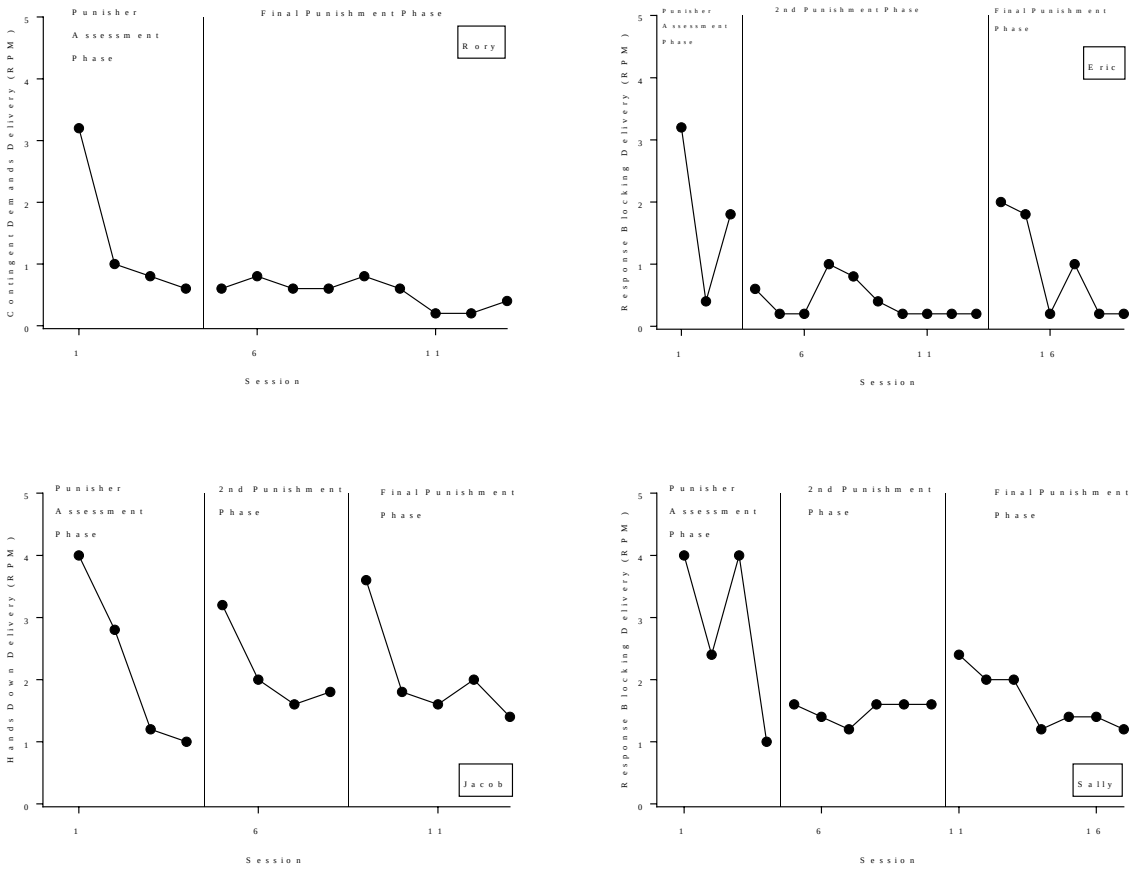

Figure 6. Rate of punisher delivery during the earlier punishment phases compared and the final punishment phase for all participants.

Appendix

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 game, puzzle, coloring). What are some items/activities you think _____likes and can engage with appropriately?

2. Describe _____'s play skills:

3. What are some other things _____ prefers?

4. What are some things _____ dislikes?

Interventions	**Has the individual 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 the participant dislikes?**	**Would you be willing include this procedure in the student's behavior program?**	**Is this procedure something teachers could implement with integrity?**
Hands down	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
Response blocking	1 2 3	1 2 3	Yes or No	Yes or No	Yes or No
Contingent demands	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	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 they 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