Identifying Reinforcers for Treating Automatically Reinforced Problem Behavior

A Thesis Presented

by

James E. Cook

The Department of Counseling and Applied Educational Psychology

In partial fulfillment of the requirements

for the degree of

Master of Science

in the field of

Applied Behavior Analysis

Northeastern University

Boston, MA

August 2011
Thesis Title: Identifying Reinforcers for Treating Automatically Reinforced Problem Behavior

Author: James E. Cook

Department: Counseling and Applied Educational Psychology

Approved for Thesis Requirements of Master of Science Degree

Eileen Roscoe

Jason Bourret

William Ahearn
Identifying Reinforcers for Treating Automatically Reinforced Problem Behavior

by

James E. Cook

B. S., West Virginia University

Submitted in partial fulfillment of the requirements for the degree of

Master of Science in Applied Behavior Analysis

in the Bouvé College of Health Sciences Graduate School

of Northeastern University, August 2011
Acknowledgements

I thank Eileen Roscoe for advising me in the development and implementation of this project and serving as my thesis chair. I also thank Jason Bourret and William Ahearn for serving on my thesis committee and providing excellent feedback. I like to thank Jeffrey Kalles for assisting me throughout the project with data collection and interobserver agreement. I thank the New England Center for Children for allowing me to conduct this study at their facilities.
Table of Contents

A. Abstract.....................................................................................................................6
B. Introduction.................................................................................................................7
C. Method.........................................................................................................................13
D. Results.........................................................................................................................20
E. Discussion......................................................................................................................23
F. References....................................................................................................................28
G. Figure 1.........................................................................................................................32
H. Figure 2.........................................................................................................................33
I. Figure 3.........................................................................................................................34
J. Figure 4.........................................................................................................................35
K. Figure 5.........................................................................................................................36
L. Figure 6.........................................................................................................................37
M. Figure 7.........................................................................................................................38
N. Figure 8.........................................................................................................................39
O. Figure 9.........................................................................................................................40
Abstract

Because clinicians cannot withdraw or terminate the reinforcer associated with automatically reinforced problem behavior, effective behavioral treatment relies on identifying potent reinforcers. To this end, clinicians may conduct preference assessments based on item selection or based on both item interaction and problem behavior. The purpose of the current study is to evaluate two different preference assessments for determining items for use in an intervention for problem behavior maintained by automatic reinforcement. Two individuals diagnosed with an autism spectrum disorder, who exhibited motor stereotypy maintained by automatic reinforcement, are participants in this study. Therapists conducted two leisure-item preference assessments, the paired-stimulus (PS) and competing items (CI) formats. Therapists replicated each assessment four times, alternating assessments in a multielement design. Results to date show that the PS yielded higher levels of stability than did the CI across replications. Implications of these results for determining items for use in NCR and DRO behavioral treatment programs are examined.
Identifying Reinforcers for Treating Automatically Reinforced Problem Behavior

Automatically reinforced problem behavior can be difficult to treat because the reinforcer is directly tied to the response. Therefore, clinicians cannot readily withhold the reinforcer, preventing extinction. For this reason, interventions that directly target automatically reinforced problem behavior often include the presentation of a stimulus that may interrupt the response-reinforcer relation. Examples of such interventions include response blocking (Lerman & Iwata, 1996), protective equipment (Moore, Wacker, & Pennington, 2004), overcorrection (Foxx & Azrin, 1973), and response interruption and redirection (Ahearn, Clark, MacDonald, & Chung, 2007).

Another type of intervention, often referred to as noncontingent reinforcement (NCR), involves the delivery of a stimulus (e.g. a leisure item) continuously or on a time-based schedule, regardless of the individual’s responding. Horner (1980) evaluated NCR by presenting participants with continuous access to leisure items that could be manipulated. As a result, lower levels of problem behavior (e.g. self-injury) and moderate levels of appropriate engagement were observed when compared to a condition without these items. Next, Horner evaluated the effects of a differential reinforcement component that involved the delivery of edibles contingent on item engagement. He found that appropriate behavior further increased and problem behavior further decreased as a result of adding this component. Although NCR alone resulted in only a moderate decrease of problem behavior, this was one of the first studies to demonstrate the potential efficacy of this treatment procedure for behavior that seemed self-stimulatory in nature. However, it remains unclear whether the participants’ problem behavior was maintained by automatic reinforcement because a functional analysis of the behavior was not conducted.
Numerous other studies have shown NCR-based treatments to be effective in treating automatically reinforced problem behavior. For example, Roscoe, Iwata, and Goh (1998) found that NCR was more effective than sensory extinction in decreasing SIB maintained by automatic reinforcement. DeLeon, Anders, Rodriguez-Carter, and Neidert (2000) and Lindberg, Iwata, Roscoe, Worsdell, & Hanley (2003) extended research on NCR by showing that it can be effectively implemented for long durations by varying the items presented.

Another treatment found to be effective in decreasing automatically-reinforced problem behavior is differential reinforcement of other behavior (DRO). In this treatment, a therapist delivers a reinforcer if problem behavior is not observed following a predetermined time interval. If the problem behavior occurs, the reinforcer is withheld. Cowdery, Iwata, and Pace (1990) evaluated DRO for reducing a boy’s automatically reinforced scratching behavior. Following a no-interaction baseline, the authors conducted a 2-min DRO, using an interval based on the longest interresponse time observed during baseline. During this condition, the therapist delivered a token that was later exchangeable for edibles and activities contingent on the participant not emitting scratching during the preceding 2-min interval. Over the course of the treatment, the DRO interval was gradually increased to 15 min.

Several researchers have subsequently evaluated DRO in treating automatically reinforced problem behavior. For example, Ringdahl et al. (2002) used DRO to treat a participant’s hand flapping that occurred only when the participant was left alone. Taylor, Hoch, and Weissman (2005) found DRO to be more effective than NCR for decreasing a participant’s vocal stereotypy. Rozenblat, Brown, Brown, Reeve, and Reeve (2009) compared two DRO schedules, one derived from the 25th percentile of baseline IRTs and the other derived from the 95th percentile. Results showed that the former schedule was more effective in decreasing a
participant’s vocal stereotypy. These findings indicate that the size of the DRO interval may affect the effects of DRO schedules for decreasing problem behavior.

A common feature of NCR and DRO interventions for automatically-reinforced problem behavior is that they require the identification of alternative reinforcers to be effective. In NCR, the stimulus used in treatment must effectively compete with the automatically reinforced behavior. In differential reinforcement procedures, the reinforcer must effectively increase some other behavior that competes with the target response. In DRO specifically, the delay or removal of the item contingent on problem behavior must function as a punisher for the target behavior to decrease. Therefore, a critical component of each of these interventions is the use of an alternative reinforcer that may effectively compete with the target behavior, increase an alternative response that competes with the target behavior, or decrease the target behavior because of its contingent removal.

In order to effectively identify a reinforcer for use during NCR and DRO, researchers often conduct systematic preference assessments, such as a paired-stimulus (PS) preference assessment (Fisher et al., 1992) or a competing items (CI) assessment (Piazza et al., 2000). In the PS preference assessment format, a therapist simultaneously presents two items. Contingent on the participant selecting one of the items, the therapist allows access to the selected item. In this format, items selected more often are considered to be more preferred, and thus more likely to function as potential reinforcers than items selected less often. In a CI assessment, the therapist presents each item singly for a prespecified duration and simultaneously records problem behavior and item interaction. Items associated with lower levels of problem behavior and higher levels of item interaction are considered to be more preferred than those associated with higher levels of problem behavior and lower levels of item interaction.
Identifying preferred stimuli through the use of preference assessments may allow for the identification of potential reinforcers for use in treatment of automatically reinforced problem behavior, improving the likelihood of developing an effective treatment. Vollmer, Marcus, and LeBlanc (1994) compared two types of NCR for reducing three participants’ automatically reinforced self-injurious behavior (SIB). During one type of NCR, the therapist continuously presented preferred stimuli identified during a PS preference assessment, and during the other type of NCR, the therapist continuously presented only nonpreferred stimuli. NCR was more effective in decreasing SIB when preferred rather than nonpreferred stimuli were used, indicating that NCR for treating automatically reinforced problem behavior may be more effective when preferred stimuli identified from a PS preference assessment are used.

Several researchers have used stimuli identified during preference assessments for use in NCR and DRO interventions. For example, authors have conducted a PS assessment to inform stimuli for use in effective NCR (Derby et al., 1995; Lindberg et al., 2003) and DRO treatments (Derby et al., 1995). In addition, authors have conducted a CI assessment to identify stimuli for use in effective NCR (Piazza et al., 2000; Ahearn, Clark, DeBar, & Florentino, 2005; Shore, Iwata, DeLeon, Kahng, & Smith, 1997) and DRO treatments (Piazza, Fisher, Hanley, Hilker, & Derby, 1996; Ringdahl, Vollmer, Marcus, & Roane, 1997). Although authors have demonstrated the effective use of NCR and DRO using preferred items identified from preference assessments, no authors to date have compared the relative efficacy of NCR or DRO when preferred stimuli are identified thorough different preference assessment methods.

Derby et al. (1995) examined the influence of preference during DRO or NCR for two participants’ problem behavior. Although the authors conducted a functional analysis of the problem behavior, the results were inconclusive. Despite this, in an effort to treat the
participants’ problem behavior, the authors conducted a modified PS preference assessment using four leisure items to identify stimuli for use in treatment. During the PS assessment, the therapist presented each stimulus 15 to 25 times. Contingent on participant approach to a stimulus, the therapist presented the selected item for 5 min or until the first occurrence of problem behavior. The experimenters collected data on percentage of selection and the latency to SIB following selection. Based on the results, two items, one associated with the highest percentage selection and the other with the longest latency to problem behavior, were subsequently used in a treatment analysis. For one participant, Trina, the authors found that continuously presenting the item associated with low percentages of selection and long latencies to problem behavior resulted in lower levels of problem behavior than when continuously presenting the item associated with high percentages of selection and short latencies to problem behavior. For the other participant, Rick, the authors compared the effects of an item associated with high percentages of selection and short latencies to problem behavior and an item associated with low percentages of selection and long latencies to problem behavior. The authors compared these stimuli in a DRO intervention in which Rick had continuous access to the item at the start of the session. Contingent on problem behavior, the therapist removed the item and represented it only following 1 s without problem behavior. Although DRO with both item resulted in lower levels of problem behavior than an ignore condition, DRO with the item associated with low percentages of selection and long latencies to SIB resulted in the lowest levels of SIB.

Shore et al. (1997) also evaluated NCR and DRO for decreasing automatically reinforced problem behavior. They used preferred leisure items identified during a single-item duration based preference assessment. During this assessment, the therapist singly presented each item for 10 min, and observers measured item engagement and problem behavior. Items associated with
high levels of item engagement and low levels of problem behavior were included during the NCR and DRO evaluations. The experimenters found that NCR resulted in high levels of item engagement and low levels of SIB relative to baseline. During the DRO treatment evaluation, across several DRO interval and duration of item access values, problem behavior maintained at high levels. The only DRO variation that resulted in changes the occurrence of problem behavior was when the DRO schedule closely resembled NCR (a 5 s DRO interval with 60 s duration of item access). Under this schedule the occurrence of SIB increased. These findings suggested that the preferred leisure items were only effective during an NCR intervention and not in a DRO intervention.

Groskreutz, Groskreutz, and Higbee (2011) compared the effects of high-preference leisure items identified from two different preference assessment methods during an NCR intervention for automatically reinforced vocal stereotypy. The experimenters conducted PS and CI assessments with the same eight leisure items. From these assessments, a high-preference item was identified from the paired-stimulus assessment (i.e., a stimulus associated with high percentages of selection) and a high-competition item was identified from the competing items assessment (i.e., an item associated with low levels of stereotypy). During the NCR treatment evaluation, the authors found that the high-competition items were more effective in suppressing vocal stereotypy than were the high-preference items.

In summary, experimenters have demonstrated that NCR or DRO can decrease automatically reinforced problem behavior. Furthermore, there is preliminary support that identifying preferred leisure items (identified from a preference assessment) for use in these treatments may enhance treatment efficacy. However, it remains unclear whether certain high preference stimuli identified from different preference assessment formats may yield different
effects during NCR and DRO treatments. For this reason, it may be helpful to evaluate the extent to which an item identified as high preference according to a PS versus a CI assessment may affect treatment of automatically reinforced behavior.

Therefore, the purpose of this study was to compare the effects of two different leisure items (one identified as high preference according to the PS method and one identified as high preference according to the CI method) during NCR and DRO treatment procedures for reducing automatically reinforced motor stereotypy.

Method

Participants

Alex was a 12-year old boy diagnosed with autism who attended a residential school for children with developmental disabilities. Alex communicated using partial sentences.

Charles was an 18-year old male diagnosed with autism who attended a residential school for children with developmental disabilities. Charlie communicated using vocal approximations and one-word sentences.

Alex and Charles were included in this study because their caregivers reported that they exhibited high levels of motor stereotypy that interfered with their academic programming.

Settings

Sessions were conducted either in a 3 m x 5 m room containing a table and two chairs or in participants’ bedrooms at the residential facility.

Response Measurement and Interobserver Agreement

During the functional analysis, the dependent variable measured was motor stereotypy for both participants. Motor stereotypy for Alex was defined as any repetitive, noncontextual movement of the body for at least 1 s or any instance of Alex placing his shoulder, hand or finger
on his ear or ears. Motor stereotypy for Charles was defined as any instance in which Charles grabs any part of his shirt between two or more digits and repetitively moves the fabric of his shirt in a back-and-forth, up-and-down, or side-to-side motion for at least 1 s. Observers recorded stereotypy using a duration measure and summarized the data as a percentage of session.

During the functional analysis and treatment analysis, we collected interobserver agreement (IOA) by comparing observers’ records. An agreement was scored when both observers recorded an occurrence or nonoccurrence of stereotypy during each 1-s interval. IOA was summarized as percentage of agreement by dividing the total number of agreements by the total number of agreements plus disagreements and multiplying this number by 100%. For Alex, IOA was calculated for 38% of FA sessions and averaged 91% (range 85 to 96%). For Charles IOA was calculated for 33% of FA sessions and averaged 92% (range 86 to 100%).

Functional Analysis

We conducted a functional analysis (as described by Iwata et al., 1982/1994) to determine the function of participants’ motor stereotypy. Conditions included alone or no interaction, attention, control, and demand, and sessions lasted 5 min each. In the alone condition (Alex only), the participant was alone in a room with no materials. In the no interaction condition (Charles only) the therapist was present in the room but did not interact with the participant nor deliver programmed consequences. In the attention condition, a therapist pretended to do work and delivered brief vocal and physical attention contingent on stereotypy. If the participant engaged in stereotypy continuously, the therapist delivered attention every 5 s. In the control condition, the participant had continuous access to high-preference leisure items (based on previous leisure item assessment), and the therapist delivered brief vocal and physical
attention on a FT 15-s schedule. There were no programmed consequences for stereotypy or any other behavior. In the demand condition, the therapist presented demands using a progressive prompting hierarchy (vocal prompt, model, manual guidance). Contingent on compliance (completion of the demand before the physical guidance prompt), the therapist delivered praise. Contingent on stereotypy, the therapist removed demand materials and turned his or her back to the participant for 15 s. If the participant emitted stereotypy during the break interval, the therapist did not extend the interval.

**Preference Assessments**

Prior to initiating the preference assessments, the experimenter interviewed a caregiver or staff person who frequently worked with each participant using the Reinforcer Assessment with Individuals with Severe Disabilities (RAISD) questionnaire (Fisher, Piazza, Bowman, & Amari, 1996). The items included in the stimulus arrays for both participants were based on the results of the RAISD. For Charles, we included seven leisure items and one control item (a piece of cardboard). Because Alex was observed to frequently exhibit pica with potential control items, we did not include a control item with him. Instead, eight leisure items were included in Alex’s stimulus array. Four edibles were identified for inclusion in a separate array for Alex consisting of both edible and leisure items. Prior to initiating the preference assessments, the therapist conducted a forced-exposure trial with each leisure item wherein the therapist modeled how to appropriately engage with the item and then provided the item to the participant for 2 min. Alex was permitted to sample each edible prior to initiating preference assessments including edibles.

The therapist conducted two preference assessments, the paired-stimulus (Fisher et al., 1992) and competing items (Piazza et al., 2000) formats. Each preference assessment format was conducted four times, and the assessments were alternated using a multielement design.
**Paired-stimulus assessment.** The assessment consisted of a number of trials in which the stimuli were presented in pairs. During each trial, the therapist placed two items in front of the participant and prompted him to “choose” one. The participant had 5 s to make a selection, defined as making physical contact with one of the two items presented. Upon making a selection, the therapist granted access to the selected item for 30 s and removed the other item. Attempts to select both items were blocked. If no selection was made after 5 s, no choice was recorded and the therapist initiated the next trial.

The dependent variable measured was selection, summarized as the percentage of trials selected. High preference stimuli were defined as those selected on a high percentage of trials. We calculated IOA by dividing the number of agreements (the same selection on a given trial) by the total number of trials and multiplying this number by 100%. For Alex, IOA was calculated for 75% of sessions and averaged 99% agreement (range, 99% to 100%). For Charles, IOA was calculated for 50% of sessions with 100% agreement.

In addition to recording selection, observers also recorded the duration of item contact and motor stereotypy during the 30-s interval of item access. These data were summarized as the percentage of time the item was available. Data are available upon request.

**Competing items assessment.** The assessment consisted of the therapist singly presenting each item. In assessments that included only leisure items in the array, the therapist presented items for 5-min sessions. In assessments that included edibles in the array, the therapist items for 30-s sessions. During each session, the therapist presented a single item continuously and did not interact with the participant. Observers measured the duration of item interaction and motor stereotypy, and these measures were summarized as percentage of session. High preference items were defined as those associated with higher levels of item interaction and
lower levels of stereotypy, whereas low preference items were defined as those associated with lower levels of item interaction and higher levels of motor stereotypy. IOA was obtained by comparing a second-by-second record of sessions scored by two observers. An agreement was defined as both observers recording an occurrence or nonoccurrence of item interaction or stereotypy for each second. We calculated IOA by dividing the total number of agreements by the total number of agreements plus disagreements and multiplying by 100%. For Alex IOA was calculated for 36% of sessions and averaged 93% for motor stereotypy (range, 85% to 99%) and averaged 97% for item interaction (range, 86% to 100%). For Charles IOA was calculated for 34% of session and averaged 96% for motor stereotypy (range, 87% to 98%) and averaged 98% for item interaction (range, 92% to 100%).

Two items, one identified as HP according to the PS (HP/PS), one identified as HP according to the CI (HP/CI), were included in the subsequent reinforcer assessment and treatment evaluation.

**Reinforcer Assessment**

We conducted a reinforcer assessment to determine whether the HP items identified by each of the preference assessments functioned as reinforcers for an arbitrary response (i.e., a button press for Alex and a “More please” mand for Charles). Baseline and reinforcement conditions were conducted, and a reversal design was used to demonstrate experimental control. Prior to the first session of each condition, we conducted a pretraining session to expose the participant to the contingencies in effect. For example, during baseline, the therapist manually guided the participant to emit the response, and then delivered no differential consequences. During the pretraining session immediately prior to the first session of the reinforcement
condition, the therapist manually guided the participant to emit the response and then delivered access to the leisure item associated with the reinforcement condition.

IOA was obtained by dividing sessions into 10-s blocks and comparing the number of target responses recorded in each block by two observers. An agreement was defined as both observers recording the target response in the same 10-s block. We calculated IOA by dividing the number of agreements by the total number of agreements plus disagreements multiplied by 100%. For Alex IOA was calculated for 35% of sessions and averaged 95% (range 92-98%). For Charles IOA was calculated for 33% of sessions and averaged 91% (range 82-95%).

**Baseline.** The participant sat at a table across from the therapist. For Alex, the button was presented on the opposite side of the table within his reach. The therapist delivered no programmed consequences for button pressing or manding. Sessions lasted 5 min.

**HP/CI Reinforcement.** Sessions were similar to baseline except that the therapist placed the HP/CI item on the table in front of the participant at the start of the session. The therapist blocked participant attempts to take the item. Contingent on each occurrence of the target response, the therapist delivered the item for 15 (Alex) or 30 (Charles) seconds. The therapist stopped the session clock when the item was delivered and restarted the session clock when the therapist collected the item. No consequences were in effect for target responses that occurred when the reinforcer was available (i.e., the duration of item access was not extended). Sessions were 5 min of session time or 10 or 20 min of total time for Charles or Alex, respectively.

**HP/PS Reinforcement.** This condition was identical to the HP/CI reinforcement condition except that the therapist placed the HP/PS stimulus on the table in front of the participant, and the therapist delivered this item contingent on the target response.
Edible Reinforcement (Alex only). This condition was similar to the other reinforcement conditions, except that the therapist placed an edible (a peppermint patty) on the table in front of Alex and delivered a small piece of the edible contingent on button pressing. As consumption time was short (often less than 1 s), session time was not paused for reinforcer consumption.

Treatment Assessment

We evaluated the effects of the HP/PS reinforcement, HP/CI reinforcement, and edible reinforcement (Alex only) using a reversal design. In addition, we compared the effects of NCR and DRO interventions using an embedded multielement design during baseline and reinforcement phases. During intervention, either DRO or NCR was in effect, and the schedule type varied across sessions in a 1:1 ratio with the DRO session preceding the NCR session.

During Alex’s treatment assessment, IOA was calculated for 33% of sessions and averaged 92% for motor stereotypy (range, 86 to 99%) and averaged 97% for item interaction (range, 88 to 100%). During Charles’s treatment assessment, IOA was calculated for 33% of sessions and averaged 99%, (range, 96 to 100%) and averaged 99% for item interaction (range, 97 to 100%).

Baseline. The therapist was present but did not interact with the participant. No items were presented and no programmed consequences were in effect. Sessions lasted 5 min.

DRO. The therapist delivered the HP item following a predetermined time interval (based on the mean IRT of stereotypy in baseline) without stereotypy. Because both Alex and Charles’s average IRT during baseline was 7 s, the DRO interval for both was 5 s.

NCR. The therapist presented the HP item at the start of the session, and the participant had continuous access to the item for the remainder of the session.
Results

Figures 1 and 2 display the results of the functional analyses conducted with Alex and Charles, respectively. In these functional analyses, motor stereotypy occurred most often during the alone or no interaction conditions, suggesting that participants’ motor stereotypy was maintained by automatic reinforcement.

Figure 3 depicts the results of the four PS and CI assessments conducted with Alex with only leisure items. Because the recorder was consistently selected on a high percentage of trials during the PS assessment, it was selected as the HP/PS item. Because the harmonica was consistently associated with high levels of item interaction and low levels of stereotypy during the CI assessment, it was selected as the HP/CI stimulus.

Although the recorder was identified as high preference in the PS assessment, it was not high preference according to the CI because it was associated with high levels of stereotypy. Although harmonica was identified as high preference during the CI assessment, it was not high preference during the PS because it was associated with only moderate percentages of selection. When comparing outcomes of the PS and CI, it is interesting to note that the PS yielded more stable results than the CI across replications.

Figure 4 depicts the results of the PS and CI assessments conducted with Alex using an array of both edible and leisure items. In the PS edibles displaced leisure items as HP. In the CI stereotypy occurred at low levels and interaction occurred at high levels for most stimuli in the array. Because the peppermint patty was selected in a high percentage of trials in the PS and was associated with high levels of interaction and low levels of stereotypy in the CI, it was selected as the HP edible for use in the reinforcer assessment.
Figure 5 depicts the results of the four PS and CI assessments conducted with Charles. Play-doh was selected as the HP/PS item because it was consistently selected on the greatest high percentage of trials. During the CI, all of the leisure items were associated with high levels of item interaction and low levels of stereotypy, with no stereotypy occurring during assessments 3 and 4. The magazine was identified as the HP/CI item because it was consistently associated with high levels of item interaction and low levels of stereotypy. By contrast, the magazine was identified as a LP item (i.e., it was associated with low percentages of selection) during the PS assessment.

Figure 6 depicts the results of the reinforcer assessment conducted with Alex. During baseline and during the HP/PS (recorder) reinforcement conditions, low levels of responding occurred. Although the recorder was identified as a high preference item during the PS because it was frequently selected over other items and although Alex frequently interacted with it during CI assessment sessions, it did not function as a reinforcer for even a relatively simple response such as button pressing. During the HP/CI (harmonica) reinforcement condition, similarly low levels of responding occurred. Although harmonica was identified as a HP item in the CI assessment because it was associated with high levels of interaction and low levels of stereotypy, it did not function as a reinforcer. During the HP edible reinforcement condition, a significant increase in responding was observed, and this effect was replicated during the subsequent HP edible reinforcement phase, suggesting that this item functioned as an effective reinforcer.

Figure 7 depicts the results of the reinforcer assessment conducted with Charles. During baseline, manding did not occur. During the HP/CI reinforcement condition, an increase in manding was observed, and these results were replicated in a subsequent HP/CI reinforcement condition. During a return to baseline, manding returned to low levels. During the HP/PS
reinforcement condition, a higher level of manding occurred than that observed in the HP/CI reinforcement condition. These results were then replicated indicating that both the HP/CI and HP/PS items functioned as reinforcers for manding, but the HP/PS item functioned as a more potent reinforcer.

Figure 8 depicts the results of the treatment assessment conducted with Alex. During baseline, stereotypy occurred at moderate-to-high and variable levels. During the HP/CI reinforcement phase, DRO produced no change in the level of stereotypy. Although a moderate decrease in stereotypy was observed during the NCR condition, levels were variable. To control for potential sequence effects, a successive series of only NCR and then only DRO sessions were conducted. Although responding initially decreased during NCR, it gradually increased to high levels. During DRO sessions, responding maintained at levels previously observed in DRO. During a reversal to baseline, stereotypy occurred at high levels. During the HP/PS reinforcement phase, no change in the level of stereotypy was observed during DRO sessions and slightly lower yet variable levels of responding were observed during NCR sessions. During the HP edible reinforcement phase, decreases in stereotypy were observed during DRO and NCR conditions, but greater decreases and more stable levels of responding were observed in the NCR condition. During a return to baseline, levels of stereotypy increased. During a replication of the HP edible reinforcement phase, significant decreases in stereotypy were again observed during only the NCR condition.

Figure 8 also depicts item interaction for Alex during the HP/CI and HP/PS reinforcement conditions. When the HP/CI item was delivered, generally higher levels of item interaction were observed in the NCR condition than in the DRO condition. Interaction maintained at high levels during the series of successive NCR sessions and occurred at lower
levels in the series of successive DRO sessions. When the HP/PS item was delivered in the DRO and NCR conditions, overall higher levels of item interaction occurred during NCR and DRO conditions, with a slightly higher level during NCR relative to DRO.

Figure 9 depicts the results of the treatment assessment conducted with Charles. During baseline, stereotypy occurred at moderate levels. During the HP/CI reinforcement phase, stereotypy decreased to zero in both DRO and NCR conditions. During a return to baseline, stereotypy increased to moderate levels. During the HP/PS reinforcement phase, stereotypy decreased to zero in both NCR and DRO conditions and remained at zero except for an increased level of responding in a single session of DRO.

Figure 9 also depicts item interaction for Charles during the HP/CI and HP/PS reinforcement conditions. When the HP/CI and HP/PS items were delivered, item interaction was high except for a single session of DRO during the HP/PS reinforcement phase in which item interaction occurred at a lower level.

**Discussion**

We found that PS and CI preference assessments yielded different outcomes and subsequently, identified different stimuli as high preference. This is not unusual as other research has found that conducting different preference assessment formats with the same participant and the same array of stimuli may identify different stimuli as high preference (DeLeon & Iwata, 1996; Fisher et al. 1992; Groskreutz et al. 2011; Hanley, Iwata, Lindberg, & Conners, 2003; Roane, Vollmer, Ringdahl, & Marcus, 1998). One factor that may account for this difference is the access time granted to the participant with each item in the different preference assessments. In the PS, following selection the participant had access to the item for 30 s; while in the CI, the participant had continuous access to a single item for 5 min. Steinhilber and Johnson (2007)
conducted two MSWO preference assessments using leisure items in which access time following selection lasted either 15 s or 15 min. They found that the assessments produced different preference hierarchies as access time was varied. The difference in access time in the PS and CI assessments may account for the difference in preference demonstrated for certain items. For instance Charles engaged with the magazine for 100% of the interval during the CI, but it was identified as a low preference item according the PS. It is possible that aspects of the magazine that may make it more reinforcing or preferred over 5 min may not be as readily accessed in 30 s. As a result, Charles may have allocated his responding in the PS towards those items that have reinforcing qualities that are more easily accessed in a shorter period of time.

We extended the preference assessment literature in a number of ways. First, we conducted multiple replications of both preference assessments to evaluate the stability of the results produced by each assessment format. For Alex, the PS assessment yielded more stable or reliable outcomes across replications than did the CI assessment, suggesting that the CI method may not yield reliable outcomes. If preference assessments are not stable, then they will not have high predictive validity. Previous researchers who have evaluated preference assessments often conduct the assessment one time with each participant (e.g., Fisher et al., 1992; Groskreutz et al., 2011; Piazza et al., 2000; Roane et al, 1998). Although these authors have demonstrated the assessment’s predictive validity by conducting a reinforcer assessment, it is unclear whether the assessment is reliable. Because previous research has shown that preferences may not be stable for all participants (Hanley, Iwata, & Roscoe, 2006) this has important implications for reinforcer identification and the treatment of automatically reinforced problem behavior. Given that preference assessment results are capable of being influenced by potential motivating operations, particularly related to the potential satiation of those stimuli included in the array.
(Gottschalk, Libby, & Graff, 2000; Hanley et al., 2006), it is possible that conducting only a single preference assessment may fail to identify the best potential reinforcer from an array, a finding that may be determined by conducting multiple exposures to a preference assessment.

Second, we evaluated the predictive validity of each of these preference assessments by conducting a reinforcer assessment. For Charles, the results of this reinforcer assessment showed that the HP/PS leisure item was a more effective reinforcer than the HP/CI item. For Alex, neither the HP/PS leisure item nor the HP/CI item functioned as a reinforcer for increasing an arbitrary response. These findings conflict with previous research that has shown that items identified as high preference according to a preference assessment function as reinforcers (e.g., DeLeon & Iwata, 1996; Fisher et al., 1992; Piazza et al., 1996; Roane et al, 1998).

Third, we compared the relative efficacy of HP items from both assessment formats for reducing stereotypy during two frequently used interventions for automatically reinforced problem behavior, NCR and DRO. The results obtained from the treatment assessment conducted with Charles thus far seem to indicate no significant difference in the effectiveness either the HP/PS or HP/CI item in reducing automatically-reinforced behavior, but replication of these results is required. During the treatment assessment for Alex, neither NCR nor DRO with HP/PS nor HP/CI items resulted in clinically significant decreases in stereotypy. One explanation for this outcome is that the high-preference leisure items identified did not function as effective reinforcers. For Charles, significant reductions in the level of stereotypy were observed when items that had been validated as reinforcers in the reinforcer assessment were delivered in NCR and DRO treatments. For Alex, given that neither leisure item functioned as a reinforcer during the reinforcer assessment, it is not surprising that these items were not effective during DRO. However, because the competing items assessment is procedurally similar to NCR, a surprising
finding for Alex was that NCR with the HP/CI item did not produce clinically acceptable reductions in problem behavior for Alex, suggesting that this assessment may not always be predictive of NCR effects. By contrast, the results of the reinforcer assessment were predictive of NCR treatment effects. More specifically, only an edible that was shown to function as a reinforcer during the reinforcer assessment, resulted in durable reductions of stereotypy during NCR. However, an edible was not effective in reducing stereotypy during DRO. This may relate to the findings of Shore et al. (1997) in which NCR was effective for reducing automatically-reinforced SIB but DRO was not. In their discussion the researchers commented that the participants’ preference for item access was abolished in the DRO treatment as it required participants to forego an immediate, readily available reinforcer provided by SIB. A similar process may be responsible for the results observed in the treatment comparison conducted with Alex.

It is also worth noting that levels of item interaction tended to be differentially higher in NCR than in DRO in the treatment comparison conducted with Alex. This may relate to the duration of item access granted in the different treatments. Item interaction may not have been as reinforcing during the shorter periods of item access granted in DRO than in the longer periods of item access granted in NCR, and as such may help account for the difference in treatment effects. This is most pronounced when the HP/CI item was delivered in NCR and DRO. In this condition, item interaction and reductions in stereotypy, although not durable, were greater in NCR than in DRO. Indications of this difference in item interaction and treatment effects may be observed in the series of preference assessments conducted. In the preference assessments the HP/CI item, the harmonica, was not identified as high preference according to the PS in which periods of item access were shorter. Periods of item access were of similar duration in the
reinforcer assessment, in which access to the HP/CI item did not prove to be reinforcing. Steinhilber and Johnson (2007) conducted reinforcer assessments in which access time to stimuli was varied. Their results demonstrated that duration of access can influence the effectiveness of a stimulus in functioning as a reinforcer. Future research may examine if reinforcement effects that vary as a result of access time vary similarly in their treatment effects as a result of access time.

Another consideration when conducting preference assessments for informing the treatment of automatically-reinforced problem behavior is whether the assessment yielded reliable outcomes. If preference assessments do not yield stable outcomes, this may have implications for treatment. Because the competing items assessment is procedurally similar to an NCR treatment, preference stability during this assessment may inform NCR treatment effects (i.e., unstable preferences may suggest that NCR will not be consistently effective). For Alex, the CI assessment did not yield stable outcomes. In addition, during the NCR condition of the HP/CI phase, low-to-moderate and variable levels of stereotypy were observed. Therefore, it is possible that the poor stability during the CI assessment may suggest that NCR with those stimuli will not be consistently effective. Lindberg et al. (2003) found that decreases in stereotypy during brief (10 min) NCR sessions did not maintain during extended session durations (120 min). When including multiple items during these extended sessions, the authors observed maintenance of NCR treatment effects. In future research, experimenters could evaluate the effectiveness of a NCR when leisure items from stable versus unstable preference assessment outcomes are obtained. In addition, one could compare the effectiveness of delivering multiple high-preference items (from an unstable assessment outcome) to delivering a single high preference item (from a stable assessment outcome) during NCR for automatically reinforced problem behavior.
References


Figure 1. The results of multielement functional analyses conducted with Alex. The data indicate the percentage of the session during which stereotypy occurred.
Figure 2. The results of multielement functional analyses conducted with Charles. The data indicate the percentage of the session during which stereotypy occurred.
Figure 3. The results of a series of four PS and CI preference assessments conducted with Alex. The left column depicts the percentage of trials an item was selected across replications of the PS preference assessment. The right column depicts the results of the CI preference assessments. The top right panel depicts the measure of stereotypy and the bottom right panel depicts the measure of item interaction across replications of the CI. Data from the CI indicate the percentage of the session during which the behavior occurred.
Figure 4. The results of a PS and CI preference assessment conducted with Alex with an array of leisure and edible items. The left column depicts the percentage of trials an item was selected in the PS assessment. The right column depicts the results of the CI preference assessment. The top right panel depicts the measure of stereotypy and the bottom right panel depicts the measure of item interaction in the CI. Data from the CI indicate the percentage of the session during which the behavior occurred.
Figure 5. The results of a series of four PS and CI preference assessments conducted with Charles. The left column depicts the percentage of trials an item was selected across replications of the PS preference assessment. The right column depicts the results of the CI preference assessments. The top right panel depicts the measure of stereotypy and the bottom right panel depicts the measure of item interaction across replications of the CI. Data from the CI indicate the percentage of the session during which the behavior occurred.
Figure 6. Results of the reinforcer assessment conducted with Alex. Data indicate the rate of button pressing.
Figure 7. Results of the reinforcer assessment conducted with Charles. Data indicate the rate of manding.
Figure 8. Results of the treatment comparison conducted with Alex. The top panel depicts the level of stereotypy across conditions, and the bottom panel depicts the level of item interaction for those conditions in which either the HP/CI item, HP/PS item, or edible were delivered.
Figure 9. Results of the treatment comparison conducted with Charles. The top panel depicts the level of stereotypy across conditions, and the bottom panel depicts the level of item interaction for those conditions in which either the HP/CI or HP/PS items were delivered.