Preference for Reinforcers of Different Efficacies with Persons with Developmental Disabilities

by

Pamela D. Kelso

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Psychology

University of Manitoba

Winnipeg

Abstract

Past research has shown that preference for a stimulus is correlated positively with reinforcer efficacy – more preferred stimuli tend to be more potent reinforcers. Researchers have recently examined the use of progressive ratio (PR) reinforcement schedules to evaluate preference. However, research on the concordance between preference and performance under a progressive ratio schedule is limited. Therefore, two experiments were conducted to extend this research. In Experiment 1 I compared the relative reinforcing efficacy of six stimuli obtained under a PR schedule to the mean preference hierarchy obtained through paired-stimulus preference assessments conducted before and after PR reinforcement sessions. In Experiment 2 I evaluated the PR schedule for quantifying the reinforcing value of three stimuli from Experiment 1 (i.e., one high, one medium, and one low preference stimulus). The results of Experiment 1 indicated that items that yielded higher breaking points and responses per minute were also more preferred. However, concordance between percent preference and breaking points and response rates for less preferred items was more variable. The results of Experiment 2 showed that high preference stimuli increased responding over baseline for all participants. However, low preference reinforcers also increased responding over baseline for some participants.

Acknowledgements

First and foremost, I would like to acknowledge my advisor, Dr. C. T. Yu. His guidance and support throughout this project and during my time at the University of Manitoba have been invaluable. I would also like to thank Dr. Toby Martin who served as a research advisor on this project and Carly Chand, Lilian Saltel, Alejandra Zaragoza Sherman, May Lee, Jade Wightman, Lindsay Arnal, Mylena Ribeiro, and Adriana Velasco Cristobal for their assistance on data collection and reliability checks. I would like to express my gratitude to my committee members, Dr. Garry Martin, Dr. Joseph Pear, and Dr. Dennis Hrycaiko, for their insight in regards to this project. Last, I would like to thank the staff and residents of St. Amant for their participation.

Table of Contents

List of Tables

List of Figures

Preference for Reinforcers of Different Efficacies with Persons with Developmental Disabilities

Having reinforcers and using them effectively is paramount to the success of behavioral interventions. Direct preference assessment is an effective and efficient method for identifying reinforcers for individuals with severe and profound developmental disabilities with limited communication skills (Hagopian, Long, & Rush, 2004; Tullis et al., 2011). Studies have shown that preference is correlated positively with reinforcer efficacy – more preferred stimuli tend to be more potent reinforcers. Several recent applied studies have examined the use of progressive ratio (PR) reinforcement schedules to evaluate preference. However, research on the concordance between preference and performance under a PR schedule is quite limited. The current study sought to extend this area of research in two experiments. Experiment 1 examined the concordance between preference obtained through preference assessments and breaking points and response rates obtained under a PR reinforcement schedule. Experiment 2 evaluated the reinforcing efficacy of high, medium, and low preference stimuli under a PR schedule. In the ensuing sections, the concept of preference and relevant preference assessment research are reviewed. The methodology and results of each experiment are then reported, followed by a general discussion.

Preference

Preference can be measured using both indirect and direct methods (Hagopian et al., 2004). Indirect methods of preference assessment involve interviewing or surveying third-party care providers about an individual's likes or dislikes. Direct methods involve presenting stimuli systematically and observing an individual's response or level of

engagement with each stimulus. The results of direct preference assessments have been shown to be more accurate than indirect methods for indentifying reinforcers (Green et al., 1988; Green, Reid, Carnipe, & Gardner, 1991; Parsons & Reid, 1990). Therefore, the focus of my review will be on direct preference assessments.

Preference has been defined as "the relative strength of discriminated operants" (Martin, Yu, Martin, & Fazzio, 2006). It can be measured based on a pattern of choosing or the frequency with which an individual engages with or approaches a given stimulus in comparison to another. Approach responses are evident when an individual moves toward or consumes a given stimulus, and engagement responses are measured based on the length of time an individual interacts with a given stimulus (Hagopian et al., 2004). When an individual is said to prefer one stimulus over another, this means that the frequency with which they have approached that stimulus is greater than that of another when two or more stimuli are present (Martin et al., 2006). Stimulus preference is often expressed as the percentage of trials a stimulus has been chosen out of the total number of trials it was presented during a preference assessment. Also, a preference ranking, or hierarchy, may be derived based on the rate of responding allocated to each stimulus.

Research on Single, Paired, and Multiple Stimulus Presentations

Various procedures have been developed to assess an individual's preferences (Cannella, O'Reilly, & Lancioni, 2005). The single-stimulus procedure involves presenting one stimulus on each trial and measuring the individual's approach response (i.e., approaching the stimulus within a predetermined time frame) (Pace, Ivancic, Edwards, Iwata, & Page, 1985). Items that have been approached with greater frequency are considered more preferred than items that have been approached less frequently. The

paired-stimulus (PS) or paired-choice procedure involves presenting two items concurrently on each trial and the individual is instructed to choose one (Fisher et al., 1992). Each item is paired with every other item. The multiple-stimulus (MS) procedure involves presenting 6-8 stimuli concurrently on each trial and all stimuli are presented on each trial (e.g., Carr, Nicolson, & Higbee, 2000; DeLeon & Iwata, 1996; Windsor, Piché, & Locke, 1994). The multiple-stimulus without replacement (MSWO) method is similar to the MS procedure, except that the stimulus selected on each trial is removed from the array on subsequent trials (DeLeon & Iwata, 1996).

The single-stimulus method has been criticized on the grounds that some individuals with developmental disabilities routinely approach all items presented, thus failing to identify a preference hierarchy. Fisher et al. (1992) compared single-stimulus and PS procedures and found that the PS procedure was more effective in determining preferences. The MS method was more efficient than the PS method in requiring fewer assessment trials, but it did not differentiate preferences as well as the PS approach (Windsor et al., 1994). The MSWO method, however, was as effective as the PS method, and more effective than the MS method, in differentiating preferences (DeLeon & Iwata, 1996). In addition, Carr and colleagues (2000) have successfully used 3 stimulus-presentation sessions in an MSWO procedure to identify reinforcers instead of 5 sessions (DeLeon & Iwata, 1996). However, the PS method is commonly used in research because the procedure ensures that each stimulus is paired with every other stimulus an equal number of times.

Research on Object, Picture, and Spoken Stimulus Presentations

Choice options are often presented in three different modes (object, picture, and spoken words) during preference assessments (Cohen-Almeida, Graff, & Ahearn, 2000; Conyers et al., 2002; de Vries et al., 2005; Northup, George, Jones, Broussard, & Vollmer, 1996; Reyer & Sturmey, 2006; Schwartzman, Yu, & Martin, 2003). The object mode involves presenting the actual items during a preference assessment (e.g., food or activity items), whereas the picture mode involves presenting pictures of the items instead of the items themselves. In the spoken mode, the items are named or described vocally and sequentially (e.g., Conyers et al., 2002). In all modes, the individual receives the chosen item or activity.

Northup et al. (1996) compared spoken and picture preference assessments with verbal children. Results showed that for three of the four participants, vocal and picture presentations were equally effective in identifying high and low preference stimuli, and both procedures were more effective than third-party reinforcer surveys. Cohen-Almeida and colleagues (2000) demonstrated similar results with vocal and object presentations. They found that for 4 of the 6 participants, vocal and object presentations produced similar high-preference items. Not surprisingly, vocal presentations were more efficient in both studies.

Conyers et al. (2002) found that an individual's discrimination skills predicted the effectiveness of stimulus modes in preference assessments. In their research, participant's discrimination skills were determined using the *Assessment of Basic Learning Abilities* (ABLA) prior to conducting preference assessments using object, picture, and spoken presentations. Results indicated that: (1) participants who had demonstrated an auditory-

visual discrimination were able to select their most preferred option in all three modes; (2) participants who had demonstrated a visual-visual quasi-identity matching discrimination were able to select their most preferred option in object and picture modes, but not in the spoken mode; and (3) participants who had demonstrated a simple visual discrimination were able to select their most preferred option in the object mode only. Reyer and Sturmey (2006) and de Vries et al. (2005) have reported similar findings with both high and low preference work tasks and leisure activities, respectively.

Research on Preference and Reinforcer Effectiveness

Following a preference assessment, the most preferred stimuli have often been shown to be positive reinforcers (e.g., Carr et al., 2000; DeLeon, Frank, Gregory, & Allman, 2009; Graff & Larsen, 2011; Jerome & Sturmey, 2008; Mangum, Fredrick, Pabico, & Roane, 2012; Piazza, Fisher, Hagopian, Bowman, & Toole, 1996; Reed et al., 2009; Roscoe, Iwata, & Kahng, 1999). Reinforcer assessments typically involve identifying a target response (e.g., pressing a microswitch), conducting a no-reinforcement baseline, and administering each stimulus being evaluated as a consequence for the target response in alternating conditions. For example, the most preferred item identified by the preference assessment is presented following switch pressing in each condition and the least preferred item is presented following switch pressing in another condition. If the rate of responding in one reinforcement condition is higher than baseline and higher than the other reinforcement condition, then one may conclude that the stimulus is a more potent reinforcer relative to the stimulus in the other condition. Others have evaluated reinforcer efficacy in an ABAB design, where A is a no-reinforcement baseline condition and B is a reinforcement condition, prior to

conducting the preference assessment (e.g., Lee, Yu, Martin, & Martin, 2010). A stimulus is a positive reinforcer if the response rates during reinforcement conditions are higher than those found during the baseline conditions.

Schedule Requirements and Preference

A fixed-ratio (FR) schedule involves presenting a stimulus following the occurrence of a predetermined number of responses that remains constant throughout the experimental session. An FR 1 schedule (every response is followed by reinforcement) is often used to evaluate whether preferred stimuli identified in preference assessments are reinforcers. For example, Carr et al. (2000) reinforced correct target behaviors using high, medium, and low preference items on an FR 1 schedule in alternating conditions following a MSWO preference assessment. Lee et al. (2010) used an FR 1 schedule in an ABAB design to quantify reinforcer efficacy before conducting their preference assessments.

Some studies have examined how different schedule requirements influence preference. Tustin (1994) showed that the potency of reinforcers that are concurrently available might change due to increasing schedule requirements. He examined preference for reinforcers under different schedule requirements using a demand function (number of reinforcers earned at different ratio requirements) and a work function (number of responses emitted at different ratio requirements). For example, in one participant, the number of reinforcers earned for one of the reinforcers (auditory stimulation) declined as the response requirement increased from FR 1 to FR 20, while the number of reinforcers earned for the alternative reinforcer (visual) increased even though the response requirement remained constant at FR 5 for that reinforcer. The results were replicated for

this participant with visual stimuli and attention as reinforcers. In another participant, similar patterns (decline in preference for one reinforcer and increase in preference for another) were observed when the response requirement was increased for both reinforcers. DeLeon, Iwata, Goh, and Worsdell (1997) extended this work by studying the effects of increasing ratio requirements on preference between two different food items and between a food and a non-food item (leisure activity). When a food and a non-food item were concurrently available and each could be earned by pressing a panel at FR 1, participants earned each item approximately equally often. When FR requirements were increased across phases (FR 2, FR 5, FR 10, and FR 20), the number of each item earned remained approximately the same. However, when two food items were concurrently available, preference for one item increased as the FR requirement increased. That is, for two food items that are equally preferred at low schedule requirements, preference may change favoring one item when schedule requirements are increased.

The effects of schedule requirement have also been examined using a PR schedule. A PR schedule is similar to an FR schedule except that the number of responses increases systematically within a single session (Hodos, 1961). For example, in a PR schedule with a step size of 1, the number of responses required for the first reinforcement is 1 and increases by 1 for each successive reinforcement (i.e., FR 1 for the first reinforcement, FR 2 for the second, FR 3 for the third, and so on). The session is terminated after a pre-specified period of no response or after a predetermined session time has been reached (Roane, 2008). The duration of no response for session termination

(e.g., 1 min) is referred to as the *break-point criterion*, and the highest response ratio completed is referred to as the *breaking point* (Stafford & Branch, 1998).

Roane, Lerman, and Vorndran (2001) demonstrated that high preference stimuli identified using PR schedules functioned more effectively in decreasing problem behaviors than stimuli identified as low preference. Penrod, Wallace, and Dyer (2008) further showed that under the PR schedule, the difference in response rates maintained by the high preference stimuli versus the low preference stimuli was larger at higher ratios relative to lower ratios. Jerome and Sturmey (2008) also demonstrated that access to highly preferred caregivers produced higher rates of responding under PR schedule arrangements versus response rates for less preferred caregivers in individuals with developmental disabilities.

Francisco, Borrero, and Sy (2008) compared concurrent and single presentations of high and low preference stimuli under a PR schedule. During the concurrent condition, the high preference stimulus produced higher response rates than the low preference stimulus. During the single presentation condition, the low preference stimulus produced response rates comparable to the high preference stimulus during the concurrent condition. Glover, Roane, Kadey, and Grow (2008) also investigated the use of PR schedules to compare two reinforcers (one high and one low preference item) under single and concurrent presentations and found that all three of their participants responded more for the high preference item in both concurrent and single presentations.

In addition to being more effective in differentiating between low and high preference stimuli, PR schedules could be a potentially efficient method for assessing preferences (Roane et al., 2001). In addition, PR break points provide information about

the amount of work a reinforcer will support as opposed to another reinforcer or stimulus (DeLeon et al., 2011). PR schedules also identify the limit of responding or the amount of responding an individual will engage in before ceasing to respond. In contrast, FR schedules do not provide information about the individual's threshold of responding for a particular reinforcer. When using a PR schedule, the reinforcement ratio is increased incrementally to the point that the individual stops responding, thus providing us with information regarding the maximum amount of responding supported by that reinforcer.

Statement of the Problem

In preference studies using PR schedules, most have included only low and high preference stimuli (e.g., Francisco et al., 2008; Glover et al., 2008; Roane et al., 2001). DeLeon and colleagues (2009) used stimuli corresponding to three points on a preference hierarchy in their investigations of PR schedules. The authors showed that stimuli ranked as high preference following a PS preference assessment generally resulted in higher breaking points as compared to moderate and low preference stimuli, and that moderately preferred stimuli generally resulted in higher breaking points than low preference stimuli. Reed et al. (2009) evaluated PR performance across 6 items of different preference values, with one participant. The authors found that the two high preference stimuli generally resulted in higher breaking points than low preference stimuli. However, one of the low preference stimuli produced breaking points similar to high preference stimuli. Thus, it would be important to examine the correspondence between PR schedule performance (breaking points and response rates) and preference with more participants. Also, previous research has suggested that reinforcer potency derived from reinforcer assessments conducted under an FR 1 schedule corresponds well with preference from PS

assessments (Lee et al., 2010). As such, it would be important to examine whether a PR schedule would show similar characteristics. Therefore, the purpose of Experiment 1 was to evaluate the concordance between preference derived from PS assessments of 6 items and preference for those items presented under a PR schedule. Experiment 2 evaluated the potency of the high, medium, and low preference items identified by the PS assessments in Experiment 1, as reinforcers using a PR schedule.

EXPERIMENT 1

Method

Experiment 1 evaluated the concordance between preference identified by two PS assessments and breaking points and response rates obtained under a PR schedule with a step size of 1 and a 1-minute break-point criterion. The PS procedure was used to identify 6 items with different preference values. Each stimulus was then used to reinforce microswitch pressing using a PR schedule for three sessions. Two preference assessment sessions were conducted for each participant, one before and one after the PR sessions.

Participants and Setting

Four persons with developmental disabilities participated in this study. Participant 1 was a 43-year-old male, Participant 2 a 45-year-old male, Participant 3 a 54-year-old female, and Participant 4 a 34-year-old male. Participants 1 and 2 had both passed up to level 4 (two-choice visual quasi-identity conditional discrimination) of the ABLA and Participants 3 and 4 had passed up to level 3 (two-choice simple visual discrimination). Participants were recruited from a community resource and residential facility that provides for the needs of individuals diagnosed with a developmental disability. All participants had a diagnosis of severe or profound intellectual disabilities with limited or

no speech. Ethics approval was received from the University of Manitoba Psychology/Sociology Research Ethics Board before the study began. Written informed consent for participation was obtained from the Substitute Decision Maker of each participant and the participant's assent to take part in the study was evaluated based on their willingness to work with the Experimenter at each contact.

All sessions were conducted individually in an assessment room at the St. Amant Research Centre. The assessment room was furnished with a table and two chairs. The experimenter and participant sat at opposite sides of the table facing one another during all sessions. An observer was present during some sessions to assess reliability.

Preference Assessment Items and Equipment

Items used during reinforcer and preference assessments consisted of small bite-sized edibles for Participants 1, 2, and 4 and leisure items for Participant 3 due to dietary restrictions. Each participant's caregiver was asked to nominate a list of items for each participant, and items were drawn from this pool for reinforcer assessments. Items that the participant was known to dislike were not included. A stopwatch was used to time trial intervals.

For reinforcer assessments, a 6-cm diameter round microswitch was used for Participants 1, 2, and 4. A 12-cm diameter round microswitch was used for Participant 3 as she had difficulty pressing the smaller switch. The microswitches required 2-3 g of force to activate and produced an audible "click" when depressed. The microswitch was connected to a portable computer through an X-Keys® USB Switch Interface. A computer program, written in Python Programming Language, recorded the occurrence of each switch press. The program also alerted the experimenter via an audible "beep" and a

written message on the computer screen when each ratio had been completed during the PR schedule.

Response Definitions and Interobserver Reliability Checks

Preference assessments. During PS preference assessments (described below), the target responses on each assessment trial included one of the following: (a) a rejection response, defined as pushing a stimulus away, within 8 s after the stimuli were presented; (b) an approach response, defined as pointing to or touching a stimulus, without rejecting it, within 8 s after the stimuli were presented; or (c) neither a rejection nor an approach response occurring within 8 s after the stimuli were presented. During an interobserver reliability check, an observer independently recorded a participant's response on each preference assessment trial. A trial was scored as an agreement if both the observer and experimenter recorded the same response; otherwise, it was scored as a disagreement. Percent agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100% (Martin & Pear, 2011; Miltenberger, 2011). Interobserver reliability checks were conducted for Participants 2, 3, and 4 during 50%, 100%, and 100% of preference assessment sessions. Reliability checks were not conducted for Participant 1 because an observer was not available for either of the two preference assessment sessions. The mean percentage of agreement across all observed sessions for the other participants was 100%.

Progressive ratio reinforcement. During PR sessions (described below), the target behavior was defined as the participant pressing the microswitch. Since each response was recorded automatically by the computer program, no interobserver reliability checks on switch pressing were conducted. However, the recording equipment

and the computer program were checked for accuracy before and approximately half way during the experiment, and the equipment recorded responses correctly on both occasions.

Preference Assessment Procedure

Preference for 6 stimuli from the list of items nominated by each participant's caregiver was assessed using a PS procedure (Fisher et al., 1992). During an assessment, two stimuli were presented concurrently on each trial. Across trials, each stimulus was paired with every other stimulus for two trials in counterbalanced positions. Preference for each item was defined as the percentage of trials an item was chosen based on the number of trials that item was available.

On each trial, two items were placed side by side, approximately 10 cm apart, on the table in front of the participant. The participant was instructed to look at each item in succession, and then asked to select one. If the participant approached one item within 8 s of the request, he/she was given the selected item immediately to consume (for food items) or for engagement for 15 s (for leisure items). If the participant approached neither item within 8 s, the items were removed from the table and the next trial was presented. If the participant approached both items simultaneously, he/she was blocked gently, the items were removed from the table for 5 s, and the trial was repeated. The percentage of trials each item was chosen out of trials that item was available was computed as an index of preference.

Progressive Ratio Reinforcement Procedure

The 6 stimuli identified by the first PS assessment were evaluated, one at a time, using a PR reinforcement schedule for three sessions. Sessions were separated by at least

24 hours and began with the participant sitting at a table with a microswitch within arm's reach. A predetermined session termination length of 30 minutes was imposed and sessions were terminated if 1 minute elapsed with no target response (break-point criterion), with the exception of Participant 3 in which case sessions were terminated when she indicated that she was "all done". At the beginning of each session, the experimenter modeled the switch pressing response and physically guided the participant to engage in the behavior once. Afterwards, the experimenter began the session by instructing the participant to press the microswitch. The instruction was repeated once every 60 s during the session. Praise and the stimulus being evaluated were provided to the participant following each switch press. The response ratio for reinforcement began with 1 response and increased by 1 after the completion of each ratio (e.g., FR 2 for the second reinforcement, FR 3 for the third reinforcement, and so on). The switch was removed from the table during the reinforcement interval and reintroduced when the participant had finished eating or engaging with the stimulus. Praise for an appropriate behavior other than the target response was given at 30 s intervals throughout the session (e.g., nice sitting). However, because Participant 3 was highly responsive to and easily distracted by interactions with the experimenter, no instruction to press the switch or praise for non-target responses was given for her throughout the session.

Time spent consuming the food item or interacting with the item during reinforcement was subtracted from the session time and responses per minute were computed as an index of response rate. The highest response ratio completed during each session was the breaking point.

Procedural Integrity

Preference assessments. Before data collection began, all researchers and observers achieved a criterion performance of 100% correct for one practice session for preference assessment. During a procedural integrity check for preference assessments, an observer independently assessed the experimenter's adherence to the procedures by scoring whether the following target behaviors occurred correctly on each trial: placement of two items, instruction for the participant to look at each item, instruction for the participant to select one item, and provision of the correct consequence (i.e., provision of the selected item for consumption, removal of items after no approach response, or blocking the approach to both items simultaneously). The percentage of steps carried out correctly per session was computed. Procedural integrity checks were conducted during 50% of all sessions for Participants 2 through 4. The mean percentage of steps carried out correctly across observed sessions for all participants was 100%. Procedural integrity checks were not conducted for Participant 1, unfortunately, because an observer was not available for the two preference assessments.

PR reinforcement. Before data collection began, all researchers and observers achieved a criterion performance of 100% correct for one practice session for PR reinforcement. During a procedural integrity check, an observer independently assessed the experimenter's adherence to the procedures by scoring whether the following target behaviors occurred correctly during a session: modeled and guided the target response once at the beginning of the session, provided the correct consequence after each ratio had been completed (stopped the timer, removed microswitch, presented the reinforcer),

provided an instruction to press the microswitch once per minute during the session (except for Participant 3), and provided praise for an alternate behavior every 30 s during the session (except for Participant 3). The percentage of steps carried out correctly per opportunity was computed. Procedural integrity checks were conducted during 33%, 33%, 100%, and 94% of all 18 sessions (3 sessions x 6 items) for Participants 1, 2, 3, and 4, respectively. The mean percentage of steps carried out correctly across observed sessions for all participants was 99.5% ranging from 99% to 100%.

Results and Discussion

Figure 1 shows the percent preference during the first and second PS assessments for each item, the mean preference across the two assessments, and the correlation between the two assessments for each participant. Across all participants, the three top ranked or most preferred items during the first preference assessment remained the three top ranked items during the second preference assessment, and no item changed by more than 2 ranks. Correlations between the two preference assessments were high for all participants (Pearson r = .84, .87, .91, and .73 for Participants 1, 2, 3, and 4, respectively). Therefore, the mean preference between assessments was used as the comparison to evaluate the breaking points and response rates during PR sessions.

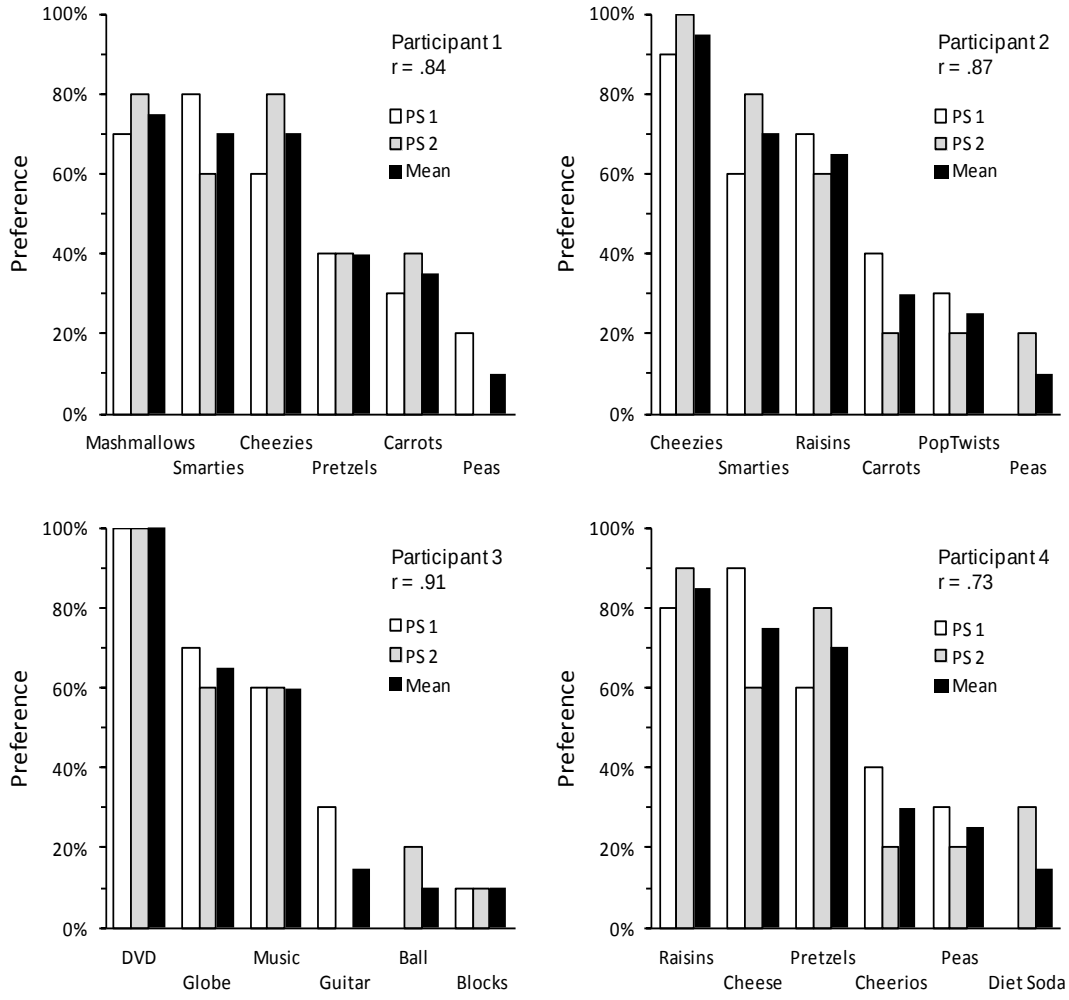

Figure 1. Percent preference during the first (white bar) and second (grey bar) paired-stimulus (PS1 and PS2) assessments for each item, the mean preference (black bar) across the two assessments, and the correlation (*r*) between the two assessments.

Table 1 shows the Pearson correlations between breaking point and response rate during each PR session across the 6 items. For all participants, high correlations were observed between breaking points and responses per minute, ranging from .56 to 1.00, with 9 of the 12 correlations being statistically significant ($p < .05$). That is, higher breaking points were associated with higher response rates.

Figure 2 displays the mean breaking points and responses per minute for Participants 1 (top row) and 2 (bottom row) for each of the six stimuli during the first PR reinforcement session (left graph), the average for the first and second sessions (middle graph), and the average for all three PR sessions (right graph). Mean percent preference (circles and dashed lines) of the two PS assessments are plotted against the left y-axis in all graphs. Breaking points (triangles) and response rates (squares) are plotted against the right y-axis in each graph. Results for Participants 3 and 4 are shown in Figure 3. Stimuli were ordered from the highest to the lowest preference values along the x-axis. Correlations between preference and breaking point across the items and between preference and response rate across the items are shown within each graph.

Concordance between preference and response rates was generally weak after the first PR session (graphs in the left column), and increased after the second and/or the third PR sessions, except for Participant 2. After the first PR session, correlations ranged from -.04 (Participant 4, Figure 3) to .74 (Participant 3, Figure 3) and a mean of .28 across all participants. Concordance between preference and response rates improved slightly after two PR sessions (graphs in the middle column), with correlations ranging from -.07 (Participant 2, Figure 2) to .68 (Participant 3, Figure 3) and a mean of .41

Table 1

Pearson Correlations between Breaking Points and Response Rates across the 6 Items

	Participant 1	Participant 2	Participant 3	Participant 4
Progressive Ratio Session 1	.79*	1.00***	.95**	.95**
Progressive Ratio Session 2	.83*	.99***	.81*	.82*
Progressive Ratio Session 3	.69	.87*	.70	.56

* $p < .05$, ** $p < .01$, *** $p < .001$

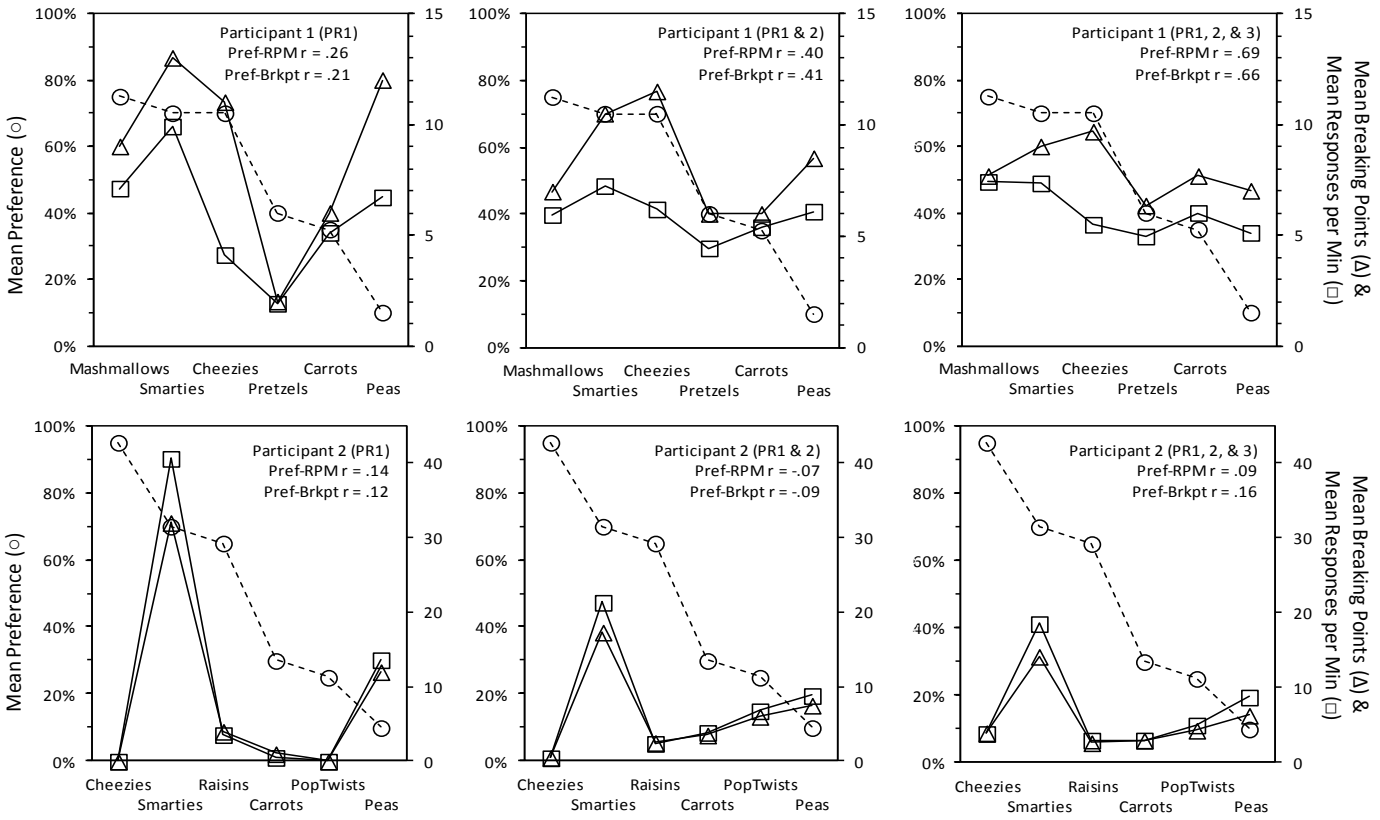

Figure 2. Breaking points (Brkpt) (triangles, right y-axis), and response rates (RPM) (squares, right y-axis) for each item for Participants 1 (top row) and 2 (bottom row) are shown during the first progressive ratio reinforcement session (PR1, left column). Mean breaking points and mean response rates for the first and second PR sessions are shown in the middle column, and mean breaking points and mean response rates for all three PR sessions are shown in the right column. Mean percent preference (circles, left y-axis) of the two paired-stimulus assessments is shown in all graphs for comparison.

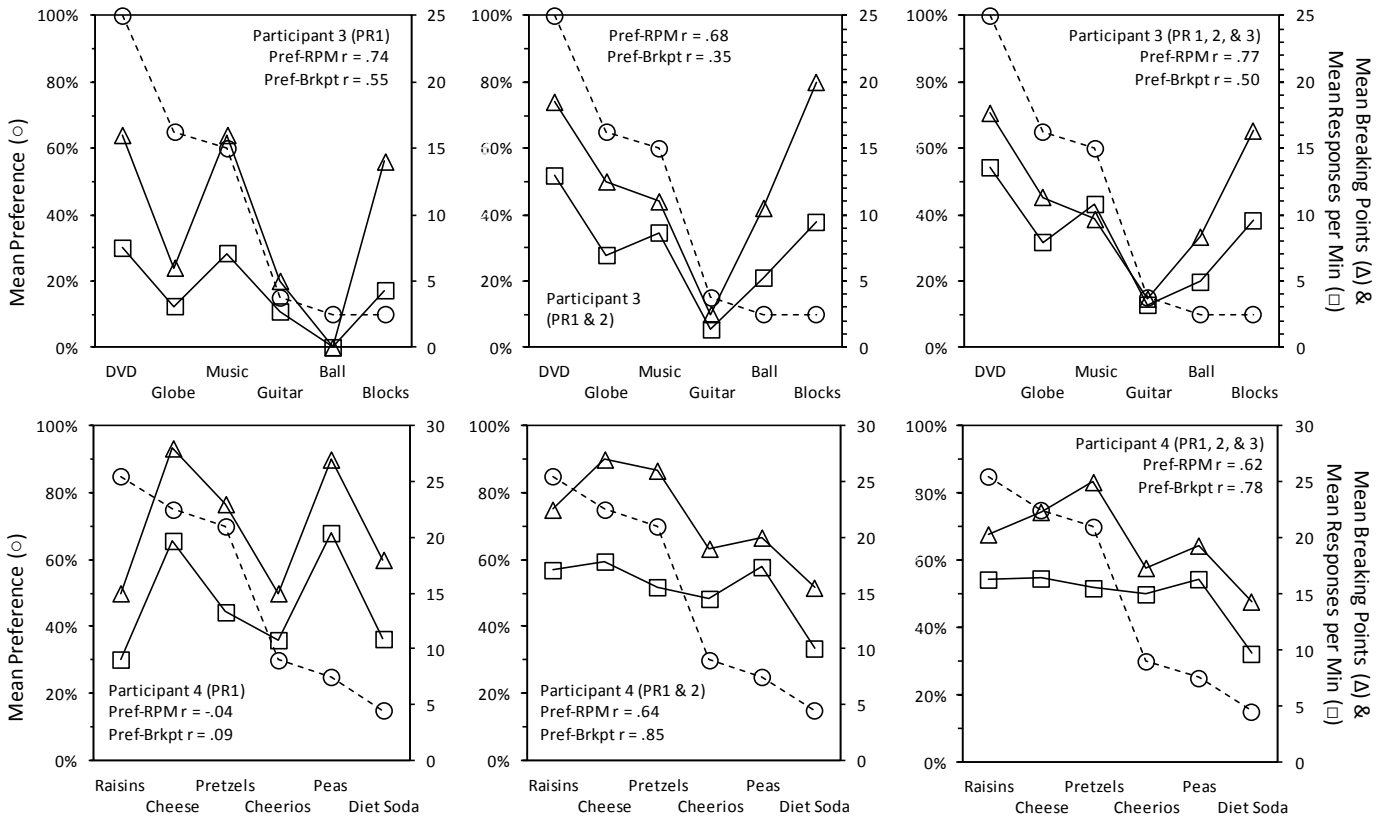

Figure 3. Breaking points (Brkpt) (triangles, right y-axis), and response rates (RPM) (squares, right y-axis) for each item for Participants 3 (top row) and 4 (bottom row) are shown during the first progressive ratio reinforcement session (PR1, left column). Mean breaking points and mean response rates for the first and second PR sessions are shown in the middle column, and mean breaking points and mean response rates for all three PR sessions are shown in the right column. Mean percent preference (circles, left y-axis) of the two paired-stimulus assessments is shown in all graphs for comparison.

across all participants. Concordance improved further after three PR sessions (graphs in the right column), with correlations ranging from .09 (Participant 2, Figure 2) to .77 (Participant 3, Figure 3) and a mean of .54 across all participants. Participant 2 was the exception to this trend.

Concordance between preference and breaking points was also weak and showed a similar trend to that described above for response rates, except again for Participant 2. After the first PR session (graphs in the left column), correlations ranged from .09 (Participant 4, Figure 3) to .55 (Participant 3, Figure 3) and averaged .24 across all participants. Concordance improved slightly after two PR sessions (graphs in the middle column), with correlations ranging from -.09 (Participant 2, Figure 2) to .85 (Participant 4, Figure 3) and a mean of .38 across all participants. Concordance improved further but only slightly after three PR sessions (graphs in the right column), with correlations ranging from .16 (Participant 2, Figure 2) to .78 (Participant 4, Figure 3) and a mean of .53 across all participants.

The similar trends observed for response rates and breaking points were not surprising given that response rates and breaking points were highly correlated (Table 1). Although concordance improved as the number of PR sessions increased, the improvement was modest after 3 sessions. Additional PR sessions might have yielded further improvement.

For Participants 1, 3, and 4, it appears that the response rates/breaking points for the less preferred items were higher than expected relative to the more preferred items. This finding has been reported by others (Francisco et al., 2008; Graff & Larsen, 2001; Lee et al., 2010; Mangum et al., 2012; Penrod et al., 2008; Roscoe et al., 1999; Taravella,

Lerman, Contrucci, & Roane, 2000). Although this was also true for Participant 2, his overall results were somewhat puzzling. His response rate/breaking point for the second most preferred item (Smarties®) was the highest among the 6 items and his responding for the other items was generally low and not well differentiated. Most surprisingly, he barely responded at all to his most preferred item (Cheezies®) during the first two PR sessions and only infrequently during the third session, even though that item was his most preferred item during the two preference assessments. Since the two preference assessments were conducted immediately before and after the PR sessions, respectively, this would suggest that his preference might have changed during the time when the PR sessions were being conducted.

The demand function (number of reinforcers earned at each ratio value) and the work function (number of responses emitted at each ratio value), as described by Tustin (1994), were examined for each item. Because the results for individual items were variable, the 6 items were aggregated into two groups of high (top 3) and low preference (bottom 3) items to facilitate visual inspection. Figure 4 shows the number of reinforcers earned (top graph) and number of responses emitted (bottom graph) at each ratio value for the high and low preference items for each participant. Participant 1 began by earning the maximum number of available reinforcers for both the high and low preference items at FR 1 and FR 2. However, number of low preference reinforcers earned began to decline at FR 3 and reached zero at FR 13. Number of high preference reinforcers earned did not decline until FR 6. The rate of decline (slope), once it began, was fairly similar for both high and low preference items. The difference is also reflected in the work

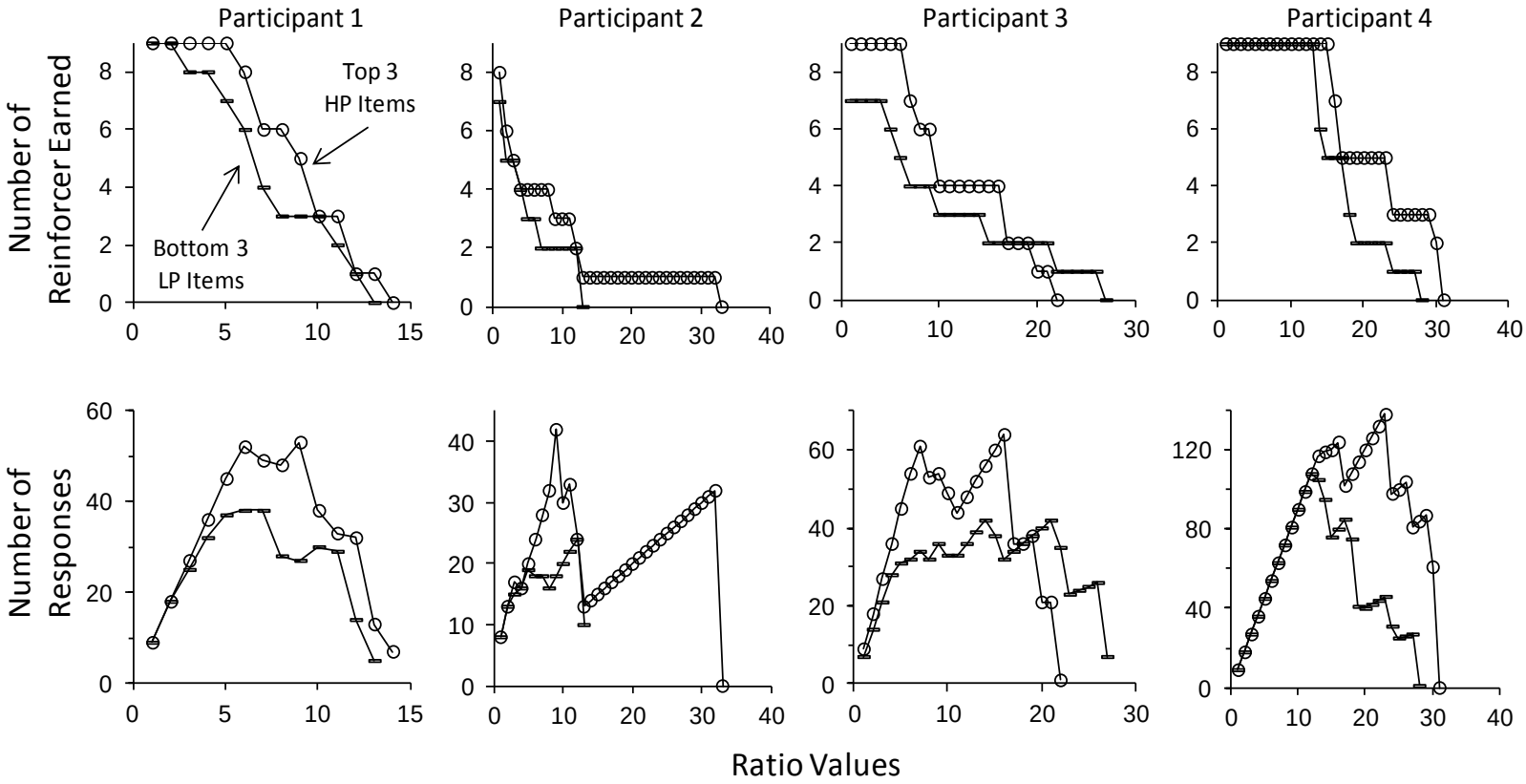

Figure 4. Number of reinforcers earned (top row) and number of responses (bottom row) emitted at each ratio value aggregated for the top 3 high preference (HP) items and for the bottom 3 low preference (LP) items for each participant.

function (bottom graph) in that responding was similar for both high and low preference items at low schedule ratios, favored the high preference items as ratio values increased, and was followed by a decline for both high and low preference items. Note that decline in responding occurred earlier for low preference items (at FR 8) than for high preference items (at FR 10), but a larger decline occurred for the high preference items. The difference in responding between high and low preference items was smaller beyond FR 10. Participant 2 began by earning a slightly higher number of available reinforcers for the high preference items at FR 1. Number of reinforcers earned began to decline at FR 2 for both high and low preference items. More reinforcers were earned and more responses were emitted for high preference items than low preference items when ratio values were between FR 4 and FR 11. The difference disappeared after FR 11, except for one high preference item that reached a breaking point of 32 in one of the PR sessions. The demand function for Participant 3 again showed that the maximum number of available reinforcers was earned for high preference items while the number of reinforcers earned for low preference items was lower at FR 1. Reinforcers earned declined for high and low preference items beginning at FR 7 and FR 5, respectively. More reinforcers were earned for high preference items than for low preference items when ratio values were between FR 1 and FR 16 (top row), and more responses were emitted for high preference items when ratio values were between FR 1 and FR 16 (bottom row). Beyond FR 16, the number of reinforcers earned for HP items declined to zero more quickly than the low preference items. The higher breaking point observed for the LP items was accounted for primarily by one LP item (blocks, see Figure 3). With respect to work function, responding to both high and low preference items was similar at

low FR values (FR 1 to FR 3), with HP items being slightly higher. At FR 4, number of responses to the high and low preference items began to diverge. Despite some variability in responding to the HP items, responding to the HP items remained higher than the LP items until FR 16, followed by a sharp decline at FR 17 where responding to the HP items converged with the LP items. Responding to the HP items declined to near zero quickly from FR 19 to FR 22. Responding for the LP items was relatively stable until FR 22 before it began to decrease. Unlike Participants 1 and 2, decline in responding occurred earlier for high preference items than for low preference items. Participant 4 began by earning the maximum number of available reinforcers for both the high and low preference items until FR 16 and FR 14, respectively. More reinforcers were earned and more responses were emitted for high preference items than low preference items when ratio values were between FR 13 and FR 31. Like Participants 1 and 2, a decline in responding occurred earlier for low preference items than for high preference items.

In summary, in all cases in Figure 4, there was little or no difference in reinforcers earned (except perhaps Participant 3) between high and low preference items at lower ratio values. Responding for the HP items increased, relative to the LP items, as ratio values increased and the difference in responding between the HP and LP items tended to decrease as ratio values further increased. All participants showed higher breaking points with the higher preference items (in the top figures) except Participant 3. The higher breaking point observed in the LP items for Participant 3 was the result of one LP item (Blocks, see Figure 3). The demand and work functions for individual items were too variable and did not yield consistent differences similar to the mean preference hierarchy obtained from the PS assessments. However, a more consistent difference emerged

between high and low preference items when the top 3 and bottom 3 items were aggregated.

It is worth noting that the items in this study were exposed to the three PR sessions individually, whereas previous research had used alternating conditions (e.g., Reed et al., 2009; Roane et al., 2001). Rapid alternations between items in a multi-element design may promote discrimination and produce larger differences between items. The step size of the PR schedule in this study was small and constant (i.e., ratio increased by 1 for each successive reinforcement), whereas the step size increased more quickly in previous research. For example, the ratios in the Reed et al. study were FR 1, FR 1, FR 2, FR 2, FR 5, FR 5, FR 10, FR 10, FR 20, FR 20, FR 30, FR 30). Although Experiment 1 suggests there is a weak to moderate concordance between preference derived from the PS assessment and relative reinforcer efficacy derived from the breaking point and response rate of the PR schedule, the true reinforcing efficacy of the items had not been demonstrated against a baseline without the reinforcer. Experiment 2 was conducted to evaluate the reinforcing efficacy of the stimuli.

EXPERIMENT 2

Method

The purpose of Experiment 2 was to evaluate the reinforcing efficacy of the reinforcing items identified in Experiment 1. Three items roughly corresponding to high, medium, and low preference were evaluated for each participant. Experiment 2 was conducted immediately after the completion of Experiment 1.

Participants and Setting

The same participants from Experiment 1 took part in Experiment 2. The research setting was the same as in Experiment 1.

Experimental Design, Items, and Equipment

For each participant, an ABAB reversal design (Johnston & Pennypacker, 2009; Kazdin, 1982; Martin & Pear, 2011) was used. During baseline (A phases), praise was given for switch pressing under a PR schedule with a step size of 1 and a break-point criterion of 1 minute. During the reinforcement (B phases), praise and the item being evaluated were presented for switch pressing under the same reinforcement schedule.

Based on the results of the mean preference obtained in Experiment 1, three items roughly corresponding to high, medium, and a low preference were evaluated for each participant (exceptions are noted below). The same equipment described in Experiment 1 was used to conduct the PR sessions.

Baseline Phase

The procedures during baseline sessions were identical to PR sessions conducted in Experiment 1 (see section on Progressive Ratio Reinforcement Procedure in Experiment 1) except that praise alone was given for switch pressing.

Reinforcement Phase

The procedures during reinforcement sessions were identical to the procedures used during PR sessions conducted in Experiment 1 (see section on Progressive Ratio Reinforcement Procedure in Experiment 1). That is, both praise and the item (food or leisure) were given for switch pressing.

Because of the variability of Participant 2's performance during the evaluation of the first item, steps were taken to increase discrimination between baseline and reinforcement sessions by turning the table 180 degrees during this phase for the second and third items. Moreover, the packaging for the food item being evaluated was also kept on the table within view to increase its saliency. The table manipulation (turning it 180 degrees) was added for Participant 4 for all three items as he was just beginning Experiment 2 at that time.

Reinforcer Efficacy

Reinforcing efficacy was defined as the mean percentage increase in response rates from baseline to reinforcement phases (Lee et al., 2010). This measure was derived by: (1) computing the mean baseline responses per minute across 6 sessions, using the last three sessions from each baseline (A) phase (all sessions were used if a phase contained only two sessions); (2) computing the mean reinforcement response rates across 6 sessions, using the last three sessions from each reinforcement (B) phase (all sessions were used if a phase contained only two sessions); (3) subtracting the mean baseline response rate from the mean reinforcement response rate; and (4) dividing the difference by the mean baseline response rate. For example, if the responses per minute during the last three sessions of the first baseline phase were 4, 4, and 3 and the responses

per minute during the last three sessions of the second baseline phase were 5, 8, and 6, the mean baseline response rate would be the average of these 6 sessions (30/6 = 5). If the responses per minute during the last three sessions of the first reinforcement phase were 10, 12, and 14 and the responses per minute during the last three sessions of the second reinforcement phase were 18, 14, and 20, the mean reinforcement response rate would be the average of these 6 sessions (88/6 = 14.7). Therefore, the computed reinforcer efficacy would be (14.7 − 5) / 5 x 100% or 194%.

Procedural Integrity Checks

Since the participants' responses were automatically recorded by the computer program, reliability checks on their responses were not conducted (see reliability checks on the computer equipment described in Experiment 1). However, procedural integrity checks were performed for all participants during baseline and reinforcement phases (see Experiment 1 for a description of the Experimenter's behaviors). Checks were conducted during 87%, 89%, 92%, and 100% of all sessions for Participants 1 through 4, respectively. The mean percentage of steps carried out correctly across observed sessions for all participants was 99.8% and ranged from 99.7% to 100%.

Results and Discussion

Figure 5 shows the breaking points and responses per minute per session during baseline (praise only) and reinforcement (praise plus edible) sessions for Participant 1. In Experiment 1, response rates and breaking points were highly correlated. However, response rates became stable more quickly than breaking points, thus response rate was used as the metric of reinforcing efficacy. Two high preference items (Cheezies® and Smarties® both had mean preferences of 70%) and a low preference item (peas had a

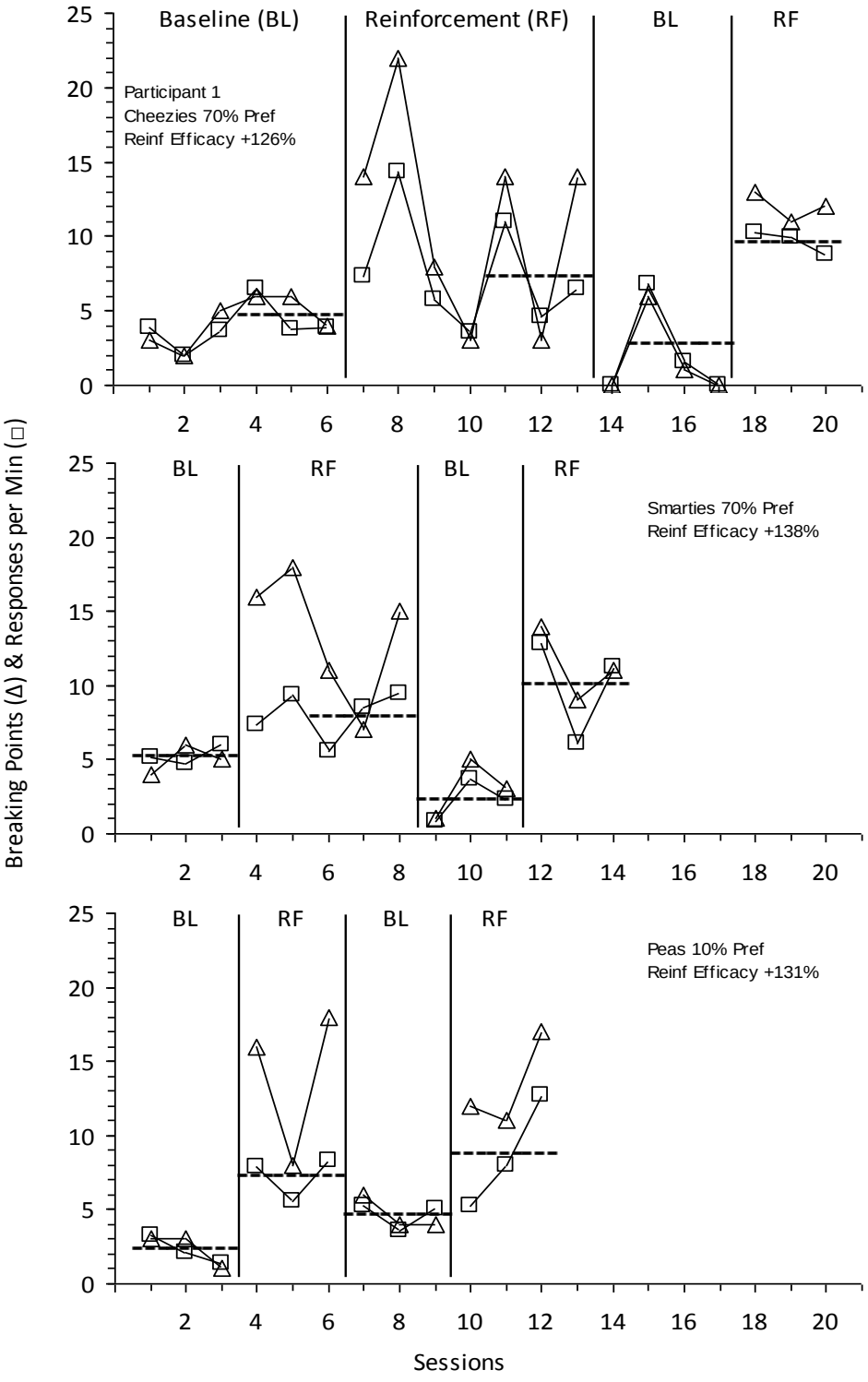

Figure 5. Breaking points (triangles) and responses per minute (squares) per session during baseline (praise only) and reinforcement (praise plus edible) sessions for Participant 1 for high and low preference items. The percent preference and computed reinforcer efficacy for each item are shown in each graph.

10% mean preference) were evaluated. The dotted horizontal line in each phase represents the mean responses per minute for the last three sessions. Despite some variability, all three items were positive reinforcers, with computed efficacies of 126% for Cheezies®, 138% for Smarties®, and 131% for peas. Peas, which was the least preferred item, was as efficacious as the two more preferred items. Responding to Cheezies® was 1.26 times higher in reinforcement phases as compared to baseline phases, while responding was 1.38 and 1.31 times higher for Smarties® and Peas, respectively.

Figure 6 shows the breaking points and responses per minute per session during baseline (praise only) and reinforcement (praise plus edible) sessions for Participant 2. Cheezies®, at 95% mean preference, Smarties®, at 70% mean preference, and peas, at 10% mean preference, were evaluated. Cheezies® was a positive reinforcer with a computed efficacy of 89% (i.e., increased responding .89 times higher than in baseline). Although the reinforcer efficacy for Smarties® was quite high (215% or 2.15 times higher than baseline), responding was variable with much overlap between baseline and reinforcement phases when all data within each phase were considered. Thus, its computed reinforcer efficacy may be questionable. Peas, the least preferred item, appeared to be a punisher in that it yielded slightly lower levels of responding during reinforcement phases than in baseline and had an efficacy of -90%. In other words, responding during baseline was .90 times higher than in reinforcement phases.

Figure 7 shows the breaking points and responses per minute per session during baseline (praise only) and reinforcement (praise plus leisure item) sessions for Participant 3. The most preferred (DVD at 100% mean preference) and the least preferred items

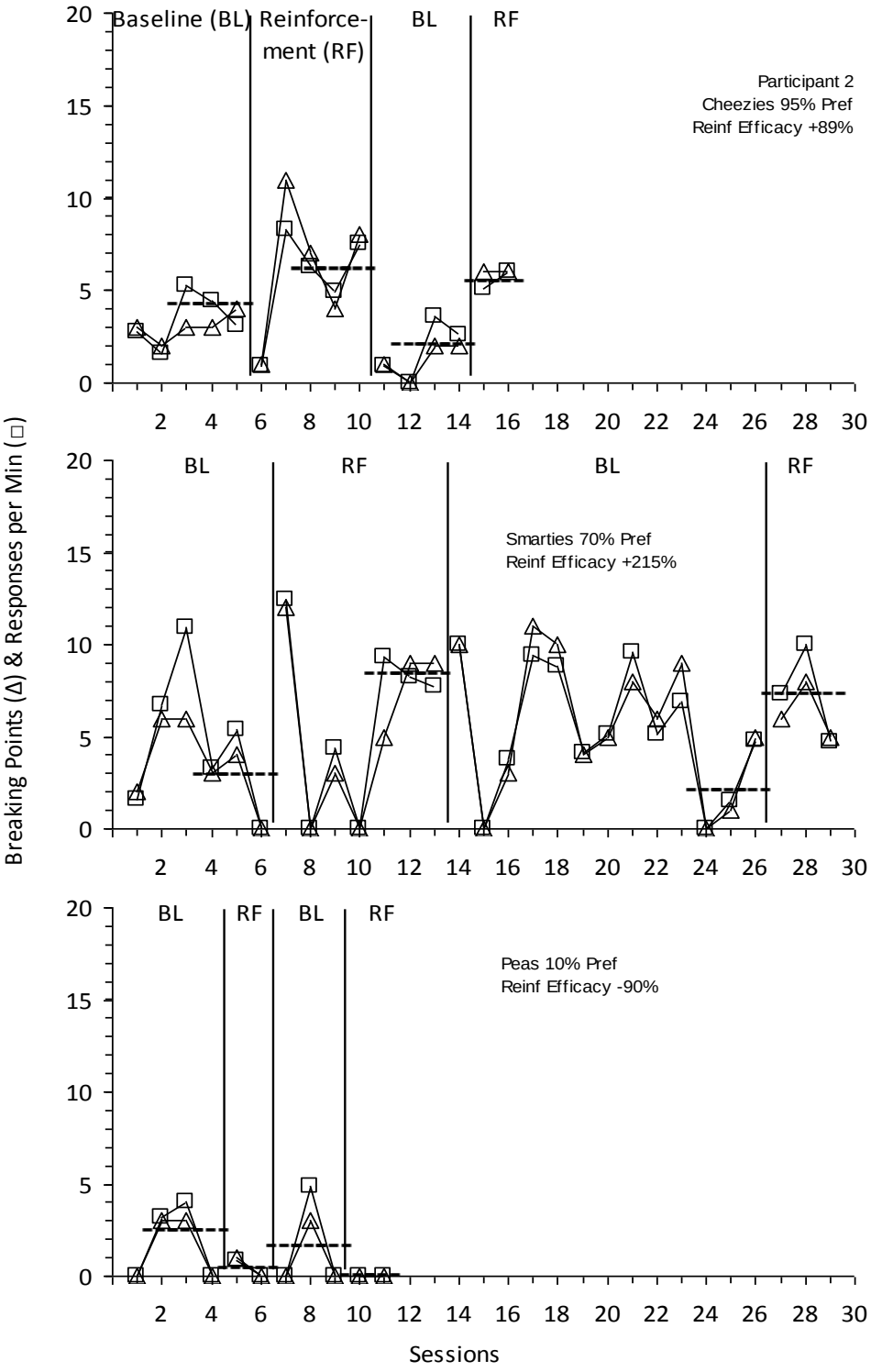

Figure 6. Breaking points (triangles) and responses per minute (squares) per session during baseline (praise only) and reinforcement (praise plus edible) sessions for Participant 2 for high, medium, and low preference items. The percent preference and computed reinforcer efficacy for each item are shown in each graph.

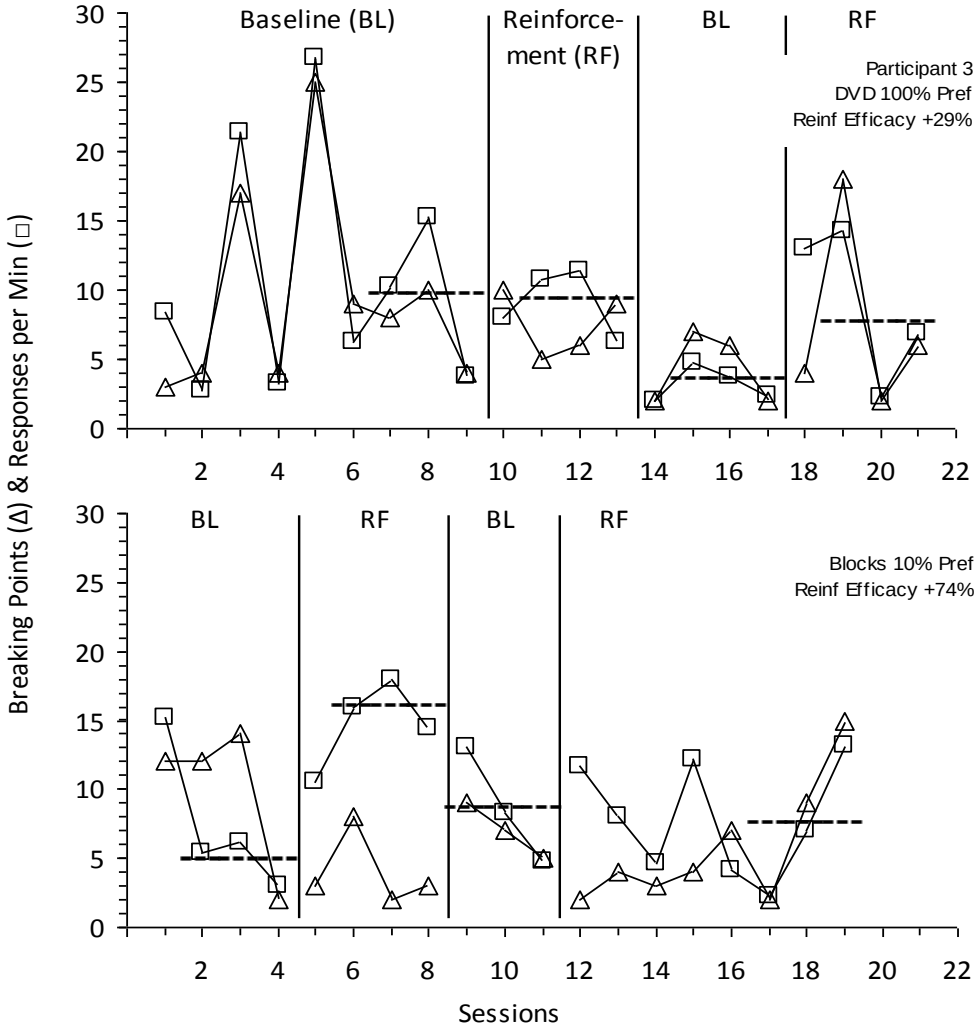

Figure 7. Breaking points (triangles) and responses per minute (squares) per session during baseline (praise only) and reinforcement (praise plus leisure item) sessions for Participant 3 for the high and low preference items. The percent preference and computed reinforcer efficacy for each item are shown in each graph.

(blocks at 10% mean preference) were evaluated. A medium preference item was not evaluated because Participant 3 fell ill and had to withdraw from the study after the first two items had been evaluated. The DVD was a positive reinforcer with an efficacy of 29% (i.e., responding during reinforcement phases was .29 times higher than in baseline). Reinforcer efficacy for DVDs was lower than expected given that this was a highly preferred item for Participant 3. However, responding during the initial baseline phase and the last reinforcement phase was highly variable. Thus, its computed reinforcer efficacy may be questionable. The reinforcer efficacy for blocks, the least preferred item, was 74% (i.e., responding during reinforcement phases was .74 times higher than in baseline), which was higher than the most preferred item, although responding was also quite variable for this item.

Figure 8 shows the breaking points and responses per minute per session during baseline (praise only) and reinforcement (praise plus edibles) sessions for Participant 4. The most preferred (raisins at 85% mean preference), a moderately preferred (Cheerios® at 30% mean preference), and the least preferred items (diet soda at 15% mean preference) were evaluated. Raisins were a positive reinforcer with an efficacy of 265% (i.e., responding during reinforcement phases was 2.65 times higher than in baseline). The reinforcer efficacy for Cheerios® was 340% (i.e., responding during reinforcement phases was 3.4 times higher than in baseline). Diet soda, the least preferred item, also functioned as a positive reinforcer and had an efficacy of 55% (i.e., responding during reinforcement phases was .55 times higher than in baseline). The most and moderately preferred items were both more efficacious than the least preferred item. However, the

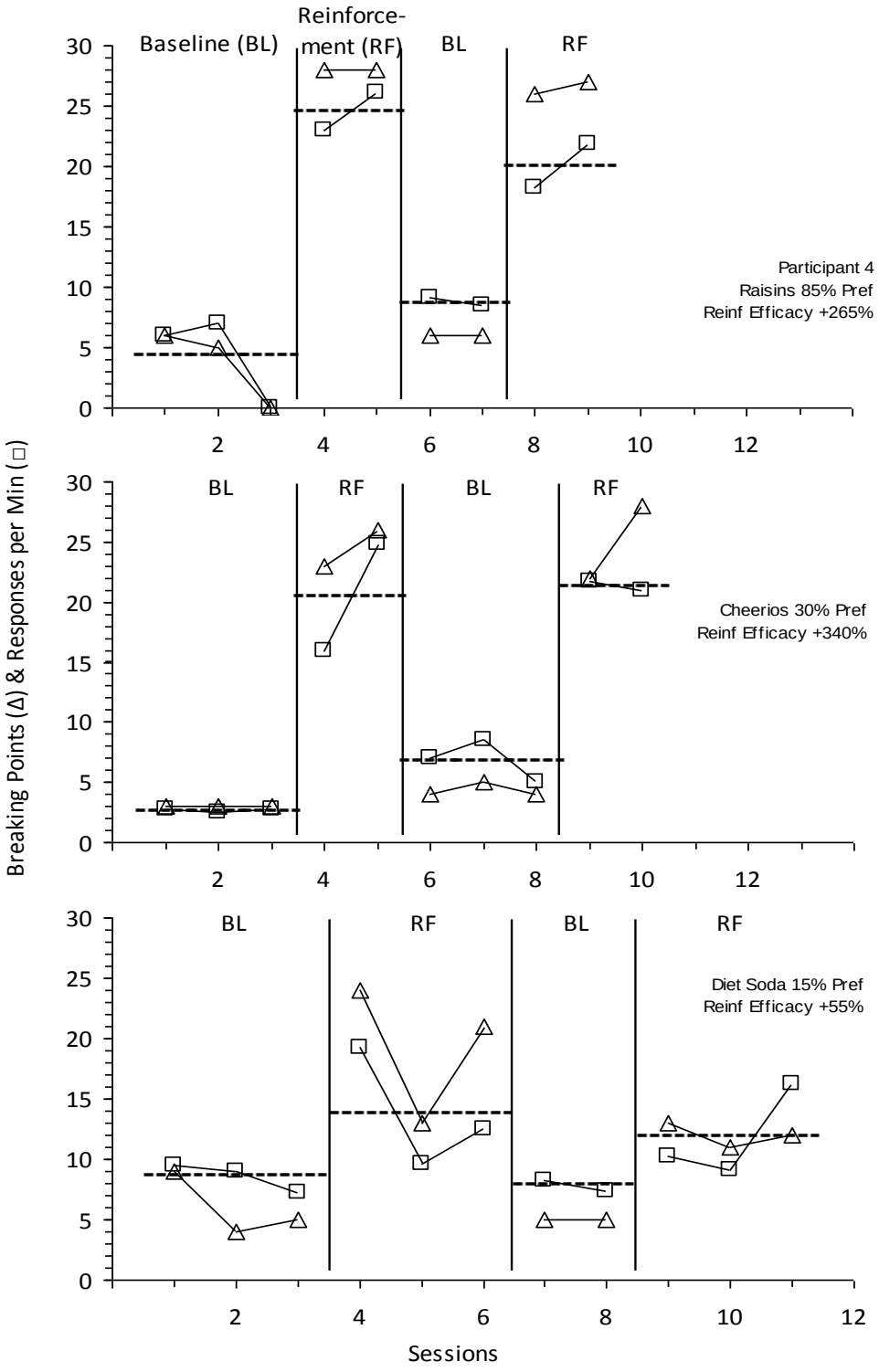

Figure 8. Breaking points (triangles) and responses per minute (squares) per session during baseline (praise only) and reinforcement (praise plus edible) sessions for Participant 4 for high, medium, and low preference items. The percent preference and computed reinforcer efficacy for each item are shown in each graph.

moderately preferred item was more efficacious than the most preferred item. Again, breaking points and response rates were quite similar across most baseline and reinforcement sessions.

The results of Experiment 2 are summarized in Table 2. For each participant, the items evaluated, their mean percent preference and preference rankings from Experiment 1, and their computed reinforcer efficacies and rankings from Experiment 2 are presented. All but one item were shown to be reinforcers. Overall, correspondence between relative preference and reinforcer efficacy was poor. The weak concordance may be due to the single-schedule arrangement. Previous studies have shown that the reinforcer efficacy of low preference items can be influenced by concurrent-versus single-schedule presentations. For example, Roscoe et al. (1999) found that when high and low preference items were concurrently available, participants responded more frequently to the high preference items. However, when the low preference items were presented as the only available option, response rates for the low preference items were as high as those observed for the high preference items during the concurrent presentation.

Table 2

Preference and Reinforcer Efficacy for Stimuli Evaluated in Experiment 2.

	Preference % (Rank)[a]	Reinforcer Efficacy % (Rank)[b]
Participant 1		
Cheezies®	70% (2.5)	126% (3)
Smarties®	70% (2.5)	138% (1)
Peas	10% (6)	131% (2)
Participant 2		
Cheezies®	95% (1)	89% (2)
Smarties®	70% (2)	215% (1)
Peas	10% (6)	-90% (3)
Participant 3		
DVD	100% (1)	29% (2)
Blocks	10% (6)	74% (1)
Participant 4		
Raisins	85% (1)	265% (2)
Cheerios®	30% (4)	340% (1)
Diet Soda	15% (6)	55% (3)

[a] Mean percent preference and ranking from the two preference assessments of the 6 items.

[b] Ranking of the items evaluated.

GENERAL DISCUSSION

In Experiment 1, the concordance between preference identified by a PS assessment and breaking points and response rates obtained under a PR schedule for six stimuli was examined. Experiment 2 evaluated the reinforcer efficacy of items roughly corresponding to items with high, medium, and low mean preference identified in Experiment 1 using a PR schedule of reinforcement.

For all participants, both administrations of the PS preference assessment produced very similar preference hierarchies and the correlation between the two assessments was high. Although a strong relationship was also observed between breaking points and responses per minute during PR sessions, their correlations with preference derived from the PS assessments were low after one PR session and increased to a moderate level after three PR sessions. This is consistent with the findings of the demand and work functions analyses, which showed higher number of reinforcers earned and higher number of responses only after the three most preferred and three least preferred items were aggregated.

Results from Experiment 2 demonstrated that for all participants, all except one item evaluated functioned as a positive reinforcer (i.e., produced increased response rate over baseline). Reinforcer efficacy, however, did not correspond with preference systematically.

Results of the current study support the findings of previous research which has shown that stimuli identified as high preference items generally function as reinforcers under PR or increasing FR schedule requirements (Francisco et al., 2008; Glover et al.,

2008; Graff & Larsen, 2011; Jerome & Sturmey, 2008; Mangum et al., 2012; Penrod et al., 2008; Reed et al., 2009). For example, results of the Penrod et al. (2008) study showed that stimuli identified as high preference increased responding over baseline for all participants. In addition, for three of the four participants in their study, high preference stimuli produced higher levels of responding over baseline as compared to low preference stimuli. In the present study, all high and middle ranked items from the preference assessments were shown to be positive reinforcers.

Results of the current study are also consistent with previous research which has demonstrated that low preference items often function as reinforcers under PR or increasing FR schedule requirements (Francisco et al., 2008; Glover et al., 2008; Graff & Larsen, 2011; Penrod et al., 2008). For example, Graff and Larsen (2011) found that when items nominated by caregivers as highly preferred by participants were included in a PS preference assessment, low preference items produced response rates in reinforcer assessments similar to high preference items. In their study, Penrod et al. (2008) also demonstrated that under increasing schedule requirements, low preference stimuli generally functioned as reinforcers for all participants. This was true in the present study except for one item (Participant 2).

A limitation of the current study is that reinforcing efficacy was only assessed using a relatively simple task (i.e., switch-pressing). Roane et al. (2001) found that while reinforcers might produce the same rate of responding for low effort tasks, different rates of responding may be produced when task effort is increased. Future research is needed to evaluate the reinforcing efficacy of stimuli using tasks that require different amounts of effort.

It has also been suggested that if items nominated as high preference by an individual's caregiver are included in preference assessments, the results of preference hierarchies and reinforcer assessments may not be consistent (Graff & Larsen, 2011). Graff and Larsen initially included items thought to be preferred by the participant in a PS preference assessment. Reinforcer assessments demonstrated that both high and low preference stimuli functioned as effective reinforcers. However, when low preference items were included in the preference assessments, results of the preference hierarchy were more consistent with results of the reinforcer assessment (i.e., high preference items functioned as reinforcers, while low preference items generated low response rates). In the current study, items that the participant was known to dislike were not included in the preference assessments. Future research should include both preferred and non-preferred items in preference assessments. In their research, Lee et al. (2010) conducted reinforcer assessments before the preference assessments to ensure that stimuli with different reinforcing efficacies would be included in preference assessments. This approach treats reinforcing efficacy as an independent variable instead of a dependent variable. Future research on progressive ratio and preference could consider this approach.

During PR sessions, each session had a predetermined length of a maximum of 30 minutes in duration in both experiments. Although a predetermined session duration, in addition to the break-point criterion, has been suggested as a termination criterion (Roane, 2008), others have suggested that such a criterion may impose a ceiling effect (Graff & Larsen, 2011; Reed et al., 2009). In Experiment 1 of the present study, Participants 1 and 2 never reached the 30-min session duration, but Participants 3 and 4 each reached this duration during 7 of the 18 PR sessions and those sessions occurred

both for high and low preference items. In Experiment 2, Participant 2 never reached the 30-min duration during any of the PR sessions, Participant 1 reached the 30-min session duration in 2 out of 24 sessions, Participant 3 reached the duration in 1 out of 20 sessions, and Participant 4 reached the duration in 5 out of 14 PR sessions. A maximum session duration is often necessary both for practical reasons (availability of participants) and to ensure that participants do not become fatigued. Some studies have used increasing step sizes (e.g., Reed et al., 2009; Roane et al., 2001) and this procedure may increase the likelihood of participants reaching a breaking point before 30 minutes. However, the procedure also has the disadvantage of reducing the resolution at higher ratio values of the PR schedule.

Future research should also consider using an objective criterion for introducing phase changes. For example, previous research has used a stability criterion defined as 3 consecutive sessions in which each session deviates from the 3-session by no more than 20% and there was no upward or downward trend in the data (Lee et al., 2010).

Another limitation to the current study is that no reliability checks were conducted for Participant 1 during PS assessment sessions because an observer was not available. While it is certainly desirable to conduct reliability checks for a sample of sessions for every participant, the mean percentage of agreement for preference assessments across all observed sessions for the other participants was 100%, suggesting that the experimenter was able to accurately record the target behaviors.

Results of the current study have several implications for clinical settings. Given that low preference items can often function as reinforcers and in some cases produce higher rates of responding than high preference items, it is important to also quantify the

reinforcing efficacy of a stimulus to ensure that a high preference item is in fact an effective reinforcer. It has also been suggested that determining the reinforcing efficacy of a stimulus may be important in clinical settings when schedules of reinforcement are made less dense (Roane et al., 2001). For example, while a variety of high and low preference reinforcers may increase responding under FR schedules of reinforcement, when schedule requirements are increased some stimuli may maintain higher levels of responding than others. Graff and Larsen (2011) noted that identifying effective reinforcers through the use of reinforcer assessments is important in settings where a variety of reinforcers are desired or when reliance on one particular reinforcer may be undesirable. For example, in a classroom setting it may not always be possible for a student to play on the computer; therefore other reinforcers may need to be utilized.

References

Cannella, H., O'Reilly, M., & Lancioni, G. (2005). Choice and preference assessment research with people with severe to profound developmental disabilities: A review of the literature. *Research in Developmental Disabilities, 26,* 1-15.

Carr, J. E., Nicolson, A. C., & Higbeee, T. S. (2000). Evaluation of a brief multiple-stimulus preference assessment in a naturalistic context. *Journal of Applied Behavior Analysis, 33,* 353-357.

Cohen-Almeida, D., Graff, R. B., & Ahearn, W. H. (2000). A comparison of verbal and tangible stimulus preference assessments. *Journal of Applied Behavior Analysis, 33,* 329-334.

Conyers, C., Doole, A., Vause, T., Harapiak, S., Yu, C.T., & Martin, G.L. (2002). Predicting the relative efficacy of three presentation methods for assessing preferences of persons with developmental disabilities. *Journal of Applied Behavior Analysis, 35,* 49-58.

de Vries, C., Yu, C.T., Sakko, G., Wirth, K.M., Walters, K.L., Marion, C., & Martin, G.L. (2005). Predicting the relative efficacy of verbal, pictorial, and tangible stimuli for assessing preferences of leisure activities. *American Journal on Mental Retardation, 110,* 145-154.

DeLeon, I. G., Frank, M. A., Gregory, M. K., & Allman, M. J. (2009). On the correspondence between preference assessment outcomes and progressive-ratio schedule assessments of stimulus value. *Journal of Applied Behavior Analysis, 42,* 729-733.

DeLeon, I. G. & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*, 519-533.

DeLeon, I. G., Iwata, B. A., Goh, H., & Worsdell, A. S. (1997). Emergence of reinforcer preference as a function of schedule requirements and stimulus similarity. *Journal of Applied Behavior Analysis, 30*, 439-449.

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.

Francisco, M. T., Borrero, J. C., & Sy, J. R. (2008). Evaluation of absolute and relative reinforcer value using progressive ratio schedules. *Journal of Applied Behavior Analysis, 41*, 189-202.

Glover, A. C., Roane, H. S., Kadey, H. J., & Grow, L. L. (2008). Preference for reinforcers under progressive- and fixed-ratio schedules: A comparison of single and concurrent arrangements. *Journal of Applied Behavior Analysis, 41*, 163-176.

Graff, R.B. & Larsen, J. (2011). The relation between obtained preference value and reinforcer potency. *Behavioral Interventions, 26*, 125-133.

Green, C.W., Reid, D.H., White, L.K., Halford, R.C., Brittain, D.P., & Gardner, S.M. (1988). Identifying reinforcers for persons with profound handicaps: Staff opinion versus systematic assessment of preferences. *Journal of Applied Behavior Analysis, 21*, 31-43.

Green, C.W., Reid, D.H., Canipe, V.S., & Gardner, S.M. (1991). A comprehensive evaluation of reinforcer identification processes for persons with profound multiple handicaps. *Journal of Applied Behavior Analysis, 4*, 537-552.

Hagopian, L., Long, E., & Rush, K. (2004). Preference assessment procedures for individuals with developmental disabilities. *Behavior Modification, 28*, 668-677.

Hodos, W. (1961). Progressive ratio as a measure of reward strength. *Science, 134*, 943-944.

Jerome, J. & Sturmey, P. (2008). Reinforcing efficacy of interactions with preferred and nonpreferred staff under progressive-ratio schedules. *Journal of Applied Behavior Analysis, 41*, 221-225.

Johnston, J. M. & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research, (3rd edition).* New York, N.Y.: Routledge/Taylor & Francis Group.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.* New York: Oxford University Press.

Lee, M. S. H., Yu, C. T., Martin, T. L., & Martin, G. L. (2010). Preference for reinforcers of different magnitudes with adults with developmental disabilities. *Journal of Applied Behavior Analysis, 43*, 95-100.

Mangum, A., Fredrick, L., Pabico, R., & Roane, H. (2012). The role of context in the evaluation of reinforcer efficacy: Implications for the preference assessment outcomes. *Research in Autism Spectrum Disorders, 6*, 158-167.

Martin, G. L. & Pear, J. (2011). *Behavior modification: What it is and how to do it, 9th Edition.* Upper Saddle River, N. J: Pearson/Prentice Hall.

Martin, T., Yu, C. T., Martin, G., & Fazzio, D. (2006). On choice, preference, and preference for choice. *The Behavior Analyst Today, 7,* 234-241.

Miltenberger, R. G. (2011). *Behavior modification: Principles and procedures, 5th edition.* Belmont, CA: Wadsworth/Cengage Learning.

Northup, J., George, T., Jones, K., Broussard, C., & Vollmer, T. R. (1996). A comparison of reinforcer assessment methods: The utility of verbal and pictorial choice procedures. *Journal of Applied Behavior Analysis, 29,* 201-212.

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

Parsons, M. & Reid, D. (1990). Assessing food preferences among persons with profound mental retardation: Providing opportunities to make choices. *Journal of Applied Behavior Analysis, 23,* 183-195.

Penrod, B., Wallace, M. D., & Dyer, E. J. (2008). Assessing reinforcer potency of high-preference and low-preference reinforcers with respect to response rate and response patterns. *Journal of Applied Behavior Analysis, 41,* 177-188.

Piazza, C. C., Fisher, W. W., Hagopian, L. P., Bowman, L. G., & Toole, L. T. (1996). Using a choice assessment to predict reinforcer effectiveness. *Journal of Applied Behavior Analysis, 29,* 1-9.

Reed, D. D., Luiselli, J. K., Magnuson, J. D., Fillers, S., Vierira, S., & Rue, H. C. (2009). A comparison between traditional economical and demand curve analyses of relative reinforcer efficacy in the validation of preference assessment predictions. *Developmental Neurorehabilitation, 12,* 164-169.

Reyer H. S. & Sturmey P. (2006) The Assessment of Basic Learning Abilities (ABLA) test predicts the relative efficacy of task preferences for persons with developmental disabilities. *Journal of Intellectual Disability Research 50,* 404–409.

Roane, H. S. (2008). On the applied use of progressive-ratio schedules of reinforcement. *Journal of Applied Behavior Analysis, 41,* 155-161.

Roane, H. S., Lerman, D. C., & Vorndran, C. M. (2001). Assessing reinforcers under progressive schedule requirements. *Journal of Applied Behavior Analysis, 34,* 145-167.

Roscoe, E. M., Iwata, B. A., & Kahng, S. W. (1999). Relative versus absolute reinforcement effects: Implications for preference assessments. *Journal of Applied Behavior Analysis, 32,* 479-493.

Schwartzman, L., Yu, C.T.., & Martin, G.L. (2003). Choice responding as a function of choice-presentation method and level of preference in persons with developmental disabilities. *International Journal of Disability, Community, & Rehabilitation, 1*, 1-7.

Stafford, D. & Branch, M. (1998). Effects of step size and break-point criterion on progressive-ratio performance. *Journal of the Experimental Analysis of Behavior, 70,* 123-138.

Taravella, C., Lerman, D., Contrucci, S., & Roane, H. (2000). Further evaluation of low-ranked items in stimulus-choice preference assessments. *Journal of Applied Behavior Analysis, 33,* 105-108.

Tullis, C. A., Cannella-Malone, H. I., Basbigll, A. R., Yeager, A., Fleming, C. V., et al. (2011). Review of the choice and preference assessment literature for individuals with severe to profound disabilities. *Education and Training in Autism and Developmental Disabilities, 46,* 576-595.

Tustin, R. D. (1994). Preference for reinforcers under varying schedule arrangements: A behavioral economics analysis. *Journal of Applied Behavior Analysis, 27,* 597-606.

Windsor, J., Piché, L. M., & Locke, P. A. (1994). Preference testing: A comparison of two presentation methods. *Research in Developmental Disabilities, 15*, 439-455.