JOURNAL OF APPLIED BEHAVIOR ANALYSIS 2012, **45,** 763–777 NUMBER 4 (WINTER 2012)

CORRESPONDENCE BETWEEN SINGLE VERSUS DAILY PREFERENCE ASSESSMENT OUTCOMES AND REINFORCER EFFICACY UNDER PROGRESSIVE-RATIO SCHEDULES

NATHAN A. CALL

MARCUS AUTISM CENTER AND EMORY UNIVERSITY SCHOOL OF MEDICINE

NICOLE M. TROSCLAIR-LASSERRE

LSUHSC HUMAN DEVELOPMENT CENTER

ADDIE J. FINDLEY AND ANDREA R. REAVIS

MARCUS AUTISM CENTER

AND

M. ALICE SHILLINGSBURG

MARCUS AUTISM CENTER AND EMORY UNIVERSITY SCHOOL OF MEDICINE

Research has suggested that a daily multiple-stimulus-without-replacement (MSWO) preference assessment may be more sensitive to changes in preference than other assessment formats, thereby resulting in greater correspondence with reinforcer efficacy over time (DeLeon et al., 2001). However, most prior studies have measured reinforcer efficacy using rate of responding under single-operant arrangements and dense schedules or under concurrent-operants arrangements. An alternative measure of reinforcer efficacy involves the evaluation of responding under progressive-ratio (PR) schedules. In the present study, 7 participants were given a single paired-stimulus (PS) preference assessment followed by daily MSWO preference assessments. After each daily MSWO, participants responded for each stimulus on a PR schedule. The correspondence between break points and preferences, as assessed by the 2 assessment formats, was examined. Results demonstrated that both preference assessments did equally well at predicting reinforcer efficacy, although the PS more consistently identified the most effective reinforcer.

Key words: preference assessment, progressive ratio, paired stimulus, multiple stimulus without replacement

---

Preference assessments serve to identify stimuli that are most likely to have sufficient reinforcing efficacy to maintain responses of clinical significance. Thus, several methods for directly assessing preferences for stimuli have been developed, evaluated, and compared, with research generally demonstrating good correspondence between assessment outcomes and reinforcer efficacy (e.g., Cannella, O'Reilly, & Lancioni, 2005; Lee, Yu, Martin, & Martin, 2010). That is, stimuli that are identified as highly preferred in a preference assessment typically function as more effective reinforcers.

Most assessment methods identify preferences based on approach or selection behaviors, and differ primarily in the manner in which the stimuli are presented. As a result of these methodological variations, each type of assessment has advantages and disadvantages for clinical applications (Hagopian, Long, & Rush, 2004). For example, because all possible

We thank Henry Roane for his assistance with this manuscript and Karen Rader, Mandy Dahir, and Valerie Volkert for their assistance with data collection.

Correspondence concerning this article should be addressed to Nathan A. Call, Marcus Autism Center, 1920 Briarcliff Rd., Atlanta, Georgia 30329 (e-mail: nathan.call@choa.org).

doi: 10.1901/jaba.2012.45-763

combinations of stimuli are presented as a pair at least once, the paired-stimulus (PS; Fisher et al., 1992) format is often considered to be a relatively exhaustive assessment format that has been shown to produce a hierarchy of preferences that corresponds well with reinforcer efficacy and that is stable over time (e.g., DeLeon & Iwata, 1996; Piazza, Fisher, Hagopian, Bowman, & Toole, 1996; Roane, Vollmer, Ringdahl, & Marcus, 1998; Windsor, Piche, & Locke, 1994). In contrast, the multiple-stimulus-without-replacement (MSWO; DeLeon & Iwata, 1996) format generally requires less time to administer (Hagopian et al., 2004). For this reason, it has been suggested that an MSWO assessment can be administered more frequently than a lengthier PS, thereby allowing clinicians to identify preferences despite any fluctuations that may occur over time (Carr, Nicolson, & Higbee, 2000; DeLeon et al., 2001).

Because many clinicians depend on identification of the most effective reinforcer possible, it is important to know if different assessment formats vary with respect to how well they identify effective reinforcers. Thus, several studies have evaluated the correspondence between preference, as identified by various assessment formats, and reinforcer efficacy (e.g., DeLeon & Iwata, 1996; Piazza et al., 1996; Roane et al., 1998). For example, DeLeon et al. (2001) compared the relative reinforcer efficacy of the items identified as highly preferred by either a single administration of a PS assessment or by daily administrations of MSWO assessments. Specifically, after conducting an initial PS assessment, the experimenters conducted an MSWO assessment on each subsequent day. On days when the highest ranked items differed between the MSWO and PS, a reinforcer assessment was conducted to determine if the two stimuli differed in their effectiveness as reinforcers. During the reinforcer assessment, participants chose between working for contingent access to the item ranked highest by the PS, the item ranked highest by the MSWO, or no item. The participants most frequently chose to work for the item ranked highest on the daily MSWO, suggesting that the daily MSWO was better than the single PS assessment in identifying the most effective reinforcer.

DeLeon et al. (2001) employed a concurrent-operants arrangement to determine the relative reinforcer efficacy of the stimuli identified as highly preferred by the PS and daily MSWO formats. A limitation of this methodology is that it can mask absolute reinforcement effects (Francisco, Borrero, & Sy, 2008). For example, even if one stimulus is less preferred than another, it may still be sufficiently reinforcing to maintain responding (e.g., Roscoe, Iwata, & Kahng, 1999). Also, the response required to indicate preference for one stimulus over another is often a very low-effort one (i.e., pointing at or consuming an item). The capacity for a stimulus to maintain such low-effort responses may not always correspond with the ability of that stimulus to maintain the higher effort responses on which reinforcement is often contingent in clinical applications (Fisher & Mazur, 1997).

More recently, several studies have used progressive-ratio (PR) schedules to evaluate reinforcer efficacy (e.g., DeLeon, Frank, Gregory, & Allman, 2009; Penrod, Wallace, & Dyer, 2008; Roane, Lerman, & Vorndran, 2001). Under a PR schedule, reinforcers are delivered after completion of increasing schedule requirements. Reinforcer assessments that have used PR schedules typically continue to escalate the schedule requirement until responding ceases. The highest schedule requirement completed (i.e., the break point) serves as a measure of reinforcer efficacy that can be interpreted as showing how much responding an individual will emit when a given stimulus is delivered as a reinforcer. Reinforcer assessments using PR schedules may be particularly well suited to evaluate the efficacy of stimuli to be used as reinforcers in clinical settings, because maximizing the amount of effort the individual

will emit per unit of reinforcement is often of paramount importance in such situations.

It remains unclear if all preference assessment formats perform equally well when it comes to the identification of effective reinforcers, as measured by performance on PR schedules. Although some studies have evaluated the correspondence between preference and reinforcer efficacy using PR schedules (e.g., DeLeon et al., 2009; Francisco et al., 2008; Glover, Roane, Kadey, & Grow, 2008; Penrod et al., 2008), most determined preference using PS assessments only. Less is known about how well preference, as measured by other assessment formats that are commonly used in clinical settings, corresponds with reinforcer efficacy, as measured by break points. Given the results of DeLeon et al. (2001), in which a daily MSWO identified reinforcers that were preferred over those identified by a single PS, it may be the case that preference, as identified by a daily MSWO, is a better predictor of performance on PR schedules over time. If this is the case, clinicians should consider this preference assessment format when identifying stimuli to be used as reinforcers for more effortful responses or responses to be reinforced on leaner schedules. Conversely, if both assessment formats correspond with performance on PR schedules equally well, then it makes sense to base the choice of assessment format on efficiency. That is, whichever preference assessment requires the least time to complete should be used. The purpose of the current study was to examine the correspondence between assessment results and reinforcer efficacy under PR schedule requirements by comparing results from a single PS and daily MSWO assessments to the break points produced for those same items during a daily PR reinforcer assessment.

METHOD

Participants and Setting

Seven individuals who had been diagnosed with developmental disabilities participated. Joe, Martin, Kyle, and Cameron participated while they attended a university-based summer program for children with autism. Joe, Martin, and Cameron were 6-year-old boys who had been diagnosed with autism. Kyle was a 5-year-old boy who had been diagnosed with autism and attention deficit hyperactivity disorder. All sessions were conducted in the morning in separate classrooms by trained doctoral students. Each room was equipped with the materials necessary for sessions, including a table, chairs, and stimuli used during the procedures. Additional items stored in the rooms were present during the experiment, but access to these items was blocked.

Jack, Edward, and Jose participated while they attended a day-treatment program for individuals with autism. Jack was an 8-year-old boy who had been diagnosed with autism. Edward was a 7-year-old boy who had been diagnosed with sensory integration disorder, developmental delay, and a seizure disorder. Jose was an 18-year-old boy who had been diagnosed with autism. Sessions for Jack, Edward, and Jose were conducted by trained bachelor's level research assistants in clinic treatment rooms (4 m by 5 m) that were equipped with the materials necessary for sessions.

When not participating in the current study, all participants spent 1 to 6 hr working on skill acquisition programming primarily focused on increasing verbal behaviors. During the experiment, items included in the preference and reinforcer assessments were unavailable to participants during the rest of their day, although no control was exerted over the availability of these items outside the time they attended their treatment programs.

Response Measurement and Reliability

Trained observers used pencil and paper to record responding during both preference assessments. Trial-by-trial data were collected on the number of times a particular item was selected, defined as touching or pointing towards one of the items. Data from the PS

were converted to the percentage of trials in which each item was selected by dividing the number of times it was selected by the total number of trials the item was presented and converting the result to a percentage. Items were then ranked by percentage for each participant. Items selected on the same percentage of trials received equivalent rankings that were based on an average. For example, if two items met criteria for the third highest percentage of selection, they each received a rank of 3.5 (i.e., the average of the 3 and 4 rankings). For the MSWO, items were ranked in the order in which they were selected.

During the reinforcer assessment, the number of target responses at each schedule requirement was also collected with paper and pencil. Target responses differed across some participants and were selected arbitrarily with one caveat: An effort was made to select target responses that would not be so effortful that the break point would be reached too rapidly during the reinforcer assessment (thereby diminishing the sensitivity of the reinforcer assessment) or so low effort that the break point would not be reached until a very high schedule requirement (thereby increasing the length of sessions to an unmanageable degree). For Edward, Joe, Martin, and Jose, the target response consisted of touching a colored card taped to the session room wall (Jose) or table top (Edward, Joe, and Martin) within easy reach of the seated participant. Each card touch was defined as an instance in which any part of the participant's hand made contact with the card. A new instance was coded only when the participant removed his entire hand from the card and subsequently replaced any part of it. For Jack and Cameron, the target response consisted of threading a bead on a string, defined as inserting a string into a bead and pulling the string out on the other side. Removing a bead from the string and replacing it counted as a new response. For Kyle, the target response consisted of pressing a button that was attached to a wooden platform. A target response was scored when the button was pressed with sufficient force to make an audible click.

A second observer independently and simultaneously collected data for an average of 92% (range, 50% to 100%) and 92% (range, 43% to 100%) of trials across participants during the PS and MSWO preference assessments, respectively. For both types of assessments, interobserver agreement was calculated by dividing the total number of agreements on item selections by the sum of the agreements and disagreements and converting the result to a percentage. The mean agreement for PS assessment responses across participants was 99%, with 100% agreement for Joe, Martin, Cameron, Jack, Edward, and Jose and 97% for Kyle. Agreement for MSWO responses across participants was 100%.

For the reinforcer assessment, a second observer independently and simultaneously collected data during 53% of sessions across participants (range, 21% to 100%). Interobserver agreement was calculated by dividing the total number of agreements by the sum of agreements and disagreements and converting the result to a percentage. An agreement was defined as both observers scoring the same number of responses at a given schedule requirement. The mean percentage agreement for target responses during reinforcer assessment sessions was 98% across participants (range, 86% to 100%).

Procedure

PS preference assessment. For each participant, a PS assessment (Fisher et al., 1992) was conducted with either six or seven items. Items were selected based on caregiver nomination or experimenter observations of participant interaction with the selected items during instructional or leisure time. Edible items were evaluated for Kyle, Martin, Edward, and Jack, and leisure items or activities were evaluated for Joe, Cameron, and Jose. Leisure and edible items were not evaluated together for any participant. Quantities of all edible items were

restricted to small portions that were typically consumed in less than 20 s. For leisure items, the duration of access to the item during reinforcer delivery was 20 s.

Each PS assessment began with exposure to each of the items by either presenting one portion of each edible item in random order or by allowing 20-s access to each leisure item. Next, two items were presented side by side, approximately 15.2 cm apart, and the participant was instructed to "pick one." Any selection response resulted in access to that item and withdrawal of the unselected item. Attempts to select both items resulted in withdrawal of both items and re-presentation. Failure to make a selection response within 5 s resulted in withdrawal and re-presentation of both items. If no selection response occurred for three consecutive presentations of the same two items, the trial was scored as though neither item had been selected and the next choice presentation began. The placement of each item (to the right or left) was randomized for each trial, including when items were re-presented after no selection or attempts to select both items. The order in which items were paired was also randomized. The next trial began after removal of the selected leisure item and the therapist saying "my turn" or after the selected edible item was consumed. The PS assessment continued until every item had been paired with every other item once.

Teaching the target response. Following completion of the PS assessment, the target response was taught. For Cameron, Joe, and Martin, the target response had already been mastered. For the remaining participants, target responses were taught using a four-step least-to-most progressive prompting procedure with a 3-s interprompt interval and a fixed-ratio (FR) 1 schedule of reinforcement. Prompts included (from least to most intrusive): opportunity for an independent response, verbal prompt, gestural prompt, and physical guidance. The therapist reinforced the target response during training by delivering 20-s access to an item identified as moderately preferred in the previous PS preference assessment (i.e., ranked 3 or 4). Each teaching session consisted of 10 trials. Mastery criterion consisted of independent emission of the target response on at least 80% of trials across two consecutive sessions.

PR assessment baseline: A baseline condition was conducted prior to the PR reinforcer assessment to demonstrate that participants did not emit the target response under extinction (DeLeon et al., 2009). This baseline was completed after acquisition of the target response and before the daily MSWO assessment and PR reinforcer assessment. The materials necessary to emit the target response were provided to the participant, who was told, "You can [emit the target response] if you want to, but you don't have to." No programmed consequences were delivered contingent on the target response. Sessions lasted until 5 min had elapsed without the participant emitting the target response or 60 min, whichever came first. Sessions were conducted until three consecutive sessions occurred with no more than one instance of the target response. One to three baseline sessions were conducted each day.

MSWO preference assessment. Following baseline, daily MSWO assessments (DeLeon & Iwata, 1996) commenced. All items evaluated in the PS assessment, with the exception of the one used to teach the target response to Edward, Jose, Jack, and Kyle, were included in the MSWO assessment. The item used as a reinforcer during training of the target response was excluded for these participants because of its recent history as a reinforcer for the target response. Daily MSWO sessions were conducted immediately after the child's arrival at the clinic to ensure a period of deprivation to stimuli included in the analysis for a period of at least 18 hr.

Prior to the assessment, the participants were briefly exposed to each of the items by either presenting one portion of each edible item or by allowing 20-s access to each leisure item, in random order. All of the items were presented

in an array in front of the participant, with the items in a straight line roughly equidistant from one another. The location of each item in the array was randomized for each trial. For participants for whom edible items were being evaluated, one portion of each edible item was presented. If leisure items were being evaluated, the actual item or a picture depicting the activity (for Jose) was presented. The participant was then instructed to "pick one." Selection responses resulted in access to the selected item. After 20 s of access to the selected leisure item, the therapist delivered the prompt "my turn" and removed the item. For edible items, the next trial began following consumption of any selected items. In each subsequent trial, all of the unselected items were presented in the array in rerandomized positions. The experimenter blocked attempts to select more than one item and rerandomized the positions of the items before repeating the trial. If no selection response occurred within 5 s, the positions of the items were rerandomized and the trial was repeated. The assessment continued until all items had been selected or three trials occurred with the same items without one of them being selected. If no selection response occurred after three trials with the same items, all of the remaining items were given an equal ranking that averaged the remaining rankings.

PR reinforcer assessment. All items from the MSWO preference assessment were included in the daily PR reinforcer assessment. Each day, one session was conducted for each item in random order until all items had been evaluated. All PR reinforcer assessment sessions were conducted immediately after the daily MSWO preference assessment for that day.

A forced exposure trial was conducted at the start of each session. For this trial, participants were physically guided to engage in the target response and then were provided access to the stimulus to be evaluated in that session and told, "This time, when you [engage in the target response], you get a [the stimulus being assessed]." Sessions began immediately after the participant had either consumed the edible item or had 20-s access to the leisure item.

Following the forced exposure, the target response resulted in delivery of one portion of the edible item or 20-s access to the leisure item according to a PR schedule. The schedule requirement doubled after each delivery of reinforcement such that the number of responses required increased from FR 1, to FR 2, FR 4, FR 8, and so on. This particular PR schedule progression was selected so that higher schedule requirements would be reached rapidly, thereby decreasing session length. Sessions were terminated after 5 min of no responding or if 60 min had elapsed (this happened very rarely), whichever came first.

The PR reinforcer assessment continued until stability was achieved based on the cumulative number of responses for each item, as reported in previous studies (e.g., Penrod et al., 2008; Roane et al., 2001; Trosclair-Lasserre, Lerman, Call, Addison, & Kodak, 2008). At the conclusion of each day of PR sessions, all of the items assessed were ranked by cumulative number of responses emitted for that item. Using these data, the reinforcer assessment continued until each participant completed a minimum of three series of the reinforcer assessment (i.e., 3 days), and the rank of cumulative responses remained stable. Stability of rank was defined as no change from the previous day in the rank of all the items based on the daily total of responses emitted for that item.

Data were examined in terms of break points obtained for each item. Break points were defined as the highest schedule requirement completed during the PR reinforcer assessment for that item on that day.

RESULTS

All responses that occurred during the PS, MSWO, and PR assessments are displayed in Figures 1 and 2. Results for the moderately preferred items that were used to teach the

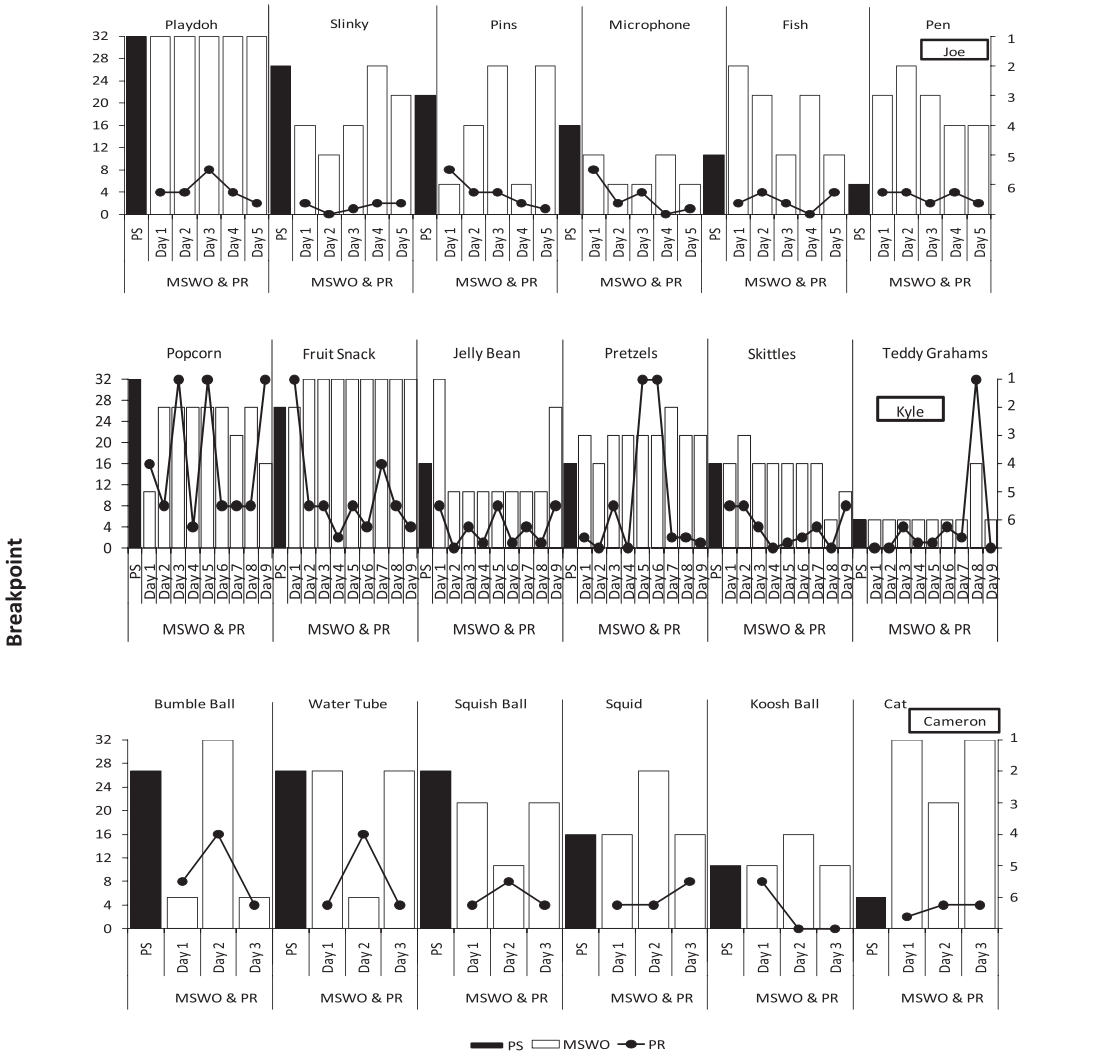

Figure 1. Results for Joe, Kyle, and Cameron. The *x* axis depicts the items and their corresponding preference assessments. The primary *y* axis depicts the break point for each item during the daily reinforcer assessment. The secondary *y* axis indicates the rank of each stimulus during paired-stimulus (PS) and multiple-stimulus-without-replacement (MSWO) preference assessments. The initial black bar for each item represents the PS preference assessment. The white bars that follow indicate the daily MSWO preference assessments. The line graph depicts the daily break point for each item.

target response to some participants were removed from the PS rankings because they were not included in the subsequent daily MSWO or PR assessments. The rankings for any items that fell below the moderately preferred item in the PS preference hierarchy were shifted upwards in the PS results accordingly. All stimuli are arranged in Figures 1 and 2 from left to right in order of decreasing preference as identified by the PS. Instances in which multiple stimuli were selected on the same number of presentations occurred for four participants in the PS assessment (Jose, Kyle, Martin, and Cameron). In general, the PS

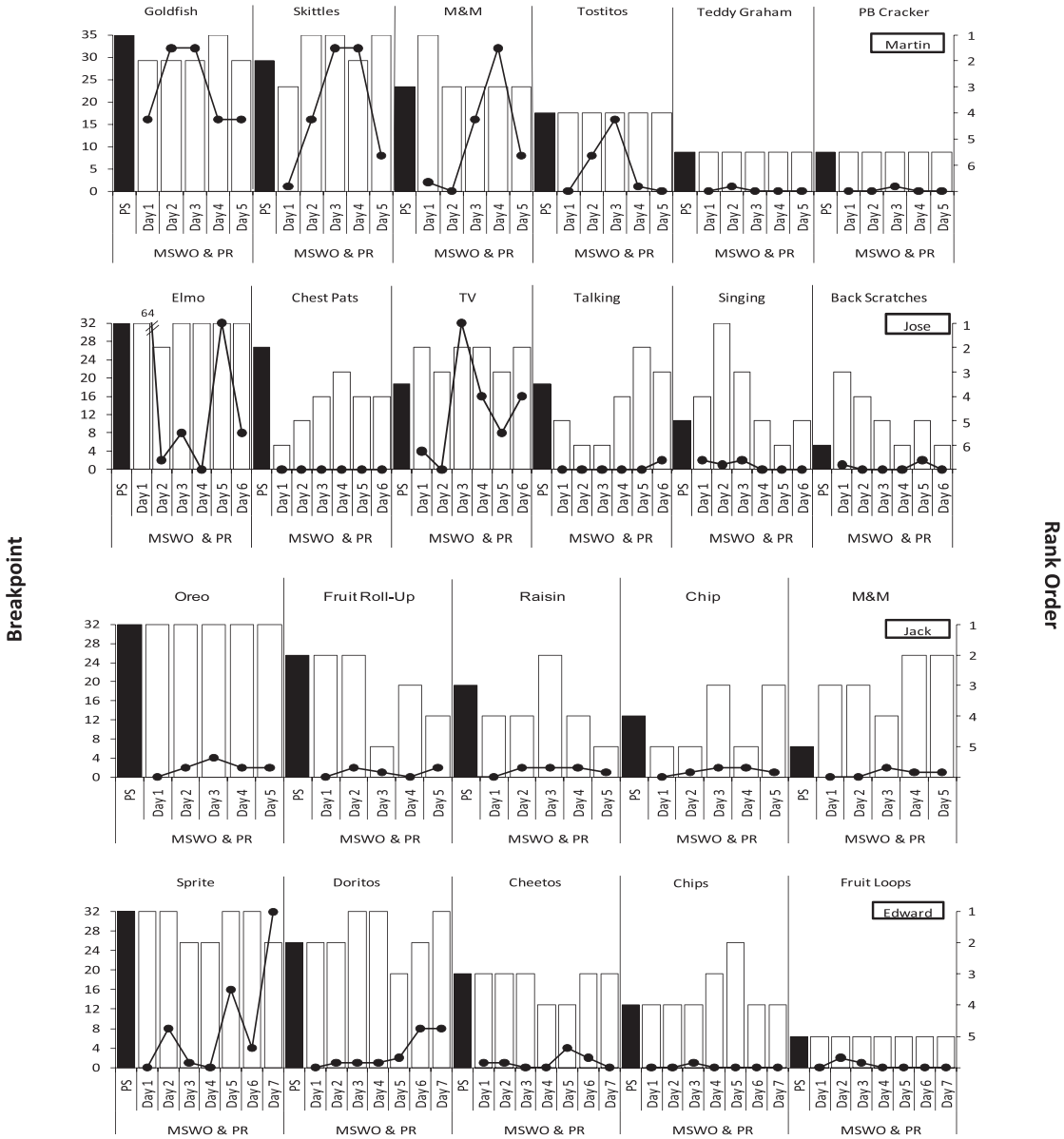

Figure 2. Results for Martin, Jose, Jack, and Edward.

results revealed a hierarchy of stimulus preferences across participants.

For all participants, three (Cameron, Martin, and Jack) to 12 (Edward) baseline sessions were required to achieve stable near-zero rates of responding for three consecutive sessions (data not shown, but are available from the first author). These near-zero rates of the target response for each participant during the extinction baseline suggest that these responses were not automatically reinforcing. Thus, any increases in the target response observed during the PR reinforcer assessment can be attributed to the reinforcing efficacy of the items delivered in those sessions (DeLeon et al., 2009).

Results of the daily MSWO are depicted in Figures 1 and 2, which show that preference rankings differed across participants in terms of stability across days. A mean Spearman's rank correlation was used as an index of stability of preferences based on the daily MSWO (Carr et al., 2000; Hanley, Iwata, & Roscoe, 2006). This analysis replicated the procedures of Hanley et al. (2006), in that the rank for each day was correlated with every other rank, and then all correlation statistics were averaged for that participant. The daily MSWO results for Martin, Kyle, and Edward met or exceeded the standard for stability of .58 used by Hanley et al. (r = .89, .76, and .82, respectively). Daily MSWO preferences for three of the remaining four participants approached but did not meet the same criterion for stability (r = .50, .52, and .54 for Jose, Jack, and Joe, respectively). However, Cameron's rankings on the MSWO were particularly unstable, producing a correlation of –.03.

Results of the daily PR reinforcer assessments appear in Figures 1 and 2 on the primary y axis. Overall, results of the reinforcer assessment were variable across participants. All participants responded at least once for every item, with the exception of Jose when responding produced chest pats. Cameron, Kyle, and Joe usually responded at least once for every item during the PR assessment sessions, as indicated by the fact that these participants had the lowest percentage of sessions with no response (11%, 15%, and 10%, respectively). In contrast, Edward, Jose, Jack, and Martin did not complete the FR 1 schedule requirement during 49%, 56%, 28%, and 37% of sessions, respectively.

Mean break points for Cameron and Kyle were 5.7 (range, 2.7 to 9.3) and 8 (range, 3.9 to 16.4), respectively. Although Joe almost always had a break point of at least 1 for all items, he also had one of the lowest overall mean break points ($M = 3$). The lowest overall mean break point was obtained by Jack ($M = 1.3$), with no break points above 4 for any item. Edward and Jose had low break points for the majority of items, but displayed elevated break points for a few items. For example, Edward had a mean break point of 8.7 for Sprite, whereas his next highest mean break point was only 3 (for Doritos). The mean break point for all of his remaining items combined was 0.6 (range, 0.1 to 1.1). Similarly, Jose had mean break points of 19 and 12.7 for Elmo and TV, respectively, but no other item produced a mean break point higher than 0.8 (the mean break point for the remaining items combined was 0.4, range, 0 to 0.8).

Martin had the highest overall mean break point ($M = 9.5$) as well as what appears to be the clearest hierarchy of break points based on visual inspection of Figure 2, with mean break points for all of the items decreasing in an orderly fashion (Goldfish, $M = 22.4$; Skittles, $M = 17.8$; M&M, $M = 11.6$; Tostitos, $M = 5.0$; Teddy Graham, $M = 0.2$; PB cracker, $M = 0.2$). The data in Figure 2 also suggest correspondence between both preference assessment formats and break points for Martin. That is, higher break points typically occurred for the most preferred items. Break points then decreased systematically across items, with almost no responding occurring for the two items that tied for least preferred on both preference assessments.

The correspondence between rankings from the preference assessments and break points is perhaps more easily seen in Figure 3, which depicts the average break point obtained across preference rankings as determined by the PS or daily MSWO. The PS break points were averaged for individual preferred items because each item had the same ranking every day due to the fact that only one PS assessment was conducted. For example, the data point for the items ranked highest by the PS for Edward represents the average of all of the break points achieved for Sprite. In contrast, rankings of an individual item as determined by the MSWO could (and frequently did) differ across days. Thus, for the MSWO, break points were

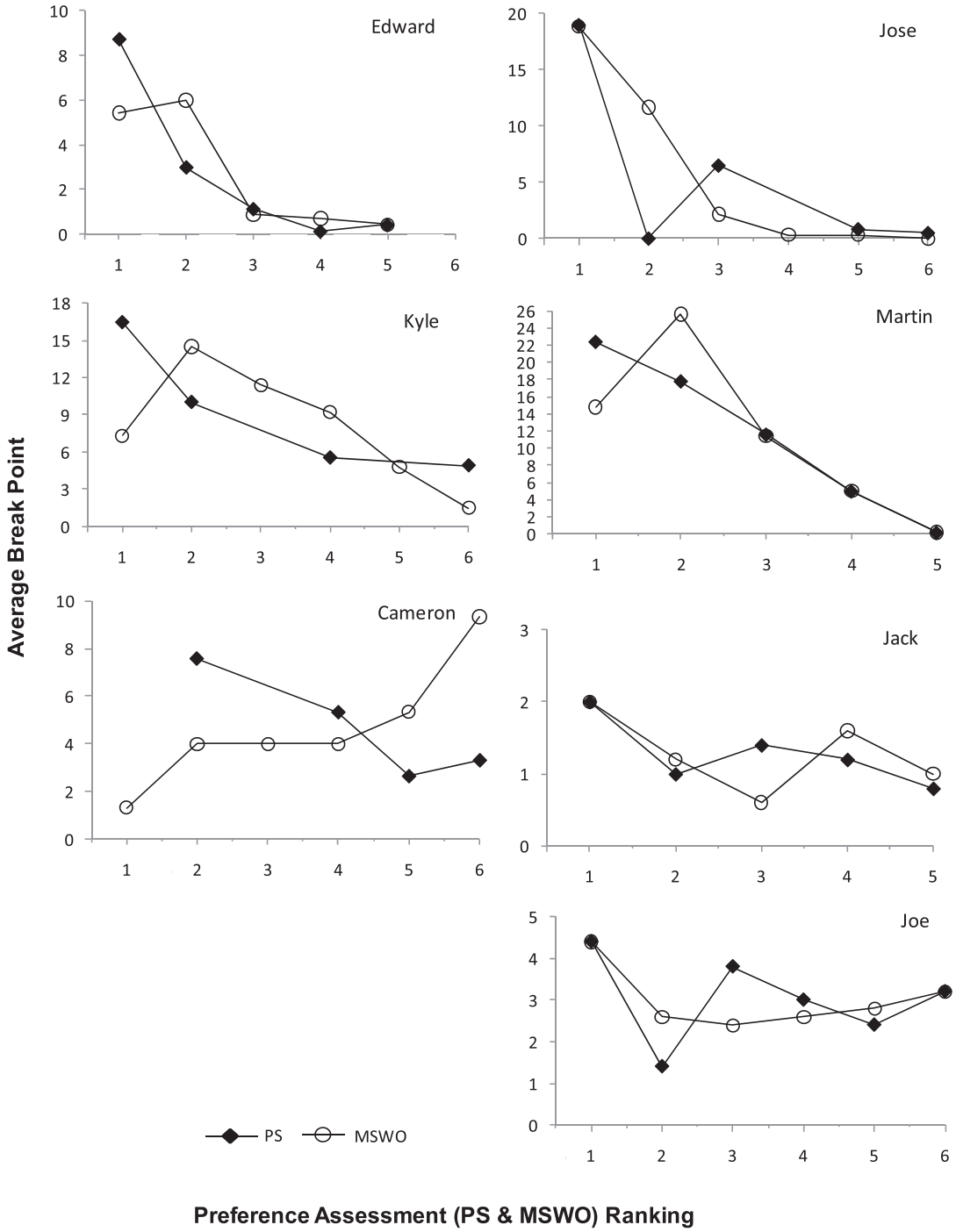

Figure 3. The preference ranking (PS), mean preference ranking (MSWO), and average break point during the reinforcer assessment (PR) for each participant.

averaged across the items that obtained the same preference ranking across days. For example, Edward's MSWO ranked Sprite as most preferred on Days 1, 2, 5, and 6 and Doritos as most preferred on Days 3, 4, and 7 (see Figure 2). Thus, the data point for the items ranked first by Edward's MSWO consists of the average of the break points for Sprite on days it was ranked most preferred and for Doritos on the days they were ranked most preferred. It should be noted that the scale of the *y* axis is adjusted to accommodate the highest mean break point for each participant to allow easier detection of differences in responding. The degree of correspondence between the assessment results and break points can be determined by the extent to which each data path shows a consistently decreasing trend.

For Edward, Kyle, and Martin, relatively high correspondence between PS rankings and break points can be readily discerned by data paths showing a decreasing trend, with no item achieving a higher break point than a higher ranked item. These results suggest that the PS did a good job of identifying not only the most effective reinforcer, but the reinforcing efficacy of stimuli across the continuum of preferences. For these participants, the MSWO rankings also generally corresponded well with average break points, with the exception of the items ranked highest by the MSWO. For these participants, the highest ranked items did not consistently produce the highest break points.

Responding by Jose and Cameron resulted in better correspondence between one preference assessment format and average break points. Results for Jose showed generally good correspondence between the rankings of the MSWO and average break points but less correspondence between the results of the PS and average break points. This low correspondence is primarily due to the fact that Jose never responded for chest pats during the reinforcer assessment. This activity was ranked second most preferred by the PS but received an averaged ranking of 4.3 on the MSWO (range, 3 to 6; see Figure 2).

For Cameron, the ranking from the PS generally corresponded well with average break points. That is, items with lower preference rankings produced lower break points, with the exception of the item ranked least preferred, which produced slightly higher break points than the item ranked 5. To some extent, this correspondence is likely a product of the fact that three items tied for the most preferred on the PS (see Figure 1). These three items, along with the next highest ranked item on the PS, produced the four highest average break points. Less correspondence was discernible between the MSWO rankings and average break points.

Responding by Jack and Joe exemplify another pattern of responding: low break points for all stimuli. Joe never produced a break point higher than 8 for any stimulus, and Jack never produced one greater than 4. These participants also generally demonstrated low correspondence between preference rankings and break points. However, the item identified as most preferred by both the PS and daily MSWO also consistently produced the highest average break points for these two participants.

To summarize, the highest ranked item according to the PS resulted in the highest break point for all participants. The highest ranked item according to the daily MSWO resulted in the highest break point for three of seven participants. In addition, for four of seven participants (Edward, Kyle, Martin, and Cameron), the preference rankings from the PS showed high correspondence with the average break point for those same items. For one participant (Jose), the preference rankings from the MSWO showed high correspondence with the average break point for those same items. For two additional participants (Edward and Kyle), the preference rankings from the MSWO showed high correspondence with the average break point except for the first ranked item. Three participants (Cameron, Jack, and Joe), displayed low response rates and low corre-

Table 1
Kendall's tau-b Showing Correlation Between Rank Order PS and Daily Break Point (PS) and Rank Order of Daily MSWO and Daily Break Point (MSWO)

Edward	PS	.445**
	MSWO	.373**
Jose	PS	.172
	MSWO	.558**
Kyle	PS	.465**
	MSWO	.392**
Martin	PS	.672**
	MSWO	.621**
Cameron	PS	.396*
	MSWO	–.231
Jack	PS	.249
	MSWO	.166
Joe	PS	.075
	MSWO	.128
Overall	PS	.354**
	MSWO	.379**

* = $p < .05$; ** = $p < .01$.

spondence between the rankings from the MSWO and average break point for those same items. Finally, two of these participants (Jack and Joe) showed similarly low correspondence between the preference rankings of PS and average break points.

A statistical analysis confirmed these interpretations of the data. The correspondence between reinforcer efficacy and preference, as determined by the relation between break points and the results of the PS and daily MSWO preference assessments, is depicted in Table 1. Kendall's tau-b was used to determine the degree of correspondence between the preference rankings of stimuli evaluated by each of the assessment formats and the break points for those same items on each day. For the daily MSWO, the correspondence was calculated between each daily preference ranking and the break point for that item on the same day. Correspondence between results of the PS and break points was determined by calculating the relation between the single preference ranking for each item and the break point for that item on each day of the PR reinforcer assessment. Analyzing the PS preference rankings in this manner replicates how PS preference assessments are commonly used in clinical applications. That is, results from a single PS assessment are often used to identify preferred stimuli that are used as reinforcers across several or more days.

Correspondence between break points and both preference assessments was statistically significant for Martin, Kyle, and Edward ($p < .01$; $p_{rep} \geq .98$). For Jose, the preference rankings of the daily MSWO were significantly correlated with break points ($p < .01$; $p_{rep} = .99$), whereas preference rankings from the PS were not. The opposite was true for Cameron. That is, rankings from the PS were significantly correlated with performance on the reinforcer assessment ($p < .05$; $p_{rep} = .94$), whereas rankings from the daily MSWO were not. For Joe and Jack, preference rankings from neither the PS nor the daily MSWO were significantly correlated with break points. Overall, when Kendall's tau-b was calculated for the combined data of all of the participants, a statistically significant, although modest, correlation was found to exist between all of the preference rankings made by each assessment format and the daily break points achieved by all participants for all of the items.

DISCUSSION

The current study demonstrated the relative degree of correspondence between preference, as identified by two formats of preference assessments that are commonly used in clinical applications (i.e., PS and a daily MSWO), and break points produced while responding on a PR schedule. For three of seven participants (Edward, Kyle, and Martin), good correspondence was observed between the results of both assessment formats and break points. For one participant (Jose), there was a significant correspondence between results of the MSWO and break points but not between the results of the PS and break points. The opposite was true for another participant (Cameron). The remaining two participants (Joe and Jack) emitted low rates of responding during the reinforcer

assessment, the results of which did not correspond well with the results of either assessment format.

Overall, the single administration of the PS seemed to be slightly better than the daily MSWO at identifying the one stimulus that was the most effective reinforcer. Therefore, when the priority is to find a single, highly effective reinforcer, clinicians may be best served by basing such judgments on the results of a single PS. However, across the entire continuum of preferences, the PS did not seem to outperform the daily MSWO with respect to correspondence between assessment results and break points. Based on this finding, when clinicians wish to identify several effective reinforcers, it seems sensible to rely on whichever assessment format takes the least total amount of time to administer. Although a single MSWO has the advantage of brevity, conducting daily MSWO preference assessments eventually will exceed the time required for a single PS.

Previous studies that examined the correspondence between preference and performance on PR schedules generally evaluated only the extreme ends of the preference continuum (Francisco et al., 2008; Glover et al., 2008; Penrod et al., 2008; Roane et al., 2001). In these studies, the reinforcing efficacy of the most preferred stimulus, or high- and low-preference stimuli, was evaluated on PR schedules. In the present study, we also evaluated the reinforcing efficacy of stimuli ranked along the entire preference hierarchy. Thus, it was possible to evaluate the degree to which decreasing preference rankings predicted orderly decreases in reinforcing efficacy. Our results showed that, for five of seven participants (Edward, Kyle, Martin, Jose, and Cameron), decreases in preference ranking from at least one assessment format were accompanied by systematic decreases in reinforcing efficacy. This finding is perhaps most important for clinical applications in which it is necessary to identify more than one likely reinforcer to avoid satiation. For the remaining participants, correspondence between preference ranking and break points seemed to be influenced by the low reinforcer efficacy of all stimuli, as indicated by the relatively low break points for all stimuli. Thus, the restricted range of reinforcer efficacy could have caused this low correspondence. These results suggest that preference ranking is a reasonable predictor of reinforcer efficacy across the preference continuum as long as the reinforcing efficacy of the assessed stimuli is heterogeneous.

Results differ from those of DeLeon et al. (2001), who showed that items identified as preferred by the daily MSWO were selected more often in a concurrent-operants arrangement than those identified by the single PS. This discrepancy may be a function of differences in the method of measuring reinforcer efficacy. DeLeon et al. examined responding under FR 1 schedules within a concurrent-operants arrangement, whereas the current study examined break points obtained while participants responded on a PR schedule within a single-operant arrangement. PR schedules are specifically designed to assess the maximum effort an individual will exert when a particular stimulus is being used as a reinforcer. The daily MSWO may indeed be superior to a single PS when it comes to identification of reinforcers for the purpose of maintaining low-effort responses. However, results of this study suggest that the superiority of the daily MSWO diminishes when it comes to identification of a reinforcer that is capable of maintaining higher effort responding. Future studies may be necessary to further clarify the relations between the results of various formats of preference assessments and reinforcer assessments.

Some caveats should be considered when interpreting these results. We did not control the latency between administration of the single PS and actual use of the stimuli assessed as reinforcers (Hanley et al., 2006), which varied as a function of how long it took each

participant to master the target response or meet the criteria to move from baseline to the reinforcer assessment. However, this latency was relatively brief (i.e., a few days). It is also possible that preferences would have shifted over more extended periods of time, thus altering the correspondence between the PS results and reinforcer efficacy. In such a scenario, correspondence between results of the daily MSWO and break points would remain consistent because the daily MSWO corrects for changes in preference.

Results also may be limited to the specific PR schedule used. In any PR schedule, the "true" break point either can be exactly the highest schedule requirement achieved or a schedule requirement that falls in the gap between the last schedule requirement completed and the next one (e.g., FR 7). Thus, the degree of specificity with which one can identify a true break point is equivalent to the size of the gap between these last two schedule requirements. Our schedule requirements increased in a multiplicative (i.e., FR 1, FR 2, FR 4, FR 8, etc.) rather than additive (i.e., FR 2, FR 4, FR 6, FR 8, etc.) fashion so that break points would be reached more rapidly. Therefore, our results may have suffered from a lack of specificity at higher schedule requirements. The extent to which this particular PR arrangement is more or less useful than other arrangements for measuring the effectiveness of reinforcers is an open question that may be worth studying in future research.

Finally, differences in the instructions provided in the baseline and PR reinforcer assessment conditions of the reinforcer assessment may have differentially influenced responding. In baseline, participants were told they were free to emit the target response but did not have to do so. In contrast, in the PR reinforcer assessment, the contingencies were stated (i.e., "when you [engage in the target response], you get [the stimulus being assessed]"). It is possible that the statement in the PR reinforcer assessment may have functioned as a prompt to engage in the target response, thereby affecting the rate of responding relative to baseline.

REFERENCES

Cannella, H. I., O'Reilly, M. F., & Lancioni, G. E. (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. doi:10.1016/j.ridd.2004.01.006

Carr, J. E., Nicolson, A. C., & Higbee, T. S. (2000). Evaluation of a brief multiple-stimulus preference assessment in a naturalistic context. *Journal of Applied Behavior Analysis*, *33*, 353–357. doi:10.1901/jaba.2000.33-353

DeLeon, I. G., Fisher, W. W., Rodriguez-Catter, V., Maglieri, K., Herman, K., & Marhefka, J.-M. (2001). Examination of relative reinforcement effects of stimuli identified through pretreatment and daily brief preference assessments. *Journal of Applied Behavior Analysis*, *34*, 463–473. doi:10.1901/jaba.2001.34-463

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. doi:10.1901/jaba.2009.42-729

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. doi:10.1901/jaba.1996.29-519

Fisher, W. W., & Mazur, J. E. (1997). Basic and applied research on choice responding. *Journal of Applied Behavior Analysis*, *30*, 387–410. doi:10.1901/jaba.1997.30-387

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

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. doi:10.1901/jaba.2008.41-189

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. doi:10.1901/jaba.2008.41-163

Hagopian, L. P., Long, E. S., & Rush, K. S. (2004). Preference assessment procedures for individuals with

developmental disabilities. *Behavior Modification*, *28*, 668–677. doi:10.1177/0145445503259836

Hanley, G. P., Iwata, B. A., & Roscoe, E. M. (2006). Some determinants of changes in preference over time. *Journal of Applied Behavior Analysis*, *39*, 189–202. doi:10.1901/jaba.2006.39-189

Lee, M. S. H., Yu, C. T., Martin, T. L., & Martin, G. L. (2010). On the relation between reinforcer efficacy and preference. *Journal of Applied Behavior Analysis*, *43*, 95–100. doi:10.1901/jaba.2010.43-95

Penrod, B., Wallace, M. D., & Dyer, E. J. (2008). Assessing potency of high- and low-preference reinforcers with respect to response rate and response patterns. *Journal of Applied Behavior Analysis*, *41*, 177–188. doi:10.1901/jaba.2008.41-177

Piazza, C. C., Fisher, W. W., Hagopian, L. P., Bowman, L. G., & Toole, L. (1996). Using a choice assessment to predict reinforcer effectiveness. *Journal of Applied Behavior Analysis*, *29*, 1–9. doi:10.1901/jaba.1996.29-1

Roane, H. S., Lerman, D. C., & Vorndran, C. M. (2001). Assessing reinforcers under progressive schedule requirements. *Journal of Applied Behavior Analysis*, *34*, 145–167. doi:10.1901/jaba.2001.34-145

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

Roscoe, E. M., Iwata, B. A., & Kahng, S. (1999). Relative versus absolute reinforcement effects: Implications for preference assessments. *Journal of Applied Behavior Analysis*, *32*, 479–493. doi:10.1901/jaba.1999.32-479

Trosclair-Lasserre, N. M., Lerman, D. C., Call, N. A., Addison, L. R., & Kodak, T. (2008). Reinforcement magnitude: An evaluation of preference and reinforcer efficacy. *Journal of Applied Behavior Analysis*, *41*, 203–220. doi:10.1901/jaba.2008.41-203

Windsor, J., Piche, L. M., & Locke, P. A. (1994). Preference testing: A comparison of two presentation methods. *Research in Developmental Disabilities*, *15*, 439–455. doi:10.1016/0891-4222(94)90028-0

Received December 23, 2011
Final acceptance July 31, 2012
Action Editor, SungWoo Kahng