Preference Stability and Reinforcer Efficacy with Preschool-Age Children

By

Courtney Rose Moore

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

Chairperson Pamela L. Neidert

Claudia L. Dozier

Jomella Watson-Thompson

Date Defended: July 22, 2015

The Thesis Committee for Courtney Rose Moore
certifies that this is the approved version of the following thesis:

Preference Stability and Reinforcer Efficacy with Preschool-Age Children

Chairperson Pamela L. Neidert

Date approved: July 22, 2015

Abstract

Reinforcement-based teaching strategies are ubiquitous in early childhood settings; however, few studies have evaluated the stability of typically developing children's preferences over time. Fluctuations in preference may lead to decrements in reinforcer efficacy. We conducted an initial paired-stimulus preference assessment (PSPA) and weekly brief multiple-stimulus-without-replacement (MSWO) preference assessments and reinforcer assessments. The purposes of the study were to evaluate the stability of children's preferences; to assess correspondence between rankings produced by PSPA and MSWO preference assessments; and to evaluate the predictive validity of PSPAs and MSWOs. We used methods for evaluating preferences and reinforcer efficacy similar to those described by Call, Trosclair-Lasserre, Findley, Reavis, and Shillingsburg (2012); however, our participants included typically developing children and we assessed preference stability over several weeks. Results indicated that although the majority of participants demonstrated unstable preferences, rankings on PSPAs and MSWOs corresponded when statistically analyzed and both preference assessment types demonstrated similar predictive validity. Idiosyncratic reinforcement effects were observed across stimuli type and participants.

Acknowledgements

First and foremost, I would like to thank my family for their unequivocal support; I would not be where I am without their love and guidance. For my dad—who taught me to relentlessly pursue my goals; for my mom—who taught me of the importance of self-care and compassion for others; and for my stepmom—who helps me balance it all.

Thank you to my advisor, Dr. Pamela Neidert, for enduring endless questions and encouraging me to continue asking more. She has helped me become a better writer and researcher. I would also like to thank Dr. Claudia Dozier, Dr. David Jarmolowicz, and Dr. Edward Morris for their dedication to my education. It has been a privilege to learn under their tutelage.

This thesis would not be possible if it were not for the support of my graduate student colleagues and undergraduate research assistants. In particular, I would like to thank Joseph Dracobly and Elizabeth Foley for their indefatigable support, both personally and professionally. I feel privileged to work alongside such talented behavior analysts.

Table of Contents

Preference Stability and Reinforcer Efficacy with Preschool Children

Reinforcement-based interventions are a valuable tool for promoting skill acquisition and reducing problem behavior in a variety of populations (Akin-Little, Eckert, Lovette, & Little, 2004; Carr & Durand, 2013; Ellison, 1997). They are also a critical part of teaching in early childhood education (ECE) settings. Reinforcement-based interventions have been used to increase appropriate behavior (Hanley, Camilleri, Tiger, & Ingvarrson, 2007), to teach sign language (Thompson, Cotnoir-Bichelman, McKerchar, Tate, & Dancho, 2007), to facilitate toilet training (Greer, Neidert, & Dozier, 2015) and to increase creative behavior (Goetz & Baer, 1973) in ECE classrooms.

The success of reinforcement-based interventions is dependent on effective identification of reinforcers. Preferences are primarily identified through indirect (e.g., surveys, interviews) or direct preference assessment (PA) methods (for review, see Rush, Mortenson, & Birch, 2010 and Cannella, O'Reilly, & Lancioni, 2005). During preference assessment interviews, the interviewer asks questions pertaining to auditory, tactile, visual, olfactory, social, and edible stimuli in order to identify potential reinforcers (e.g., Fisher, Piazza, Bowman, & Amari, 1996; Cote, Thompson, Hanley, & McKerchar, 2007). During PA surveys, the interviewer typically provides a list of potential reinforcers and asks the interviewee to identify preferred stimuli from the list (e.g., Northrup, Broussard, George, & Vollmer, 1996; Northrup, 2000; Didden & de Moor, 2004). The primary advantage of interviews and surveys are that they provide efficient and convenient methods for selecting potential reinforcers.

However, previous researchers (e.g., Northrup, 1996; 2000; Green, Reid, Canipe and Gardner, 1991) have suggested that indirect PAs may be unreliable. Northrup (2000) conducted preference assessment surveys and reinforcer assessments with twenty children diagnosed with

Attention Deficit Hyperactivity Disorder (ADHD). The authors reported that the reinforcers identified during the survey consisted of 34% true positives, 23% true negatives, 29% false positives, and 16% false negatives. The high percentage of false positives is problematic for educators because preference surveys might identify items which do not subsequently function as reinforcers.

In contrast, direct PAs are considered the most valid means of identifying potential reinforcers (Hagopian, Long, & Rush, 2004). Direct PAs involve systematic presentation of stimuli and direct observation of individual's selection or engagement with the stimuli. During direct PAs, stimuli may be presented in isolation (e.g., Pace, Ivancic, Edwards, Iwata, & Page, 1985; DeLeon, Iwata, Conners, & Wallace, 1999; Hagopian, Rush, Lewin, & Long, 2001; Roane, Vollmer, Ringdahl, & Marcus, 1998); in pairs (e.g., Fisher, Piazza, Bowman, Hagopian, Owens, & Slevin, 1992; Piazza, Fisher, Hagopian, Bowman, & Toole, 1996); or in multi-stimulus arrays (e.g., DeLeon & Iwata, 1996; Windsor, Piché, & Locke, 1994; Hanley, Iwata, Lindberg, & Conners, 2003; Layer, Hanley, Heal, & Tiger, 2008). Although direct preference assessments (PAs) have been evaluated widely for use with individuals with intellectual and developmental disabilities (IDD) (for review, see Cannella et al., 2005), very little research has been conducted on the use of direct preference assessments with young, typically developing children (Rush, Mortenson, & Birch, 2010). Early childhood educators might use direct-PA methods in order to increase the likelihood that preferred items identified in PAs function as reinforcers.

Two commonly used direct-PA methods include the paired-stimulus preference assessment (PSPA) and the multiple-stimulus-without-replacement (MSWO) preference assessment (Fisher et al., 1992; DeLeon & Iwata, 1996). During the PSPA, each item is presented with every other item in quasi-random order until every possible combination of stimuli has been presented once. The primary advantage of the PSPAs is that a rank-order preference hierarchy can be established based on the individual's selections. This may be particularly useful for teachers who might want to identify a variety of reinforcers to teach skills of varying difficulty. However, PSPAs can be time consuming because every pair combination is assessed (e.g., a 7-item PSPA will result in 21 pair combinations).

A more efficient direct PA is the MSWO (DeLeon & Iwata, 1996). During the MSWO, the individual is presented with all stimuli simultaneously. Once an item is selected, it is removed from the array on successive trials until all stimuli have been selected or no selection is made. Although MSWOs are typically conducted multiple times, Carr, Nicholson, and Higbee (2000) demonstrated that the number of stimulus presentations could be reduced from five to three while still effectively identifying preferred stimuli. Therefore, the MSWO might be used to identify preference hierarchies in a more efficient manner. The ability to conduct a valid and reliable PA quickly is valuable to early childhood educators who might face time or resource constraints. For example, an early childhood educator might be able to conduct brief MSWOs with one child while other children are engaged in small group or free choice activities.

Direct PAs might be most feasible in ECE settings where multiple teachers or caregivers are present (e.g., in states where licensing requirements require high teacher-to-student ratios) or in classrooms where children receive individualized services from specialty providers (e.g., speech therapists, occupational therapists).

Given the paucity of research on direct PA methods with young children, however; reinforcer assessments should be conducted in order to determine if a preferred item will function as a reinforcer for a given target behavior. Several researchers have examined the correspondence between preference and reinforcer efficacy with individuals with IDD (e.g., DeLeon, Frank, Gregory, & Allman, 2009; Piazza et al., 1996; Hagopian et al., 2001; Pace et al., 1985; Roane et al., 1998; Lee, Yu, Martin, & Martin, 2010; Raetz, LeBlanc, Baker, & Hilton, 2013). Although there is not a standardized method of conducting reinforcer assessments, reinforcer efficacy is typically evaluated by presenting the preferred item contingent upon an existing response in the participant's repertoire (e.g., known task, in-zone behavior, button pressing) under a low schedule requirement (e.g., fixed-ratio [FR] 1).

Evaluating reinforcer efficacy with simple, arbitrary responses and low schedule requirements might minimize the likelihood of skill deficits, response difficulty, or ratio strain masking potential reinforcement effects. Previous research by Green et al. (1991) suggested that some reinforcers are not efficacious under increasing schedule requirements; Roane (2008) proposed that progressive-ratio (PR) schedules of reinforcement (Findley, 1958; Hodos, 1961) could be used to evaluate reinforcer efficacy in applied settings. A PR schedule of reinforcement involves systematically manipulating the number of responses required to obtain access to a reinforcer (Findley, 1958). Thus, the schedule requirement serves as the independent variable when testing reinforcer potency under PR arrangements (Tustin, 1994). The last reinforced schedule requirement that the individual completes before discontinuing responding for a predetermined period of time is called the break point or breaking point (Hodos, 1961; Roane, 2008). The break point is used as measure of reinforcing efficacy.

DeLeon et al. (2009) evaluated reinforcing efficacy using PR schedules of reinforcement. After conducting PSPAs with four individuals with IDD, the authors found that 10 out of 12 comparisons resulted in larger break points for higher preferred stimuli. Thus, highly preferred items identified during PAs might be more likely to function as reinforcers even under increasing schedule requirements.

For early childhood educators, identifying a single, highly potent reinforcer might be of paramount importance. Previous researchers have found that HP items are more likely to function as reinforcers. Call et al. (2000) reported that for 100% of participants, LP stimuli were less likely to function as reinforcers than MP or HP stimuli. DeLeon et al. (2009) found larger break points for higher preferred stimuli in 10 out of 12 comparisons. Both Hagopian et al. (2001) and Piazza et al. (1996) reported that HP stimuli were more effective reinforcers when pitted against MP stimuli and that MP stimuli were more effective reinforcers when pitted against LP stimuli. Roscoe et al. (1999) also reported greater responding for access to HP items for 7 out of 8 participants. Penrod, Wallace, Dyer (2008) demonstrated that LP items would maintain responding at low response requirements (i.e., FR 1) but not under PR schedules.

Although several researchers have demonstrated correspondence between preference and reinforcer efficacy, there is evidence that preferences may change over time (e.g., Mason, McGee, Farmer-Dougan, & Risley, 1989; Zhou, Iwata, Goff, & Shore, 2001). Thus, a single administration of a preference or reinforcer assessment may not be indicative of long-term reinforcer efficacy. Mason et al. (1989) evaluated preference patterns of three children with an Autism Spectrum Disorder (ASD) using the procedures described by Pace et al. (1985) and reported that children's preferences changed from session to session. Children selected the same stimuli in 2 consecutive sessions for only 4 out of 49 sessions. However, the results of the study are limited because a) stimuli from various stimulus classes (e.g., edible, leisure, social) were presented simultaneously, b) data on preferences for individual stimuli were not presented, and

c) the authors did not systematically examine correspondence between preference stability and reinforcer efficacy.

Zhou, Iwata, Goff, and Shore (2001) also suggested that preferences for leisure stimuli may change over time. Zhou et al. (2001) conducted two modified response restriction PAs (DeLeon et al., 1999) with 22 individuals with IDD across a 12- to 20-month period. The authors used a Spearman rank-order correlation coefficient of greater than or equal to .25 to indicate stable preferences. According to their criterion, 13 participants demonstrated unstable preferences. Using a stricter criterion (i.e., .58 or .6; Salkind, 2001; Cicchetti & Sparrow, 1981), the number of participants revealing unstable preferences is even higher (i.e., 18 out of 22 participants).

Ciccone, Graff, and Ahearn (2007) evaluated preference stability for edible stimuli with eight individuals with IDD. The authors conducted two edible MSWOs six months apart and then a new set of MSWOs twelve months apart. The authors reported Spearman-rank order correlation coefficients greater than .6 for 5 out of 8 participants in Study 1 and 3 out of 8 participants in Study 2. These results extend research demonstrating unstable preferences for leisure items to edible items.

Taken together, the majority of previous research (i.e., Mason et al., 1989; DeLeon et al., 2001; Zhou et al., 2001; Ciccone et al., 2007) provides evidence that many individuals demonstrate relatively unstable preferences over time. If preferences change over time, it is likely that reinforcing efficacy may also fluctuate. Carr et al. (2000) evaluated preference stability and reinforcer efficacy with individuals with IDD. The authors conducted MSWOs and reinforcer assessments three times over a one-month period. Results indicated that 2 out of 3 participants demonstrated stable preferences (i.e., Spearman rank-order correlation coefficients above .6). The stimuli that were highly preferred were more likely to function effectively as reinforcers as compared to moderate and low preferred stimuli for all participants.

DeLeon et al. (2001) evaluated preference stability and reinforcer efficacy with five individuals with IDD. The authors conducted a PSPA and subsequent daily, brief MSWOs. When the top-preferred item identified in the daily MSWO did not correspond with the top-preferred item in the PSPA, the authors conducted a reinforcer assessment. The results indicated that the top-preferred item during the initial PSPA and daily MSWO often differed; when they did, the item identified in the MSWO was more effective as a reinforcer than the item identified in the initial PSPA. These results suggest that daily fluctuations in preference likely influence performance when the stimuli are used as reinforcers.

These findings are not surprising given that preferences are likely to be affected by a variety of factors including motivating operations (e.g., satiation, deprivation; for review, see Michael, 1982) and individual histories (e.g., pairing with preferred or non-preferred stimuli, exposure to novel stimuli). The more time that an individual has to contact new contingencies, the more likely fluctuations in preference might occur. Hanley, Iwata, and Roscoe (2006) manipulated environmental variables to produce changes in preference rankings. Specifically, the authors found that extended access to high-preferred stimuli resulted in decreases in preference rankings and pairing low-preference stimuli with other putative reinforcers increased preference rankings for these stimuli.

Therefore, preferences are likely to change over time and reinforcer assessments with PR schedules of reinforcement are likely to identify effective reinforcers, repeated evaluation of direct PA methods using PR schedules of reinforcements is warranted. Call et al. (2012) conducted repeated preference assessments and PR schedule evaluations of reinforcer efficacy with seven children with an ASD. The authors conducted an initial PSPA and then subsequently evaluated the stimuli during daily MSWOs and reinforcer assessments. Preferences were unstable for the majority of participants (i.e., 4 out of 7 participants). The highest ranked stimuli in the PSPA resulted in the highest break points for all participants; the highest ranked item in the MSWO resulted in the highest break point for three out of seven participants. Thus, the PSPA demonstrated more predictive validity than the MSWO.

Despite the evidence of preference instability and the relation between preferences and reinforcer efficacy, the majority of published research on this topic has involved individuals with IDD or adults with dementia. Very little is known about the degree to which young children's preferences fluctuate over time. Previous work in our ECE and Intervention lab at the University of Kansas has suggested that young, typically developing children may display unstable preferences. As part of the procedures in a study on toilet training, which included a differential reinforcement component, Greer, Neidert, and Dozier (in press) conducted repeated MSWOs for 22 children between the ages of 12- and 40-months old. Using a Spearman rank-order correlation coefficient of .6 to analyze the results, only three participants demonstrated stable preferences for leisure stimuli and only two participants demonstrated stable preferences for edible stimuli. These data suggest that children's preferences are likely to fluctuate over time; thus, early childhood educators might need to conduct preference assessments more frequently in order to ensure that the stimuli selected as reinforcers are still effective. However, the relation between preschool children's preferences and reinforcer efficacy is unknown because reinforcer assessments were not conducted.

Given the possibility that young, typically developing children may demonstrate unstable preferences and the paucity of research evaluating direct PAs with this population, the present study replicated and extended research by Call et al. (2012). The purposes of this study were threefold. First, we evaluated the stability of children's preferences for edible or leisure stimuli over time. Second, we evaluated the degree of correspondence between rankings on PSPA and MSWO preference assessments. Third, we evaluated the predictive validity of PSPAs and MSWOs. Implications for the use of preference assessments in ECE settings are discussed.

Method

Participants and Setting

The participants were nine children between the ages of 16-months-old and 31-months-old (M= 24.5 months). One participant's preferences (Luna) were assessed twice (i.e., repeated assessments were conducted for an extended period using one apparatus type; a second set of repeated assessments were conducted subsequently with the other apparatus type). All participants were recruited from an inclusive, educational day-care program in a child development center at a large, Midwestern university. Children in the study had no known diagnoses. There were no prerequisite skills required to participate in the study; however, children with dietary restrictions were excluded from participating in edible-item preference assessments. Children were selected based on parental consent and regular attendance (e.g., approximately less than one absence per month) as reported by classroom supervisors. Children who did not participate in a minimum of twelve consecutive reinforcer assessment sessions (i.e., two weeks of preference and reinforcer assessments) due to absences or nonparticipation were excluded from the study.

All experimental sessions were conducted in the meal area of the children's classrooms or in adjacent session rooms.

Materials

Ten stimuli were used in the edible-item assessments: Kit Kats©, M&Ms©, raisins, Goldfish© crackers, Saltine© crackers, Smarties©, Cheerios©, Cheetos©, Sour Patch Kids©, and gummy bears. Larger edible stimuli (i.e., Kit Kats©, Saltine© crackers, Cheetos©, Sour Patch Kids©, and gummy bears) were cut into small, bite-sized pieces. The experimenters attempted to select a variety of textures and tastes (e.g., salty, sweet, sour, gummy, crunchy, etc.) so as to maximize the likelihood of identifying preferred stimuli. For four participants, edible item colors were held constant (i.e., all M&Ms©, gummy bears, Sour Patch Kids©, and Smarties© were yellow) for several sessions. This was discontinued because it was unlikely to be feasible to implement in an ECE classroom; no systematic changes were observed in reinforcer efficacy before or after the change. The extra-session availability of the edible stimuli was not restricted during the course of the study; however, the majority of edible items were not typically available to the children during the school day.

Ten stimuli were used in leisure-item assessments for one participant: bubbles, magnadoodle, LeapFrog© barn, toy dog, doll, See 'N Say©, truck, dinosaur, book, and xylophone. For five participants (i.e., Benji, Luna 2015, Brett, Byron, Ollie), an iPad© was included instead of the dinosaur. The availability of these specific toys was restricted during the course of the study; however, access to similar toys in the children's classrooms or homes was possible.

During the reinforcer assessments, two identical button press apparatuses were used. For Luna 2014, Taurus, Ryan, and Darryl, a small Tupperware© dish (edible assessments only), a light, and a force-sensitive button were mounted onto a board. A power-interrupter was placed between the batteries of the light to enable the light to turn on for the duration of each button press (i.e., the light turned on when pressure was applied to the button and turned off when no pressure was applied to the button). For Byron, Benji, Brett, Luna 2015, and Ollie, a small Tupperware© dish (edible assessments only) and a force-sensitive button which produced an audible “click” noise when pressed were mounted onto a board.

During edible-item reinforcer assessments, a paper plate was placed behind each button apparatus. The plate was either empty (control) or contained approximately one cup of pieces of an edible item. During the leisure-item reinforcer assessments, the relevant leisure item was placed behind one apparatus. Two plastic bins were present to cover the buttons during the leisure item delivery for all leisure-item reinforcer assessments.

A computer (laptop or desktop with a monitor, keyboard, and mouse) using the Windows© 7 or 8 operating system was placed in an adjacent observation room and was used to record button presses. Each button was connected to a Crick© USB switch interface, which was connected to the computer via USB cable. Crick© software on the computer translated each button press into a standard letter input and !Observe© behavioral data collection software recorded each letter in real-time.

Response Measurement & Interobserver Agreement (IOA)

Preference assessments. For all preference assessments, data were collected using paper and pencil on pre-printed data sheets. Data were collected on the frequency of selection, consumption, expulsion, rejection, and no selection during the PSPAs and MSWOs. In addition, data were collected on item position during the MSWOs. Selection was defined as physical contact with the item using ones’ hand. Consumption was defined as any instance of physically manipulating a leisure item (e.g., touching the dinosaur) or an edible item passing the plane of the lips. Expulsion was defined as any instance of the edible item entering the mouth past the plane of the lips and then being moved again from the mouth past the plane of the lips. Rejection was defined as the participant pushing an item away from his or her self and not retrieving the item within 2 s. No selection was scored if selection did not occur.

A second observer collected data on at least 33% of trials per child for all PSPAs (range, 33% to 100%). A second observer collected data on at least 33% of MSWO preference assessments per child (range, 33% to 100%). During the PSPA an agreement was scored if the same item and the same responses (i.e., accept, consume, expel, reject, no selection) were scored within a trial. During the MSWOs, an agreement was scored if the same item, response, and position were scored within a trial. Exact agreement scores during the PSPAs averaged 100% for all participants. Exact agreement scores during the MSWOs averaged 100% for all participants. IOA was calculated by dividing the number agreements by the number of agreements plus disagreements and multiplying the result by 100%.

Reinforcer assessments. For reinforcer assessments, trained data collectors collected data by pressing a computer key that the !Observe© data-collection software recorded in real-time or using paper and pencil. IOA and procedural fidelity data were collected using ABC Data Pro© application on Apple iPod Touch© devices. Data were collected on the frequency of button presses (test), frequency of button presses (control), frequency of reinforcer delivery, frequency of consumption (edible), and duration of the assessment.

Button presses were defined as any response that closed the microswitch and produced illumination, including pressing the button with the arm, leg, or buttocks; sitting on the button; and rolling on the floor and hitting button. Data on control and test button presses were collected automatically on the computer by the !Observe© behavioral data collection software. Prior to each session, the buttons were tested to ensure that the automated data collection system was recording correctly. Button presses were recorded as frequency and reported as break point (i.e., the last schedule requirement completed before 1-min elapsed without responding).

Reinforcer delivery was scored using a frequency measure. During edible-item reinforcer assessments, data collectors recorded delivery when the therapist placed the edible item in the plastic container attached to the clipboard. During leisure-item assessments, reinforcer delivery was recorded when the buttons were blocked and the toy was placed within arm's reach of the child. Control delivery was recorded when the buttons were blocked (no toy was delivered).

Data collectors recorded edible reinforcer consumption using a frequency measure. Consumption was defined as anytime the child swallowed the entire piece of food within 10 s of delivery. Consumption was not scored if any portion the food was spit out and/or left unconsumed within 10 s of delivery. Consumption data were converted to a percentage of stimuli delivered that were consumed by dividing the number of consumptions by the number of deliveries and multiplying by 100%. Consumption data were scored using paper-and-pencil.

Session durations were recorded automatically by the !Observe© behavioral data collection software. The raw data outputs were manually inspected; in the event that a session was terminated incorrectly (i.e., a session continued after 60 s without responding on the test or control button), all data occurring after the break point were excluded from data analysis.

During reinforcer assessments, IOA was collected on reinforcer delivery for a minimum of 28% of sessions per participant. Reinforcer delivery IOA scores were calculated by adding the number of agreements across a 10-s interval divided by the number of agreements plus disagreements and multiplying by 100%. Reinforcer delivery agreement scores for IOA were collected on 40% of sessions for Darryl and averaged 92% (range, 89% to 94%), 29% of sessions for Luna 2014 and averaged 97% (range, 92% to 100%), 8% for Ryan and averaged

99% (range, 93% to 100%), 33% of sessions for Taurus and averaged 97% (range, 92% to 100%), 37% of sessions for Benji and averaged 98% (range, 95% to 100%), 43% of Byron's leisure assessments and averaged 98% (range, 93% to 100%), 31% of sessions for Brett and averaged 100% (range, 96% to 100%), 33% of sessions for Ollie and averaged 97% (range, 92% to 100%), 38% of Byron's edible assessments and averaged 96% (range, 77% to 100%), 43% of Luna 2015's edible assessments and averaged 91% (range, 73% to 100%), and 46% of Luna 2015's leisure assessments and averaged 98% (range, 94% to 100%).

IOA data were collected on control delivery during leisure-item reinforcer assessments for three participants (Benji, Byron, Luna 2015) during a minimum of 28% of sessions. Control delivery IOA scores were calculated by adding the number of agreements across a 10-s interval divided by the number of agreements plus disagreements and multiplying by 100%. Data were collected on 46% of sessions for Luna 2015's leisure assessments and averaged 98% (range, 92% to 100%), 37% of sessions for Benji and averaged 99% (range, 93% to 100%), and 31% of sessions for Byron and averaged 100%.

Procedural fidelity data were collected on a minimum of 96% of sessions for reinforcer deliveries. Data were calculated by dividing the number of correct reinforcer deliveries reported by the data collector divided by the number of schedule changes dictated by the response frequency multiplied by 100%. Data were collected on 98% of sessions for Darryl and averaged 94% (range, 0% to 100%); 96% of sessions for Luna 2014 and averaged 94% (range, 50% to 100%), 100% of sessions for Ryan and averaged 99% (range, 80% to 100%); 100% of sessions for Taurus and averaged 93% (range, 0% to 100%); 100% of sessions for Benji and averaged 98% (range, 67% to 100%); 100% of sessions for Byron's tangible assessments and averaged 94% (range, 43% to 100%); 100% of sessions for Brett and averaged 93% (range, 26% to

100%); 100% of sessions for Ollie and averaged 87% (range, 0% to 100%); 100% of sessions for Byron's edible assessments and averaged 93% (range, 0% to 100%); 100% of sessions for Luna 2015's edible assessments and averaged 98% (range, 67% to 100%); and 100% of sessions Luna 2015's leisure assessments and averaged 94% (range, 50% to 100%). For control delivery, procedural fidelity data were collected on a minimum of 44% of leisure-item reinforcer assessments for three participants (Byron, Benji, and Luna 2015). Data were calculated by dividing the correct number of control deliveries reported by the data collector divided by the number of schedule changes dictated by the response frequency multiplied by 100%. Data were collected on 58% of sessions for Byron and averaged 87% (range, 50% to 100%); 44% of sessions for Benji and averaged 90% (range 0 to 100%); and 100% of sessions for Luna 2015 and averaged 94% (range, 0% to 100%).

Procedure

Paired-stimulus preference assessment (PSPA). A 10-item PSPA was conducted at the start of the study, using procedures similar to those described by Fisher et al. (1992). The purpose of the PSPA was to identify two high-preference, two moderate-preference, and two-low preference stimuli for inclusion in the remainder of the study. Participants received either 15 s or 30 s of pre-session exposure to each item in the array. Each item was then randomly paired with each other item until every pair combination had been presented one time. At the start of each trial, the experimenter said, "Pick your favorite." If a selection was made within either 15 s or 30 s of the instruction, the child received 15- or 30-s access to the preferred item (leisure) or 1 piece of the preferred item (edible). If a selection was not made, a new trial with a different pair of stimuli was initiated. If participants attempted to make simultaneous selections within a trial, the experimenter blocked the attempt and represented the instruction, "Pick one." The stimuli ranked first, second, fifth, sixth, ninth, and tenth were then used in the subsequent MSWOs and reinforcer assessments. In the event that stimuli were tied in ranking, the experimenters examined the individual item pairings during the PSPA to determine which item was selected when the stimuli were pitted against one another. If the item's rankings remained tied after evaluating individual pairings, the experimenter selected randomly between the tied stimuli. For one participant (Darryl), several items were not selected on any trials during the PSPA. Thus, the MP stimuli were those ranked 3 and 4 during the PSPA.

Multiple-stimulus-without-replacement preference assessment. At the beginning of each week, a single MSWO preference assessment was conducted using procedures similar to those described by DeLeon & Iwata (1996). Participants received 15 s or 30 s of pre-session exposure to each item. Following pre-session exposure, all six stimuli were presented simultaneously and the participants were instructed to, "Pick one." Selections resulted in 15- or 30-s access to the leisure item or one piece of the edible item. The selected item was then removed from array, item order was rearranged, and remaining stimuli were presented. If participants attempted to make simultaneous selections within a trial, the experimenter blocked the attempt and represented the instruction, "Pick one." If no selection was made within 15 s or 30 s of presentation, all stimuli were removed and represented once. If no selection occurred on the subsequent trial, all stimuli were removed and the session was terminated.

Reinforcer assessment. Subsequent to each MSWO, a reinforcer assessment session was conducted with each item used in the MSWO. Reinforcer assessment sessions were conducted within six days of the MSWO. Edible-item reinforcer assessments were conducted at least 30 min after either a scheduled snack or meal and at least 30 min after arrival. Reinforcer assessments were conducted in a multi-element design with no more than three consecutive sessions in a row. Each session was followed by a brief (i.e., 3- to 5-min) break and each item was evaluated once per week. Session rooms contained two button press apparatuses placed on the floor and experimental materials; no other items were present in the session rooms. Button press apparatuses were placed equidistant from each wall in the session room. The position of the item was semi-randomly alternated such that no item was presented on the same side more than twice in a row. The therapist was also semi-randomly rotated such that no therapist ran more than two consecutive sessions with one participant in order to decrease the likelihood of experimenter-specific effects on responding.

Each session began with exposure to response-consequence relation. The experimenter instructed the participant to sit in a position that was equidistant from each button and told the participant, “Try this one and see what happens” while pointing to the button. If necessary, the experimenter provided a model or physical prompt to ensure that the participant contacted the response-consequence relation.

Following pre-session exposure, the participant was prompted to sit equidistant from each button. The session began when the experimenter said, “You can do whatever you want.” The experimenter interacted with the participant as they would in the classroom (e.g., verbal, physical, and facial attention were provided approximately every 30 s to 1 min). Test and control button presses were placed on independent, identical PR schedules. Leisure-item reinforcer assessments were conducted identically to edible-item reinforcer assessments, except that for the leisure assessments, the buttons were blocked for 15 s or 30 s during reinforcer delivery.

Button presses were reinforced on a progressive-ratio (PR) schedule of reinforcement that increased by a step size of one after each delivery of reinforcement (i.e., FR1, FR2, FR3, etc.). If an error of commission occurred (i.e., reinforcer was delivered early) the schedule requirement was reset to the previous requirement and no schedule change occurred. For example, if the schedule requirement was FR3, and a reinforcer was delivered after two responses, the participant would be required to press the button three more times before a reinforcer would be delivered. If an error of omission occurred, (i.e., the requirement was met and no reinforcer was delivered) a schedule change still occurred. The session was terminated after 1 min without responding or after a maximum of 10 min, whichever occurred first. For five participants (Benji, Byron, Brett, Luna 2015, Ollie), experimenters were instructed to ensure that participants did not gain access to stimuli outside of their normal classroom materials immediately following sessions in order to decrease the likelihood that extra-session access to preferred stimuli influenced responding within session (i.e., less responding within session could result in shorter session durations and decreased latency to obtaining other preferred stimuli outside of session). Three timers were used: one was used to time reinforcer delivery during leisure-item assessments, one was used to record session duration, and one was used to record periods of non-responding such that session termination could be determined.

A concurrent-operant arrangement was used during the reinforcer assessments, with one task resulting in access to the item and one task resulting in no programmed consequence. The only difference between edible and leisure assessments was that the task was made unavailable anytime the schedule requirement was met on either button. Previous research by Roscoe, Iwata, and Kahng (1999) suggested that concurrently available reinforcers may mask the reinforcing effects of less preferred stimuli; thus, reinforcers were never assessed simultaneously. Given that we did not conduct a baseline, this arrangement allowed us to assess whether button pressing was more likely to occur if leisure or edible stimuli were delivered contingent upon button pressing.

Data Analysis

All statistics were calculated using the IBM Statistical Package for the Social Sciences 22© (SPSS) on a Windows© 7 or 9 operating system. All graphs were created using Prism© graphing software.

Preference stability. Kendall's coefficient of concordance (W) was calculated in order to evaluate preference stability during weekly MSWO rankings. Strong stable preferences were defined as $W \geq .800$.

PSPA and MSWO correspondence. Kendall's tau-b correlation coefficients (τ_B) were calculated in order to evaluate the correspondence between initial PSPA rankings and daily MSWO rankings.

Visual analysis of PSPA and MSWO correspondence was also conducted (see Figures 7 through 11). Because the rating-scale anchors differed between PSPAs and MSWOs, stimuli were categorized as highly preferred (HP), moderately preferred (MP), and low preferred (LP). HP stimuli were defined as stimuli ranked first or second during the PSPA or first or second during the MSWO. MP stimuli were defined as stimuli ranked in the middle of the hierarchy (i.e., fifth or sixth for most participants; third and fourth for Darryl; fifth for Taurus; third and sixth for Lucy) during the PSPA or third or fourth during the MSWO. LP stimuli were defined as stimuli ranked ninth, tenth, or no selection during the PSPA or fifth, sixth, or no selection during the MSWO. Strong correspondence was defined as an item staying within the same category (i.e., HP, MP, or LP) between the initial PSPA and greater than or equal to 80% of subsequent MSWOs. Moderate correspondence was defined as an item staying within the same category between the initial PSPA and between 60% and 80% of subsequent MSWOs. Poor correspondence was defined as an item staying in the same category between the PSPA and less than 60% of subsequent MSWOs.

Predictive validity. Predictive validity was evaluated using statistical and visual (see Figures 1 to 6) analysis. Kendall's tau-b correlation coefficients were calculated in order to evaluate the predictive validity of each PA type. For MSWOs, each rank was compared to the relative break point for the corresponding reinforcer assessment. For the PSPAs, each item was compared to the relative break point for each reinforcer assessment in which it was presented.

Figures 1 to 6 depict predictive validity for each PA type. The relative break point was calculated by subtracting the break point of the control button from the break point of the test button within each session. In some cases, break point on the control button exceeded the break point on the test button, resulting in negative values. The PSPA break points reflect the average relative break point for each item, because the PSPA was only conducted once per item (i.e., if crackers were ranked 1 during the PSPA, the break point depicted is the average relative break point across all sessions where crackers were present). The MSWO break points reflect the average relative break point for each item according to its rank; because an item's rank could (and often did) change between assessments, the break points for each item ranked 1 were averaged (i.e., if cracker was ranked as number 1 on Week 1, 7, and 9, and Sour Patch Kids© were ranked as number 1 on Weeks 2, 3, 4, 5, 6, and 8, the break points for crackers and Sour Patch Kids© on these weeks were averaged).

Using visual inspection, good predictive validity was defined as a general decreasing trend in mean relative break point across the hierarchy and at least four out of six stimuli demonstrating lower break points across the preference hierarchy.

Preference stability and reinforcer efficacy. Preference stability and reinforcer efficacy were evaluated using visual analysis (see Figures 7 to 11). For ease of visual inspection, relative breakpoints are plotted in Figures 12 to 14. Reinforcing efficacy was defined as a relative break point above zero

Results

Preference Stability

Table 1 depicts Kendall's W values for each participant. Unstable preferences were defined as Kendall's W values equal to or above .800. Seven participants demonstrated unstable preferences across assessments (Byron Leisure, Byron Edible, Luna 2014, Ollie, Benji, Luna 2015 Leisure, and Ryan).

MSWO and PSPA Correspondence

Table 2 depicts Kendall's τ_B for each participant. 8 out of 11 comparisons demonstrated statistical significance ($p \leq .01$) suggesting good correspondence between PA types. Three comparisons (i.e., Luna 2014, Luna 2015 Edible, and Ollie) demonstrated poor correspondence between assessment types. Overall, these results suggest that initial PSPA rankings and weekly MSWOs tended to correspond.

When edible stimuli were categorized as HP, MP, and LP, correspondence between PSPAs and MSWOs across all stimuli was only observed for one participant (Darryl). However, strong correspondence was observed for the HP items for 4 out of 5 participants for whom edible-item assessments (Luna 2015 Edible, 83%; Darryl, 94%; Ryan, 100%; Taurus, 83%). Strong correspondence was only observed for MP items for 2 out of 5 participants for whom edible assessments were conducted (Taurus, 75%, Darryl 94%). Strong correspondence was also observed for the LP items for 3 out of the 5 participants (Luna 2015 Edible, 83%; Darryl, 100%, Taurus, 92%).

When leisure stimuli were categorized as HP, MP, LP, strong correspondence between PSPAs and MSWOs for HP or MP stimuli was not observed for any participant according to the definition. Strong correspondence was observed for the LP items for 3 out of 6 participants (Benji, 80%; Brett, 100%; Ollie, 83%).

Predictive Validity

Table 3 depicts the τ_B correlation coefficients for each participant and PA type. Both PA types showed good predictive validity for two participants (Brett and Darryl). For both participants, the PSPA had slightly better predictive validity (i.e., $p \leq .01$) than the MSWOs (i.e., $p \leq .05$). The PSPA showed better predictive validity for one participant (Ryan). The MSWO showed better predictive validity for one participant (Taurus). Both assessment types showed poor predictive validity for seven out of eleven (63%) participants.

Table 4 depicts descriptive statistics of the break points during both preference assessment types. Overall, higher ranked items resulted in higher break points compared to low ranked items; however, considerable variability across break points was observed. Items that were not selected during preference assessments yielded the lowest break points overall with less variability in the break points observed across sessions.

The MSWO identified the stimulus yielding the highest mean break point in 2 out of 11 assessments; the PSPA identified the stimulus yielding the highest mean break point in 4 out of 11 assessments. The MSWO identified the top-two stimuli yielding the highest mean break point in 3 out of 11 assessments; the PSPA identified the top-two stimuli yielding the highest mean break point in 4 out of 11 assessments. The PSPA identified the lowest ranked stimulus yielding the lowest break point in 5 out of 11 assessments and the MSWO identified the lowest ranked stimulus yielding the lowest break point in 4 out of 11 assessments. Thus, the MSWO might better identify top-preferred stimuli that function as potent reinforcers.

Figures 1 through 6 depict the predictive validity of the MSWO and PSPAs for each participant. The preference rankings are scaled to the x-axis and the average break point is scaled to the y-axis. Open circles denote the average preference ranking during the PSPA and closed circles denote preference ranking during the MSWO. The degree to which each PA predicted reinforcing efficacy can be determined by the extent to which the average break points show a decreasing trend.

Visual inspection suggested that both PA types demonstrated good predictive validity for three participants (Darryl, Brett, and Ryan). 6 out of 11 participants demonstrated an overall decreasing trend across the preference hierarchy for at least one PA type (Byron Edible, Byron Leisure, Brett, Darryl, Ryan, and Luna 2015 Leisure). The MSWO demonstrated better predictive validity for two assessments (Byron's edible assessments and Byron's leisure assessments). Both preference assessments yielded poor predictive validity with the majority of participants.

Preference and Reinforcer Efficacy

Figures 7 through 11 depict the preference rankings and break points of button presses for each participant. The preference ranking during the initial PSPA is scaled to the right y-axis and denoted by black bars. The preference rankings for the weekly MSWOs are scaled to the right y-axis and denoted by gray bars. The break point of button presses is scaled to the left y-axis and depicted by circles. Open circles denote the break point on the control button and closed circles denote the break point on the button that yielded the preferred item.

For ease of visual inspection, 12 through 14 depict the preference rankings and relative break points for each participant. The graphing conventions are the same as Figures 7 through 11; however, the relative break point for each item within session is scaled to the left y-axis and denoted by closed circles. The relative break point was calculated by subtracting the break point of the control button from the break point of the test button within each session. Reinforcing efficacy was evaluated by visually examining relative break points across sessions for a given item.

During edible assessments, three participants demonstrated a decreasing trend across the preference hierarchy. Darryl demonstrated the highest break points for HP items; reinforcing efficacy decreased in an orderly fashion across the preference hierarchy. LP stimuli did not function as reinforcers for Darryl. A similar pattern was observed for Taurus; however, top- and moderately preferred items functioned similarly as reinforcers.

Ryan also demonstrated a decreasing trend across the preference hierarchy; however, no items consistently functioned as reinforcers. It is possible that a) responding was not under the control of the experimental contingencies b) the light produced by button presses was more reinforcing than access to the edible item c) responding on both buttons was adventitiously reinforced at some point due to the PR schedule, when switching between buttons.

Two participants demonstrated no hierarchical relation between rank and reinforcing efficacy (e.g., Byron and Luna 2015). Both participants demonstrated that all items functioned similarly as reinforcers. Interestingly, both participants demonstrated behavioral patterns that were indicative of learning over time. These data highlight a limitation in the present study that call the validity of the initial results into question and indicate the importance of conducting sessions until stable results are observed.

Although the results of the edible assessments were generally similar to those reported by previous researchers (i.e., preference rankings tended to correspond with reinforcing efficacy or all items functioned as reinforcers), results from the leisure assessments yielded a different pattern of responding. Most participants pressed both the test and control buttons during reinforcer assessments. For ease of visual inspection, Figures 12 through 14 depict the relative break point within each session. None of the participants demonstrated consistent reinforcement effects across the stimulus hierarchy; most participants demonstrated inconsistent reinforcement effects overall.

Byron and Benji both demonstrated that no items consistently functioned as reinforcers. Three participants (Luna 2014, Luna 2015, and Ollie 2015) demonstrated reinforcement effects that were unrelated to preference rankings (i.e., the items that functioned as reinforcers were not necessarily ranked highest). For example, See'NSay©, LeapFrog© Farm, and bubbles consistently functioned as reinforcers for Ollie's button pressing despite being ranked as moderately and low preferred. Only one participant demonstrated that the top-preferred item functioned as reinforcer during the majority of sessions (Brett; iPad©).

In summary, preference rankings were predictive of reinforcing efficacy for the majority of participants for whom edible assessments were conducted; however, preference rankings were not predictive of reinforcing efficacy for any participants for whom leisure assessments were conducted.

Discussion

Preference Stability

Results of this study indicated that 63% of participants demonstrated preference patterns that did not meet our criterion for stability. During the edible assessments, three of five participants demonstrated stable preferences. During the leisure assessments, only one of six participants demonstrated stable preferences.

Although the methods of evaluating stability have differed across studies, these findings were generally consistent with previous research. Previous researchers have reported that between 33% and 100% of participants demonstrated unstable preferences. The following percentages of participants demonstrating unstable preferences have been reported by previous researchers: 100%, 73%, 59%, 20%, 38% (Study 1) and 63% (Study 2), 58%, 86% (leisure items) and 90% (edible items); and 33% (Mason et al., 1989; Zhou et al., 2001; DeLeon et al., 2001; Raetz et al. 2013, Ciccone et al., 2007; Call et al., 2012; Greer, Neidert, & Dozier, 2015; Carr et al., 2003).

Given the differences in evaluating and defining stability across studies, Hanley et al. (2006) attempted to equate previous research on preference stability by assigning ranks to preferred items, calculating Spearman's rank order correlation coefficients, and defining stability as $r = .58$. Using this criterion, the authors combined findings by Mason et al. (1989), Carr et al. (2003), and Zhou et al. (2001) and reported that preferences were unstable for 71% of participants. Including Hanley et al.'s (2006) findings, unstable preferences were observed for 61% of participants across the four studies. The present study closely replicated these findings, with 63.3% of participants demonstrating unstable preferences over time.

In addition to replicating previous research on preference stability, these data expand upon previous research in several ways. First, few researchers have evaluated preference stability with young, typically developing children. Preference stability might be particularly pertinent to evaluate with young, typically developing children because of the speed of developmental changes that occur in early childhood. For example, children learn to walk and talk in early childhood; as they do, they have more opportunities to contact new reinforcement contingencies in their environment, thus increasing the likelihood that fluctuations in preference may occur as they contact new sources of reinforcement. The high percentage of unstable preferences observed in the present study suggests that young children's preferences are likely to fluctuate in early childhood. The variety and rapidity of developmental changes occurring in early childhood also highlight areas for future research. For example, the present study did not attempt to evaluate which developmental changes might correlate with changes in preference. Future researchers might conduct systematic assessments to evaluate children's skill levels across various domains (e.g., verbal skills, motor skills) and evaluate whether changes in scores correlate with changes in preference. It would be interesting to evaluate whether changes in vocal-verbal skills correlate with preference stability. Future researchers might attempt to replicate research by Northrup (2000) who evaluated children's verbal reports of preferences against reinforcer efficacy. If children can accurately identify preferred items that subsequently function as reinforcers, more frequent preference assessments might be feasible in early childhood settings.

Second, this study addressed limitations present in previous research. For example, this study added extended research on preference stability (e.g., Ciccone et al., 2007; Greer et al., 2015) by conducting systematic reinforcer assessments. This study also addressed a limitation present in research by Mason et al. (1989) who presented items from different stimulus classes simultaneously. Research by DeLeon, Iwata, and Roscoe (1997) demonstrated that leisure items were displaced from preference assessments when leisure and edible items were presented in the same arrays. Therefore, items were presented singularly in order to decrease the likelihood that stimuli from one stimulus class (e.g., edibles) might displace otherwise effective reinforcers (e.g., leisure items) when presented simultaneously, resulting in false negatives. This information is helpful for early childhood educators who might be more interested in finding a variety of effective reinforcers than determining the degree to which one reinforcer is more efficacious than another.

Third, this study replicated and extended the Call et al. (2012) study to a novel population over a longer period of time (i.e., weeks rather than days). These findings extend the generality of findings on preference stability by demonstrating similar results with a novel population and different experimental arrangement than previous studies.

Although these findings add to previous research on preference stability, they are limited in several ways. First, the results only provide descriptive information on the prevalence of unstable preferences. The passage of time provides opportunities for behavior to contact new environmental contingencies, thus resulting in potential changes in preference. In the present study, "preference" refers to the extent to which an individual selects an item during an assessment and "stability" refers to the tendency of the same items being selected over time. Thus, empirical evidence that many children display unstable preferences provides evidence that preferences do change, but further examination of why and how to change preferences (e.g., Hanley et al, 2006) with typically developing children is warranted.

Although the present methodology did not allow researchers to assess *why* preferences became stable or unstable, it still might be useful as a diagnostic tool for identifying *when* preferences become stable or unstable. Some degree of preference instability is developmentally appropriate (i.e., indicative of a child's willingness to sample new stimuli in their environment). Further, restricted, repetitive patterns of behavior, interests, or activities are used as a diagnostic criterion to identify children diagnosed with an autism spectrum disorder (ASD) (American

Psychiatric Association, 2013). Conducting repeated PAs might allow practitioners to identify children who demonstrate extremely rigid preferences such that interventions can be implemented.

PSPA and MSWO Correspondence

Results of this study indicated that for the majority of participants (73%), rankings on PSPAs tended to correspond with rankings during MSWOs. Previous findings evaluating PSPAs and MSWOs correspondence have yielded mixed results. DeLeon & Iwata (1996) reported that 6 out of 7 participants demonstrated strong correspondence between MSWOs and PSPAs. However, DeLeon et al. (2001) reported poor correspondence between an extensive PSPA and daily brief MSWOs. Lanner, Nichols, Field, Hanson, and Zane (2009) reported mixed findings: only the highest and lowest preferred items were matched across assessments.

We conducted visual analysis of correspondence between categories of stimuli (i.e., HP, MP, and LP) during MSWOs and PSPAs. According to our criteria, only one participant demonstrated strong correspondence across all stimuli. Interestingly, this participant also showed stable preferences and strong predictive validity across both assessment types. One explanation for the results is that this participant's preferences were more resistant to environmental changes overall, and thus neither changes occurring over time nor differences in the assessments strongly affected preference rankings.

It is also possible that we identified a more disparate preference hierarchy with this participant. One important consideration when evaluating HP, MP, and LP stimuli is that although both PAs produce a preference hierarchy, the hierarchies produce rankings on an ordinal scale. That is, the distance in relative preference between items ranked 1 and 2 might be more discrepant than the distance between items ranked 2 and 3, even though the ranks are separated by the same numerical distance. This might explain why HP and LP stimuli are more likely to correspond during repeated assessments.

The majority of participants for whom edible assessments were conducted demonstrated good correspondence for HP and LP items. During the leisure assessments, half of the participants demonstrated good correspondence for LP items. Thus, the present research partially replicates results by Lanner et al. (2009) in demonstrating that HP and LP items are generally more likely to correspond when assessed under MSWO and PSPA arrangements than MP items.

In order to compare these results to those reported by Call et al. (2012), we applied the same criteria to the preference rankings reported by the authors. Results indicated that only two participants demonstrated good correspondence according to our definition (i.e., Martin demonstrated 90% correspondence for HP items, 90% correspondence for MP items, and 80% correspondence for LP items; Edward demonstrated 93% correspondence for HP items and 85% correspondence for LP items). Both participants also demonstrated stable preferences according to visual inspection and a general decreasing trend in reinforcing efficacy across the hierarchy.

One limitation to the present analysis is that statistical analysis appeared to yield different results than visual analysis. However, the definition of correspondence was arbitrarily defined; it is possible that other ways of defining correspondence might have yielded different results. Because we evaluated the initial PSPA against subsequent MSWOs, generally stable preferences would have to be observed in order to meet criterion for correspondence. An alternative method of examining correspondence would be to evaluate a single PSPA against a single MSWO. However, the primary purpose of evaluating correspondence was to evaluate which assessment demonstrated better predictive validity when the PSPA and MSWO rankings differed.

Predictive Validity

The third purpose of the present study was to evaluate which assessment better predicted reinforcing efficacy when PSPA and MSWO rankings did not correspond (e.g., when preferences changed). Previous research on the predictive validity of PSPAs and MSWOs has yielded mixed results. DeLeon et al. (2001) found that the PSPA resulted in better predictive validity than the MSWO. Their findings suggested the brief assessments (e.g., MSWOs) should be conducted prior to acquisition tasks in order to increase the likelihood that fluctuations in preference or reinforcing efficacy are accounted for.

However, DeLeon & Iwata (1996), Lanner et al. (2009), and Call et al. (2012) found that the PSPA and MSWO yielded similar predictive validity. Call et al. (2012) reported that PSPAs and MSWOs were equally effective at identifying preferences across the preference hierarchy. Four participants demonstrated statistically significant correlations between reinforcing efficacy and each PA type; one participant demonstrated statistically significant correlations between reinforcing efficacy and the MSWO; and one participant demonstrated statistically significant correlations between reinforcing efficacy and PSPA rank (Call et al., 2012).

The present study partially replicated these findings. For participants for whom edible assessments were conducted, both preference assessments showed good predictive validity for two participants; the PSPA demonstrated better predictive validity for one participant and the MSWO demonstrated better predictive validity for one participant. However, poor predictive validity was observed for the majority of participants. It is possible that the differences between studies might be due to differences in methodology (i.e., concurrent-operant arrangement vs. single-operant arrangement), participant characteristics (i.e., young, typically developing children versus individuals with IDD or dementia), or due to similar reinforcing efficacy across the preference hierarchy. For example, some participants demonstrated that all items functioned similarly as reinforcers (e.g., Byron's edible assessments). For these participants, poor predictive validity across the preference hierarchy is not necessarily problematic. In fact, these results suggest that either assessment might be used to identify a variety of effective reinforcers for use in ECE contexts.

Although visual analysis yielded somewhat discrepant results from statistical analysis (with the exception of data for Brett and Darryl), this is likely a product of different methods of defining predictive validity. For example, Byron's leisure MSWO did not meet statistical significance, but demonstrated good predictive validity when visually examined. This could be explained by the low and similar levels of reinforcing efficacy across the lower items in the preference hierarchy.

In summary, the majority of previous research and results of this study indicated that both assessments showed similar predictive validity. Thus, when early childhood educators are interested in identifying a variety of effective reinforcers, either PA can be used.

When early childhood educators are interested in identifying a single, highly potent reinforcer, the predictive validity of the highest preferred stimulus in each PA type needs to be evaluated. Previous research by DeLeon et al. (2001) suggested that highest ranked stimuli in the PSPA resulted in the highest break points for all participants and the highest ranked item in the MSWO resulted in the highest break point for 3 out of 7 participants. Call et al. (2012) reported that the highest ranked item in the PSPA resulted in the highest break point for all participants and the highest ranked item in the daily MSWO resulted in the highest break point for 3 out of 7 participants. Thus, the PSPA demonstrated more predictive validity than the MSWO in both studies.

The present study partially replicated these findings. The highest ranked stimulus in the PSPA resulted in the highest mean relative break point for two participants; the highest ranked stimulus in the MSWO resulted in the highest mean relative break point for four participants. The lowest ranked stimulus in the PSPA resulted in the lowest mean relative break point for five participants; the lowest ranked stimulus in the MSWO resulted in the lowest mean relative break point for three participants. Results of the present study suggested that the MSWO was more likely to demonstrate good predictive validity for the top-preferred items.

With the exception of the assessments which showed poor predictive validity, these results support Call et al.'s (2012) recommendation that when attempting to identify a variety of reinforcers of varying reinforcing efficacy, both assessments show similar predictive validity. However, the present study suggested that the MSWO may be more effective at identifying a single, potent reinforcer. Thus, early childhood educators might use the brief, MSWO to identify top-preferred items that can be used in ECE settings.

Although Call et al. (2012) suggested that it may take longer overall to conduct multiple MSWOs, it is possible that the cumulative duration of the assessments is not of practical relevance for early childhood educators. Although it make take longer cumulatively to conduct daily or weekly MSWOs, it may be more feasible for an educator to implement a 5-min MSWO once per day than one 20-min PSPA. Additionally, participant characteristics (e.g., scanning skills, developmental level, choice-making skills, vocal-verbal skills, occurrence of problem behavior) might also influence which PA type is most useful for early childhood educators. For example, the relation between children's choice making skills and the number of items in an array might affect the predictive validity of MSWOs.

Further, conclusions regarding predictive validity are limited due to the unclear results from the leisure-item assessments. There are possible explanations for the differences observed between edible and leisure assessments. First, deprivation from food was systematically programmed (i.e., 30 min of deprivation from any food items), thus the present study may have capitalized upon establishing operations for food as reinforcers. Further, although access was not systematically controlled, participants rarely received access to the specific edible items during the course of their school day, thus further increasing the likelihood of deprivation from these items. In contrast, similar items from the leisure assessments might have been available in the participant's classrooms. Previous research indicates that access to items affects performance (e.g., Klatt, Sherman, & Shield, 2000; Roane et al., 2005; Kodak et al., 2007; Vollmer & Iwata, 1991; Roane, Call, & Falcomata, 2005; Gottschalk, Libby, & Graff, 2000). For example, Vollmer & Iwata (1991) demonstrated that deprivation from stimuli for 15 min resulted in little engagement during subsequent assessments; in contrast, deprivation of one to four days resulted in higher levels of engagement in subsequent assessments.

Similarly, access to items after sessions has been shown to affect subsequent responding; Gottschalk et al. (2000) reported that establishing operations continued to affect responding even 24 to 72 hours after exposure. In the present study, immediate post-session access to special leisure items was controlled and the exact items used in the reinforcer assessments were unavailable for the duration of the study; however, similar leisure items might have been available in the participants' classrooms or homes. Future researchers might systematically program deprivation in order to increase the likelihood of obtaining a reinforcement effect. However, even if the observed results are due to insufficient establishing operations, these data have important implications for early childhood educators. Specifically, items are unlikely to function as reinforcers if they are available at other times during participant's days. Thus, early childhood educators might strategically select putative reinforcers that they can control access to.

If the observed effects are due to reinforcer potency, in addition to programming deprivation, future researchers might manipulate various dimensions of the stimulus (e.g., immediacy, magnitude). In the present study, one explanation for why the leisure items might not have functioned as reinforcers was due to insufficient magnitude (i.e., duration of delivery). The access duration in this study was relatively short (i.e., 15 or 30 s). Previous researchers (i.e., Trosclair-Lasserre, Lerman, Call, Addison, & Kodak, 2008) have suggested that magnitude may influence reinforcing efficacy especially at higher schedule requirements. Thus, future researchers might replicate the present leisure-item assessments with varying access durations for preferred items and evaluate whether reinforcing efficacy is observed.

It is also possible that undifferentiated responding was a product of features of the experimental task (i.e., buttons with lights or sounds). Young children typically play with toys that light up or make noise when pressed; thus, the task might have been automatically reinforcing. The buttons might have been more reinforcing than access to the leisure items because the buttons were both novel to the children and restricted outside of the experimental sessions. Further, if the light or sound reinforced button presses, the button presses were always reinforced on a dense schedule of reinforcement (i.e., FR1). These features of the task could have increased the likelihood that the button presses were automatically reinforcing.

It is interesting that differentiated responding was observed across the majority of participants for whom edible assessments were collected even though the sound or light was still present during these assessments; however this may have been due to the potent reinforcing efficacy of edible items. DeLeon et al. (1997) demonstrated that edible items displaced leisure items on eleven out of fourteen assessments, providing evidence that edible items might be more potent reinforcers than leisure items when presented simultaneously. Bojak & Carr (1999) also suggested that edibles might function as more potent reinforcers because individuals have a longer history with food than leisure items.

A final possibility is that the concurrent-operant arrangement hindered discrimination for some participants in the study; thus responding may not have come under control of the experimental arrangement. Several features of the experiment were designed to facilitate discrimination. First, every session began with pre-session exposure to the response-consequence relation. Second, the item was visibly present behind the button which resulted in access. Third, the schedule requirements increased fairly slowly (i.e., a step size of one). However, sessions were rapidly alternated, run in close succession (i.e., 3- to 5-min breaks), and run in the same session rooms; this may have hindered discrimination of the experimental contingencies. Rotating the stimulus position also might have resulted in carry-over effects between sessions (i.e., a history of responding and obtaining reinforcement on the test button might have resulted in button presses on control when the stimulus position was switched between sessions).

Discrimination failures and carry-over effects are a well-known limitation to the multi-element design. One way to mitigate this problem would be to replicate the methodology used by Call et al. (2012) where baseline data collection demonstrated that the target response would not occur in the absence of reinforcement. Researchers could then implement a reinforcer assessment using a single-operant arrangement and PR schedule of reinforcement.

Importantly, two participants demonstrated delayed reinforcement effects which appeared indicative of learning over time. These participants were the only two participants for whom both edible- and leisure-item assessments were conducted. Thus, it is possible that a combination of poor reinforcing efficacy and features of the experimental arrangement hindered discrimination. Conducting all phases until stability was observed might have mitigated problems related to learning over time.

The lack of differentiated responding observed during the majority of leisure-item assessments highlights a major limitation to the present study. An additional limitation of the experimental arrangement was the use of an arbitrary, low-effort response. This response was selected to ensure that all participants would be able to engage in the target response regardless of developmental level and to minimize the likelihood of skill deficits, response difficulty, or ratio strain masking potential reinforcement effects. The PR schedule was used to increase the likelihood that items identified as reinforcers during the assessment would serve as reinforcers in other contexts. However, future researchers should examine reinforcer efficacy using a variety of different responses. It is possible that different results would have been obtained with more complex tasks or tasks requiring greater response effort. It is important to evaluate socially significant responses that are of interest to educators and parents (i.e., skill acquisition, communication, etc.).

Another limitation of the present study is that experimental procedures were modified during the study. Changes only occurred during the first set of preference and reinforcer assessments (i.e., Luna 2014, Taurus, Ryan, and Darryl). All modifications were made in order to address limitations to the methodology based on observations by graduate teaching assistants who supervised preference and reinforcer assessments. These changes included: changing access delivery from 30 s to 15 s; discontinuing extra-session reinforcement trips; and equating the colors of food items. Specifically, access delivery was changed from 30 s to 15 s during both preference and reinforcer assessments for Luna 2014 during weeks 4 through 6 and access delivery was changed and the color of edible items was also held constant (i.e., all yellow) during weeks 4 through 6 for Ryan and Taurus. Both changes were made during weeks 4 through 6 with Darryl, and access duration remained at 15s for the remainder of all assessments. None of the modifications appeared to systematically affect the results; however, it is crucial to avoid changing features of the experimental arrangement in order to ensure internal validity. The results for these participants must be interpreted with caution, given that the access time differed across initial PSPAs and subsequent MSWOs. No modifications were made during the course of the study for the second set of reinforcer and preference assessments conducted (i.e., Byron's edible-item assessments, Byron's leisure-item assessments, Benji, Brett, Luna's 2015 edible-item assessments, Luna's 2015 leisure assessments, Ollie).

An additional limitation to the present study is evident when evaluating data for Byron and Luna 2015; both participants demonstrated differentiated patterns of responding toward the final weeks of reinforcer assessments. The present study did not include stability criteria. Previous researchers have used a stability criterion in which response rates had to deviate less than 20% from the mean rate of three sessions for three consecutive sessions (Lee at al., 2010). Using a stability criterion would be especially important in the present study given that DeLeon et al. (2001) suggested that brief assessments might become more reliable after repeated administrations. The same might be true of reinforcer assessments for the participants in this study. These data also highlight the possibility that preference instability might also be produced by reinforcer methodology (i.e., initially unstable preferences might actually be product of the experimental arrangement).

A final limitation to the present study was that preference assessments were only conducted for up to ten weeks. Future research should examine preference stability and reinforcer efficacy for a longer period of time in order to determine the maximum amount of time that can elapse between preference assessments while still effectively identifying reinforcers.

Despite the limitations, these data suggest that direct preference assessments are effective tools for identifying preferences of young, typically developing children. Further, these data confirmed previous research suggesting that individuals may demonstrate changes in preference over time, suggesting the need for repeated preference and reinforcer assessments with young children. The present methodology provides one method for evaluating rigid preferences, which may be developmentally inappropriate.

The results also indicated that both PSPAs and MSWOs demonstrate good correspondence and similar predictive validity when assessing edible-item preference. Further, the MSWO might better predict the stimulus that will function as the most efficacious reinforcer. Future research should continue to evaluate the predictive validity of PAs with young, typically developing children in order to increase the likelihood that reinforcement-based interventions will be successful in ECE settings.

References

Akin-Little, K., Eckert, T., Lovett, B., & Little, S. (2004). Extrinsic reinforcement in the classroom: Bribery or best practice. *School Psychology Review, 33,* 344–362.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

Bojak, S.L. & Carr, J.E. (1999). On the displacement of leisure items by food during multiple-stimulus preference assessments. *Journal of Applied Behavior Analysis, 32*, 515–518.

Call, N.A., Trosclaire-Lasserre, N.M., Findley, A.J., Reavis, A.R., & Shillingsburg, A.M. (2012). Correspondence between single versus daily preference assessment outcomes and reinforcer efficacy under progressive ratio schedules. *Journal of Applied Behavior Analysis. 45*, 763–777.

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),* 1–15.

Carr, J., Nicholson, A., & Higbee, T. (2000). Evaluation of a brief multiple-stimulus preference assessment in a naturalistic context. *Journal of Applied Behavior Analysis, 33,* 353–357.

Cicchetti, D.V. & Sparrow, S.A. (1981). Developing criteria for establishing interrater reliability of specific stimuli: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiencies, 86*, 127–137.

Ciccone, F.J., Graff R. B., & Ahearn W.H. (2007). Long-term stability of edible preferences in individuals with developmental disabilities. *Behavioral Interventions*, *22*, 223–228.

Carr, E.G. & Durand, V.M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis, 18,* 111-126.

Cote, C.A., Thompson, R.H., Hanley, G.P., & McKerchar, P.M. (2007). Teacher report and direct assessment of preferences for identifying reinforcers for young children. *Journal of Applied Behavior Analysis, 40*, 157–166.

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*(*4*), 519–533.

DeLeon, I.G., Iwata B.A., Conners J., & Wallace M.D. (1999). Examination of ambiguous stimulus preferences with duration-based measures. *Journal of Applied Behavior Analysis, 32*, 111–114.

DeLeon, I.G., Iwata, B.A., & Roscoe, E.M. (1997). Displacement of leisure reinforcers by food during preference assessments. *Journal of Applied Behavior Analysis, 30,* 475–484.

Didden, R. & de Moor, J. (2004). Preference assessment in toddlers with mild developmental and physical disabilities: A comparative study. *Journal of Developmental and Physical Disabilities*, *16*, 107–116.

Ellison, D. (1997). Compliance training decreases maladaptive behaviors in two people with developmental disabilities: Report of five years' treatment. *Behavioral Interventions, 12,* 183–194.

Findley, J.D. (1958). Preference and switching under concurrent scheduling. *Journal of the Experimental Analysis of Behavior, 1,* 123–144.

Fisher, W., Piazza, C., Bowman, L. & Amari, A. (1996). Integrating caregiver report with a systematic choice assessment. *American Journal on Mental Retardation, 101,* 15–25.

Fisher, W., Piazza, C., Bowman, L., Hagopian, L., Owens, J., & Slevin, I. (1992). A comparison of two approaches for identify reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis, 25,* 491–498.

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*, *24*(3), 537–552.

Greer, B.D., Neidert, P.L., & Dozier, C.L. (in press). A component analysis of toilet-training procedures recommended for young children. *Journal of Applied Behavior Analysis*.

Goetz, E.M. & Baer, D.M. (1973). Social control of form diversity and the emergence of new forms in children's blockbuilding. *Journal of Applied Behavior Analysis, 6(2),* 209–217.

Hagopian, L.P., Long, E.S., & Rush, K.S. (2004). Preference assessment procedures for individuals with developmental disabilities. *Behavior Modification*, *28*(*5*), 668–677.

Hagopian, L.P., Rush, K.S., Lewin, A.B., & Long E.S. (2001). Evaluating the predictive validity of a single stimulus engagement preference assessment. *Journal of Applied Behavior Analysis, 34,* 475–485.

Hanley, G.P., Camilleri, A., Tiger, J., & Ingvarrson, E. (2007). A method for describing preschoolers' activity preferences. *Journal of Applied Behavior Analysis, 40,* 603–618.

Hanley, G.P., Iwata, B. A., Lindberg, J.S., & Conners J. (2003). Response-restriction analysis: I. Assessment of activity preferences. *Journal of Applied Behavior Analysis, 36*, 47–58.

Hanley, G.P., Iwata, B.A., & Roscoe, E.M. (2006). Some determinants of changes in preference over time. *Journal of Applied Behavior Analysis, 39(2)*, 189–202.

Hodos, W. (1961). Progressive ratio as a measure of reward strength. *Science*, *134(3483),* 943–944.

Hursh, S.R. (1980). Economic concepts for the analysis of behavior. *Journal of the Experimental Analysis of Behavior*, *34(2),* 219–238.

Hursh, S.R. (1984). Behavioral economics. *Journal of the Experimental Analysis of Behavior*, *42(3),* 435–452.

Lanner, T., Nichols, B., Field, S., Hanson, J., & Zane, T. (2009). The clinical utility of two reinforcement preference assessment techniques: A comparison of duration of assessment and identification of functional reinforcers. *The Behavior Analyst Today, 10,* 456–466

Lannie, A., & McCurdy, B. (2007). Preventing disruptive behavior in the urban classroom: Effects of the Good Behavior Game on student and teacher behavior. *Education and Treatment of Children, 30,* 85–98.

Layer, S.A., Heal, N.A., Hanley, G.P., & Tiger, J.H. (2008). Determining individual preschoolers' preferences in a group context. *Journal of Applied Behavior Analysis, 41,* 25–37.

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.

Mason, S.A., McGee, G.G., Farmer-Dougan, V., & Risley, T.R. (1989). A practical strategy for ongoing reinforcer assessment. *Journal of Applied Behavior Analysis*, *22*, 171–179.

Michael, J. (1982). Distinguishing between discriminative and motivational functions of stimuli. *Journal of the Experimental Analysis of Behavior, 37,* 149–155.

Northup, J. (2000). Further evaluation of the accuracy of reinforcer surveys: a systematic replication. *Journal of Applied Behavior Analysis, 33,* 335–338.

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(2),* 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.

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.

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.

Raetz, P.B., LeBlanc, L.A., Baker, J.C. & Hilton, L.C. (2013). Utility of the multiple-stimulus without replacement procedure and stability of preferences of older adults with dementia. *Journal of Applied Behavior Analysis, 46,* 765–780.

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., 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.

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.

Rush, K.S., Mortenson, B.P., & Birch, S.E. (2010). Evaluation of preference assessment procedures for use with infants and toddlers. *International Journal of Behavioral Consultation and Therapy, 6*, 2–16.

Salkind, N.J. (2001). Statistics for people who think they hate statistics. London: Sage.

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*(2), 203–220.

Thompson, R., Cotnoir-Bichelman, N., McKerchar, P., Tate, T., & Dancho, K. (2007). Enhancing early communication through infant sign training. *Journal of Applied Behavior Analysis, 40,* 15–23.

Tustin, R.D. (1994). Preference for reinforcers under varying schedule arrangements: A behavioral economic 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(6)*, 439–455.

Zhou, L., Iwata, B.A., Goff, G.A., & Shore, B.A. (2001). Longitudinal analysis of leisure-item preferences. *Journal of Applied Behavior Analysis, 34(2),* 179–184.

Table 1. Preference Stability

Participant	***W***
Benji (Leisure)	.634
Brett (Leisure)	.873
Byron (Leisure)	.357
Byron (Edible)	.497
Darryl (Edible)	.887
Luna 2015 (Leisure)	.657
Luna 2015 (Edible)	.822
Luna 2014 (Leisure)	.121
Ollie (Leisure)	.162
Ryan (Edible)	.768
Taurus (Edible)	.857

Table 2. Correspondence between MSWOs and PSPAs.

Participant	$\boldsymbol{\tau_B}$
Benji (Leisure)	.480**
Brett (Leisure)	.850**
Byron (Leisure)	.435**
Byron (Edible)	.396**
Darryl (Edible)	.796**
Luna 2015 (Leisure)	.458**
Luna 2015 (Edible)	.748
Luna 2014 (Leisure)	.017
Ollie (Leisure)	.230
Ryan (Edible)	.651**
Taurus (Edible)	.826**

** $p \leq .01$

Table 3. Predictive Validity of MSWOs and PSPAs.

Participant	**MSWO (τ_B)**	**PSPA (τ_B)**
Benji (Leisure)	-.070	-.070
Brett (Leisure)	-.335*	-.377**
Byron (Leisure)	-.089	.003
Byron (Edible)	-.122	.016
Darryl (Edible)	-.557**	-.612**
Luna 2015 (Leisure)	-.109	-.213
Luna 2015 (Edible)	-.240	-.217
Luna 2014 (Leisure)	.074	-.271
Ollie (Leisure)	-.035	.192
Ryan (Edible)	-.129	-.257*
Taurus (Edible)	-.269*	-.260

* = $p \leq .05$; ** = $p \leq .01$

Table 4. Descriptive Statistics.

	MSWO	**PSPA**
Break point for highest	M = 4	M = 4.91
preferred item (Rank 1)	Range, -3 to 21	Range, -8 to 21
Break point for low preferred	M = 1.46	M = 1.82
item (Ranked 6 or 10)	Range, -6 to 20	Range, -5 to 18
Break point for items not	M = .5	M = .11
selected (NS)	Range, 0 to 3	Range, 0 to 1

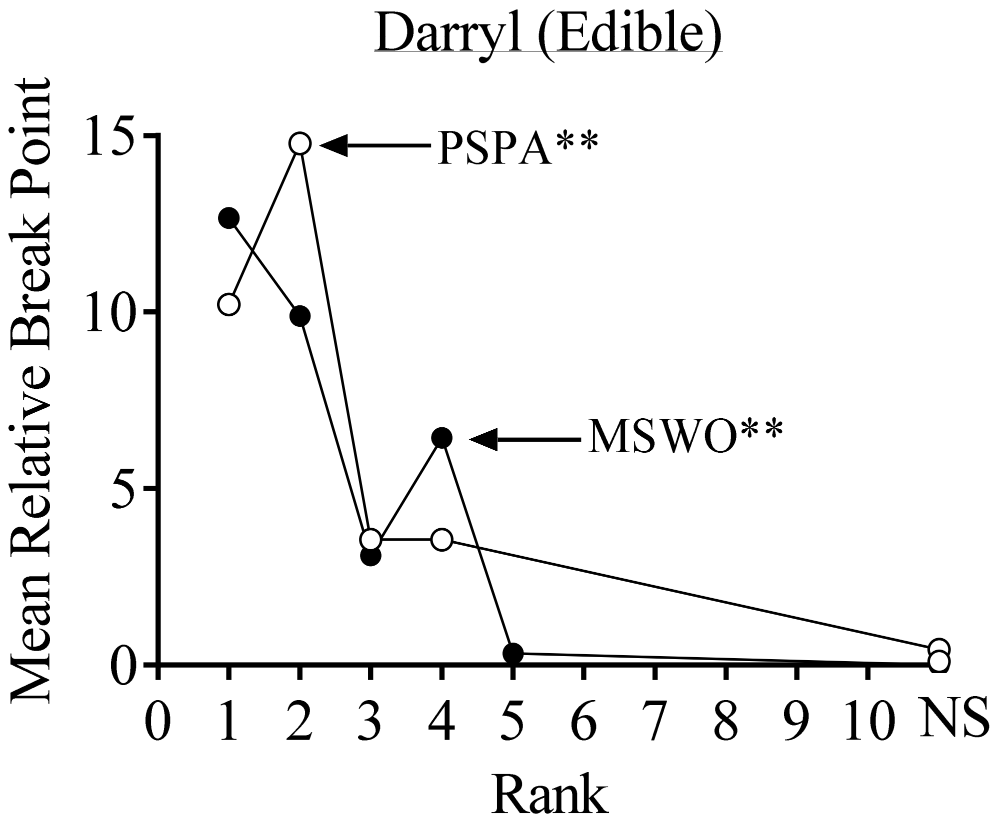


Figure 1. Mean relative break point and preference ranks for Darryl. ** = p ≤ .01;
* = p ≤ .05

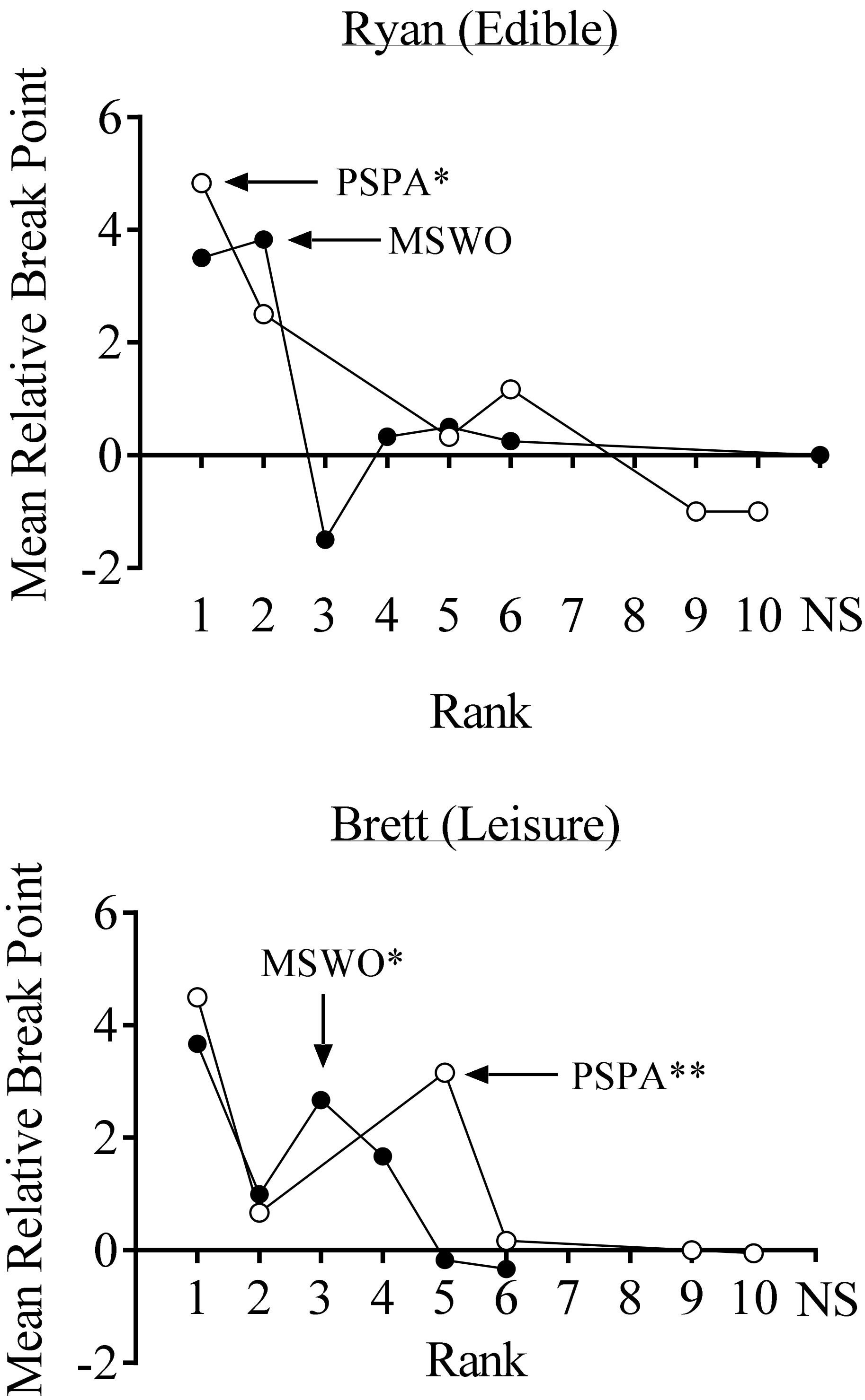
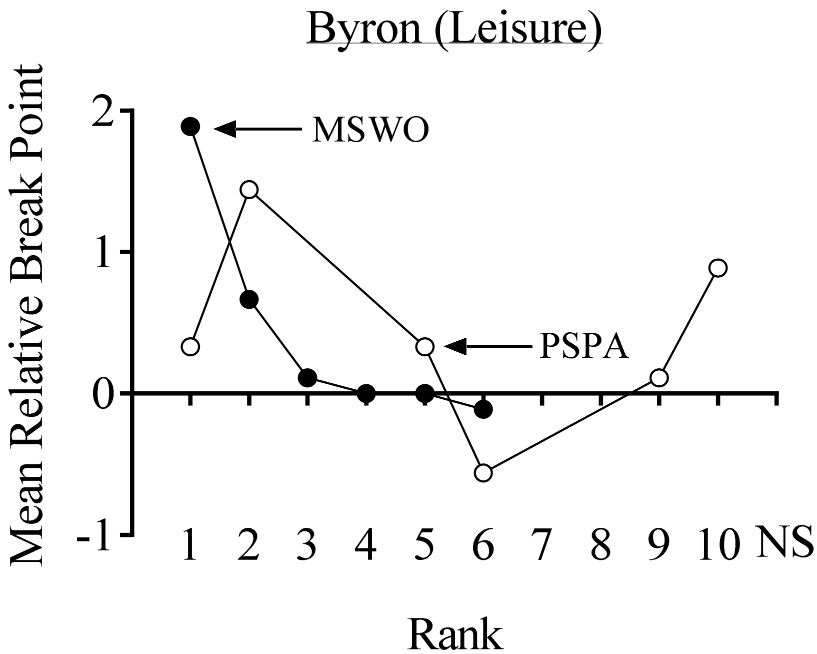


Figure 2. Mean relative break point and preference ranks for Ryan and Brett. ** = $p \leq .01$; * = $p \leq .05$



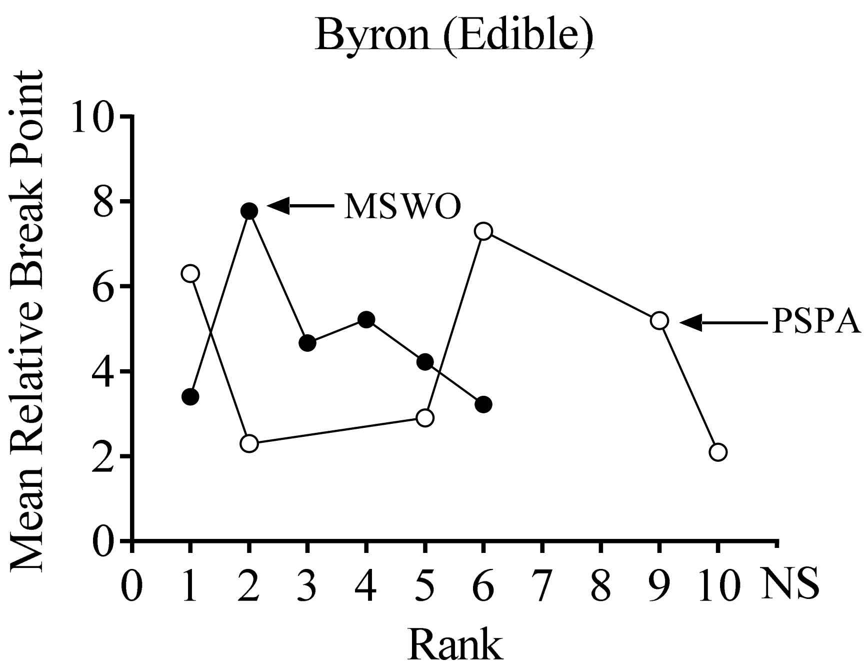
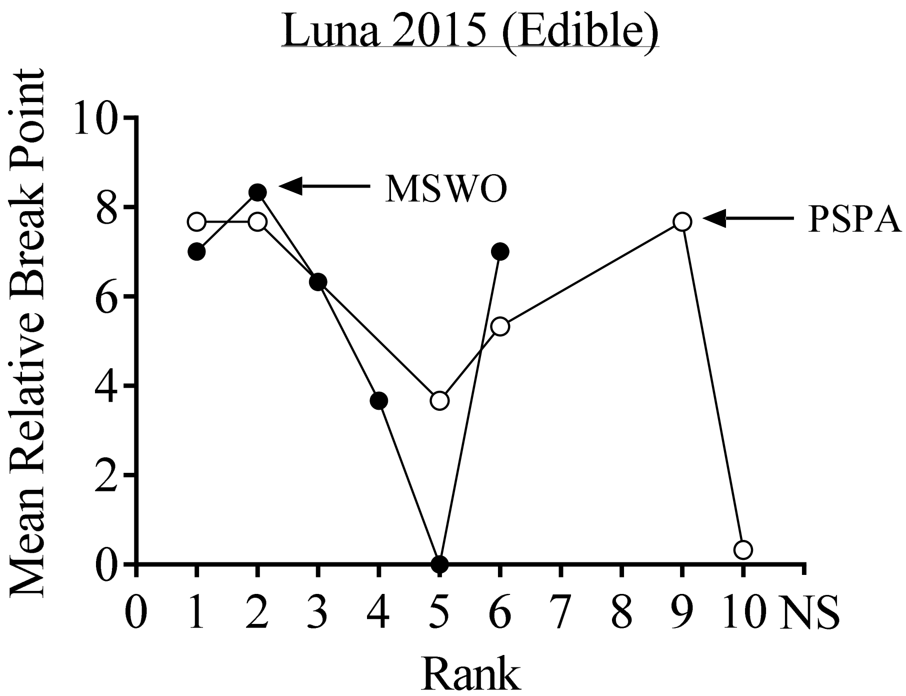


Figure 3. Mean relative break points and preference ranks for Byron. ** = $p \leq .01$; * = $p \leq .05$



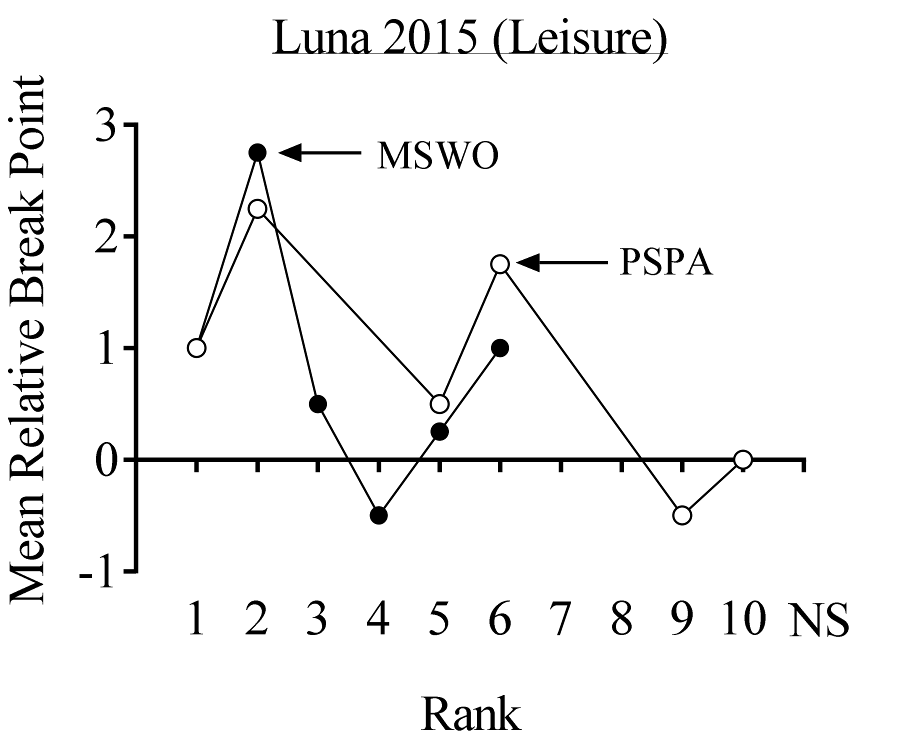


Figure 4. Mean relative break points and preference ranks for Luna 2015. ** = p ≤ .01;
* = p ≤ .05

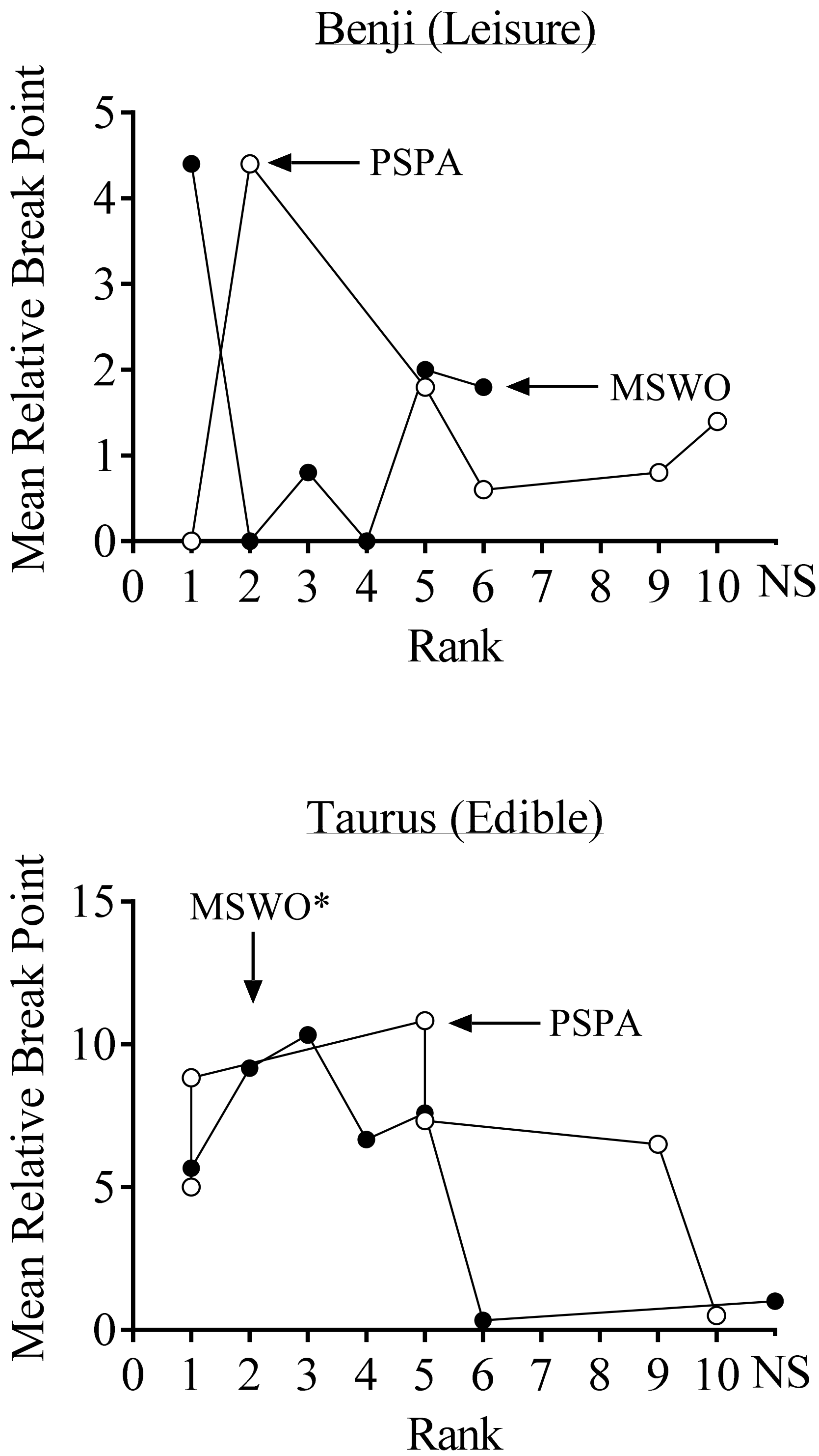


Figure 5. Mean relative break point and preference ranks for Benji and Taurus. ** = $p \leq .01$; * = $p \leq .05$

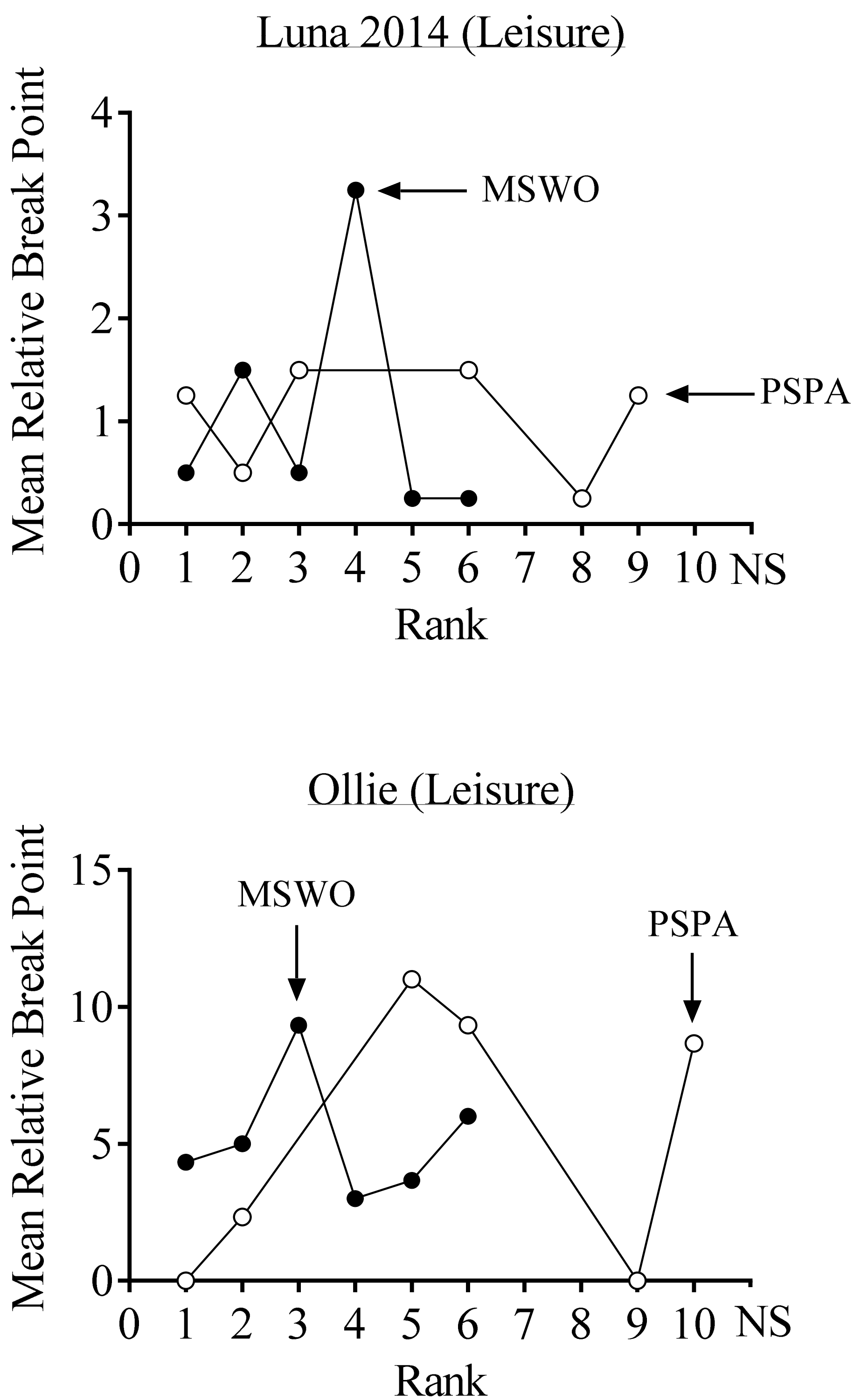


Figure 6. Mean relative break points and preference ranks for Luna 2014 and Ollie. ** = $p \leq .01$; * = $p \leq .05$

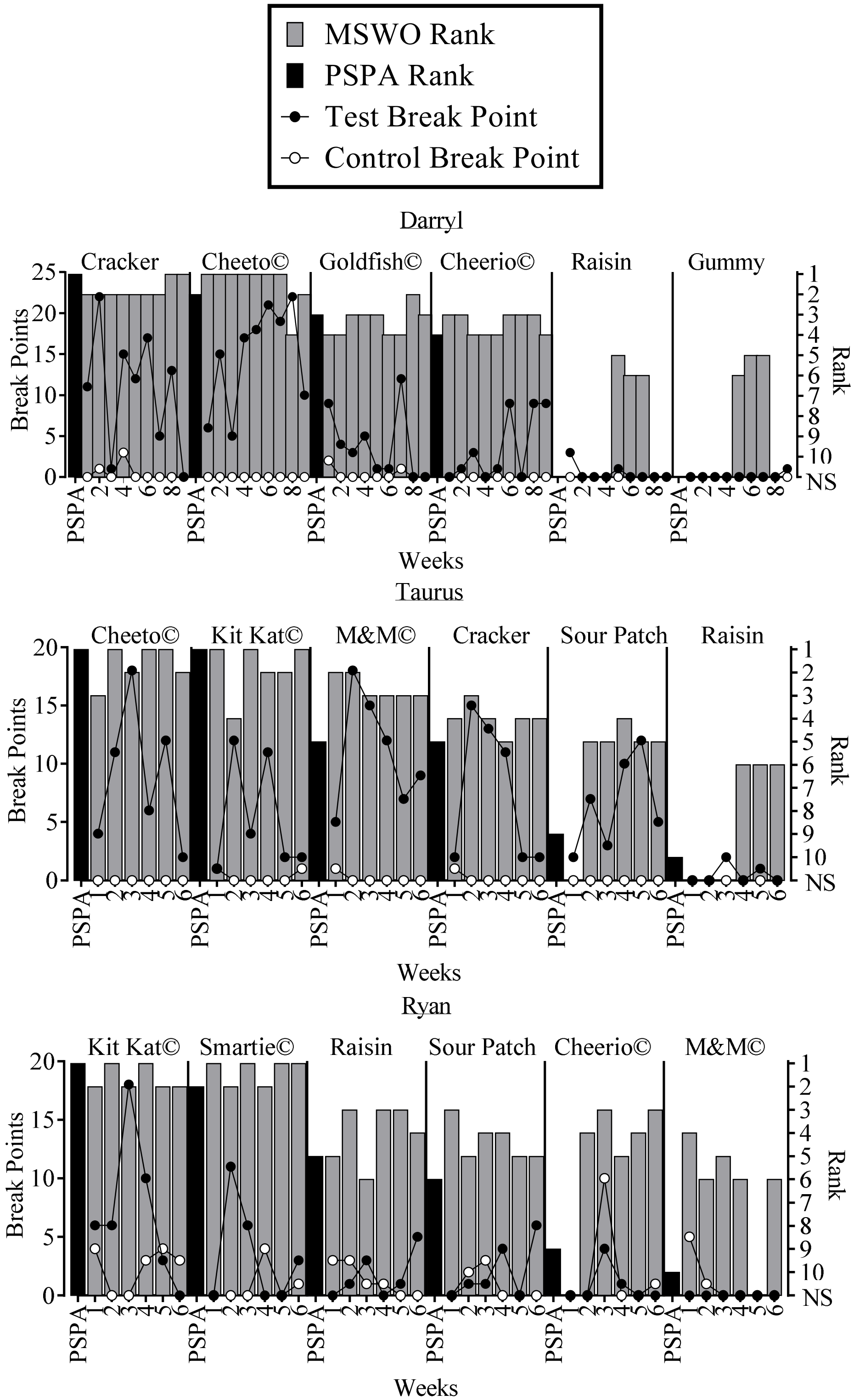


Figure 7. Break points and preference ranks for Darryl, Taurus, and Ryan.

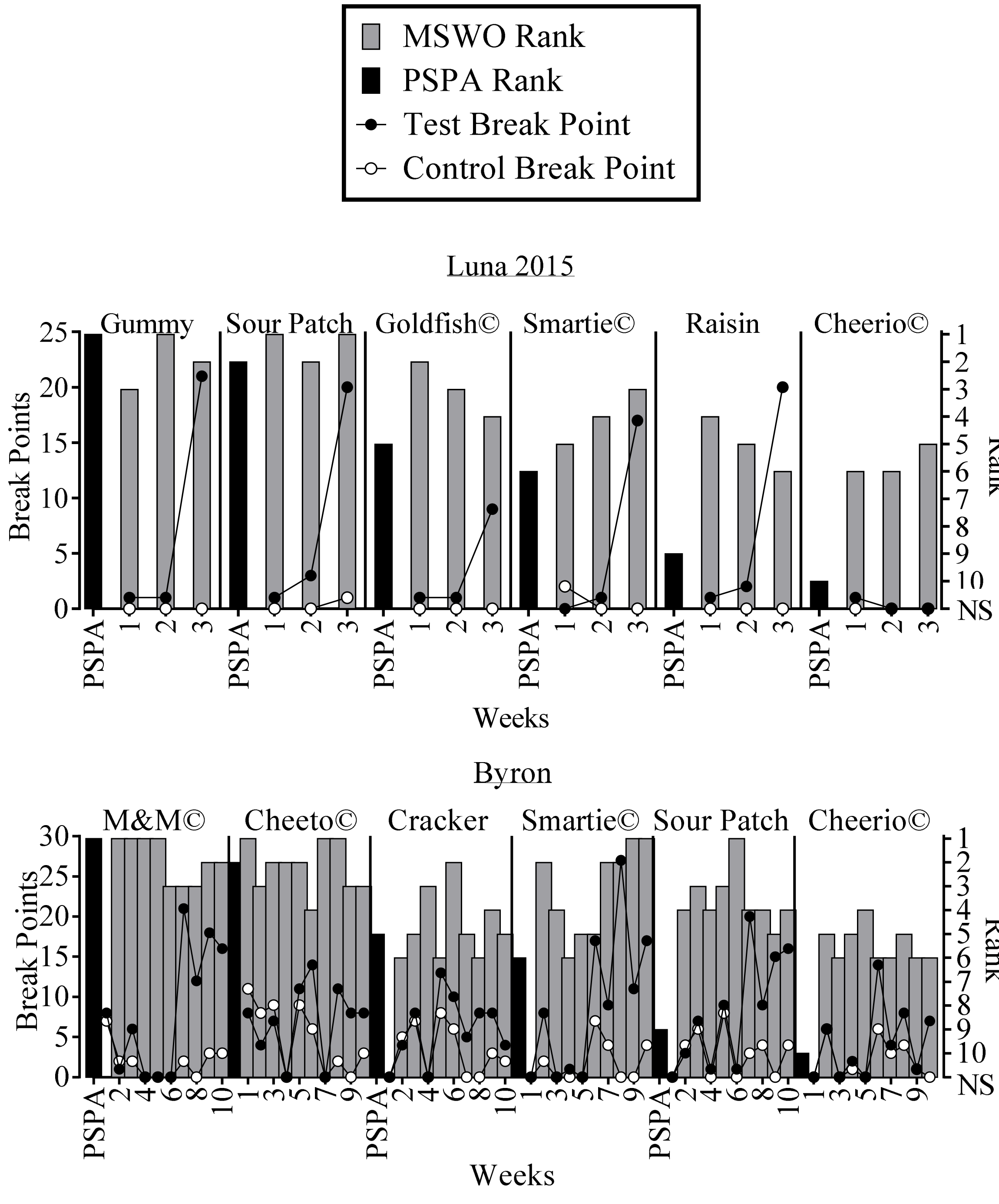


Figure 8. Break points and preference ranks for Luna 2015 and Byron.

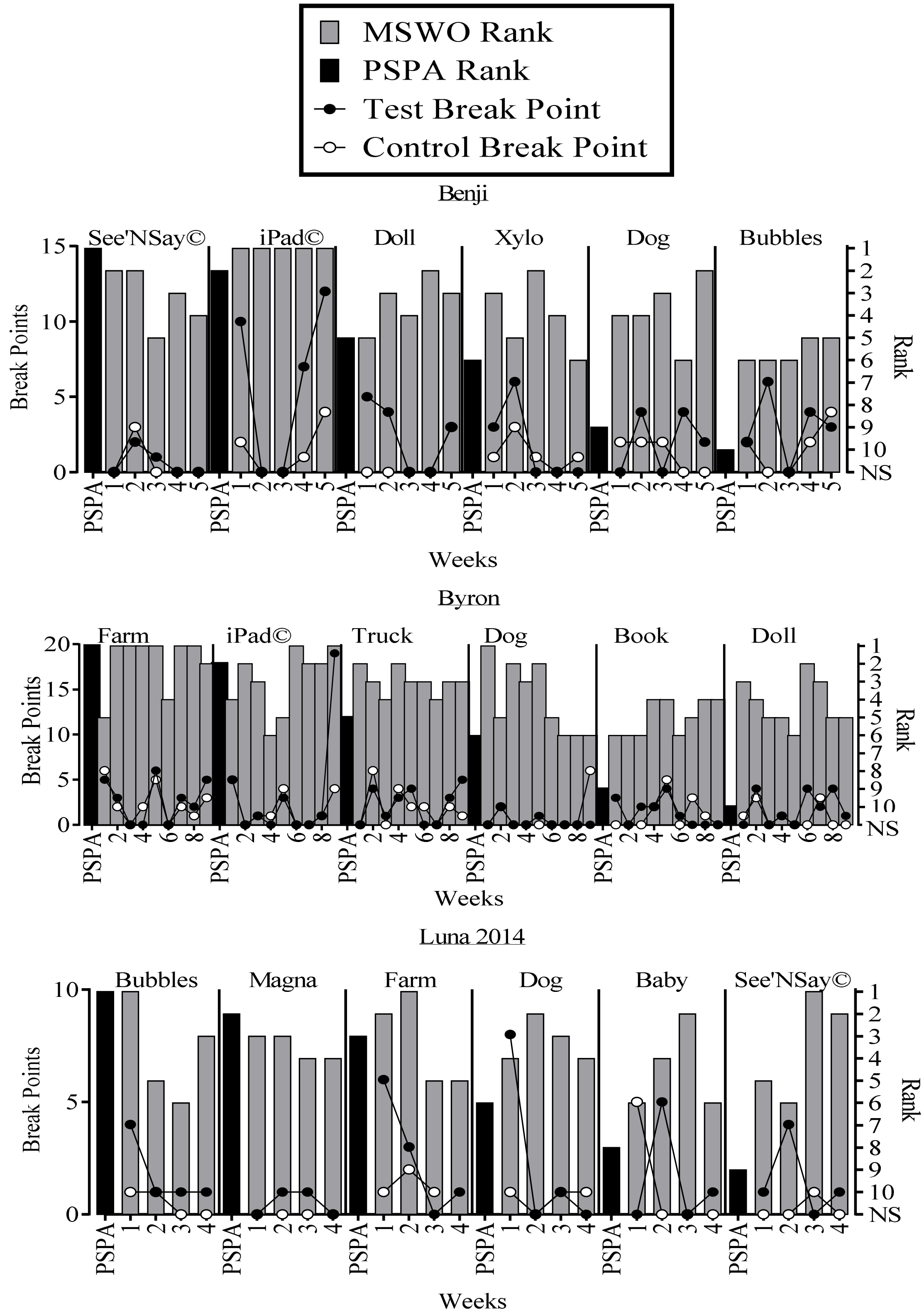


Figure 9. Break points and preference ranks for Benji, Byron, and Luna 2014.

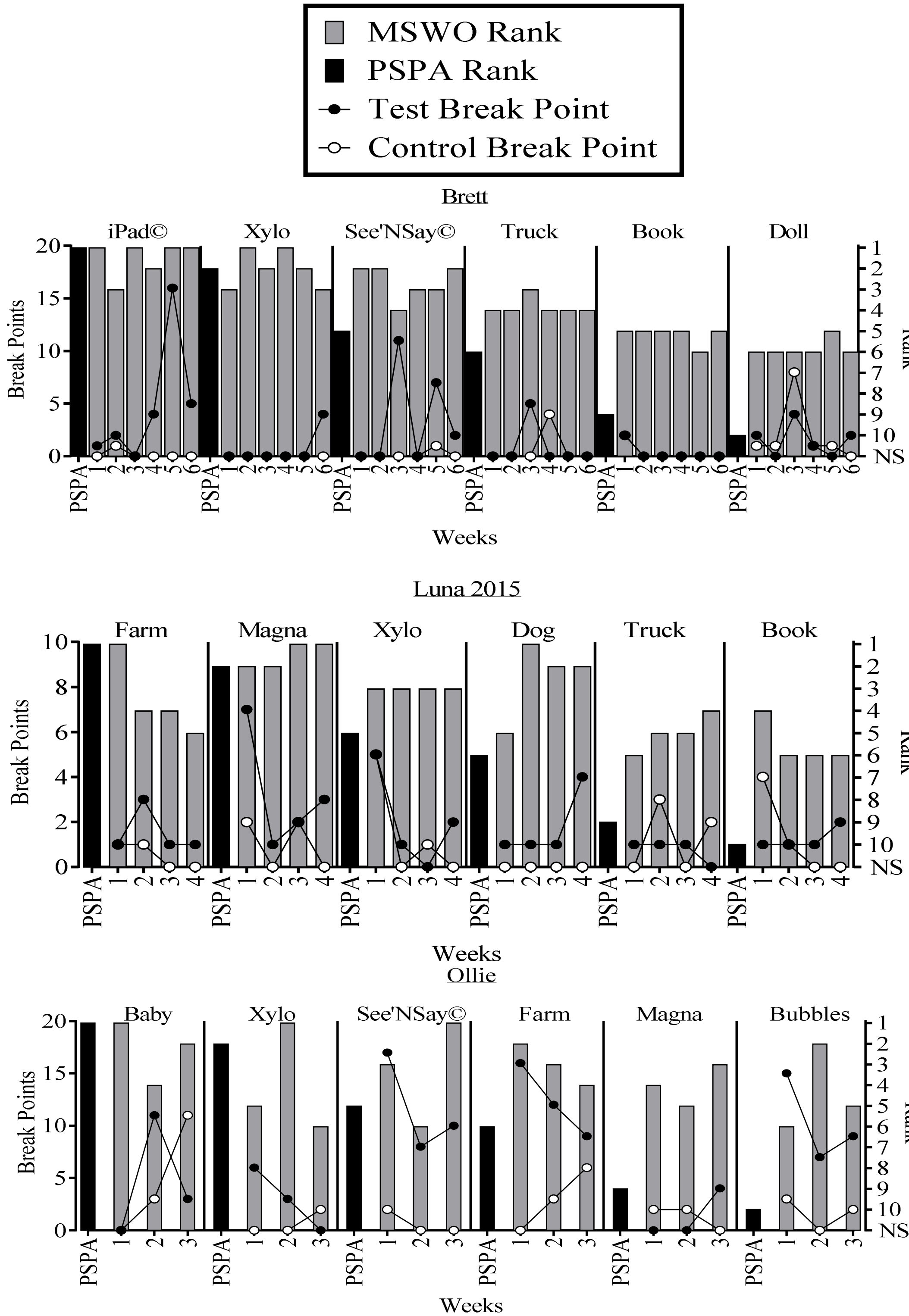


Figure 10. Break points and preference ranks for Brett, Luna 2015, and Ollie.

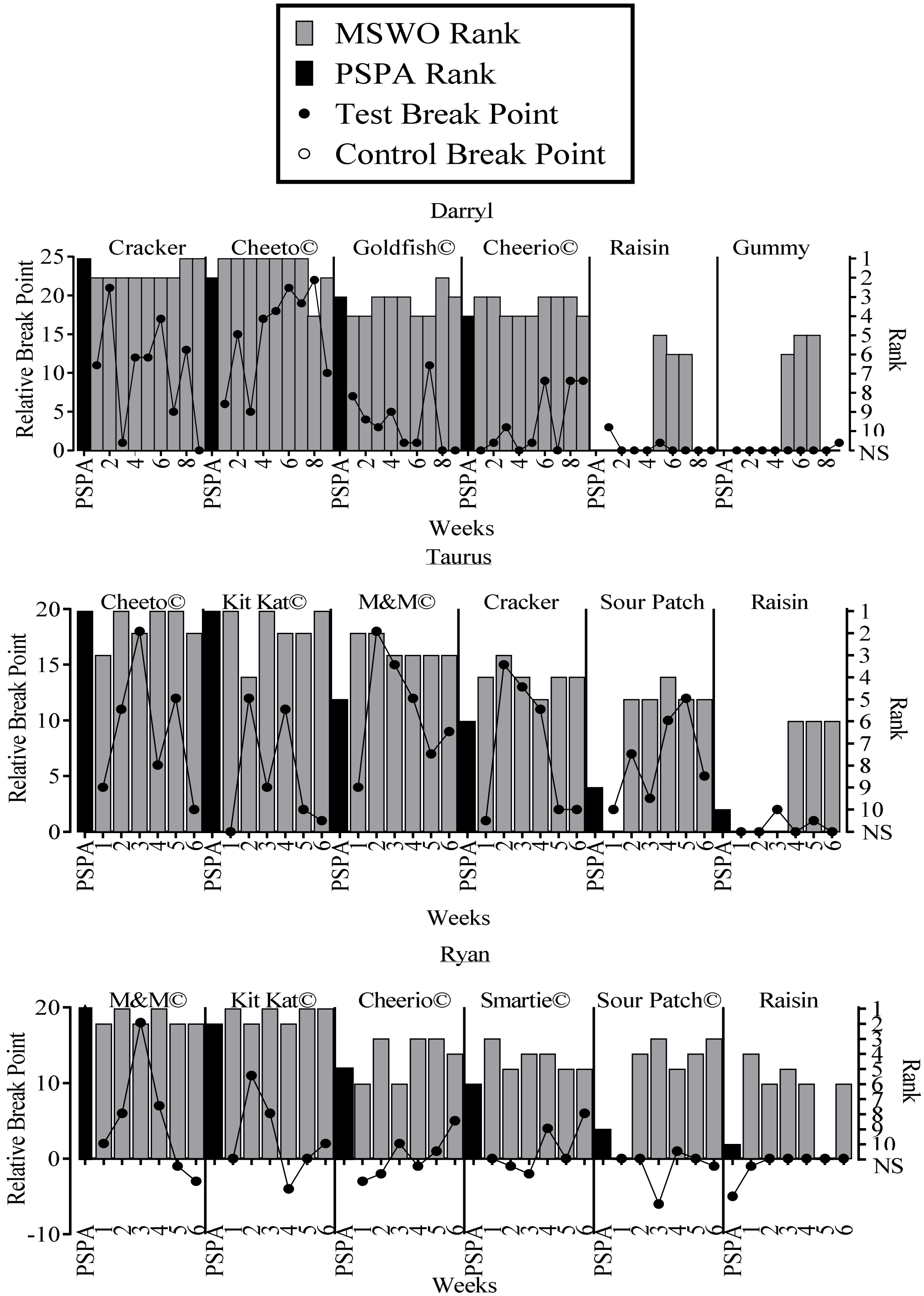


Figure 11. Relative break points and preference ranks for Darryl, Taurus, and Ryan.

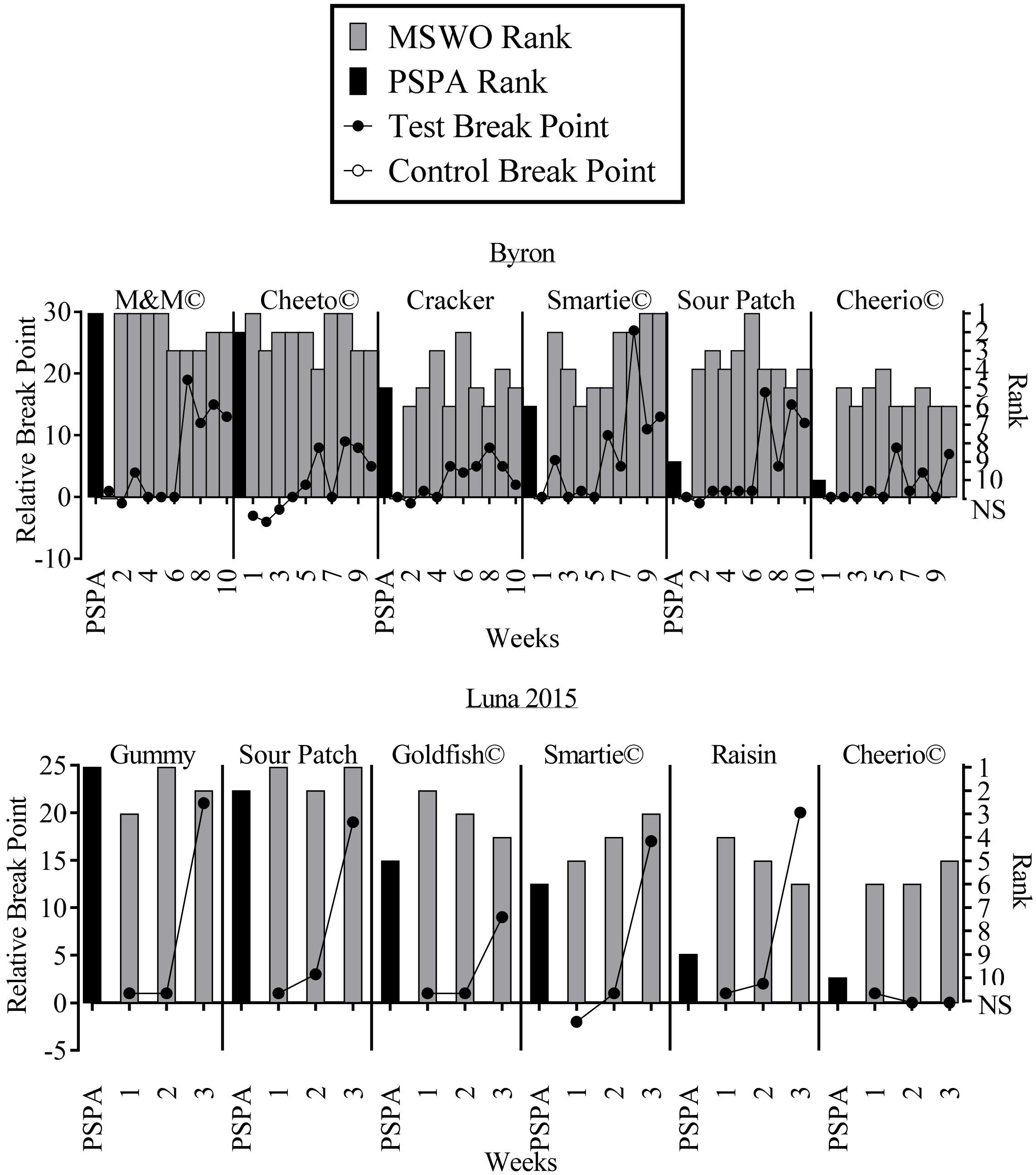


Figure 12. Relative break points and preference ranks for Byron and Luna 2015.

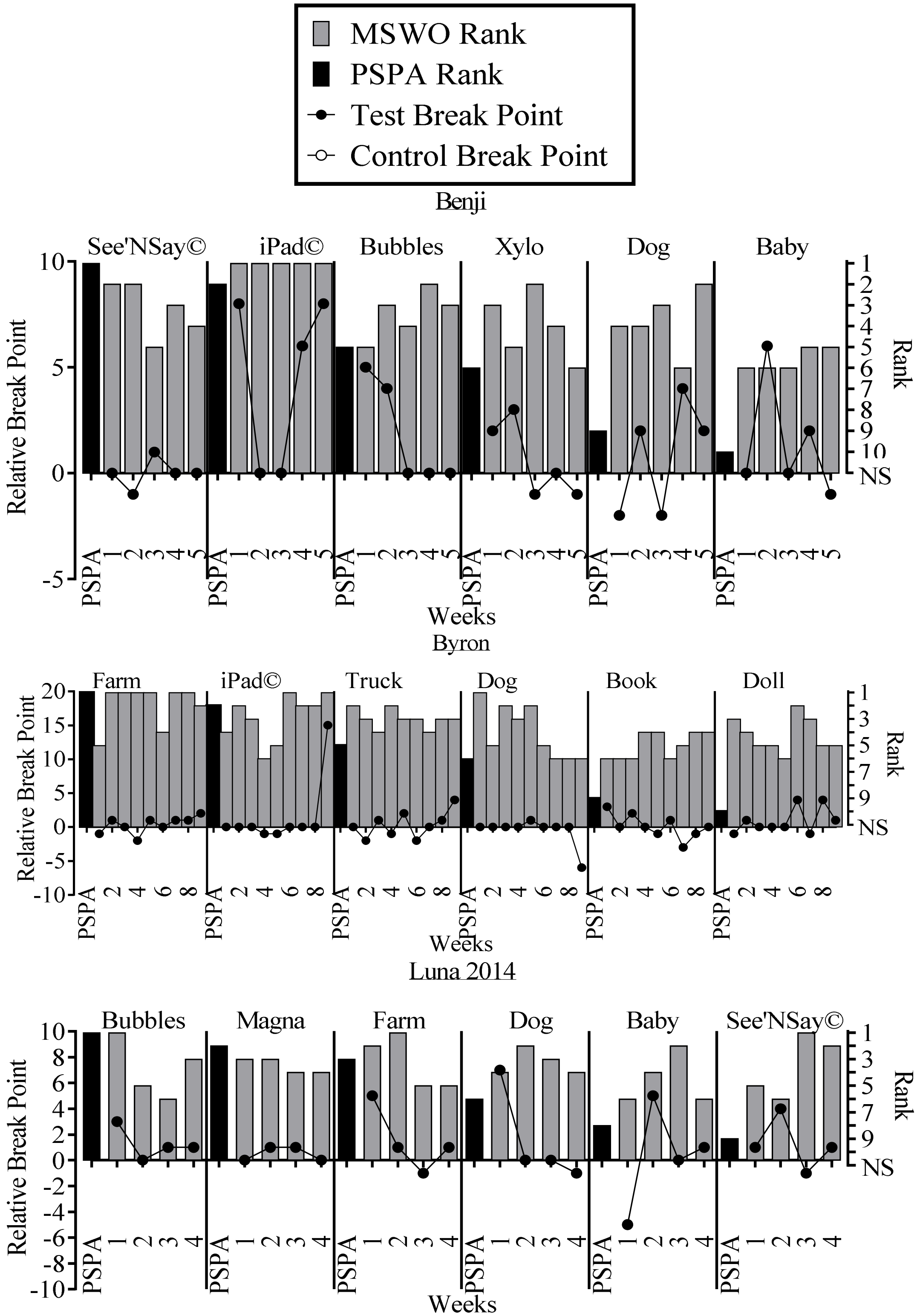


Figure 13. Relative break points and preference ranks for Benji, Byron, and Luna 2014.

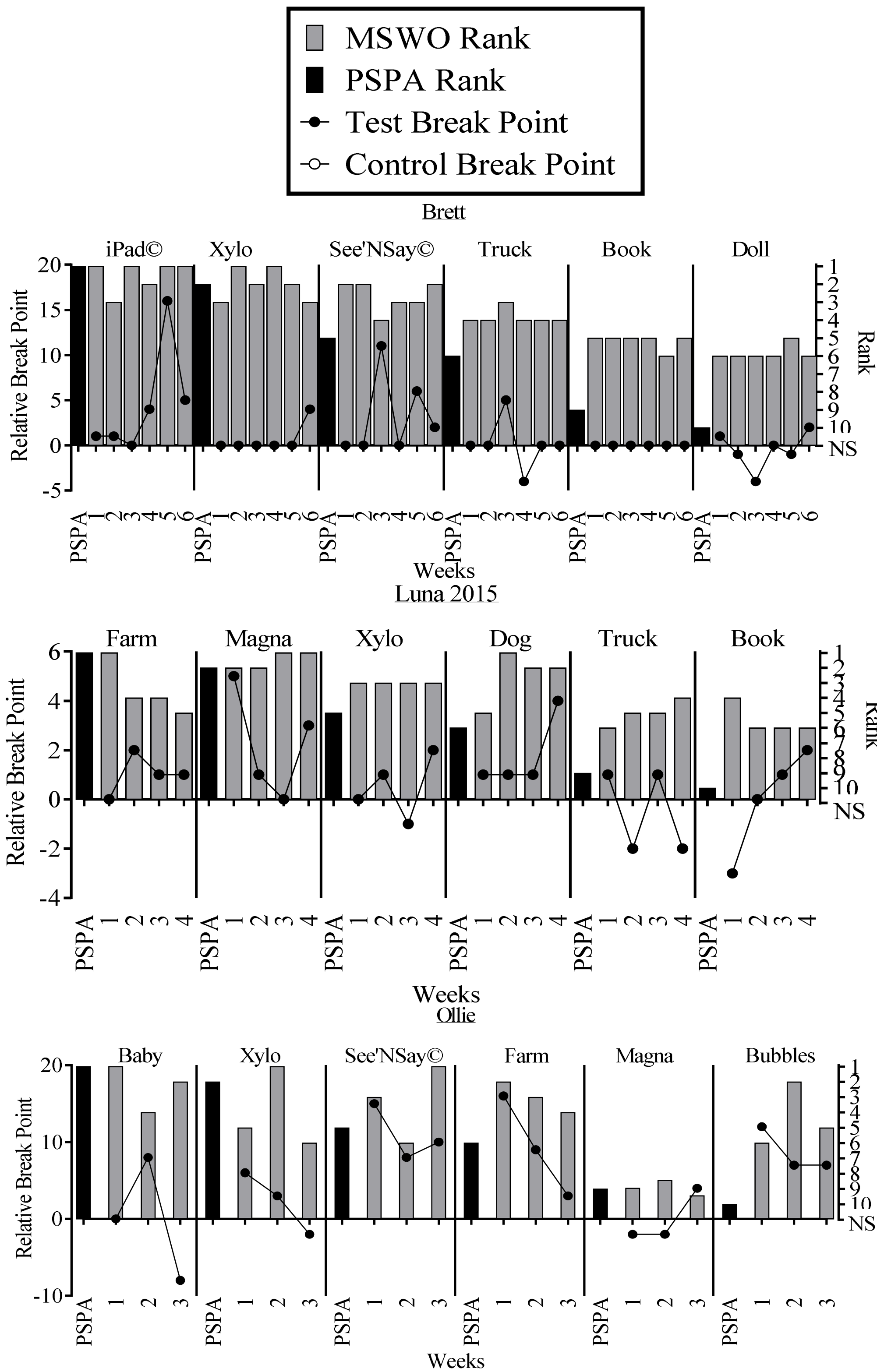


Figure 14. Relative break points and preference ranks for Brett, Luna 2015, and Ollie.