JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1996, **29,** 519–533 NUMBER 4 (WINTER 1996)

EVALUATION OF A MULTIPLE-STIMULUS PRESENTATION FORMAT FOR ASSESSING REINFORCER PREFERENCES

ISER G. DELEON AND BRIAN A. IWATA

THE UNIVERSITY OF FLORIDA

We compared three methods for presenting stimuli during reinforcer-preference assessments: a paired-stimulus format (PS), a multiple-stimulus format in which selections were made with replacement (MSW), and a multiple-stimulus format in which selections were made without replacement (MSWO). Results obtained for 7 participants showed moderate to high rank-order correlations between the MSWO and PS procedures and a similar number of identified reinforcers. In addition, the time to administer the MSWO procedure was comparable to that required for the MSW method and less than half that required to administer the PS procedure. Subsequent tests of reinforcement effects revealed that some stimuli selected in the PS and MSWO procedures, but not selected in the MSW procedure, functioned as reinforcers for arbitrary responses. These preliminary results suggest that the multiple-stimulus procedure without replacement may share the respective advantages of the other methods.

DESCRIPTORS: assessment, developmental disabilities, reinforcer preference

Several methods have been developed by which therapists can identify stimuli that might function as reinforcers for individuals with developmental disabilities. These procedures have included a variety of formats, such as caregiver interviews (Green et al., 1988; Windsor, Piche, & Locke, 1994), single-stimulus approach methods (Pace, Ivancic, Edwards, Iwata, & Page, 1985; Smith, Iwata, & Shore, 1995), and arrangements that involve selection from among concurrently available stimuli (Fehr, Wacker, Tresize, Lennon, & Meyerson, 1979; Paclawskyj & Vollmer, 1995).

One example of a procedure using concurrently available stimuli was described by Fisher et al. (1992). Their paired-stimulus method involved presenting two stimuli simultaneously and allowing participants to choose one. Assessment continued until each item had been paired with every other item. In this procedure, reinforcement effects were predicted based on relative preference among stimuli: Greater preference was attributed to stimuli that were selected on a high percentage of the trials in which they were available. Fisher et al. compared this method to that developed by Pace et al. (1985), in which relative preference was based on the percentage of trials in which a participant approached items that were presented singly. After assessing the same 16 items using both methods, two classes of stimuli were compared in a concurrent-operants arrangement: (a) those selected on at least 80% of trials in both types of assessments (high-high stimuli), and (b) those selected on more than 80% of the single-stimulus trials but selected on 60% or less of the paired-stimulus trials (sp-high stimuli). Results revealed that all 4 participants distributed their responding more towards the response option that produced high-high stimuli as a consequence, suggesting that the paired-stimulus procedure identified reinforcers more accurately than did the single-stimulus approach method. Based

This research was supported in part by a grant from the Florida Department of Health and Rehabilitative Services. We thank Julie Burke, Sonya Fischer, Han-Leong Goh, SungWoo Kahng, Dorothea Lerman, Jodi Mazaleski, Bridget Shore, and April Worsdell for conducting various aspects of the study.

Reprints may be obtained from Brian Iwata, Psychology Department, The University of Florida, Gainesville, Florida 32611.

on these results, the authors suggested that, given an 80% cutoff for approach responding, the single-stimulus method may overestimate preference for some stimuli.

A variation of the paired-stimulus method, involving selection from among more than two concurrently available stimuli, was recently described by Windsor et al. (1994). These authors compared the Fisher et al. (1992) paired-stimulus (PS) method to a multiple-stimulus (MS) method in which the entire array of six items was available on each trial. Typically, each procedure identified the same most preferred item. However, the MS procedure required less overall time (i.e., fewer trials) to administer. A six-item comparison using the MS presentation method required five administrations consisting of 10 trials each (50 trials), whereas the PS method required five administrations consisting of 30 trials each (150 trials) to assess the same six stimuli. Mason, McGee, Farmer-Dougan, and Risley (1989) suggested that treatment effects might be enhanced as a function of conducting reinforcer assessments prior to each session. Thus, one advantage of briefer assessment procedures might be to facilitate more frequent sampling, thereby allowing therapists to accommodate idiosyncratic shifts in client preference.

However, the PS method produced more consistent results across sessions. As measured through Kendall's rank coefficient of concordance, the PS method resulted in a mean coefficient of .631 relative to the .486 produced by the MS method. This finding suggests that although the MS method required less administration time, the more consistent results produced by the PS method may actually make it more time efficient. That is, if consistent results are not produced by the MS method, then several administrations may be required to determine clear and stable preferences, whereas the more consistent results produced by the PS method may indicate that stable preferences can be determined in fewer, or even single, sessions.

In addition, the PS method resulted in a more distinct ranking of the items than did the MS method. That is, because the most preferred items were not available on each PS trial, participants chose from among less preferred items on some proportion of the trials. By contrast, some participants never selected some of the available stimuli in the MS procedure because they always selected the same (most preferred) stimuli. If many stimuli remain unselected, no differentiation regarding relative preference is possible, leading to a conclusion that those stimuli would be ineffective as reinforcers. Of the 48 stimuli assessed for the 8 participants by Windsor et al. (1994), eight (16.7%) were never selected in the MS procedure even though all stimuli were previously identified as "liked items" by caregivers. As such, the MS method may be prone to the production of false negatives—items that may function effectively as reinforcers if tested directly but whose potential is obscured by the continuous availability of the most preferred items.

In this study, we describe a third procedure that was developed as an attempt to combine the best features of the PS format (Fisher et al., 1992) with those of the MS format (Windsor et al., 1994). It is a variation of the MS format in which selections are made without replacement (MSWO): When an item is chosen from the array, it is unavailable during the next presentation. This procedure could be repeated until either all items have been selected or until a criterion is reached indicating that no more selections will be made. The procedure essentially requires individuals to choose among the less preferred items, a feature that was responsible for the more distinct rankings in the PS method. Furthermore, by yoking the number of trials to that in the MS format used by Windsor et al. (1994),

it can be determined if equally useful information can be gathered in a similarly brief assessment. Thus, the present study compared a preference procedure involving multiple-stimulus presentation and selection without replacement to the multiple-stimulus method described by Windsor et al., using the paired-stimulus procedure as the standard for comparison. Comparisons between the two MS procedures included rank-order correlations with the PS procedure, consistency of rank orders across sessions, time of administration, and number of potential reinforcers identified. In a second experiment, stimuli that were never selected during the MS procedure, but were selected on some proportion of trials during the MSWO and PS procedures, were directly tested for their efficacy as reinforcers. In this fashion, we attempted to determine if these stimuli were false negatives (based on the MS assessment) or false positives (based on the MSWO and PS methods).

EXPERIMENT 1: COMPARISON OF PREFERENCE ASSESSMENT METHODS

Method

Participants, Setting, and Materials

Seven adults with profound developmental disabilities participated in the study. All lived at a state residential facility for persons with developmental disabilities and were selected for participation because they had a number of behavioral deficits and could benefit from the identification of additional reinforcers. Rupert was a 26-year-old male with Down syndrome. He walked with an unsteady gait, responded to simple requests, and displayed a few simple signs. Rita was a 25-year-old woman with Cornelia deLange syndrome. She was ambulatory but displayed no receptive or expressive language skills. Bessie, a 43-year-old woman, had a limited but functional vocal repertoire, although she engaged in frequent echolalia. Jeremy was a 43-year-old man who displayed no expressive language skills but who could follow a limited number of requests. Jack was a 39-year-old man who had a limited expressive repertoire, but who responded well to a variety of verbal requests. Carlos was a 45-year-old man who showed no receptive or expressive language skills and had a slight visual impairment. Max was a 32-year-old man with limited communicative skills who displayed periodic episodes of aggressive and disruptive behavior.

All sessions were conducted in one of several therapy rooms at a day-treatment program on the grounds of the facility. The same room was used throughout the study for each participant. Each room contained a table, two chairs, and materials used during the course of the study. Prior to each session, participants were instructed or prompted to sit in one of the chairs; the experimenter sat in the other. Seven items per participant were chosen for presentation during each assessment. The majority of the items were arbitrarily selected by the experimenters without prior knowledge of the participant's preference for those items. A few additional items were selected based on casual observation of preferences and caregiver opinion of preferred and nonpreferred items.

Response Measurement and Reliability

For all methods, a selection response was recorded when the participant made physical contact with one of the presented items. The participant had 30 s to select an item. If the participant made contact with more than one item, the first item contacted was recorded as the selection. If no item was selected within the 30-s period, the trial ended. The procedures following a no-selection trial varied across presentation methods (see below). When a selection was made, the trial ended after the participant received 30-s ac-

cess to the item (leisure stimuli) or after the participant had completely consumed the item (edible stimuli). With a few exceptions, all participants were able to interact independently with the stimuli being assessed. (The exceptions involved turning on electrically operated leisure items. Experimenters assisted the participants in these cases.) In addition to recording stimulus selections, observers also recorded session duration for 66.7% of the sessions. Duration was measured from the moment a participant was first instructed to select on the first trial until the last item was consumed.

Observers recorded selections on data sheets that were customized for each procedure. For most sessions, the experimenter also served as the observer. On 36.2% of the sessions, a second observer independently recorded selections for purposes of interobserver agreement. Agreements were defined as both observers having recorded the same selection or no selection for each trial. Interobserver agreement was computed by dividing agreements by agreements plus disagreements and multiplying by 100%. Observers disagreed on only three trials throughout the study, yielding an interobserver agreement score of 98.9%.

Procedures

Prior to the beginning of the first session, participants were given a sample of each of the edible items and were given 30-s access to each of the leisure items. Subsequently, participants were exposed to one or two assessment sessions per day.

Multiple stimulus without replacement (MSWO). For this assessment procedure, each session began with all items sequenced randomly in a straight line on the table, about 5 cm apart. While a participant was seated at the table approximately 0.3 m from the stimulus array, the experimenter instructed the participant to select one item. After a selection was made, the item was either removed from the immediate area (leisure item) or was not replaced (food item). Prior to the next trial, the sequencing of the remaining items was rotated by taking the item at the left end of the line and moving it to the right end, then shifting the other items so that they were again equally spaced on the table. The second trial then followed immediately. This procedure continued until all items were selected or until a participant made no selection within 30 s from the beginning of a trial. In the latter case, the session ended and all remaining items were recorded as "not selected."

Multiple stimulus with replacement (MS). MS sessions were conducted in a fashion identical to that described above, with one exception. After the end of each trial, the item just selected was returned to the array (in the case of leisure items) or was replaced in the array by an identical item (in the case of edible stimuli).

Paired stimulus (PS). A similar procedure was followed during this assessment, except that only two items were presented during each trial, and the session continued until each item had been paired with every other item (21 total trials per session). Stimulus pairings followed a predetermined order, such that the same stimulus was never presented on two consecutive trials. Stimuli were randomly positioned (left or right side) on each trial. In contrast to both multiple-stimulus methods, failure to select an item (one of the pair) did not terminate the session but simply produced the next trial.

Experimental Design

Each participant was exposed to five consecutive sessions of each procedure, for a total of 15 sessions. The order of procedures varied across participants: MSWO, PS, and MS (Rupert, Rita, and Max); MSWO, MS, and PS (Bessie and Jack); MS, PS, and MSWO (Jeremy); and PS, MSWO, and MS (Carlos).

Results and Discussion

The primary dependent variable consisted of a percentage score indicating the number of times an item was selected over the number of trials during which the item was presented. These percentages are shown in Figure 1 for all participants. From a clinical perspective, the most important outcome of the assessments is the top-ranked stimulus. For 4 of the 7 participants (Bessie, Jeremy, Carlos, and Max), all three assessment methods identified the same reinforcer as the most highly preferred. With some variations in the exact rankings, the MSWO procedure matched the top three ranked items of the PS procedure for 4 of 7 participants (Rupert, Bessie, Rita, and Max) and two of the top three ranked items for the remaining participants. A similar result was found for the MS procedure (top three matches for Rita, Jeremy, Carlos, and Max; two of top three matches for Rupert and Bessie). For all stimuli that were selected at least once during all three assessments, rank discrepancies across assessments typically involved only minor deviations. The largest discrepancy found for any participant was coffee for Jack, which was ranked first by the MSWO procedure and fourth by the PS procedure. No other stimulus (when selected at all) was separated by more than two rank positions across assessments.

The number of stimuli selected at all during each assessment method, and hence identified as a potential reinforcer, can also be determined from Figure 1. For all participants, the MS method produced more unselected items than did the MSWO or PS methods; in two cases (Bessie and Jack), only two items were selected throughout the entire MS procedure. This same pattern was evident when the data are combined for all participants. For both the PS and MSWO methods, at least 90% of the items were selected at least once across participants. For the MS procedure, only 24 items were ever selected across participants. Stated differently, 25 items were never selected in the MS procedure.

A quantitative index of correspondence across stimulus rankings produced by the assessment procedures is provided in Table 1, which shows the Kendall rank-order correlation coefficients between the MS and PS procedures and between the MSWO and PS procedures.[1] All correlations were positive and, in many cases, the correlation was fairly high. Notable exceptions occurred for Jack, for whom the MS/PS and MSWO/PS coefficients were .24 and .52, respectively. In addition, the correlation between the MS and PS procedures for Bessie, .48, was low. Mean correlations for the MS/PS procedures and MSWO/PS procedures were .61 and .72, respectively. For 5 of the 7 participants, coefficients were higher for the MSWO/PS correlation than for the MS/PS correlation. The exceptions were Jeremy and Carlos.

Consistency of rankings was measured using Kendall's correlation of concordance. Table 2 shows the correlation coefficients for each assessment method and the means of the coefficients across participants. Overall, both the MSWO and PS methods produced similar, moderate to high across-session correlations, with respective means of .81 and .83. The coefficients were highly significant for all 7 participants. By contrast, the MS procedure resulted in somewhat lower correlations ($M = .56$). Although the MS correlations were at least moderately high for most participants, the selection patterns by 3 participants (Bessie, Jeremy, and Jack) produced correlations lower than .50.

Measurements of the time required to conduct each procedure revealed that the entire five-session assessment required a mean

[1] Given the number of correlation coefficients calculated in this study, which may inflate the per-experiment error rate, all interpretations based on correlational data should be viewed cautiously.

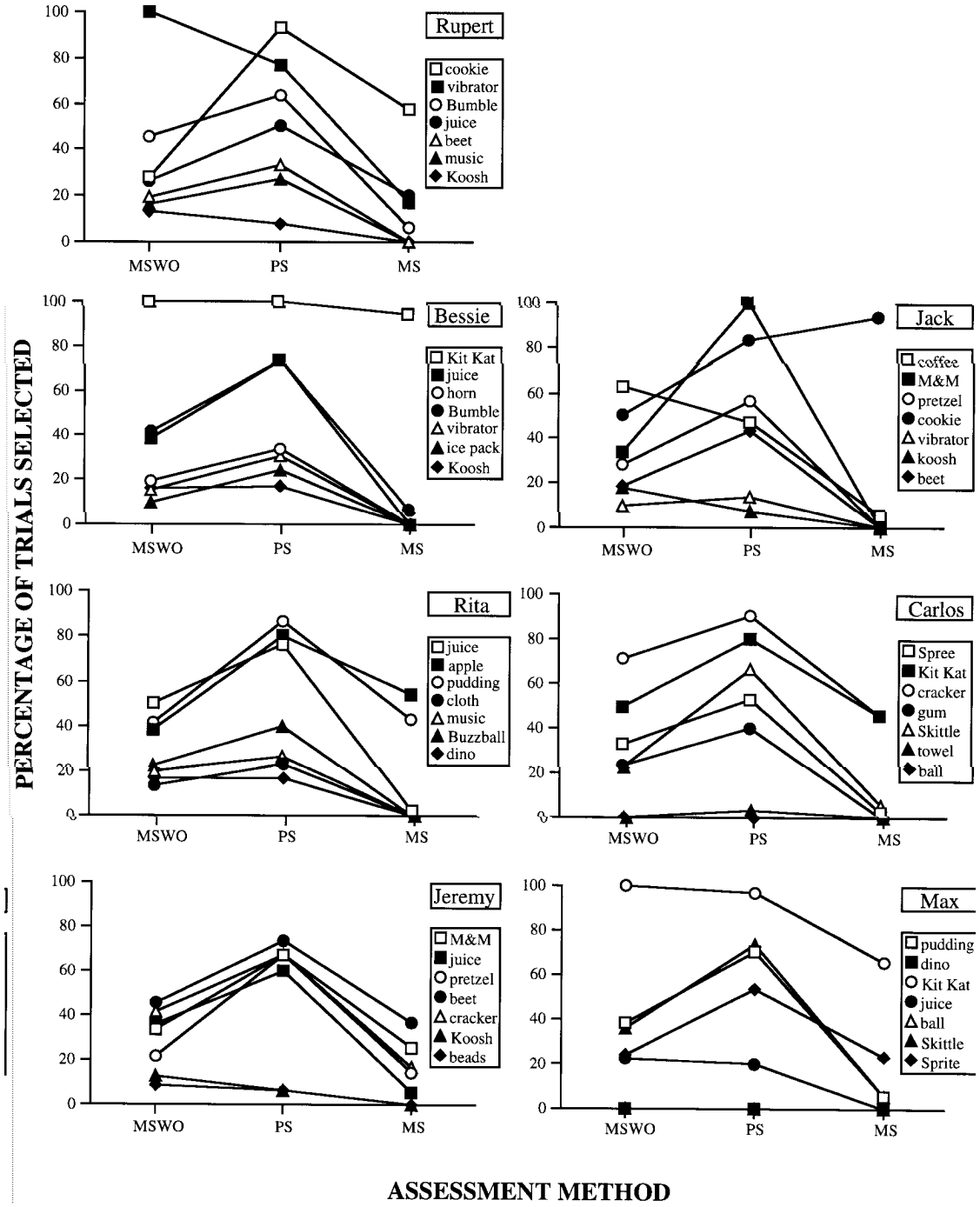

Figure 1. Percentage of trials on which stimuli were selected, when available, by 7 participants for all three assessment methods: multiple stimulus without replacement (MSWO), paired stimulus (PS), and multiple stimulus with replacement (MS).

Table 1
Kendall Rank-Order Correlation Coefficients Between the Multiple Stimulus with Replacement (MS) and Multiple Stimulus Without Replacement (MSWO) Procedures with the Paired Stimulus (PS) Procedure

Participant	MS-PS correlation τ ($N = 7$)	MSWO-PS correlation τ ($N = 7$)
Rupert	.67*	.76*
Bessie	.48	.76*
Rita	.62*	.71*
Jeremy	.81**	.62*
Jack	.24	.52
Carlos	.81**	.78*
Max	.62*	.86**
Group *M*	.61	.72

* $p < .05$. ** $p < .01$.

Table 2
Kendall Rank Coefficients of Concordance Across Administration Rankings for the Multiple Stimulus Without Replacement (MSWO), Paired Stimulus (PS), and Multiple Stimulus with Replacement (MS) Procedures

Participant	MSWO correlation W ($N = 7$)	PS correlation W ($N = 7$)	MS correlation W ($N = 7$)
Rupert	.845**	.79**	.75**
Bessie	.85**	.84**	.42
Rita	.58**	.79**	.64**
Jeremy	.76**	.64**	.47*
Jack	.87**	.91**	.42
Carlos	.87**	.92**	.65**
Max	.90**	.90**	.56*
Group *M*	.81	.83	.567

* $p < .05$. ** $p < .01$.

of 16.5 min, 21.8 min, and 53.3 min for the MS, MSWO, and PS procedures, respectively. These group means are highly representative of data for individual participants. That is, for 6 of the 7 participants, the MS assessment required less time to complete than did the MSWO assessment, which in turn required less time than did the PS assessment. The exception was Carlos, for whom the MSWO assessment required 21.9 min to complete, whereas the MS assessment required 24.6 min. For only 1 participant did either of the two multiple-stimulus procedures take more than half the time to complete than did the PS assessment. This result occurred for the MSWO assessment with Rita, which took 26.0 min relative to the 48.1 min required to complete the PS assessment. However, the average time to complete the MSWO assessment may be somewhat underestimated due to the termination of sessions. As noted previously, sessions for both multiple-stimulus methods were terminated if no selection was made within 30 s. This never occurred during an MS session, but occurred in 18 of the 35 sessions (51.4%) of the MSWO sessions across participants. Of the 18 terminated sessions, five were terminated during the final trial and the remaining 13 were terminated during the sixth trial. By omitting these sessions and recalculating the average assessment time for the MSWO assessments, an estimated mean assessment time of 24.1 min is obtained.

Overall, the three assessment formats produced similar results in identifying the most preferred stimuli. Likewise, similarity across the entire array is reflected in the generally moderate to high, but always positive, correlation coefficients across procedures. However, the MSWO and PS procedures generally produced more consistent rankings across administrations. Both multiple-stimulus procedures required less time to administer than the PS assessment. As stated previously, a procedure that provides similar information in less time affords therapists the opportunity to conduct frequent reinforcer assessments and perhaps to tailor treatment and training programs to idiosyncratic shifts in preference. It is curious that the MSWO procedure generally required slightly more time to complete than the MS assessment, given that fewer stimuli were involved across successive trials in the former. Anecdotal observations during the two procedures provide a plausible explanation. Specifically, in the MS procedure, given that the most high-

ly preferred items were always available, participants often selected those items immediately. By contrast, the most preferred items were not always available in the MSWO procedure, and participants seemed to take more time to scan the array. An accumulation of the brief periods required to scan less preferred items seemed to account for the slight increase in time.

Based on the assumption that frequent selection of an item during an assessment is predictive of that item's reinforcing efficacy, the MSWO and PS procedures predicted that more items in the arrays would function as reinforcers. Of the 49 items evaluated for all participants with the MS procedure, 25 were never selected by participants. According to the MS procedure then, these stimuli would not be predicted to function as reinforcers. By contrast, 21 of those same items (84%) were selected some proportion of the time in both of the other procedures. One item not selected in the MS procedure, M&Ms®, ranked first for Jack in the PS procedure. These items may in fact function as reinforcers, but that sort of prediction (using the MS procedure) might have been obscured by the continuous availability of a small number of highly preferred items.

EXPERIMENT 2: EVALUATION OF REINFORCEMENT EFFECTS

A second experiment was conducted to verify predictions about stimuli that were never selected in the MS procedure. Reinforcement effects were examined by arranging a schedule of contingent delivery for four of these items to determine if they could support levels of responding above baseline. All four items had been selected some proportion of the time in the other two assessment methods. Failure of these items to increase rates of responding on arbitrary tasks over baseline levels would support a conclusion that the MSWO and PS procedures are more prone to the production of false positives than is the MS procedure (i.e., items would be identified as potential reinforcers when, in fact, they do not function as such). Alternatively, increases in responding would suggest that the items can function as reinforcers and that the MS procedure is more prone to produce false negatives.

Method

Participants and Settings

Experiment 2 included 4 participants from the first experiment: Bessie, Rupert, Jack, and Carlos. Each had selected an item during the MSWO and PS procedures that was not selected in the MS procedure. The items tested were (a) fruit juice (approximately 2 oz per delivery in a small paper cup) for Bessie, (b) beets (one quarter of a 1/4 in. thick slice) for Rupert, (c) peanut M&M® candy (one per delivery) for Jack, and (d) chewing gum (one quarter of a stick per delivery) for Carlos. Sessions were conducted in the same rooms used to conduct the initial assessments. These rooms usually contained the same tables and chairs described previously. The sole exception was Jack, for whom a microswitch panel (described below) was mounted on a small movable cart.

Materials and Response Definitions

Different materials were used for each participant, depending on the nature of the response. Bessie placed game pieces (red or black checkers) into a Connect Four game. This involved selecting a game piece from a box and inserting it through the small slot at the top of the game board. Rupert pressed an ink stamper onto a standard-size legal pad. Only responses that left a visible mark were recorded. Jack pressed a response panel that activated a microswitch. The panel was made of plastic and measured 12.7 cm by

20.3 cm. Depression of the panel activated a small red light in the center of the response panel. Only responses that produced activation of the light were recorded. Finally, Carlos picked small wooden blocks off the table and placed them into a plastic bucket.

Response Measurement and Reliability

All sessions lasted 10 min. Data were collected by the same experimenters who administered the assessments in Experiment 1 with the use of hand-held computers. The computers divided each session into 10-s bins. During 34.5% of the sessions, two observers independently collected data on participants' responding. Interobserver agreement was calculated on an interval-by-interval basis by dividing the smaller number of responses observed during an interval by the larger number. These fractions were summed across intervals and then divided by the total number of intervals in the sessions. Using this method, the mean agreement score was 98.3% (range, 91.7% to 100%).

Procedures

Preexperimental training. Bessie and Rupert each had previous histories with their respective responses, thus requiring no training. The other 2 participants, Jack and Carlos, required initial training to shape the response. For Jack, a simple prompt (i.e., "Jack, press this panel") and modeled response were sufficient to get him to begin pressing the panel. This prompt was delivered once at the beginning of each session throughout the experiment. For Carlos, it was necessary to shape the response with the use of modeling and reinforcement. At the beginning of each session and every 30 s thereafter for the duration of each session, an experimenter modeled dropping a block into the bucket. If Carlos imitated the experimenter, a small piece of Kit Kat® candy was delivered. Kit Kat® ranked first or second in all three assessments for Carlos.

Baseline. When participants reliably emitted the responses, baseline sessions began. Participants were seated in front of the response materials and, when required, were prompted to emit the response. No other interactions occurred between the participant and the experimenter. For Carlos, baseline began when the modeling and reinforcement procedure had increased rates of responding to a stable rate. As such, the baseline for evaluation of the chewing gum for Carlos was at the same time an extinction phase relative to the previous training.

Fixed-ratio schedule of reinforcement. After stable rates of responding were observed in baseline, each participant was exposed to a fixed-ratio (FR) 1 schedule of delivery of the item in question. During sessions, the experimenter sat across the table from the participant (or on the other side of the cart from Jack). Contingent upon each response, the experimenter delivered the relevant item by placing it on a plate in front of the participant (Jack and Carlos), on the table in front of the participant (Bessie), or directly into the participant's hand (Rupert).

Experimental Design

For Rupert and Jack, an A-B-A reversal design was implemented, in which A was baseline and B consisted of the FR 1 schedule. For Carlos, a B1-A-B2-A reversal design was used, in which A was baseline, B1 was response training using the candy, and B2 was the FR 1 schedule of delivery of chewing gum. Thus, following the FR 1 phase, 3 of the 4 participants experienced a reversal to baseline. For the final participant, Bessie, no changes in response rates occurred as a result of FR 1 delivery of juice. Therefore, a reversal seemed unnecessary. However, to determine if response rates had not increased due to a ceiling effect, a single-session probe was attempted on the 6th day of the FR 1 condition using Kit Kat® candy as the reinforcer.

Results and Discussion

Figure 2 shows the results for all 4 participants. FR 1 delivery of juice had little effect on Bessie's performance: Response rates during baseline and FR 1 averaged 0.48 per minute and 0.66 per minute, respectively. However, when candy was made contingent upon the same response, an immediate increase in responding to 3.2 per minute was observed. FR 1 delivery of beets produced a slight increase in Rupert's responding. His mean rate during both baseline conditions was 1.29 per minute. During the FR 1 phase, responding increased slightly to a mean of 1.86 per minute. Jack's responding increased noticeably during FR to a mean of 4.54 per minute from a mean of 0.78 per minute during both baseline phases. Finally, Carlos' responding also increased substantially. During initial training using candy, his rate of dropping blocks in a bucket rose to 0.96 per minute for the last seven sessions. After what appeared to be an extinction burst during the first two sessions of baseline, rates then decreased to zero for five continuous sessions. This was followed by an increase in responding to a mean of 0.86 per minute during FR 1 gum and a subsequent decrease during the return to baseline.

Thus, for 3 of the 4 participants, items that had never been selected during the MS procedure but had been selected on some proportion of the trials during the MSWO and PS procedures produced increases in responding when delivered on a contingent basis. Juice failed to produce increases for Bessie, and we determined that failure was not due to a ceiling effect because delivery of candy for the same response resulted in a substantial increase. Only modest increases were observed with Rupert, but the extent to which rates of stamping could have increased is not clear. Rupert typically released the rubber stamp following each response, and it may be that rates could not have considerably increased above those observed in baseline. Unfortunately, no highly preferred stimuli were tested, a preparation that may have shed light on the potential for increase. Nevertheless, substantial increases above baseline rates were observed for both Jack and Carlos. These results suggest that, in at least some cases, items that remain unidentified as reinforcers in the MS procedure do function effectively as reinforcers, and that the MSWO and PS procedures more readily identify those items as reinforcers.

GENERAL DISCUSSION

For the 7 participants in this study, an assessment format in which stimuli were presented in an array and selections were made without replacement (MSWO) identified more reinforcers than a similar procedure in which stimuli were placed back into the selection array (MS). In addition, the MSWO procedure produced results that were similar in terms of overall ranks and consistency of ranks to a procedure in which stimuli were presented in pairs (PS), but in substantially less time. If the similarities between the rankings produced by the MSWO and PS procedures can be deemed acceptable, the MSWO procedure appears to share the respective advantages of the other two procedures.

The extent to which failure to identify reinforcers occurs in the MS procedure is difficult to gauge from these results. Here, it occurred for three of the four items evaluated, but only those four items were tested, and no information is available for items that were not selected in the MS procedure by the other 3 participants. In addition, only one item was tested per participant, and that item was selected because it was the item not selected in the MS procedure that had the highest combined rank from the other procedures. Thus, it is not clear how items less highly ranked might have performed if sub-

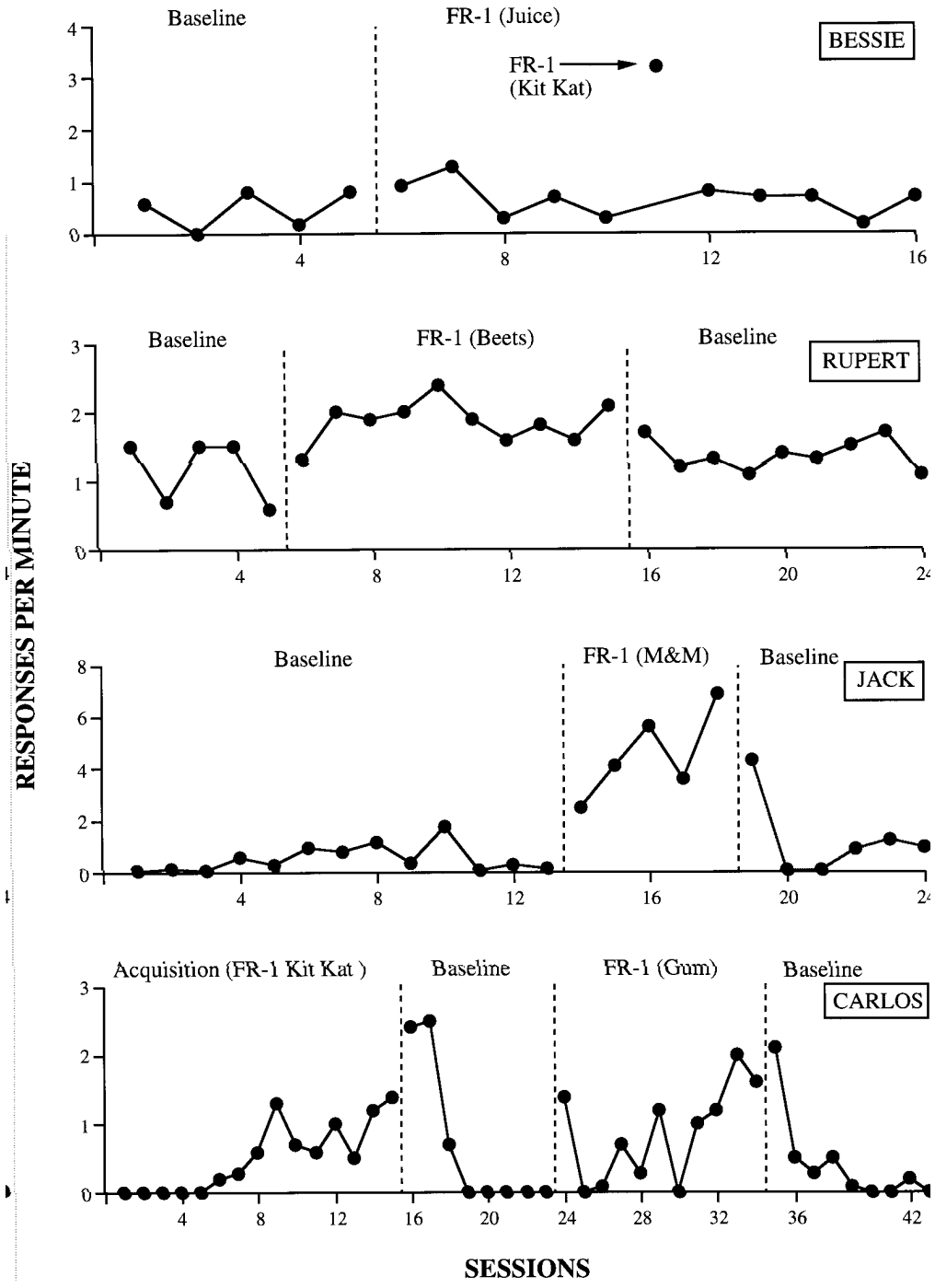

Figure 2. Results of reinforcer potency evaluations for four stimuli selected by participants during the PS and MSWO assessments but not during the MS assessment.

jected to reinforcer testing, and a possible limitation of both the MSWO and PS procedures is that both are too inclusive. Several of the items chosen less frequently in these procedures may not maintain the levels of behavior maintained by the more frequently chosen items. Thus, if one's goal is to find one or two highly potent reinforcers, the MS procedure may more effectively weed out stimuli of questionable utility. However, in terms of evaluating the assessments strictly by how many reinforcers are identified, we suggest that the critical demonstration is not how often failure to identify occurs, but that it occurs at all. If so, and given the results of Experiment 2, we conclude that the MSWO and PS procedures appear to identify more stimuli that are at least minimally reinforcing than does the MS procedure. This feature of the MSWO and PS procedures may be beneficial when attempting to identify a variety of potential reinforcers for individuals reported to have few known reinforcers.

We previously suggested that the MS assessment is prone to the production of false negatives due to the continuous availability of the most preferred items. However, Windsor et al. (1994) used a nearly identical MS procedure that resulted in far fewer items never being selected. Across their 8 participants, only 8 of 48 items were never selected in the multiple-stimulus procedure (all eight were selected at least once in the paired-choice procedure). By contrast, across our 7 participants, 25 of 49 items remained unselected. Several explanations for this discrepancy are possible. In the present study, more items were used per person, perhaps making it more likely that a nonreinforcing item was included in the array. Furthermore, all the items used by Windsor et al. (1994) were selected by caregivers as items that the participants seemed to enjoy. No such criterion was placed upon the array in the present study, making the array more likely to include items that were not reinforcing, thereby producing a higher proportion of nonselections.

Alternatively, the discrepancy might be partially attributable to differences in the assortment of stimuli used in the two studies. That is, Windsor et al. (1994) used all food items, whereas we used a combination of both food and leisure items. Seventeen of the 25 items (68%) never selected in the present study were leisure items, and only 24 of 41 (58.5%) leisure items were selected at all. It may simply be the case that for this population, food items, as a class of reinforcers, are more potent than the stimuli we identified as leisure items. In addition, preferences for leisure items may simply be more idiosyncratic than for food items, and the leisure items used in the present study simply were not items that the participants found reinforcing. These speculations suggest that if therapists are particularly concerned with identifying reinforcing leisure items, food items may have to be omitted from the array.

A general preference for food over nonfood items may help to explain another difference between our results and those reported by others. Specifically, the PS procedure of Windsor et al. (1994) failed to yield any stimulus that reached the 80% selection criterion used by Fisher et al. (1992) to define a preferred item. By contrast, Fisher et al. found at least two stimuli per participant that met this criterion, and all but 1 participant in the present study selected one or more items on at least 80% of trials during the PS assessment. Because Windsor et al. used food items exclusively, whereas Fisher et al. and we used more diverse arrays consisting of both food and nonfood items, two tentative conclusions follow from these observations. First, outcomes obtained using the PS, and perhaps the MSWO, procedure may vary as a function of the composition of the array. That is, when all the items are

similar and highly preferred (as may be the case with food rather than nonfood items), less distinct rankings might emerge. Alternatively, when food items are included with nonfood items in mixed arrays, results of the present study suggest a high preference for food such that selection percentages for food reach higher levels at the expense of nonfood items, resulting in some food items reaching 80% selection. In the present study, all items that reached 80% selection in the PS assessment were food items.

A second conclusion is that it is currently difficult to set a cutoff criterion for the prediction of effective reinforcers. Again, Fisher et al. (1992) defined items that reached 80% selection in the PS procedure as high-preference items, and this criterion was based on that set by Pace et al. (1985) for the single-stimulus presentation method. However, if selection percentages can vary as a function of the array composition, stimuli falling far short of 80% selection may function at least modestly as reinforcers (the stimuli tested in the second experiment, gum for Carlos and beets for Rupert, were selected on 40% and 33.3%, respectively, of the trials during which they were available in the PS procedure). A selection criterion of 80% would have resulted in the rejection of these stimuli as reinforcers. The difficulty of setting a percentage criterion for the MSWO procedure is further compounded by the nature of the measurement. That is, given a perfectly consistent pattern of selection in which one item is always chosen first, another always chosen second, and so on, the second highest ranked item will be chosen on only 50% of the trials in which it was available (never on the first trial, always on the second trial, and unavailable thereafter), thus falling below the 80% criterion set for other assessment methods. It is noteworthy that, for Bessie, the item tested (juice) failed to maintain rates of behavior above those observed in baseline, yet this item ranked third in the MSWO assessment and second in the PS assessment.

Additional research is needed to more clearly determine not only the predictive validity of outcomes obtained from reinforcer-preference assessments but how those outcomes should best be interpreted. Initial steps in this direction have been taken recently by Piazza, Fisher, Hagopian, Bowman, and Toole (1996), who reported that the three highest ranked items or activities from arrays of at least 12, as selected through a PS assessment, appeared to function more effectively as reinforcers than did stimuli chosen from the middle or bottom of the rankings, when compared in a concurrent-operants arrangement. In that study, reinforcer potency was predicted, and later verified, simply by ranking the items based on relative preference. Thus, until more is known about how the use of specific assessment procedures or stimuli might actually determine (i.e., set limits on) the manner in which participants' selections are made, relative ranking might be preferable to a percentage criterion in making predictions about reinforcer efficacy, although rankings may also be subject to procedural influences. That is, the findings of Fisher et al. may be specific to preference assessments that involve relatively large stimulus arrays. For smaller arrays, results may have to be interpreted more cautiously, as evidenced by the data obtained for Bessie in the second experiment, which showed that juice (ranked second and third highest in the PS and MSWO procedures, respectively) did not function as a reinforcer in a free-operant arrangement. This cautionary note is particularly relevant for multiple-stimulus formats, because the array size may be practically limited by the number of items that can be presented simultaneously before position biases, related to the effort of having to reach further for items on the edges of the array, are expressed.

REFERENCES

Fehr, M., Wacker, D., Tresize, J., Lennon, R., & Meyerson, L. (1979). Visual, auditory, and vibratory stimulation as reinforcers for profoundly retarded children. *Rehabilitation Psychology, 26,* 201–209.

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

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

Mason, S. M., 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.

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

Paclawskyj, T. R., & Vollmer, T. M. (1995). Reinforcer assessment for children with developmental disabilities and visual impairment. *Journal of Applied Behavior Analysis, 28,* 219–224.

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.

Smith, R. G., Iwata, B. A., & Shore, B. A. (1995). Effects of subject-versus experimenter-selected reinforcers on the behavior of individuals with profound developmental disabilities. *Journal of Applied Behavior Analysis, 28,* 61–71.

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

Received January 22, 1996
Initial editorial decision March 23, 1996
Revisions received June 20, 1996; July 8, 1996
Final acceptance July 24, 1996
Action Editor, Wayne W. Fisher

STUDY QUESTIONS

1. Describe the paired-stimulus (PS) and multiple-stimulus (MS) assessment formats. What appears to be the main advantage of the MS format and what is a potential limitation?

2. How does the authors' MSWO format differ from the MS format described by Windsor et al. (1994), and what are its potential advantages over the MS and PS formats?

3. The authors compared the PS, MS, and MSWO procedures in the first experiment. What were the dependent measures, and what results were obtained?

4. What was the purpose of the second experiment and what criteria were used in selecting the stimuli?

5. Describe the basic design used in the second experiment and the results that were obtained.

6. Results of the second experiment suggested that the MS assessment may be prone to the production of false negatives, yet Windsor et al. (1994) used a nearly identical MS procedure that resulted in far fewer items never being selected. How did the authors account for these discrepant findings?

7. Based on information contained in the article, describe several ways in which either the assessment methodology itself or the stimuli selected for presentation may affect outcomes obtained during reinforcer preference assessments.

8. Given the results of this study, indicate which assessment format would be preferred for the following purposes, and explain your selection(s).

(a) Frequent (e.g., daily or weekly) selection of a single potent reinforcer.

(b) Frequent selection of several reinforcers that might be varied to prevent satiation.

(c) Infrequent selection of several reinforcers.

Questions prepared by Iser DeLeon and Jana Lindberg, University of Florida