# EVALUATION OF THE MULTIPLE-STIMULUS WITHOUT REPLACEMENT PREFERENCE ASSESSMENT METHOD USING ACTIVITIES AS STIMULI

EDWARD J. DALY, III, NIKKI J. WELLS, AND MICHELLE S. SWANGER-GAGNÉ

UNIVERSITY OF NEBRASKA–LINCOLN

JAMES E. CARR

WESTERN MICHIGAN UNIVERSITY

AND

GINA M. KUNZ AND ASHLEY M. TAYLOR

UNIVERSITY OF NEBRASKA–LINCOLN

The current study examined the accuracy of the multiple-stimulus without replacement (MSWO) preference assessment for identifying preferred common classroom activities as reinforcers with children with behavioral disorders. The accuracy of predictions from the MSWO regarding high, medium, and low stimulus preference was tested by providing contingent access to activities for completing math problems within an independent seatwork format. Overall, there was an interaction effect between preference ranking (high, medium, or low) and number of problems completed. The results confirm and extend previous findings regarding the accuracy of predictions with the MSWO. The findings also reveal, however, some individual differences that may account for instances in which student behavior did not conform to predictions of stimulus preference assessments.

DESCRIPTORS:   math problem completion, preferences, reinforcer, preference assessment

————————————

Stimulus preference assessment methods have a strong empirical basis (e.g., DeLeon & Iwata, 1996; Fisher et al., 1992; Pace, Ivancic, Edwards, Iwata, & Page, 1985; Piazza, Fisher, Hagopian, Bowman, & Toole, 1996) and can assist behavior analysts in identifying a variety of stimuli that can serve as potential reinforcers for use in adaptive or educational training programs or in behavior-reduction programs. Preference assessments are more useful when a choice of more than one stimulus is offered at a time (Fisher et al.), in that providing a choice among stimuli permits differential response allocation that results in the identification of a hierarchy of potential positive reinforcers (i.e., some stimuli are selected to the exclusion of others).

When presenting a choice among stimuli in a preference assessment, those stimuli can be concurrently presented in pairs (Fisher et al., 1992) or in an array of multiple stimuli. For example, DeLeon and Iwata (1996) compared two multiple-stimulus methods, one in which stimuli were replaced after participant choice and one in which the stimuli were not replaced. Their results indicated that the latter method (multiple-stimulus without replacement; MSWO) identified a range of potentially reinforcing stimuli in a relatively efficient manner. Further research on the MSWO assessment method has confirmed and extended these findings (Carr, Nicolson, & Higbee, 2000; Higbee, Carr, & Harrison, 2000).

Although much of the extant research has evaluated the use of preference assessments among individuals with developmental disabilities, prior research has examined the utility of preference assessment methods for identifying

potential reinforcers beyond those individuals with low-incidence disabilities. Preference assessment methods have been examined with adults with schizophrenia (Wilder, Ellsworth, White, & Schock, 2003), children with attention deficit hyperactivity disorder (Northup, George, Jones, Broussard, & Vollmer, 1996; Northup, Jones, Broussard, & Jones, 1995), and children with behavior disorders (Paramore & Higbee, 2005). One characteristic of these populations is that they tend to have relatively sophisticated verbal repertoires, which may have benefits for stimulus preference assessments. For instance, Cohen-Almeida, Graff, and Ahearn (2000) found that young adults who emitted spoken stimulus choices produced similar results to approach responses (i.e., touching stimuli) emitted by those same participants. Furthermore, they found that using spoken choices in preference assessments was less time consuming than using choices that required motor responses.

Although spoken choice of preference, via reinforcer surveys or participant nomination methods, may overidentify items that function as reinforcers with higher functioning populations (i.e., many nominated items do not increase the future probability of behavior when presented contingently; Northup, George, et al., 1996), spoken expression of preference in a stimulus-choice format has been shown to identify high- and low-preference reinforcers accurately for some populations (Northup, 2000; Northup et al., 1995, 1996). As an example, Paramore and Higbee (2005) conducted MSWO assessments to identify edible items that could be used to improve the classroom on-task behavior of adolescent students with behavior disorders. Participants were allowed to make stimulus choices with either words or motor responses (i.e., reaching for an item). The results largely confirmed the predictions generated by the MSWO assessments regarding low-, medium-, and high-preference stimuli, further supporting the accuracy of MSWO assessments and extending the research to verbal expression of preference in a stimulus-choice format.

Students with behavior disorders have greater academic performance deficits than their peers (Anderson, Kutash, & Duchnowski, 2001; O'Shaughnessy, Lane, Gresham, & Beebe-Frankenberger, 2002). In fact, to be eligible to receive special education services according to federal law, students with behavior disorders must display academic underachievement (Mooney, Epstein, Reid, & Nelson, 2003). Thus, improving academic responding and achievement should be a salient target of educational programs for students with behavior disorders. MSWO assessments might prove useful to educators in identifying stimuli to be used as a part of educational programs with these students. Teachers sometimes object, however, to some classes of reinforcing stimuli as being unnatural, contrived, or potentially unhealthy (e.g., edible items, video games) for use in classrooms. By contrast, teachers may be less likely to object to reinforcement programs in which students receive contingent access to more typical school activities (e.g., access to the gym, computer, library, games). Thus, the current study examined whether the results of MSWO assessments using activities and privileges reported to be acceptable by classroom teachers could be used to increase math problem completion in a small-group instructional setting.

Besides targeting academic responding and using activities or privileges as reinforcing consequences, the present study also directly compared the reinforcing effects of low-, medium-, and high-preference activities using a concurrent-schedules arrangement in which students could choose the reinforcing consequence associated with task completion (e.g., choosing between the low- and medium-preference activities). It was expected that students would choose the relatively more preferred consequence over the less preferred consequence (i.e., high preference over medium

preference, medium preference over low preference, and low preference over control) as predicted by the MSWO assessment and that they would complete more math problems to obtain higher preference activities than for lower preference activities.

## METHOD

### Participants and Setting

Four African-American children, who had been classified as students with behavior disorders by their school district, participated in the study. Tiana, Eddie, Dion, and Star were all 9 years old at the beginning of the study and in fourth grade. To qualify for special education services as a student with behavior disorders in their local school district, students had to display situationally inappropriate behavior in school that deviated significantly from peers and that interfered significantly with educational performance (Nebraska Department of Education, 2000). The diagnosis was based on a comprehensive multidisciplinary evaluation that included behavioral observations; behavior rating scales; analysis of the setting; and assessment data regarding social-affective, academic achievement, and developmental functioning. All students spent portions of their day in a special education classroom in which the special education teacher either provided remediation for skills taught in the regular education classroom or taught skills directly based on the district curriculum and the students' existing skill levels.

The stimulus preference assessment and reinforcer evaluation were conducted in the school psychologist's office located in the school. All students were seated at one circular table within the room. The experimenter sat at the table with the students while an independent observer stood behind her.

### Materials

*Stimulus preference assessment.* Prior to the preference assessments, a list of 10 possible activities was generated based on perceived teacher acceptability and feasibility for use in the school setting. The experimenter gave the list to the students' special education teacher and asked the teacher to mark which activities she believed would be acceptable and appropriate for use as potential reinforcers in the classroom. From the original list of 10 items, eight were selected for the assessment based on teacher preference and feasibility of administration during experimental sessions. Items included time playing with a Gameboy, going to the library, walking around the school, drawing on paper, playing cards with the experimenter, playing in the gym, listening to the radio, and using the computer. Thus, eight activity cards (12.7 cm by 20.3 cm) contained the written name of one activity centered in the middle of each card and were presented during each assessment.

*Worksheets.* Math worksheets containing single-digit multiplication problems were used to evaluate potential reinforcers for Tiana, Eddie, and Dion. Worksheets with single-digit addition problems with sums to 18 were used for Star. The probes were generated from a Web site (Wright, 2006).

Crossword puzzles and word-search puzzles also were used during the reinforcement phase of the investigation to serve as a concurrent alternative academic-based activity for which there were no explicit academic demands (i.e., to control for the occurrence of academic behavior in the absence of any programmed reinforcement contingency). Crossword puzzle sheets contained 10 puzzles on a page. Word-search puzzles contained approximately 100 letters (10-letter rows by 10-letter columns) and a 10-word vocabulary bank. Puzzles were taken from a children's book containing crossword and word-search items. Thus, throughout all sessions, students could choose to do puzzles as an alternative to doing math problems. All participants did complete puzzles at one point or another during experimental sessions, but not at levels that interfered with targeted academic behavior.

## Stimulus Preference Assessment

*Response definition and measurement.* The dependent measure consisted of a stimulus selection response, which the observer recorded when a student chose an activity card from a horizontal array of up to eight cards. Observers recorded stimulus selections based on a verbal statement relevant to one of the available activities (e.g., saying, "Gameboy") or pointing or gesturing toward a specific activity card.

*Procedure.* Three separate but procedurally identical MSWO preference assessments were conducted with each student. At the beginning of the session, the experimenter randomized the order of all eight activity cards and then aligned them in a horizontal array before the student. The experimenter told the student that he or she would be able to choose activities for which he or she might like to do schoolwork. Following each selection, the experimenter removed the chosen activity card and rearranged the remaining activity cards so that all the cards to the student's left of the chosen card were shifted one place to the right and the furthermost card on the student's right was moved to the furthermost place on his or her left. This procedure was followed to reduce the possibility that selections might be made based on the position of a card. For example, after the student made the first selection from an array of eight cards and the examiner shifted the activity cards, seven cards remained on the table and the next selection was made. This procedure was followed until there was only one card remaining on the table, with each selection trial removing one more activity card from the array. Participants emitted a selection response on all trials, eschewing the need for further experimenter prompting. Students were not provided access to the activities following each selection. All three sessions were completed on separate days within 10 days.

The experimenter recorded the student's selection in rank order (1 to 7) for the seven activity pairings until the student selected seven activities, at which time the experimenter ranked the last activity as 8 and terminated the session. For each student, the median rank for each activity across the three sessions was chosen as an indication of individual preference for that activity. High-preference activities were the two activities with the two highest median rankings for a particular student. The two activities with the two lowest median rankings were classified as low preference for that student, and the two activities with the fourth and fifth highest median rankings were classified as medium preference.

## Reinforcer Evaluation

*Response definition and measurement.* During the reinforcer evaluation, the primary dependent measure was the number of math problems completed accurately. Math problem completion was calculated for each student during each session. For a problem to be scored as correct, all correct digits had to be aligned in the appropriate columns. Observers also recorded student choice during the reinforcer evaluation. Before each session, students could choose one activity (high, medium, or low preference) or to "do nothing" as the activity they could engage in for meeting a criterion for number of problems completed during each session. The do-nothing activity served as a control response to examine whether students were motivated by the opportunity to escape work or engage in another activity besides math. Alternately, in the event that a student chose activities randomly, selection of the do-nothing activity was as likely as any other condition to be chosen and therefore served as a control condition for undifferentiated response allocation. All sessions were 5 min in length.

*Baseline.* Students met as a small group and were offered the opportunity to complete math worksheets, crossword puzzles, word-search puzzles, or to do whatever they pleased for 5 min. Although not explicitly stated, allowing the students to do whatever they pleased functioned as the do-nothing activity in

baseline. The experimenters walked around the room to supervise student behavior. There were no programmed contingencies in place for work completion, and students completed no problems in three sessions. Students did complete some crossword and word-search puzzles during baseline sessions.

*Reinforcement.* Activities (high, medium, low, or do nothing) were presented in a concurrent-schedules arrangement (Kazdin, 1982). Three conditions (four-choice, three-choice, and two-choice) were randomly alternated until all conditions had been conducted five times. In the four-choice condition, students chose from among one high-preference activity, one medium-preference activity, one low-preference activity, and the opportunity to do nothing (i.e., the control alternative). The three-choice sessions included one medium-preference activity, one low-preference activity, and the control contingency. The two-choice sessions included one low-preference activity and the control contingency.

Fifteen sessions (three per condition) were conducted for each student during small-group lessons. During the sessions, all 4 students were present in the room with the experimenter and an independent observer. At the beginning of the session, students were offered a choice of which reinforcer they would receive contingent on completion of a criterion number of math problems. Students also were given the choice of completing puzzles instead of the math problems. It should be noted that the students were not allowed to change the selected activity once the experimenter initiated the instructional task.

Before each session, the experimenter randomly selected one activity from each preference level to serve as the student's alternative for each level of preference. For example, the experimenter selected either playing in the gym or using the computer as the high-preference activity for Eddie. Two, three, or four stimulus cards (corresponding to the choice arrangements described above, with each card representing one activity or the do-nothing alternative) were displayed before each student to choose from, and activities chosen served as the consequence for subsequent math problem completion. The experimenter then told students they were free to choose completing math problems or puzzles for 5 min. At the end of the 5-min session, the experimenter collected the students' work, calculated math problem completion rates for each student, compared the performance to the selected criterion measure (described below), and provided feedback regarding whether each student earned access to his or her chosen activity. Finally, students who met the criterion number of problems were allowed 7 min of access to the chosen activity (time to transition to and from activities was not included in the 7 min). Students were not allowed to change their choices during the activity.

If a student chose to do nothing, he or she was allowed to do whatever he or she pleased (e.g., talk to a friend) for 7 min under the general supervision of the experimenter. The student was not prompted in any way to engage in any particular activities if he or she chose the do-nothing activity, and students never chose to continue working when this option was selected. Finally, if the student did not meet the performance criterion for math completion, he or she was escorted back to the classroom.

*Reinforcement contingency.* Criteria for earning access to the activities were established based on math problem completion rates obtained during academic probes conducted individually with each student prior to the initiation of baseline. During these academic probes, students completed math problems in the presence of an experimenter for 5 min on three separate days. Unlike the baseline condition, (a) only 1 student was present with the experimenter at a time; (b) the experimenter requested explicitly that students do problems in the individual sessions, whereas students had the choice not to

do problems during baseline and experimental sessions; and (c) no concurrent activities were available (i.e., no puzzles). No programmed consequences were applied during the academic probes.

The results of the academic probes yielded five criteria for reinforcement. Four of the criteria were equal to the highest number of problems completed by each student during the academic probes (64, 90, 104, and 129 correct responses), whereas the fifth criterion was the median of all students' highest performances (97 correct responses). The criterion for reinforcement for all students varied randomly from one reinforcement session to the next. Specifically, at the end of each reinforcement session, the experimenter randomly chose one of five criteria from a box. The purpose of varying reinforcement criteria from session to session was to capitalize on the performance-enhancing effects of indiscriminable contingencies while accommodating individual differences in students' skill proficiency (Freeland & Noell, 2002). Attainable performance criteria were necessary to assure that students could actually earn access to the activities. The contingency, which was based on their actual performance, increased the likelihood that students would contact the reinforcement contingencies over the course of the experiment.

*Interobserver agreement.* During the stimulus preference assessment, the experimenter and an independent observer recorded all selections made by the students in every session for purposes of obtaining interobserver agreement data. Interobserver agreement was calculated by dividing the number of agreements on stimulus selections by the number of agreements plus disagreements and converting the result to a percentage. The agreement between the experimenter and independent observer was 100% across all sessions for all student selections. For the reinforcer evaluation, an independent observer independently scored the math worksheets after the study was completed. Point-by-point agreement was calculated by dividing the total number of agreements for correctly and incorrectly completed math problems by the number of agreements plus disagreements and converting the ratio to a percentage. Mean agreement was 99% when results were aggregated across all student responses (range, 96% to 100%). Agreement data were not collected on student choice during the reinforcer evaluation.

*Treatment integrity.* During the stimulus preference assessment, an independent observer recorded the percentage of steps implemented correctly according to a protocol that outlined the procedures to be followed (available from the first author). Procedural integrity was 100% for all sessions. For all experimental sessions during the reinforcer evaluation, an independent observer recorded the percentage of steps implemented correctly according to a protocol that outlined the procedures to be followed (available from the first author). Treatment integrity was 100% for all sessions across students.

## RESULTS

During the initial three MSWO assessments, the high-preference activities for Tiana were playing in the gym and using the computer; medium-preference activities included going to the library and listening to the radio; and low-preference activities included walking around the school and drawing on paper. For Eddie, high-preference activities included playing in the gym and using the computer; medium-preference activities included walking around the school and listening to the radio; and low-preference activities included going to the library and drawing on paper. For Dion, high-preference activities included playing in the gym and using the computer; medium-preference activities included drawing on paper and playing cards with the experimenter; and low-preference activities included walking around the school and listening to the radio. For Star,

high-preference activities included playing cards with the experimenter and using the computer; medium-preference activities included playing with a Gameboy and playing in the gym; and low-preference activities included walking around the school and going to the library.

The frequency of correctly completed problems during baseline and the three conditions of the reinforcement condition for each student are displayed in Figure 1. Problem completion rates and percentage of sessions in which students received reinforcement (by condition) appear in Table 1. All students demonstrated clear increases in performance relative to baseline when reinforcement contingencies were introduced, regardless of the activities that were available. The baseline condition resulted in no problem completion across all students. Although all students completed problems during the various reinforcement conditions, they varied in the number of problems completed and the consistency with which they completed problems across two-choice, three-choice, and four-choice conditions.

Tiana's performance (Figure 1, top) was fairly consistent across all three conditions, in that she completed an equivalent number of problems across conditions and generally chose the highest preference activity that was available. Specifically, she always chose to work for the high-preference stimulus in the four-choice condition, the medium-preference activity in the three-choice condition, and the low-preference activity in the two-choice condition (i.e., when the only alternative was to do nothing). Tiana met the criterion for reinforcement in 100% of the sessions, and her math computation levels were very similar across all three levels of preference choice, both in terms of the mean number of problems completed per level of preference ($Ms$ = 139, 137, and 134 problems for the high-, medium-, and low-preference stimuli, respectively) and variability (range of standard deviations, 23 to 32) (Table 1).

For the remaining 3 students, there was an apparent interaction between condition and problem completion. Eddie (Figure 1, second panel) always chose to work for the high-preference activity when it was available (i.e., in the four-choice condition). By contrast, when the high-preference activities were not available (i.e., in the three-choice and the two-choice conditions), he occasionally chose to do nothing rather than the medium- and low-preference activities. Overall, he chose the most highly preferred activity that was available on 12 of the 15 sessions (80%). In addition, his problem completion was lower in the two-choice condition than in the other reinforcement conditions. He met the criterion for reinforcement during 100% of the four-choice and three-choice sessions, but met the criterion in only 40% of the sessions during the two-choice condition (Table 1). His mean math problem completion was highest in the four-choice condition ($M$ = 174 problems), followed by the three-choice condition ($M$ = 149 problems), and the two-choice condition ($M$ = 87 problems).

Dion (Figure 1, third panel) worked for the highest preference activity that was available on 10 of the 15 sessions (67%) overall. In the four-choice condition, he chose the high-preference activities three times and the low-preference activity twice. In the three-choice condition, he chose the medium-preference activities two times and the low-preference activity three times. Finally, in the two-choice condition, he chose the low-preference activity in every session. He met the reinforcement criterion during 100% of the four-choice and three-choice sessions (Table 1) and during 80% of the two-choice sessions. His mean number of problems completed was highest and had the least amount of variability in the four-choice condition ($M$ = 133 problems). Problem completion levels were higher in the three-choice condition ($M$ = 124 problems) than in the two-choice condition ($M$ = 113 problems).
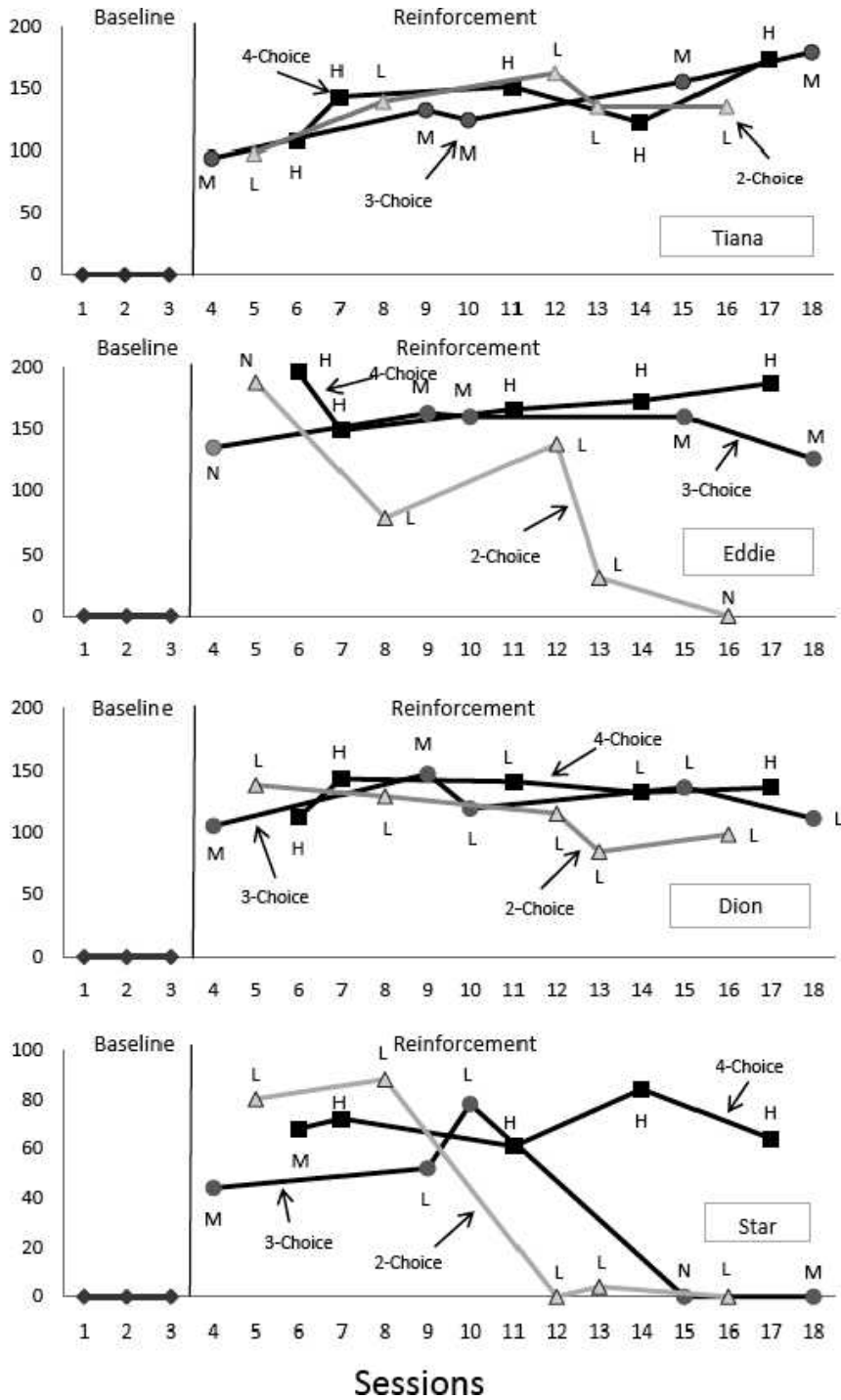
Figure 1.   Number of correctly completed problems across four-choice, three-choice, and two-choice sessions during the reinforcer evaluation for all students. Selection of high-, medium-, low-preference reinforcer or do nothing is indicated for each session by H, M, L, or N, respectively, next to the data point.

Table 1

Problem Completion Rates (Means and Standard Deviations Rounded to Whole Numbers) and Percentage of Sessions Reinforced

| Student Condition | Activity preference choice | | | | |
| --- | --- | --- | --- | --- | --- |
| | High | Medium | Low | Control | % reinforced |
| Tiana | | | | | |
| Four-choice | 139 (25) | | | | 100 |
| Three-choice | | 137 (32) | | | 100 |
| Two-choice | | | 134 (23) | | 100 |
| Eddie | | | | | |
| Four-choice | 174 (18) | | | | 100 |
| Three-choice | | 152 (17) | | 135 | 100 |
| Two-choice | | | 82 (54) | 94 (133) | 40 |
| Dion | | | | | |
| Four-choice | 130 (16) | | 136 (6) | | 100 |
| Three-choice | | 126 (30) | 122 (13) | | 100 |
| Two-choice | | | 113 (22) | | 80 |
| Star | | | | | |
| Four-choice | 70 (10) | 68 | | | 0 |
| Three-choice | | 22 (31) | 65 (18) | 0 (0) | 20 |
| Two-choice | | | 34 (45) | | 20 |

Star (Figure 1, bottom) chose the highest preference item that was available on 11 of the 15 sessions (73%) overall. In the four-choice condition, she chose the high-preference activities four times and the medium-preference stimulus once. In the three-choice condition, she chose the medium- and low-preference activities twice each and the do-nothing consequence once. Finally, in the two-choice condition, she chose the low-preference item in every session. She never met the criterion for reinforcement in the four-choice sessions, met the criterion in 20% of the three-choice sessions, and met criterion in 20% of the two-choice sessions (Table 1). Consequently, her mean number of problems completed was substantially lower than those of the other students. Star did complete more problems on average during the four-choice ($M = 70$ total problems) than during the three-choice ($M = 35$ total problems) or two-choice ($M = 34$ total problems) sessions.

## DISCUSSION

The results of the current study support the utility of MSWO stimulus preference assess-ments in identifying reinforcing activities for use in a typical instructional format (i.e., small-group independent seatwork). Overall, the highest preference alternative that was available in any given choice arrangement was chosen 80% of the time (48 of 60 sessions). In the four-choice condition, students chose the high-preference activities 85% of the time (17 of 20 sessions). In the three-choice condition, students chose the medium-preference activities 65% of the time (13 of 20 sessions). In the two-choice condition, students chose the low-preference activity 90% of the time (18 of 20 sessions).

It is interesting that the low-preference activity was chosen most often when it was the highest preference item available. This condition, however, had only two choices: the low-preference activity or the control (do nothing) activity. Thus, it is possible that a specified activity, even if it is a lower preference activity, was more preferred than no specified activity. In other words, the absence of more reinforcing activities might have established the lower preference activity as a more potent reinforcer. This finding is similar to previous

outcomes in which low-preference stimuli have functioned as effective positive reinforcers under specific circumstances (e.g., Roscoe, Iwata, & Kahng, 1999; Taravella, Lerman, Contrucci, & Roane, 2000).

Additional support for the utility of MSWO assessments in identifying high-preference activities that function as effective reinforcers may be found in the interaction between the various reinforcement conditions and the number of problems completed. In general, the students completed more problems in the four-choice condition than in the other two choice conditions. The combination of increased number of choices and access to higher preference activities may have altered student motivation to complete problems. For example, Eddie and Dion earned the activity 100% of the time in the four-choice and three-choice sessions, but earned it less often in the two-choice sessions (40% and 80%, respectively). These patterns in the data raise questions for future research about whether providing choices among multiple reinforcers may be effective at maximizing the reinforcing value of those reinforcers. That is, it is possible that the choice associated with each of the conditions might have augmented the reinforcing value of the selected activities to some degree (e.g., Fisher, Thompson, Piazza, Crosland, & Gotjen, 1997). Thus, offering choices among potentially reinforcing activities may increase the effectiveness of those reinforcers when used in educational programs.

When using the results of stimulus preference assessments to generate potential reinforcers for academic tasks, differences in target-skill proficiency and the way in which reinforcement criteria are selected may play an important role in how accurately assessment results predict future behavior on those tasks. In the current study, although there was some variability in student responding, all students demonstrated similar levels of problem completion across at least two different conditions. Also, the students met the activity criteria for most of the sessions. It should be noted that Star completed substantially fewer problems than the other 3 students. One possibility for these disparate results is that Star earned fewer reinforcers due to the application of a single criterion to all students. That is, Star's math completion rates were highly variable, particularly in the two-choice and three-choice conditions. Given the discrepancy in performance between Star and her peers from the beginning, the contingency established based on students' prior performance in the academic probes may have produced ratio strain (Cooper, Heron, & Heward, 2007). Future research should investigate the parameters within which predictions based on activity preferences are most accurate, with variations in task difficulty (i.e., response effort), reinforcement criteria, and number and types of reinforcing activities that are presented being salient variables for such analyses.

Although the do-nothing contingency was intended as a control condition, the operant mechanism that occasioned choice of this option is unknown. Specifically, this alternative was provided to control for differential response allocation (i.e., students who selected alternatives in a random manner would be as likely to select this alternative as any others). However, the do-nothing contingency may have functioned as either a negative reinforcement contingency (i.e., students could escape school tasks), a positive reinforcement contingency (i.e., unspecified positive reinforcers that were not programmed by the experimenter were possibly available), or a combined positive and negative reinforcement contingency (if motivation fluctuated from one session to another). Thus, it is unknown how the choice of the control condition affected student responding. Likewise, the current study employed randomly available student activities that were selected by the experimenter before each session (e.g., the experimenter chose which of the two high-preference activities would be available). It is

possible that the variability of access to these activities could have affected response allocation. Conclusions regarding choice of the do-nothing alternative, as well as all choices during the reinforcer evaluation, should be interpreted with caution because interobserver agreement data were not collected on choice during the reinforcer evaluation.

A unique feature of this study was the use of a small-group instructional format for reinforcer evaluation. One limitation of the study is that, although the study was conducted in a typical classroom arrangement (i.e., small-group), it was not actually conducted in a classroom. Future research on methods of stimulus preference assessments should also examine the accuracy of predictions of these assessments in natural settings.

The small-group instructional format of the experimental arrangement may have introduced another important and likely source of variability in performance that affected the outcomes. It was noted that the choices made by Dion and Star might have been influenced by the activity choices of their peers. For example, Tiana's selection of gym time as a high-preference activity may have influenced Star's choice of medium-preference activities (i.e., gym time) during sessions in which those activities were simultaneously available. In future investigations, social variables could be controlled either by allowing students to discuss choices prior to the stimulus preference assessment or through careful selection of activity options prior to presenting them to students when they are in the presence of other students. The reinforcing value of some activities may be over- or underestimated by stimulus preference assessments if they do not take into account the individuals who may be available to engage in the activity with the students. In future studies, researchers could examine possible interaction effects of these variables by analyzing session-by-session student choices as a function of other student choices.

The current study extends research on the effects of stimulus choice arrangements within a concurrent-schedules arrangement during reinforcer evaluations (Piazza et al., 1996). The arrangement used in the current study permitted a direct test of hypotheses regarding the relative reinforcing effects of the activities through direct comparisons of high-, medium-, and low-preference activities in a naturalistic context using typical classroom reinforcers. Varying the number of choices per condition permitted the evaluation of each level of activity preference in which one alternative was always more preferred than other concurrently available options. Given the consistency of the current results along with previous research showing that low-preference stimuli may have reinforcing properties under some conditions (Roscoe et al., 1999; Taravella et al., 2000), researchers and practitioners should not overlook the potential for low-preference items to serve as reinforcers when there are constraints on other, relatively more preferred stimuli. Finally, future research should attempt to identify whether an individual analysis is necessary for each student or whether preference assessment outcomes or group contingencies can be developed that accommodate a diversity of preferences among multiple students.

## REFERENCES

Anderson, J. A., Kutash, K., & Duchnowski, A. J. (2001). A comparison of the academic progress of students with EBD and students with LD. *Journal of Emotional and Behavioral Disorder, 9,* 106–115.

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

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

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). New York: Macmillan.

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29,* 519–532.

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.

Fisher, W. W., Thompson, R. H., Piazza, C. C., Crosland, K., & Gotjen, D. (1997). On the relative reinforcing effects of choice and differential consequences. *Journal of Applied Behavior Analysis*, *30*, 423–438.

Freeland, J. T., & Noell, G. H. (2002). Programming for maintenance: An investigation of delayed intermittent reinforcement and common stimuli to create indiscriminable contingencies. *Journal of Behavioral Education*, *11*, 5–18.

Higbee, T. S., Carr, J. E., & Harrison, C. D. (2000). Further evaluation of the multiple-stimulus preference assessment. *Research in Developmental Disabilities*, *21*, 61–73.

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

Mooney, P., Epstein, M. H., Reid, R., & Nelson, J. R. (2003). Status and trends in academic intervention research for students with emotional disturbance. *Remedial and Special Education*, *24*, 273–287.

Nebraska Department of Education. (2000). *Rule 51: Regulations and standards for special education programs* (Title 92, Nebraska Administrative Code, Chapter 51). Lincoln: Author.

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. (1996). A comparison of reinforcer assessment methods: The utility of verbal and pictorial choice procedures. *Journal of Applied Behavior Analysis*, *29*, 201–212.

Northup, J., Jones, K., Broussard, C., & George, T. (1995). A preliminary comparison of reinforcer assessment methods for children with attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis*, *28*, 99–100.

O'Shaughnessy, T. E., Lane, K. L., Gresham, F. M., & Beebe-Frankenberger, M. E. (2002). Students with or at risk for learning and emotional-behavioral difficulties: An integrated system of prevention and intervention. In K. L. Lane, F. M. Gresham, & T. E. O'Shaughnessy (Eds.), *Interventions for children with or at risk for emotional and behavioral disorders* (pp. 3–17). Boston: Allyn & Bacon.

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.

Paramore, N. W., & Higbee, T. S. (2005). An evaluation of a brief multiple-stimulus preference assessment with adolescents with emotional-behavioral disorders in an educational setting. *Journal of Applied Behavior Analysis*, *38*, 399–403.

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*, 1–10.

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

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

Wilder, D. A., Ellsworth, C., White, H., & Schock, K. (2003). A comparison of stimulus preference assessment methods in adults with schizophrenia. *Behavioral Interventions*, *18*, 151–160.

Wright, J. (2006). *Math worksheet generator*. Retrieved September 1, 2004, from http://www.interventioncentral.org