2006

Evaluation of cost effective preference assessments for use in general education settings

Jennifer Luella Resetar
Louisiana State University and Agricultural and Mechanical College

EVALUATION OF COST EFFECTIVE PREFERENCE ASSESSMENTS FOR USE IN GENERAL EDUCATION SETTINGS

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by
Jennifer Luella Resetar
B.A., University of Louisiana at Monroe, 2001
M.A., Louisiana State University, 2003
December 2006

Acknowlegements

I would like to extend my deepest appreciation and thanks to Dr. George H. Noell for being an excellent mentor and friend. Without his tireless guidance and support, this journey would not have been possible. I would also like to thank Dr. Frank Gresham, Dr. Mary Lou Kelley, and Dr. Kristin Gansle for being members of my dissertation committee and offering invaluable input. Finally, I wish to thank all of those individuals who helped me carry out this study including: Brian Esteve, Luella Resetar, Kashunda Williams, and Robert Wright.

Table of Contents

List of Figures

Abstract

This study investigated the effectiveness of various preference assessments when used in a general education setting. Three separate experiments were conducted to elucidate the usefulness of various forms of preference assessments with the elementary general education population. The first experiment compared the outcomes of a teacher survey, teacher ranking, child survey, and brief multiple stimulus without replacement (MSWO) preference assessment. A Spearman rho correlation found that the indirect assessments either did not or only weakly correlated with the preference assessments. The second experiment utilized a single subject alternating treatments design to compare the reinforcing effectiveness of items identified as the most preferred via the brief MSWO preference assessment and teacher ranking. Reinforcers were assessed using a single operant design that required the child to answer math problems. The average number of digits correctly answered did not significantly differ between the preference assessment reward condition and the teacher ranking condition for all 4 participants. Overall, the number of digits correctly answered was significantly greater in the experimental conditions than the no reward condition. Finally, the third experiment compared the outcomes of a one session, three session, and five session MSWO assessment. A Spearman rho correlation coefficient showed that the five session and three session assessments were strongly related, the one session and three session assessments were strongly correlated, and the one session assessments had a moderate to high correlation with the five session assessments. It was also found that the participant's preferences changed over time.

Introduction and Review of Literature

One of the major goals of applied behavior analysis and teaching is to increase the occurrence of socially significant behaviors (Baer, Wolf, & Risley, 1968). In classroom situations, there are many behaviors that a school psychologist or teacher may wish to increase. For example, he or she may wish to increase reading and math fluency. In managing the classroom, he or she may wish to increase appropriate responding and on task behaviors. In order to increase targeted behaviors, effective reinforcers commonly need to be identified (Ivancic, 2000; Sulzer-Azaroff & Mayer, 1986).

There is a great deal of literature supporting the effect of positive reinforcement on increasing academic skills. It has been demonstrated that contingent reward and instruction both work to increase the number of words correctly read per minute (Noell, Freeland, Witt, & Gansle, 2001; Noell et al., 1998). Rewards in the form of praise and tokens have been shown to be an important part of some effective interventions such as classwide peer tutoring (Delquadri, Greenwood, Whorton, Carta, & Hall, 1986; Greenwood, Arreaga-Mayer, Utley, Gavin, & Terry, 2001). Tokens have also been used successfully to increase correct word recognition and math fact responding (Pavchinski, Evans, & Bostow, 1989). Freeland and Noell (1999) found that both intermittent and continuous rewards selected from a "goody box" contingent on answering more digits correct increased the number of digits correctly answered for math problems.

The Individuals with Disabilities Education Act (IDEA) of 1997 states that an individualized education program (IEP) team should use positive behavioral interventions to address children's problem behaviors (Drasgow & Yell, 2001). The idea of least restrictive treatment supports reinforcement based strategies as a first method of intervention (Jacob &

Hartshorne, 2003). Home-based reinforcement of school behavior has been effectively used to increase a variety of appropriate behaviors and decrease a host of inappropriate behaviors (Atkeson & Forehand, 1979; Barth, 1979). In addition, a review of the differential reinforcement literature supports the effectiveness of these reinforcement procedures when used in classrooms (Sulzer-Azaroff & Mayer, 1986).

Considering the many applications of reinforcement to increase significant school related behaviors, it is important that effective positive reinforcers are identified for use. A positive reinforcer is a stimulus that occurs after a behavior and results in an increase in the future probability of that behavior. Reinforcers are defined by their effect on behavior. If an item is applied or removed after a behavior, but the probability of that behavior does not increase in the future, that item is not a reinforcer. When rewards to be used in an intervention are selected arbitrarily, one runs the risk of implementing an ineffective intervention (Sulzer-Azaroff & Mayer, 1986).

Three general methods to identify reinforcers and preferred items have been suggested: indirect assessments, preference assessments, and reinforcer assessments (Fisher & Mazur, 1997; Hagopian, Long, & Rush, 2004; Ivancic, 2000). Each method has its own advantages and disadvantages. Reinforcer assessments are the most definitive way to determine whether or not a stimulus will reinforce behavior, but they also take the most time and expertise to carry out. Preference assessments do not take as long to complete as reinforcer assessments, but they only identify preferred stimuli. They do not demonstrate that the stimulus itself will increase the future probability of a behavior (Fisher & Mazur, 1997). Finally, indirect assessments, such as surveys and interviews, take the least amount of time to

administer, but studies examining their correlation with reinforcer assessments have not been promising (Hagopian et al., 2004).

Indirect Assessments

Indirect assessments of reinforcers or preferred items usually involve client or caregiver interviews or surveys (Hagopian et al., 2004). These methods of assessments have been developed in order to quickly identify potential reinforcers. Several surveys and interviews have been constructed to identify reinforcers for clients with mental retardation (Fisher, Piazza, Bowman, & Amari, 1996; Matson et al., 1999) and typically developing children (Cautela & Brion-Meisels, 1979; Fantuzzo, Rohrbeck, Hightower, & Work, 1991; Keat, 1974). Child or caregiver nomination has also been considered as a possible method for efficiently selecting items for use in interventions (Green, Reid, Canipe, & Gardner, 1991; Green et al., 1988; Northup, Jones, Broussard, & George, 1995).

Development of Reinforcer Surveys

Several reinforcer surveys have been constructed with the intent of identifying reinforcers for children. When working with children, professionals often wish to implement interventions to increase target behaviors. These surveys were designed to help professionals readily select items that are preferred by individual children for use as potential reinforcers (Cautela & Brion-Meisels, 1979; Fantuzzo, Rohrbeck, Hightower, & Work, 1991; Keat, 1974).

One of the first surveys developed to identify rewards for children was the Survey Schedule of Rewards for Children (Keat, 1979). This survey was based on the Reinforcement Survey Schedule constructed by Cautela and Kastenbaum (1967) for use with adults. Keat presents the survey and offers guidelines for its use, but never describes how it was

constructed. Also, there is no mention of internal consistency, reliability, or validity data. Therefore, the quality of this instrument has been questioned (Cautela & Brion-Meisels, 1979).

In 1979, Cautela and Brion-Meisels developed the Children's Reinforcement Survey Schedule to improve upon Keat's (1974) survey. This new survey was constructed for use with children in kindergarten through sixth grade. All reward categories were developed via a large-scale study with 300 children in which each child was asked to name five things they liked best. Three forms were made. Forms A and B are for children attending kindergarten through third grade, and each form contains 25 items. Children rate each item on a three-point scale, which is displayed using pictures related to the item. Children are read each item and instructed to circle the corresponding picture. Form C was developed for children attending fourth through sixth grade and contains 80 items, which were also rated using a three-point scale. All forms can be administered individually or in groups and take about 20 to 35 minutes to complete.

Cautela and Brion-Meisels (1979) administered the Children's Reinforcement Survey Schedule to 141 students attending grades kindergarten through sixth. Test-retest correlations were significant for all forms when administered three weeks later. The authors listed three uses for the schedule: assessment, research, and intervention development. No validity data were reported, so the treatment utility of this survey is currently unknown.

Matson et al. developed a preference assessment scale for individuals with severe and profound mental retardation in 1999. Development was based on data from 185 individuals with severe or profound mental retardation. Ninety-two items were selected for analysis based on review of the literature and expert nomination. The caregivers of the 185

participants were interviewed and asked about how preferred they believed each of the 92 items were for each individual. For items to remain in the scale, they had to be endorsed by at least 25% of the sample. After inclusion criteria were examined, several items were discarded or reworked and the final item pool consisted of 60 items in four categories: edibles, tangibles, activities, and sensory.

In the second study carried out by Matson et al. (1999), internal consistency, inter-rater reliability, and test-retest reliability were examined. Participants were 100 individuals with severe or profound mental retardation. Staff members who had worked with the participants for at least six months were interviewed. Cronbach's alpha produced a coefficient of .94 for internal consistency. Split-half reliability was also significant with a coefficient of .90. Total scale inter-rater reliability was .99. Finally, test-retest reliability for two assessments taken three weeks apart ranged from .48 to .64 across scales. Internal consistency of the scale was excellent, and appropriate for decision making. Scale reliability was considered adequate by the authors. Scale validity was not assessed, however, so no claims can be made about the scale's ability to identify effective reinforcers.

Many reinforcer surveys have been developed so that reinforcers can easily be identified. Two of the surveys detailed above reported adequate reliability for use. None of the scales previously mentioned were compared to actual reinforcer assessments. Without treatment utility data, the effectiveness of these types of scales remains unknown.

Validity of Child Nomination of Reinforcers

Some authors have taken the research on reinforcer nomination a step further by experimentally examining the ability of these methods to identify actual reinforcers (Northup, 2000; Northup, George, Jones, Broussard, and Vollmer, 1996; Northup et al.,

1995). Northup (2000) noted that it is often assumed that verbal children can identify their own reinforcers, and he tested this assumption in several studies using various methods. It is important to determine whether simpler methods of reinforcer identification such as surveys do indeed correlate with more intensive methods such as experimental assessment of reinforcers. If so, this would provide a quicker and easier way to identify reinforcers for use.

A comparison of a verbal forced-choice questionnaire, child nomination, and direct observation for identifying highly preferred reinforcers for children with attention deficit hyperactivity disorder (ADHD) was carried out by Northup et al. (1995). Ten children between the ages of five and eight participated, nine males and one female. All participants met DSM-III-R criteria for ADHD. For the nomination assessment, each child was shown five toys and asked which was their favorite. The forced-choice questionnaire considered all possible pairings of the five toys. For each pair, the child was asked which toy they would rather play with. Then, each toy was ranked based on the number of times it was selected. Finally, a 10-minute direct observation was carried out. During this assessment, the child was placed in an observation room that contained each of the five toys. The child was told to do whatever he or she wanted until the experimenter returned. Toys were ranked based on the percentage of intervals in which the child played with them.

A 10-minute simultaneous treatments design was used to determine the relative reinforcement value of the toys rated as preferred by each preference assessment method. Each preferred toy was placed on a different table containing identical academic tasks. A control table contained the academic task, but no toy. The child was told that if he or she wanted to earn playtime with a toy on a table, he or she should do the work at that table. The child was allowed to switch tables and was also told that he or she could do nothing. At the

end of the assessment, children were allowed at least two minutes of playtime with each toy associated with the academic tasks chosen.

The authors reported that identified toy preferences were highly variable across assessment methods. All three methods agreed for only one child. The overall agreement between observation and nomination resulted in a .3 correlation. Observation and forced-choice outcomes also produced a .3 correlation. There was a .4 agreement between nomination and forced-choice. The nomination method was the least likely to identify a reinforcer, with agreement between it and the simultaneous treatment reinforcer assessment being .4. Agreement between the forced-choice and observation assessments and the simultaneous treatment reinforcer assessment was comparable, with agreements of .7 and .6 respectively. Northup et al. concluded that the treatment utility of these different assessment methods might not be equivalent for children with ADHD. They suggested future research consider further development and evaluation of verbal reinforcer assessment methods and replication with more typical children.

Northup et al. (1996) compared the utility of a reinforcer survey, a verbal stimulus choice questionnaire, and a pictorial choice questionnaire for identifying reinforcers for children with ADHD. Participants were four children, two males and two females, between the ages of six and nine. All children met DSM-III-R criteria for a diagnosis of ADHD. A revised form of the Child Reinforcement Survey was used to identify fifteen stimuli as potential reinforcers for each child. The stimuli were then organized into five categories: edibles, tangibles, activities, attention, and negative reinforcement. A control category was also added that contained an item from the five categories that was rated as “not at all” liked on the survey. The survey consisted of nine stimuli from the five categories. Each stimulus

was named and the child was asked to report whether they liked it a lot, it a little, or not at all. A percentage score was then calculated for each category by dividing the total sore of ranking by the total possible score of each category. High preference categories were those scoring 75% or greater. The verbal stimulus choice questionnaire paired each category with every other category and asked the child to report which category of stimuli they would rather receive. A percentage score was calculated by dividing the number of times a category was chosen by the number of times it was presented. High preference categories were those scoring 75% or greater. Finally, a pictorial stimulus choice questionnaire was administered. It was identical to the verbal choice questionnaire, except it utilized token coupons to represent categories instead of verbal labeling. The token coupons were different colors and had symbols representing the category they stood for.

After the preference assessments were completed, a reinforcer assessment was carried out to determine which categories of stimuli reinforced work. The child was presented with a coding task and the baseline number of items completed was determined. Then, the child was presented with each of the five coupons separately and told that he or she could earn as many coupons as he or she wanted if he or she coded a criterion number of squares determined by baseline performance. Finally, all preference assessments were readministered to determine reliability.

Results showed that overall, the verbal and pictorial stimulus choice questionnaires had greater utility, with total accuracy being 70% for the former and 80% for the latter. They more readily distinguished between high and low preference items than the survey, which reported more false positives for high preference items. The survey accurately distinguished between high and low preference items for 55% of the total sample of participants.

Agreement across administrations was 65% for the survey, 60% for the verbal stimulus-choice method, and 80% for the pictorial stimulus choice method.

The authors conclude that surveys may not accurately differentiate high and low preference items, and verbal or pictorial choice methods are more likely to correspond with reinforcer assessments. They also caution that since categories were investigated, rather than individual stimuli, it remains unclear whether all items in the category were reinforcers or if one item within a category was particularly potent or weak. The authors suggest future investigation of verbal preference assessments completed by parents or teachers.

In 2000, Northup carried out a systematic replication of the 1996 study conducted by Northup and colleagues. Files of 20 children who had attended a summer program for ADHD in the past five years were reviewed. The accuracy of a reinforcer survey was evaluated by comparing it to the results of a concurrent operants reinforcer assessment. A 42-item survey, with seven items each representing six categories, was administered. Items were ranked by children as being liked not at all, a little, or a lot, and a percentage score was calculated for each category. High preference categories were those with a score of 75% or greater. The reinforcer assessment began with a baseline measure of the number of simple math problems completed without reinforcement. During the reinforcer assessment, seven token coupons were available, six representing a reinforcer category and one control. The child was told that if he or she completed a criterion number of problems determined in baseline, he or she could choose a coupon. Coupons were replaced after being chosen, and students could answer as many problems as they wanted. Finally, a return to baseline was conducted to determine any lasting reinforcement effects associated with the token coupons. A comparison of items identified as preferred by the reinforcer survey and items found to be reinforcers by the

concurrent operants reinforcer assessment found that the total accuracy of the reinforcement survey was 57%, which replicated the earlier results of Northup et al. (1996). True positive accounted for 34% of the items identified. True negative accounted for 23% of the items identified. False positives accounted for 29% of the items identified, and false negatives accounted for 13% of the items identified. Again, the author warns that presenting the items within categories may have obscured individual effects of weak or potent stimuli. The author concludes that the reinforcer survey added little information beyond chance with this population.

Overall, studies that have investigated the validity of child nominations and surveys have found that these surveys identify items that increase the future probability of a target behavior at about chance levels. Although these methods are easy to use and often implemented, their utility has not been proven. There are several limitations noted in this research that may have affected outcomes. First, the effectiveness of categories of reinforcers was assessed rather individual items. Second, coupons were used to represent items instead of presenting the actual items. This also delayed reinforcement. Third, all children sampled in these studies had diagnoses of ADHD. Finally, only child nominations and surveys were assessed. No teacher or parent surveys were considered, even though these are often used in practice (Northup, 2000; Northup et al., 1996; Northup et al., 1995).

Validity of Caregiver Nomination

Some research has considered the validity of caregiver reinforcer surveys and nomination. Several studies have considered the effectiveness of caregiver nomination of items to be used as reinforcers (Fisher et al., 1996; Green et al, 1991). Other studies have compared the rankings of item preference produced by caregiver surveys and preference

assessments (Green et al., 1991; Green et al., 1988). Finally, studies have evaluated the reinforcing effectiveness of items identified using caregiver surveys versus those selected using preference assessments (Green et al., 1991; Green et al., 1988).

In 1996, Fisher et al. tested the effectiveness of a caregiver report for predicting client preferences. Participants were six children with severe destructive behavior and diagnoses of severe or profound mental retardation. The primary caregiver of each child also participated. First, each caregiver ranked a set of standard stimuli based on believed child preference. Then, the Reinforcer Assessment for Individuals with Severe Disabilities (RAISD) structured interview was used to help caregivers generate a list of potential reinforcers themselves. After receiving this information from caregivers, two paired-choice assessments were conducted using methods similar to those in Fisher et al. (1992). All stimuli to be assessed were paired with one another, and client approach was measured. The first assessment considered the standard stimuli that the caregiver was asked to rank, and the second assessment evaluated the stimuli generated by the caregiver with the help of the RAISD.

During phase two of the experiment, a concurrent operants reinforcer assessment was conducted to compare the most highly preferred stimuli found through the two paired-choice assessments. Reinforcer effectiveness was examined using a reversal design. Again, the procedure used was similar to that in the Fisher et al. (1992) study. Stimuli associated with each separate choice assessment were placed in separate squares or chairs, and the total duration of time spent in each square or chair was the dependent measure.

Results revealed that caregiver rankings of the standard stimuli did not correlate with the results of the paired-choice assessment for standard stimuli, but the caregiver stimuli identified via the RAISD did correlate with the associated preference assessment. Visual

analysis of reinforcer assessment data showed an increase in total duration of in square or in chair behavior for the RAISD caregiver selected stimuli. The authors suggest that using a structured interview, such as the RAISD may help identify reinforcing stimuli. They speculate that one reason these results were obtained is that caregivers are better at ranking stimuli that they have seen the participant interact with.

Green et al. (1988) compared reinforcers identified based on staff opinion with those identified via a preference assessment. In experiment one, seven profoundly mentally retarded, nonambulatory individuals ages 12 to 34 participated. Twelve stimuli were chosen for assessment. Using a single operant method, each stimulus was presented a total of 36 times, and approach and avoidance to the stimulus was observed to determine reinforcer preference. When a stimulus was approached, it was made available for five seconds. The stimuli were ranked for each student according to the average percentage of approach behaviors across assessment sessions.

Green et al. (1988) found that five of the students approached at least one of the stimuli in the preference assessment 80% or more of the time. At least five staff members completed a survey for each student, with 35 staff members participating. The staff opinion survey assessed staff perceived student preferences for the 12 stimuli using a one to five rating scale. The value of each item was averaged across staff ratings and then ranked based on average scores. Spearman's rank order correlation found no statistically significant relationship between the survey and preference assessment rankings. The results suggest that these two types of assessments do not identify the same stimuli as preferred.

In experiment two of the Green et al. (1988) study, stimuli identified as highly preferred based on staff opinion and preference assessments were evaluated for reinforcing

effectiveness. The five participants from experiment one who approached at least one stimulus 80% of the time during the preference assessment participated. Four groups of stimuli were evaluated: those identified as high preference by both the preference assessment and the staff opinion survey, those identified as high preference by the preference assessment and low preference by the staff opinion survey, those identified as low preference by the preference assessment and high preference by the staff opinion survey, and those identified as low preference by the preference assessment and low preference by the staff opinion survey. The dependent variable was the level of prompt required by the student to perform a target skill upon request. When the student performed the task at the least intrusive prompt level that was required for the behavior at baseline, the related stimulus being tested was presented.

In this second study, Green et al. (1988) found that at least one of the stimuli ranked as highly preferred by the preference assessment resulted in behavior change for all five students. For four students, stimuli from the high preference and high staff opinion group were associated with the highest responding. For one student, stimuli from the high preference and low staff opinion group were associated with the highest responding. Stimuli ranked low by the preference assessment, regardless of their rating on the staff opinion survey, did not result in a behavior change. In discussing their findings, the authors state that preference rankings based on staff opinion do not agree with the results of preference assessment rankings. In addition, staff ratings of items did not predict their effectiveness as reinforcers as well as the preference assessment.

Green et al. (1991) replicated and extended the work of Green et al. (1988) by further evaluating the reinforcing effectiveness of items identified through staff opinion surveys and

preference assessments. Six individuals participated in experiment one. As in the earlier study, 12 stimuli were selected for assessment, and the target behavior for preference was approach to the stimulus. Staff opinion surveys assessed all 12 items using a five-point scale. An average of five staff members completed a survey for each student. A Spearman rank correlation identified a significant relationship between the two assessment methods for two students, with coefficients of .49 and .58. For the other four students, no significant correlation was found between the two methods, which replicated the Green et al. (1988) findings. The authors state that since no correlations were found for the majority of the students, staff opinion is not a reliable predictor of student preference.

In their second experiment, Green et al. (1991) considered the reinforcing value of the four groups of stimuli evaluated in their 1988 study. The four groups of stimuli considered were high preference/high opinion, high preference/low opinion, low preference/high opinion, and low preference/low opinion. A baseline assessment was conducted in which the effects of a graduated prompt sequence or cue on a target behavior was evaluated. During each of the four contingency conditions, the associated stimulus was provided contingent on a designated level of the target behavior. For all of the students who had preferred at least one stimulus, a change in behavior occurred when a highly preferred stimulus based on preference assessment was presented. Stimuli that were ranked low preference through the preference assessment did not result in a behavior change. Whether or not items were identified by the opinion survey did not seem to matter. Again, these results replicated the findings of the Green et al. (1988) study in that items identified as highly preferred through preference assessments functioned as reinforcers.

Green et al. (1991) conducted a third experiment in which they examined preferences for stimuli not considered in the first assessment. They identified individualized potential reinforcers for six students through speaking with caregivers. Then, they performed another single operant preference assessment with these items using the same method as in experiment one. They wanted to determine whether they would find many additional reinforcers using this method of item selection and assessment. They found a highly preferred stimulus for two of the students, one of which did not prefer any of the stimuli during the first assessment. The authors conclude that it may be beneficial to add staff identified items to the common pool of those normally assessed.

Studies considering the effectiveness of caregiver nomination of items for use as reinforcers have been promising. Fisher et al. (1996) found that when caregiver nomination was guided by the RAISD structured interview, the items identified correlated with single operant preference assessments and were shown to reinforce behaviors during a concurrent operant reinforcer assessment. Green et al. (1991) found the caregiver nomination of items increased the number of items identified that acted as reinforcers. It appears that caregiver nomination of items to test can aid in the selection of items tested in preference and reinforcer assessments. Also, when caregiver nomination is guided by a structured interview, these studies suggest that a preference assessment may not be necessary since it has been shown to identify the same items. Further replication will determine whether this is indeed the case.

Caregiver surveys are often used to identify reinforcers (Green et al., 1991; Green et al., 1988). Green et al. (1988) found no correlation between staff rankings and preference assessments, but Green et al. (1991) found a correlation for two of six participants. In both

studies, the authors also tested the reinforcing effectiveness of items identified via a preference assessment and a caregiver survey. They concluded that items identified as high preference by the preference assessments always acted as reinforcers while those identified by the caregiver survey only did some of the time. The usefulness of caregiver surveys remains unclear due to the conflicting findings reported. In addition, these studies only considered institutional staff ratings for individuals with severe or profound disabilities. Surveys completed by teachers and parents and for typically developing populations have yet to be evaluated.

Correspondence between Child and Teacher Nomination

Several studies have investigated how well teacher nomination or use of rewards corresponds with child survey ratings. It is of interest whether teachers are using rewards that their students find reinforcing. It is also important to determine whether teachers can identify items that their class or a child in their class indicates is preferred (Caffyn, 1987; Fantuzzo et al., 1991; Jacob, Daly, King, & Cheramie, 1984).

Jacob et al. (1984) investigated the accuracy of teacher predictions of student reward preferences as measured by the Children's Reinforcement Survey Schedules Form C. Forty-nine students and 20 teachers participated. All students were either in the fifth or sixth grade. The 80 item survey asks students to rate how much they like certain activities or items on a three point scale. The survey was given to each student and his or her teacher. The teacher was asked to complete the survey as he or she believed the student would. The Pearson product-moment correlation coefficient calculated was .32 when averaged across all 49 student and teacher pairs. When only items rated as "Like" or "Like Very Much" were considered, the relationship between teacher and student choices was much higher at .84. The

authors believe the mean correlation reveals that teachers were moderately successful in selecting reinforcers that the students chose. It is important to note that survey items did not only consider classroom related rewards but also measured non-school rewards. Since the effects of non-school rewards on individual children may not be observed by teachers, accuracy of reporting for these items would not be expected to be as high as for classroom related rewards. The authors concluded that it is important that students are involved in selecting rewards for individual interventions.

In 1987, Caffyn considered the attitudes of English students and teachers towards various rewards and punishments. A questionnaire was utilized which asked 510 students from age 13 to 15 and 99 teachers to rate how successful various items would be in different contexts. Then, they were asked to list the two items they believed would be most successful in each context. Rewards considered included: special treats, special certificates, credits or merits, praise in multiple situations, five minutes of free time, a parent note home, and a favorable report card. Overall, students' and teachers' agreement of reward effectiveness varied across items and contexts.

Fantuzzo et al. (1991) developed the Child Reinforcement Survey for use in a study that examined how often teachers use rewards, whether there are differences in the rewards used across grades, what types of rewards children prefer, and whether teachers use the rewards that their students prefer. Forty-eight teachers from five different schools participated. Ninety-eight children from the participating teachers' classes also participated. Teachers taught grades one through five, and students attended grades two through five. Teachers completed a questionnaire about their use of categories of reinforcers. They also rated how effective they believed certain categories of reinforcers to be and how often they

use certain rewards to increase certain types of behaviors. Children were interviewed using the Child Reinforcement Survey.

The Child Reinforcement Survey is individually administered and considers 36 possible reinforcing items. Items chosen for inclusion in the survey were identified through a review of the research and by asking a group of teachers to list rewards. Children respond whether they like each item a little or a lot. Results for teacher questionnaires showed that 92% of teachers reported using rewards from two or more categories of reinforcers.

Results showed that ninety-four percent of teachers reported that they used rewards to improve conduct and homework completion. Eighty-one percent said that they used rewards to improve reading behavior, while 73% use rewards to improve math behavior. Teachers in the lower grades reported using significantly more categories of rewards than teachers in the upper grades. Child preference for rewards was idiosyncratic. There was no significant correlation between teacher use of rewards and child preferences. The authors state that it appears that teachers are using rewards to improve behaviors, but they are not attempting to match student preferences.

The research shows that teachers report using a variety of rewards fairly often, but correlations between teacher reward choice and child preference have ranged greatly (Caffyn, 1987; Fantuzzo et al., 1991; Jacob et al., 1984). A study that examined the relationship between teachers' use of rewards and their classes' preference found no correlation (Fantuzzo et al., 1991). Another study revealed that agreement between teachers and students varied across items and contexts (Caffyn, 1987). One study found a weak correlation between items identified as preferred between pairs of teachers and students, but when only highly preferred items were considered, there was a strong correlation between the two

(Jacob et al., 1984). No studies have compared teacher reward nominations to preference assessments or tested teacher reward nominations through reinforcer assessments. The literature regarding teacher's ability to predict student preferences remains inconclusive (Caffyn, 1987; Fantuzzo et al., 1991; Jacob et al., 1984).

In sum, several indirect assessments of potential reinforcers have been developed for both typically developing and disabled populations. Reliability data for several survey instruments are promising, but validity data are yet to be published (Cautela & Brion-Meisels, 1979; Matson et al., 1999). Correlations between child surveys and reinforcer assessments have generally found weak or no correlations, but limitations of this research bar any firm conclusions (Northup, 2000; Northup et al., 1996; Northup et al., 1995). The research considering the relationship between caregiver nominations or surveys and preference assessments has produced mixed results (Fisher et. al, 1996; Green et al., 1991; Green et al., 1988). Finally, the literature examining the relationship between teacher and student selected rewards has produced varied and inconclusive results. In addition, teacher surveys have yet to be compared to direct assessments. (Jacob et al., 1984).

Based on the current state of research examining indirect assessments, further investigations are warranted. First, the relationship between child surveys, teacher surveys, and preference assessments has not been established. Second, it is important to determine whether items identified by teachers as highly preferred act as reinforcers. Finally, the effectiveness of teacher-identified items as reinforcers has not been compared to items selected via a preference assessment.

Direct Assessments – Preference and Reinforcer Assessments

Most investigations of preferred reinforcers contain two phases: a preference assessment and a reinforcer assessment. The goal of a preference assessment is to assess an individual's preference for potential reinforcers (Fisher & Mazur, 1997). The majority of preference assessments utilize an approach based measure (Hagopian, Long, & Rush, 2004). Four main forms of preference assessments have been proposed: a single operant or single stimulus procedure, a forced-choice or paired-choice procedure, a multiple-stimulus (MS) or group procedure, and a multiple-stimulus without replacement (MSWO) procedure (DeLeon & Iwata, 1996; Fisher et al., 1992; Hagopian et al., 2004; Pace, Ivancic, Edwards, Iwata, & Page, 1985; Windsor, Piche, & Locke, 1994).

As Fisher and Mazur (1997, p. 396) explained, "The purpose of a reinforcer assessment is to evaluate stimuli that have been identified as being preferred to determine whether they actually function as reinforcers (i.e., verifying reinforcer function)." Two types of reinforcer assessments are most often reported in the literature. The single operant reinforcer assessment evaluates the absolute reinforcing effectiveness of each item separately (Pace et al., 1985). The concurrent operants reinforcer assessment evaluates the relative reinforcing effectiveness of several items at one time (Fisher et al., 1992; Northup, 2000; Northup et al., 1995). These assessments are often used in the literature to determine the treatment utility of preference assessments (Fisher et al., 1992; Paclawskyj & Vollmer, 1995; Piazza, Fisher, Hagopian, Bowman, & Toole, 1996).

Single Stimulus and Paired Choice Comparisons

Multiple studies have compared the efficiency and effectiveness of the single stimulus and paired choice preference assessments (Fisher et al., 1992; Paclawskyj &

Vollmer, 1995; Piazza et al., 1996). In the single stimulus procedure, each stimulus is presented separately and approach is the dependent measure (Pace et al., 1985). The paired-choice procedure calls for pairing all stimuli to be assessed with each other, and then a choice between the two is made when the individual approaches one of the stimuli (Fisher et al., 1992). Experimenters have attempted to determine whether one method has more treatment utility than the other (DeLeon & Iwata, 1996; Fisher et al., 1992; Higbee et al. 2000; Paclawskyj & Vollmer, 1995; Piazza et al., 1996; Windsor et al., 1994).

Pace et al. (1985) were the first to develop and test a systematic behavioral preference assessment procedure. Six individuals between the ages three and 18, with profound mental retardation participated in their first experiment. Sixteen stimuli with various characteristics were chosen for assessment. During the preference assessment, a single stimulus was presented, and participant approach was measured. When a stimulus was approached, it was offered to the participant for five seconds. If there was no approach, the stimulus was removed, and the participant was prompted to sample the stimulus. Then, a second probe of the same stimulus was conducted. If the participant did not approach the stimulus within five seconds, a new stimulus was assessed. If approach occurred, the stimulus was made available to the participant for five seconds. Eight sessions of 20 trials each were carried out for each participant. All six participants differentially approached the assessment stimuli, and four of the participants approached several of the stimuli on 80% or more of the trials. The authors concluded that this was a useful method for identifying preferred items.

In Pace et al.'s (1985) second experiment, stimuli were assessed for reinforcing effectiveness using a reversal design. Stimuli approached on 80% of the trials during

experiment one were considered preferred, and stimuli approached on 50% or less of the trials were considered non-preferred. During baseline, the therapist presented a request for the participant to exhibit an adaptive behavior, but no consequence was offered. In the preferred condition, a request was made and a preferred stimulus was offered contingent on compliance within five seconds. In the non-preferred condition, a request was made and a non-preferred stimulus was offered contingent on compliance within five seconds. Results revealed that preferred stimuli increased the occurrence of adaptive behaviors relative to baseline and the non-preferred condition for five of the six participants. The authors concluded that the single stimulus assessment was effective in identifying reinforcing stimuli for these individuals.

Fisher et al. (1992) compared the single operant preference assessment devised by Pace et al. (1985) to a forced-choice method. Four individuals with severe or profound developmental disabilities participated. The 16 stimuli chosen for assessment in the Pace et al. (1985) study were also used in this investigation. For the single stimulus assessment, the 16 stimuli were assessed for approach one at a time. During the forced-choice assessment, all 16 stimuli were paired with one another for assessment. Pairs were presented in a random order, and client approach was measured. If the participant approached an item, access was given to that item for five seconds as in the Pace et al. (1985) study. Approach to both stimuli was blocked. If the participant did not approach either item, the client was allowed to sample both stimuli for five seconds, and then they were presented once again. If a stimulus was approached, it was given to the client for five seconds. If no stimulus was approached, the next trial began with a new pair of stimuli. Overall, the single stimulus condition produced

more preferred stimuli than the forced-choice assessment. The forced-choice assessment produced greater differentiation among items.

In the second phase of the Fisher et al. (1992) experiment, the reinforcing effects of two types of stimuli were compared: stimuli approached on at least 80% of trials on both the single stimulus and forced-choice assessments (high-high condition) and stimuli approached on at least 80% of trials on the single stimulus assessment and 60% or less of trials on the forced-choice assessment (SP-high condition). During baseline, the amount of time the participant spent in a chair or square was measured, but no consequences were provided. During treatment, high-high stimuli were placed in one square or chair and SP-high stimuli were placed in the other square or chair. When the client exhibited in chair or in square behavior, the stimulus in that chair or square was delivered to the participant for five seconds. The amount of time spent in the square or in the chair was the dependent measure. For all four participants, the duration of in chair or in square behavior for chairs or squares with a high-high stimulus was significantly greater than for those with SP-high stimuli. The authors interpret this as indicating that the forced-choice assessment has good concurrent validity, and that the single stimulus preference assessment is prone to false positives.

The generality of paired-choice assessments was further established by Paclawskyj and Vollmer (1995). They compared the predictive validity of a single stimulus preference assessment and a forced-choice preference assessment for four male students with visual impairments. Teachers identified six items for assessment. For the single stimulus preference assessment, physical guidance was used to prompt students to explore the presented item for three seconds. Then, physical guidance was retracted and approach behavior was measured. If the student approached the item, the item was made available for five seconds. If the

student did not approach the item, physical guidance was reinstated. Then, the trial was repeated.

For the forced-choice procedure, all items were paired with each other for presentation. The experimenter physically guided the student to touch the left and then right item for three seconds each. Then, the experimenter removed his or her hand and approach behavior was measured. Approach consequences were the same as in the single stimulus procedure. Each stimulus was presented 10 times in a random order during each assessment.

In the second phase of the experiment, a reinforcer assessment was carried out to compare items identified through each method. They used a combination multiple baseline reversal design to compare baseline performance with two conditions for three participants from the first phase of the experiment. The two conditions compared a single stimulus high preference/paired-choice low preference item and a paired-choice high preference/single stimulus low preference item. For each child, a target behavior was selected from the child's curriculum, and compliance behavior was measured.

Results showed that the paired-choice procedure produced greater differentiation of items than the single stimulus procedure. In the single stimulus high preference/paired-choice low preference phase compliance rate declined or remained low for two participants. For three participants, compliance rate increased in the paired-choice high preference/single stimulus low preference phase. The authors concluded that the paired-choice procedure produced more differentiation and identified more reinforcing items than the single stimulus procedure for individuals with visual disabilities.

Piazza et al. (1996) set out to further investigate the validity of paired or forced-choice preference assessments. Four males with severe or profound developmental

disabilities and severe destructive behavior participated. They used the RAISD structured interview to identify potential reinforcers. The identified items were then assessed using a paired-choice assessment similar to that conducted in the Fisher et al. (1992) study. Item approach was measured. The three items approached most frequently were considered high preference stimuli, the three items approached closest to the median number of times were considered middle preference stimuli, and the three items chosen the least were considered low preference stimuli.

A concurrent operants reinforcer assessment similar to that in the Fisher et al. (1992) study was used to compare the reinforcing effectiveness of the three different levels of preferred items. A mini choice assessment was completed before each reinforcement assessment in which the child was allowed to select which two stimuli would be compared in that session. During the mini choice assessment, each stimulus was paired with every other stimulus, and the two selected most frequently were used in the session. Three different choices were available during each session. Two chairs or squares contained items representing one of the three levels of preference, and one chair or square was a control and contained no item. Different phases were initiated to compare all three levels to each other.

Data revealed that the stimuli identified as high preference during the paired-choice assessment functioned to reinforce in chair or in square behavior for all four participants. Stimuli determined to be moderately preferred functioned as reinforcers for two of four participants, and stimuli identified as low preference did not function as a reinforcer for any of the participants. In discussing the results, the authors state that the outcome of the paired-choice assessment was shown to predict relative reinforcer effectiveness.

For the most part, the literature comparing the treatment utility of single stimulus preference assessments to paired choice preference assessments has found the paired choice method to be superior (Fisher et al., 1992; Paclawskyj & Vollmer, 1995; Piazza et all, 1996). The single stimulus method has been shown to identify reinforcers (Pace et al., 1985), but, when compared to the paired choice method, it does not identify as many reinforcers (Paclawskyj & Vollmer, 1995). Also, the single stimulus procedure has been found to produce more false positives when compared to reinforcer assessment results (Fisher et al., 1992). Generally, the paired choice method better differentiates among items and is more likely to select items that will act as reinforcers (Fisher et al., 1992; Paclawskyj & Vollmer, 1995; Piazza et al., 1996). Finally, it must be noted that all of this research has been conducted with developmentally disabled individuals serving as participants. These findings may be limited to this specific population.

Paired Choice and Multiple Stimulus Comparisons

Several comparisons have been made between paired choice, MS, and MSWO preference assessments (DeLeon & Iwata, 1996; Higbee et al. 2000; Windsor et al., 1994). As noted above, the paired choice procedure pairs all stimuli to be assessed with one another and approach is measured (Fisher et al., 1992). The MS or group procedure presents all stimuli to be assessed at once and stimuli are ranked by the number of times they are approached (Windsor et al., 1994). Finally, in the MSWO procedure, all stimuli to be assessed are presented at once, but when a stimulus is chosen it is removed from the array of choices. Stimuli are ranked based on the percentage of times an item is chosen (DeLeon & Iwata, 1996). The MS procedures were developed as more cost effective procedures for examining relative preferences. Studies have considered how well these methods correlate

with paired choice methods and their comparative treatment utility (DeLeon & Iwata, 1996; Higbee et al. 2000; Windsor et al., 1994).

Windsor et al. (1994) were the first to compare a paired-choice presentation with a MS presentation preference assessment. Eight adults with severe or profound disabilities participated. Staff listed foods they believed each participant liked, and these items were used as stimuli in the assessments. In the group presentation, all six foods identified by the staff were randomly placed equal distances apart in an array. Participants were asked which food they wanted on 10 separate trials. In the paired presentation, all six foods were paired with one another, and only two foods were presented at a time for a total of 30 trials. The food that the participant attempted to grasp or eat was scored as that selected for both methods of presentation. Ranking was based on the total number of each food item selected in each presentation. In addition, a staff member was asked to rank a student's preference for the foods assessed using a six-point scale.

Results showed that both the paired and group presentation methods identified differential preferences. A Chi square test found a significant difference in item selection for all participants using the paired presentation and all but one participant using the multiple-stimulus presentation. A Kendall rank order correlation coefficient of .75 indicated that paired and group rankings correlated across learners. The paired presentation produced more reliable results across trials, with an average correlation of .63 across administrations. The group presentation was less consistent across trials, with an average correlation of .49 across administrations. Multiple-stimulus presentations required less time to administer (M = 7 min) than paired presentations (M = 16 min). The mean correlation between staff rankings and preference assessment rankings showed a low to moderate relationship at .45, and for four of

the eight students, the same food item was ranked as most preferred by staff and preference assessments. The authors conclude that both presentation methods produce valuable information. They end by saying that the group presentation has an advantage in being more time efficient, but the paired presentation was more consistent.

DeLeon and Iwata (1996) extended the research on multiple-stimulus (MS) presentations by comparing two types of group presentations to a paired-choice method. They proposed a variation of the MS presentation method, MSWO, which easily ranks preference of items. Seven adults with profound developmental disabilities participated in the first experiment. Seven stimuli were chosen for assessment. The MSWO procedure began with all items randomly placed an equal distance apart on a table. The participant was seated at the table and told to choose one item. After an item was selected, that item was removed from the array of choices. This continued until all items were selected or the participant did not respond for 30 seconds. The MS procedure was carried out similar to the MSWO procedure, but after an item was chosen, it was not removed from the array. The paired-choice method paired all items together for a total of 21 trials. The participant was to choose one of the items from each pair. Each procedure was conducted on five separate occasions. The dependent variable for all methods was the percentage of times an item was chosen.

For four of the seven participants, all three presentation methods identified the same item as the most preferred. The MS procedure produced more unselected items than the MSWO or paired-choice methods. Kendall rank order correlations resulted in higher correlation coefficients between the MSWO and paired-choice assessment than for the MS and paired-choice assessment for five of the seven participants. Both the MSWO and paired-choice methods produced high across session correlations, indicating consistency. The

MSWO procedure produced a mean correlation of .81, and the paired-choice procedure resulted in a mean correlation of .83. The MS procedure resulted in a lower mean correlation across sessions at .56. The paired-choice procedure (M = 53.3 min) took the most time to carry out. The MSWO (M = 21.8 min) and MS (M = 16.5 min) procedures required much less time. Based on these results, all methods resulted in similar preferred items being selected, but the MSWO and paired-choice procedures produced more consistent rankings. Finally, both MS procedures were more time efficient than the paired-choice method.

In Deleon and Iwata's (1996) second experiment they evaluated the reinforcing effects of an item identified in both the paired-choice and MSWO assessments, but not in the MS assessment. They sought to determine whether the paired-choice and MSWO assessments produced false positives or the MS procedure produced false negatives. Four participants from their first experiment were included in this study. A reversal ABA design was utilized. Baseline measures of target responses were taken. Then, measures of responding on a fixed ratio one (FR1) schedule of the item were taken. Finally, a return to baseline was conducted. Target responses were simple and differed based on participant.

For three of the four participants, items that had been selected by the paired-choice and MSWO assessments, but not the MS assessment, acted as reinforcers. The authors indicated that these results show that in some cases items that are not identified by the MS procedure may act as reinforces. They conclude that the MSWO and paired-choice procedures identify more possible reinforcers. The authors state that additional research should be conducted to determine the predictive validity of outcomes obtained from preference assessments.

Higbee, Carr, and Harrison (2000) carried out further evaluation of the predictive validity of the MSWO preference assessment. Participants were nine adults with severe or profound mental retardation. Seven stimuli were selected for presentation for each participant using the RAISD structured interview. A MSWO procedure was utilized. All stimuli were randomly placed equal distances apart on a table. The participant was then asked which item they wanted most. The item that the participant first touched was recorded as the one selected and subsequently removed from the table. Stimuli were ranked according to the percentage of times they were chosen when available. Three assessment sessions were carried out.

Next, a reinforcer assessment was conducted. A baseline measure of microswitch activation was taken. Then, the reinforcer evaluation began. A FR schedule based on baseline performance was put in place. The four stimuli ranked as most preferred were each separately delivered contingent upon the criterion number of microswitch activations. Sessions were randomly alternated in a multi-element design among stimuli. Finally, baseline was reinstated and the number of responses recorded. Results showed that the stimulus ranked as most highly preferred based on the preference assessment acted as a reinforcer in six of the nine participants. Therefore, the authors concluded that the MSWO procedure does produce valid reinforcers for most individuals.

Overall, the research has shown that both the MS and MSWO assessments take less time to administer than the paired assessment. The paired assessment and the MSWO preference assessment have been shown to both produce consistent results across trials and sessions, but the MS assessment is less consistent. The MSWO preference assessment is also more highly correlated with the paired assessment than the MS assessment. In addition, the MS assessment tends to produce more unselected items than the paired and MSWO

assessments. Based on these findings, the MSWO appears to be the most efficient and valid of the three methods (DeLeon & Iwata, 1996; Higbee et al. 2000; Windsor et al., 1994). Even with these clear findings, it must be remembered that these assessments were only evaluated using individuals with developmental disabilities, and these findings may not generalize to other populations.

Brief Preference Assessments

One topic of debate is how many sessions and trials are necessary to produce reliable and valid MSWO preference assessment results. In 1996, DeLeon and Iwata utilized a five-session procedure, which has since become generally accepted as standard. Carr, Nicolson, & Higbee (2000) have since introduced a brief three session method. It is important for practitioners that the most cost effective procedure be identified so that treatment can be implemented as soon as possible (Carr et al., 2000; Graff & Ciccone, 2002).

Carr et al. (2000) tested the validity of a brief MSWO procedure consisting of three sessions. Three children with autism participated. Eight items were selected for assessment from parent and therapist nominations. Then, a brief MSWO preference assessment was conducted. The procedures were similar to those in the DeLeon and Iwata (1996) study, except only three sessions were conducted with each participant rather than five sessions. The number of times a stimulus was chosen was divided by the number of trials in which it was available and then multiplied by 100 to obtain a percentage. Percentages were then ranked one to eight, with one being the item most preferred and eight being the item least preferred.

Next, a brief reinforcer evaluation was done. The items ranked as first, fourth or fifth, and eighth by the preference assessment were considered. A target behavior was chosen

based on the child's current curriculum. First, a baseline measure of the target behavior was taken. No contingency was in place during baseline. Then, all the high preference, medium preference, and low preference stimuli were each tested separately during two probe sessions each. Depending on the contingency in place, the corresponding item was provided to the participant for 10 seconds on a FR1 schedule. Finally, eight additional MSWO assessments were conducted over a period of four weeks for each participant to evaluate preference stability over time and the correspondence between the results of the first MSWO session and all three MSWO sessions.

Based on visual analysis, it was determined that the high preference stimulus produced higher responding than baseline, the medium preference stimulus, and the low preference stimulus for all three participants. The MSWO and reinforcer assessments were completed in less than one hour for each participant. Two of the participants showed stable preferences across sessions, but one participant's preferences were variable across sessions. Spearman rank correlations showed strong relationships between the one session MSWO assessment and the three session MSWO assessment, with correlations of .85, .74, and .89 for each participant. The authors report that the current study presents support for use of a brief MSWO assessment, but they caution that they did not compare the three session MSWO assessment to the standard five session procedure. They also remind the reader that the one session MSWO assessment findings were not subjected to a reinforcer assessment. Therefore, the authors cannot make any claims about its predictive validity.

In 2002, Graff and Ciccone extended the research begun by Carr et al. (2000). They considered how many sessions and trials were necessary to produce valid and reliable results using a MSWO procedure. Fifteen students attending a school for children with autism,

developmental disabilities, and behavior disorders and ranging in age from seven to twenty-one participated. A MSWO preference assessment was conducted using seven stimuli. Seven sessions of seven trials each were conducted for each participant, and stimuli were ranked based on the percentage of approach responses during the total number of trials. Next, a reinforcer assessment was conducted with four of the participants.

An ABAB design was utilized to evaluate the effects of highly preferred stimuli on a button press response. In analyzing the data, the authors first considered whether the same item would have been identified as most preferred if less sessions were conducted. When only five sessions with three trials were used, 22 of 27 cases resulted in the same item being ranked as most preferred as when seven sessions with seven trials were conducted. When three assessment sessions consisting of seven trials were carried out, 19 of the 27 cases resulted in the same item being ranked as most preferred as when seven sessions with seven trials were conducted. When all seven sessions were conducted, but only three trials were carried out for each, the same item was ranked as most preferred as when seven trials were used for analysis for 25 of 27 cases. Items identified as highly preferred using a five session, three trial MSWO assessment were found to increase responses for all four participants. The authors conclude that assessments can be shortened by requiring less trials (three) within the standard five-session assessment. They believe the results show this to be a better alternative than using a brief assessment of three sessions with all seven trials.

The optimum number of sessions and trials for a MSWO assessment remains unclear. Carr et al. (2000) reported that the brief three session method identified items that acted as reinforcers and that a one-session method correlated well with the three-session method. Even though the brief method has been shown to identify reinforcers, Graff and Ciccone

(2002) remained skeptical since it had not been compared to the standard five-session method. They compared the highest ranked items produced by the extended seven sessions, five sessions, and three sessions. Unlike Carr et al. (2000), they concluded that a five-session method was the most cost effective rather than a three-session assessment. They did not report whether the three-session method identified reinforcers. Unfortunately, they also failed to consider a one session method altogether. Finally, they did not consider overall rank order correlations but only looked at the highest preferred item. As of now, there is no generally accepted standard for which method is most effective and efficient.

Stability of Preferences over Time

Some experimenters have assessed the stability of preference assessment results over time. If preferences are stable, then assessments need to be carried out less often. If they to change across time, then assessments may need to be carried out frequently. It is important to determine this so that treatment can be optimized (Carr et al., 2000; Green et al., 1991; Mason, McGee, Farmer-Dougan, & Risley, 1989).

Mason et al. (1989) examined the efficacy of a daily mini-assessment of reinforcer preferences. Three preschoolers with autism participated. Teaching sessions consisted of body part identification trials. Incorrect responses resulted in a prompt sequence, and correct responses were praised. The primary dependent variable was maladaptive behavior. Other dependent variables were correct responding and out of seat behavior. First, a baseline was initiated. During baseline, teachers chose items that served as reinforcers. Next, a preference assessment was conducted using the Pace et al. (1985) procedure. The items that were approached on 80% of the trails were considered to be preferred, and were included in the daily pre-session mini-assessments.

The mini-assessments consisted of a daily pre-session presentation pairing each of the items determined to be preferred from the full preference assessment. The child was told to pick one item from each of the pairs, and all of the items that were chosen were used as reinforcers during the upcoming session. The following sessions were the reinforcer assessment phase. Finally, a post intervention reinforcer assessment was conducted.

Results showed that the daily reinforcer assessment produced decreases in maladaptive behavior from baseline percentages for each child. More moderate reinforcer effects were also seen on correct responding and out of seat behaviors. Data from pre and post reinforcer assessments revealed that child preferences changed across a one-month period, and preferences were idiosyncratic across children.

In discussing the results, Mason et al. (1989) assert that ongoing reinforcer assessments were able to identify reinforcers that significantly decreased maladaptive behavior. They believe one of the strengths of ongoing mini-assessments is their efficiency for identifying reinforcers in a short period of time. Also, the authors feel that this study shows that child preferences do change over time and that there is a need to assess reinforcers often.

In the fourth experiment carried out by Green et al. (1991), they considered the durability of preferences over time. The preferences of 12 participants were assessed over time periods ranging from four to twenty-eight months. Preference assessments were conducted using the single operant method as in experiment one. A Spearman rank correlation indicated that preferences were pretty consistent over time for this population. Statistically significant correlations were found between the two assessments for 11 of the 12

participants. The authors indicated that this demonstrates the durability of preferences identified by preference assessments with this population.

Results concerning the stability of preference assessment outcomes remain inconclusive. Very little research has been done in this area. In addition, most studies concerning consistency have utilized a single stimulus preference assessment. This limits the degree to which one can say these results hold for other preference assessment methods. More research investigating the consistency of preference choices for both individuals and groups over different periods of time is warranted (Carr et al., 2000; Green et al., 1991; Mason, McGee, Farmer-Dougan, & Risley, 1989).

To summarize the preference and reinforcer assessment literature, several methods have been developed and tested (Pace et al., 1985; Fisher et al., 1992; DeLeon & Iwata, 1996). It has been found that the paired choice presentations are generally more consistent and produce greater differentiation than single stimulus presentations (Fisher et al., 1992; Paclawskyj & Vollmer, 1995; Piazza et all, 1996). All group procedures have been reported to take less time to carry out than the paired choice method, but the MS method does not correlate well with the paired choice method and results in more unselected items. The MSWO correlates well with the paired choice procedure and has greater consistency across sessions and trials when compared to the MS method (DeLeon & Iwata, 1996; Higbee et al., 2000; Windsor et al., 1994). Some brief versions of the MSWO preference assessment have been proposed, but it is currently unclear if they are as reliable and valid as standard versions (Carr et al., 2000; Graff & Ciccone, 2002). In addition, the stability of preferences across time remains unknown (Carr et al., 2000; Green et al., 1991; Mason, McGee, Farmer-Dougan, & Risley, 1989).

Purpose and Rationale

The major purpose of this study was to find fast and effective methods for identifying potential reinforcers for children attending general education classes. Reinforcers are a significant part of many classroom interventions (Sulzer-Azaroff & Mayer, 1986). Much of the current literature has examined preference and reinforcer assessment methods for use with individuals with severe or profound disabilities (Hagopian, 2000). It is crucial that cost effective methods of preference and reinforcer assessment are identified that can be used in general education settings. It is also important that these methods are simple and efficient so that busy professionals can utilize them easily (Ivancic, 2000; Matson et al., 1999).

The first experiment in this study compared the results of four types of preference assessments: teacher rankings, teacher surveys, child surveys, and a brief MSWO direct assessment of preferences. The brief MSWO procedure was chosen based on research carried out by Carr et al. (2000) that found it was able to identify preferred stimuli that acted as reinforcers. In 2000, Northup conducted a group study comparing a concurrent operants reinforcement assessment using token coupons to a child survey. He found that the child preference survey had a predictive accuracy of 57% when compared to the reinforcer assessment for children with ADHD. In this experiment, preference surveys were compared to a brief MSWO preference assessment using a tangible items preference assessment, rather than a reinforcer assessment using coupons. This experiment extended the past research examining preference surveys.

This study also examined child preference surveys as well as teacher preference surveys and teacher preference rankings for typically developing children with no diagnoses attending general education. Often, teachers choose the reinforcers for intervention (Sulzer-

Azaroff & Mayer, 1986), and it would be helpful to know whether they can select specific items that a child prefers. As has been noted by previous researchers (Matson et al., 1999) a survey that identifies potential reinforcers could be very helpful for busy professionals such as school psychologists and teachers. If it were to correlate with the brief MSWO procedure, it could save a lot of time when identifying reinforcers for use in classroom interventions. Alternatively, if it does not correlate, we can save wasted time spent on inefficient reinforcers.

Green et al. (1988) and Green et al. (1991) compared the reinforcing effectiveness of items identified as highly preferred by a single stimulus preference assessment and by a caregiver opinion survey and found that only items identified as highly preferred by the direct preference assessment increased a target behavior for participants with severe or profound mental retardation. In the second experiment of this study, the goal was to compare the absolute reinforcing effectiveness of items identified as highly preferred by the brief MSWO procedure and items identified as highly preferred by the teacher survey. It was possible that items identified as highly preferred by the teacher survey and not the brief MSWO procedure would still act as reinforcers. This experiment differed from those carried out by Green et al. (1988) and Green et al. (1991) in that general education teachers completed the survey rather than institution staff. In addition, the students were not diagnosed with any psychiatric disorder or developmentally disabled. Also, the items identified by the survey were compared to items identified via a brief MSWO assessment rather than a single stimulus assessment. This experiment attempted to elucidate whether or not differences in reinforcer effectiveness would be seen based on whether the item was selected by the child through direct assessment or by the teacher through indirect assessment.

Also, the reinforcer assessment used a socially relevant response (solving math problems) rather than a response such as button presses (Higbee et al., 2000).

The final experiment also focused on making preference assessments more efficient for children in general education. Carr et al. (2000) tested the predictive validity of a brief MSWO preference assessment, and found that it selected items that acted as reinforcers for all three participants. They also found strong relationships between outcomes based on a one-session and a three-session preference assessment, but they did not compare either to the standard MSWO preference assessment of five sessions. Graff and Ciccone (2002) extended this research and compared a three session MSWO assessment to a five session MSWO assessment and a lengthier seven session MSWO assessment, but they failed to consider a one session MSWO assessment. In addition, they only examined the item identified as most preferred rather than the rankings of all items. This experiment compared the rankings of all items identified using the standard (DeLeon & Iwata, 1996) MSWO preference assessment (five sessions), the brief MSWO preference assessment (three sessions), and a mini MSWO preference assessment (one session).

Mason et al. (1989) showed the importance of conducting daily mini-assessments for intervention efficacy. All three of their participants showed changes in preferences across a one-month period. Alternatively, Green et al. (1991) found that preferences in their sample remained fairly stable over time. The second purpose of the third experiment was to determine the stability of preferences over time in a group of typically developing children. The stability of preferences affects how often assessments need to be administered to have treatment utility.

Experiment I: Comparison of Preference Assessment Methods

This experiment compared the outcomes of several preference assessment methods using a group design and Spearman rho correlations. A teacher survey, teacher ranking, child survey, and brief MSWO preference assessment were compared. The dependent measure was item preference rankings.

Methods

Participants

Twenty children attending general education classes in grades kindergarten through second at a lower elementary school in the southeastern United States participated. Twelve participants were female and eight participants were male. Six children were in kindergarten, seven children were in first grade, and seven children were in second grade. None of the children had any diagnoses or were receiving any special services. All children were Caucasian and ranged in age from 5 to 9 years.

Twenty teachers selected one student to participate for whom they wished to identify reinforcers. A permission form was sent home to the parents of the selected children (Appendix A). The form detailed the purpose and procedures of the experiment and asked for informed consent. After informed consent was received for each child, teachers were given a consent form detailing their role in the study (Appendix B). Each child gave his or her assent to participate during the first meeting (Appendix C).

Nineteen of the teacher participants were female, and one was male. All twenty teacher participants were Caucasian. Years of teaching experience ranged from 1 to 28 years, with a mean of 13 years. Fifteen of the teachers held Bachelors degrees, 2 teachers held Masters degrees, and 3 teachers held Masters degrees plus 30 hours.

Settings and Materials

The child participants' teachers completed teacher surveys and rankings during a faculty meeting (Appendix D and E). Child preference surveys (Appendix F) and preference assessments (Appendix G) were completed in a quiet room located within the school building. Items were selected for assessment based on nomination by a teacher not participating in the study. All selected items were either tangible or edible. Examples of items include: colorful pencils, erasers, stickers, small toy dinosaurs, Hershey Kisses™ candies, Mini Snickers™ candy bars, Goldfish™ cheese crackers, and animal crackers. See Appendix G for a full list.

Data Collection and Dependent Variables

Teacher Survey Scores. A teacher survey constructed for this study was used to obtain responses. The survey consisted of 20 items representing 20 different stimuli. Teachers rated each item according to how much he or she believed the child would like to receive it using a five point scale: (1) very much like, (2) like, (3) indifferent, (4) dislike, (5) very much dislike.

Teacher Ranking Scores. A teacher survey constructed for this study was used to obtain responses. The survey consisted of 20 items representing 20 different stimuli. Teachers ranked items 1 to 20, with 1 being the item he or she believed the child would like to receive the most and 20 being the item he or she believed the child would like to receive the least.

Child Survey Scores. A child survey constructed for this study was used to obtain responses. The survey consisted of 20 items representing 20 different stimuli. The child indicated how much he or she would like to receive an item based on a five point rating scale

(very much like – open smile face, like – smile face, do not like or dislike – straight face, dislike – frown face, very much dislike – angry frown face), with different levels represented by faces. Experimenters scored items as 1 to 5.

Preference Assessment Scores. Twenty items were assessed, with a possibility of 20 item choices. Each position of item choice was associated with a different number of points. The item chosen first in the multiple-stimulus without replacement procedure received 20 points. The item chosen second received 19 points. The item chosen third received 18, and so on. The twentieth item chosen received one point. If the child did not choose an item, then that item received no points. Since three separate sessions were conducted, points accrued for each item were totaled across sessions and divided by three to obtain an average. Items were ranked based on the number of points received.

Reliability. Two observers simultaneously but independently scored 25% of the child surveys and preference assessments on separate data sheets. Interscorer agreement was calculated by dividing all rank agreements by the sum of rank agreements and disagreements multiplied by 100. Interscorer agreement for the child surveys was 100%, and interscorer agreement for the preference assessments was 99.75%.

Procedure

Teacher Survey. The instructions given for the survey were as follows, "Following are some items that might be used to reward a child for appropriate behavior. Please circle the number corresponding to how much you believe this child would like to receive the stated reward on a scale of 1 to 5: 1 representing very much like, 2 representing like, 3 representing indifferent, 4 representing dislike, 5 representing very much dislike."

Teacher Ranking. The instructions given for ranking will be as follows, "Following are some items that might be used to reward a child for appropriate behavior. Please rank these items 1-20, with 1 representing the reward that you believe the child would like to receive the most and 20 representing the reward that you believe the child would like to receive the least."

Child Preference Survey. The child survey included the same items as the teacher survey and ranking. All instructions and items were read to the child. Instructions for the child survey were, "I'm going to name some things that kids sometimes get in school. I want to know how much you like each of these things. After I name each thing, I'm going to ask you to show me how much you like each item. I want you to point to the face that shows how you feel about the item. (The experimenter will point to each item as explaining.) You might like the item very much, and that's this very happy smiley face with a grin. You might just like the item, and that's this smiley face with a smile. You might not really like it but you don't dislike it, and that's the straight face. You might dislike it, and that's this frown face. You might dislike it very much, and that's the angry frown face." Then, the experimenter read each of the items. After reading each item, the experimenter asked the child which face represented how he or she would feel about receiving that item and marked the corresponding face indicated.

Preference Assessment. The preference assessment utilized a multiple-stimulus without replacement (MSWO) method. Before beginning the assessment, all items listed in the surveys were brought in and laid out in a random array on the table about five centimeters apart in a semi-circle. The child was seated in front of the items at the table, and the instructions were read. The child then chose an item from the array and received the item.

The child was told to place the item in a sandwich bag with his or her name on it that he or she brought back to the classroom with him or her. That item was not replaced in the array after it had been chosen. The last item on the child's left was moved so that it was in the position of the last item on the child's right. All remaining items were readjusted so that all items were once again an equal distance apart. Then, the child was prompted to choose again. The order in which the items were chosen was recorded by the experimenter. Sessions ended when the child selected all items or stated that he or she did not like any of the remaining items. In addition, if the child did not select an item after 60 seconds, the session was to be terminated, but this never occurred. If any items were remaining, they were scored as receiving zero points. The preference assessment was conducted using the brief MSWO procedure, which consisted of three sessions.

Data Analysis

Spearman rho correlation coefficients were used to determine the relationships between items identified by the teacher survey, teacher ranking, child survey, and brief MSWO preference assessment. The child survey and teacher survey scores were already in a 1 to 5 format. The teacher rankings were put into a 1 to 5 format by dividing the 20 items by 4 to make 5 groups of 4 items each. Items ranked 1-4 were coded as a 1, items ranked 5-8 were coded as a 2, items ranked 9-12 were coded as a 3, items ranked 13-16 were coded as a 4, and items ranked 17-20 were coded as a 5. The brief MSWO preference assessment's ranking points were averaged across the 3 sessions. The points were put into a 1 to 5 format by distributing them into groups. All items with 15 or more points were coded as a 1, all items with less than 15 points and 10 or more points were coded as a 2, all items with less than 10 points and 5 or more points were coded as a 3, all items with less than 5 points and 1

point or more were coded as a 4, all items with less than 1 point were coded as a 5. This enabled all preference assessment formats to be compared.

Results

A Spearman rho rank order correlation was run between the preference assessment, child survey, teacher survey, and teacher rankings for each item. The relationship between the preference assessment and child survey was .36 on average. The relationship between the preference assessment and teacher survey produced an average correlation coefficient of .03. An average correlation of .14 was found between the preference assessment and teacher rankings. The mean relationship between the child survey and teacher survey was .15, and the mean relationship between the child survey and teacher rankings was .12. Finally, the average relationship between the teacher survey and the teacher rankings was the highest at .38. In addition, separate Spearman rho rank order correlations were run for each item, producing 400 correlation coefficients. About 20 significant correlations were found out of the 400 run, which would be expected due to chance.

Discussion

Experiment one was carried out in order to determine if there was a relationship between preference assessments, child surveys, teacher surveys, and teacher rankings for identifying children's preferred items. If a significant relationship were found between the preference assessment and one of the indirect assessments, it could save professionals time when attempting to identify a child's preferred items for use in reinforcement based interventions. Unfortunately, this experiment found that, overall, the average correlations between the different types of assessments were low, ranging from .03 to .38.

The preference assessment and child survey had a weak relationship at .36. In 2000, Northup conducted in a group experiment in which he compared a reinforcer assessment to a child survey. Northup (2000) found that child surveys could predict reinforcers only 57% of the time. This study showed that child surveys are also weak when used to select preferred items.

Extending the literature, this study also considered the relationship between preference assessments and teacher endorsed items. Teachers filled out both a teacher survey and a teacher ranking considering all twenty of the items in the preference assessment. There was almost no correlation between the preference assessment and teacher survey, which produced the lowest relationship at .03, and the mean correlation between the preference assessment and teacher rankings, was also very low at .14. These findings are similar to that of Green et al. (1988) in which they found that there was no relationship between institutional staff surveys and preference assessments.

Like the correlations between the preference assessments and teacher endorsed items, the correlations between the child survey and teacher responses were low. The mean correlation between the child survey and teacher survey was .15, and the relationship between the child survey and teacher ranking was .12. This replicates previous studies that have found either no correlation (Caffyn, 1987) or a low correlation (Jacob et al., 1984) between child surveys and teacher surveys of child preferred items. The lack of relationship between the surveys in this study was expected after examining the raw data; there was an obvious absence of variability among items. Children rated most items as either highly preferred or highly unpreferred, and teachers rated almost all items as highly preferred.

The highest correlation was between the teacher survey and teacher rankings at .38; this isn't surprising since they were both filled out by the same teacher. It is surprising, however, that the relationship is fairly weak. When examining the surveys, one can note that the teachers usually ranked most items as highly preferred and the rankings forced the teachers to consider the items in relation to one another, therefore, producing a larger range of rankings.

Several limitations of this study are worth noting. First, the preference assessment sessions and child survey were completed in the same sitting for most children. No child assessment lasted longer than 15 minutes, but this still may have caused fatigue and affected item rankings. Second, child assessments occurred at different times of the school day. Some assessments were before lunch and others were after lunch. This may have affected the child's rating of edible items in the array. Third, it is possible that a child may have been unfamiliar with one or more of the items in the array. Due to time constraints, the experimenter did not allow the children to sample each item before assessment. This may have caused some children to avoid certain items or choose particular items based on their novelty. Fourth, twenty items were considered. Typically, about seven items are considered in multiple stimulus preference assessments (DeLeon & Iwata, 1996). The vast array of items may have been difficult to attend to and scan for some children, although this problem was not visibly noted by the experimenter. Finally, the sample size was rather small, with twenty participants.

There are many directions future studies in this area can take. This study was one of the first to consider the relationship between preference assessments, child surveys, and teacher surveys for typically developing children without any diagnoses. Psychologists are

often asked to develop reinforcement based interventions for this population, and it is important that we find the most efficient method for determining preferred items for use in these interventions. Future studies should continue to consider this population and should aim for a larger sample size than those previously considered. Finally, the relationship between preference assessments, child surveys, and parent surveys is yet to be tested. It is possible that, since parents theoretically know their child better and spend more time with their child than the child's teacher, parents may be better able to select items that are highly preferred by their child.

In sum, this study found that item rankings produced by preference assessments did not correlate highly with those produced by child surveys or teacher surveys or rankings. The implication of this finding is that teachers may not identify items that children prefer, and children may not verbally select items that they would physically choose when presented with these items. It appears that one may need to carry out a preference assessment in order to determine what items children prefer before choosing items for use in an intervention.

Experiment II: Evaluation of Identified Preferred Items' Effectiveness as Reinforcers

This experiment used a single subject alternating treatments design to compare the reinforcing effectiveness of the most highly preferred item identified via the brief MSWO preference assessment and the most highly preferred item identified via teacher ranking. The reinforcer assessment utilized a single operant design and required the child to answer math problems. The dependent measure was the number of digits correctly solved in two minutes.

Methods

Selection Criteria

Fourteen first and second grade teachers each identified one child who they believed exhibited a performance deficit in math and could benefit from math practice. After possible participants were identified, a permission form was sent home to the parents of the children (Appendix H). The form detailed the purpose and procedures of the assessment and experiment and asked for informed consent. All child participants assented to participation during experiment one (Appendix C).

All fourteen children participated in experiment one and then were screened for performance deficits using grade level math curriculum based measures (CBMs). Each child was brought into a quiet room by the experimenter and seated at a table. The child was given a grade level math probe and told that he or she could complete as few or as many math problems as he or she desired. After two minutes passed, the probe was taken from the child and scored. Then, the child was given a second probe consisting of similar math problems at the same grade level. Again, he or she was told that he or she could complete as few or as many math problems as he or she desired. This second probe was taken from the child and scored after two minutes had passed. Finally, the child was given a third probe consisting of

similar subtraction problems. The child was told that if he or she answered more problems correctly than he or she did on his or her first or second try, he or she would be able to select an item out of a treasure chest. Then, he or she was given two minutes to answer as many questions as he or she could. If his or her score increased 20% or more during the reinforced trial, he or she was judged to have a performance deficit and was included in this experiment (Witt, 2002). If his or her score did not increase by at least 20%, he or she did not meet criteria for a performance deficit and was not included in this experiment. Four children qualified for participation.

Next, children were selected from those identified as exhibiting performance deficits for whom the teacher ranking and brief MSWO preference assessment identified different highly preferred items. All four children again qualified for participation. The highest ranked item by the teacher was never ranked greater than 4th by the child during the brief MSWO preference assessment. Rankings ranged from 4 to 17. The highest ranked item by the child during the brief MSWO preference assessment was never ranked greater than 5th by the teacher ranking. Rankings ranged from 5 to 17.

Participants and Setting

Four children from experiment one participated, three females and one male. Two of the females and the one male were attending first grade, and one female was attending second grade. All children were Caucasian. None of the children was receiving any special services or had any diagnoses. Treatment and assessment were conducted by a trained graduate student in a quiet room within the school building.

Data Collection and Dependent Variables

Math Fluency. This dependent measure was the number of digits correctly answered in two minutes during each grade level math probe. Math probes were curriculum-based measures consisting of grade level subtraction problems. Multiple forms of math probes containing similar problems were constructed for use and administered in a random order.

Reliability. Two scorers independently scored the number of digits correctly answered on 44% of the occasions. Interscorer agreement was calculated by dividing all agreements by the sum of agreements and disagreements multiplied by 100. Overall interscorer agreement was 92%.

Experimental Design

An alternating treatments design was used to evaluate the effects of three separate conditions (the brief MSWO preference assessment rewards condition, the teacher ranking rewards condition, and the control) on the number of digits correctly completed in two minutes. The sequence of conditions was counterbalanced. No more than three sessions were carried out per day for each participant.

Procedure

Fluency Assessment. Before baseline, the child was given three grade level subtraction math probes during screening. The child's median score of digits correctly answered in two minutes was taken to indicate his or her current fluency for grade level math. This score was used as the criterion for reward during the experiment. For Kailey, this was 17 digits correct. For Heidi, this was 14 digits correct. For Caleb, this was 7 digits correct. Finally, for Emma, this was 15 digits correct.

Baseline. When this condition occurred, the child was told his or her baseline score, but rewards were not discussed or provided. Then, the child was given a two minute timed grade level subtraction math probe. At the end of the probe, the child was told whether or not he or she beat his or her score, but he or she was not offered a reward.

Preference Assessment Rewards Condition. When this condition occurred, the child was told his or her criterion score and that if he or she beat this score, he or she would get to receive a reward. Then, the reward that the child was able to earn was shown to the child briefly. The reward used during this condition was the item that was identified as most preferred by the child during the brief MSWO preference assessment in experiment one. Next, the child was given a two minute timed grade level subtraction math probe. At the end of the probe, the child was told whether or not he or she beat his or her score and was given the reward if he or she did beat his or her score.

Teacher Ranking Rewards Condition. In this condition, the child was told his or criterion score and that if he or she beat this score, he or she would receive a reward. Then, the reward that the child was able to earn was shown to the child briefly. The reward used during this condition was the item that the teacher ranked as that which he or she believed the child would like to receive the most in experiment one. Next, the child was given a two minute timed grade level subtraction math probe. At the end of the probe, the child was told whether or not he or she beat his or her score and was given the reward if he or she did beat his or her score.

Results

Data were analyzed by visual inspection and comparing the mean number of digits correctly answered across conditions. Consulting Figures 1-4, one can see that, overall, the

teacher ranking reward and preference assessment reward conditions produced more digits correctly answered in two minutes than the no reward condition. This can also be seen by looking at mean digits correct per minute across the study for each participant.

In Figure 1, it can be seen that at first Kailey did not discriminate between the three conditions. Toward the end of the experiment, one can see that Kailey often completed fewer digits correct under the no reward condition, although two points do cross over the teacher reward. The difference between the preference assessment reward and teacher ranking reward is unclear and undifferentiated.

The mean digits correct across the study for Kailey were: 17.5 in the no reward condition, 24 in the teacher ranking reward condition, and 28.6 in the preference assessment reward condition. Examining the mean digits correct across the study reveals that overall Kailey completed the most digits correct under the preference assessment reward condition. In addition, she completed 6.5 more digits correct on average in the teacher ranking condition than the no reward condition, and 11.1 more digits correct on average in the preference assessment reward condition than the no reward condition. However, examination of the graphed data suggests a lack of differentiation between the teacher condition and the child condition.

Heidi's data can be examined in Figure 2. As with Kailey, there was little discrimination in the beginning of the experiment. There is little difference between the child condition and the other two conditions. A difference can be seen between the no reward condition and the teacher condition, with the teacher condition producing more digits correct.

Heidi's mean digits correct across the study were: 18.3 for the no reward condition, 23.3 for the teacher ranking reward condition, and 22.8 for the preference assessment reward

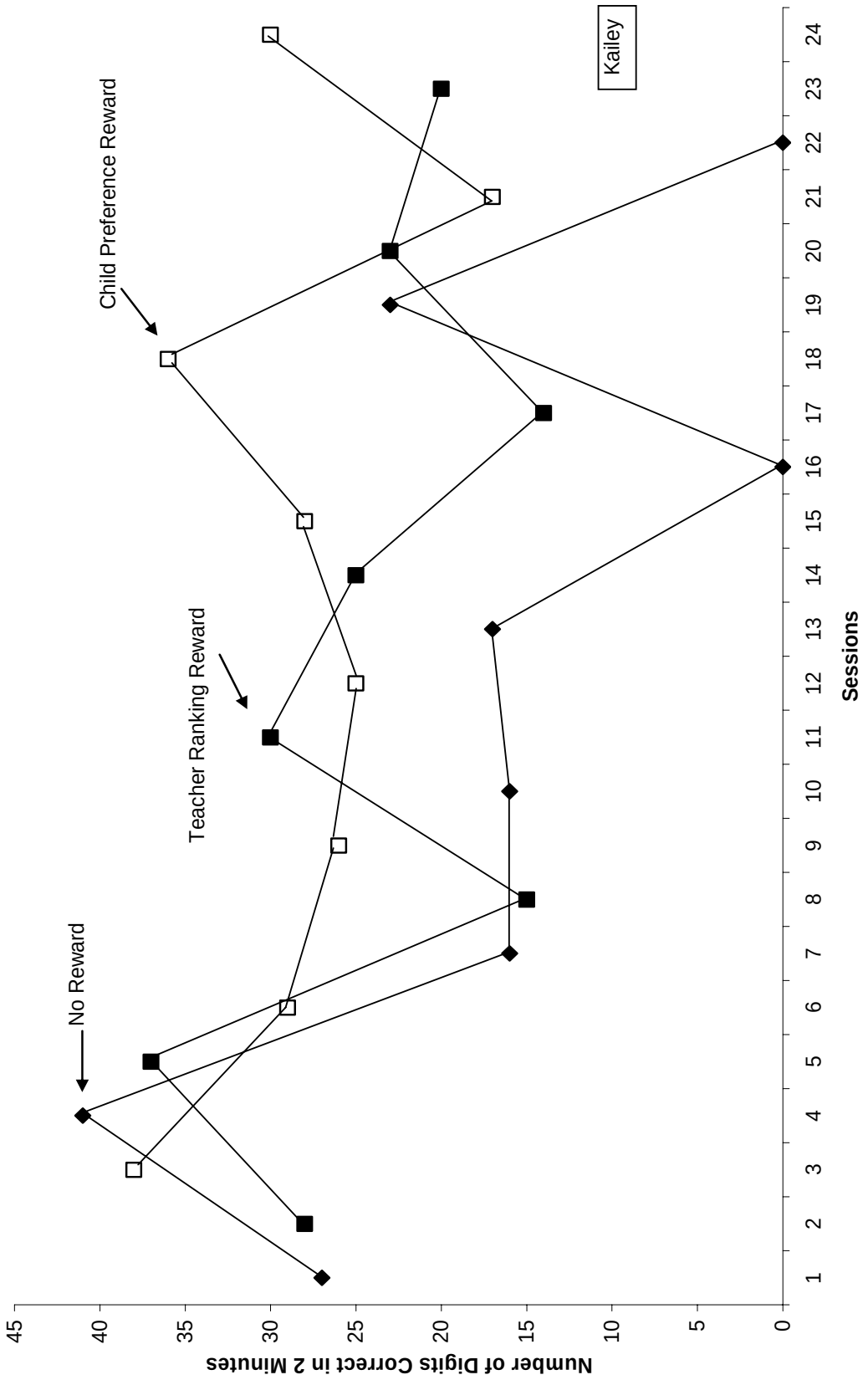

Figure 1. Number of Digits Correct in Two Minutes for Kailey.

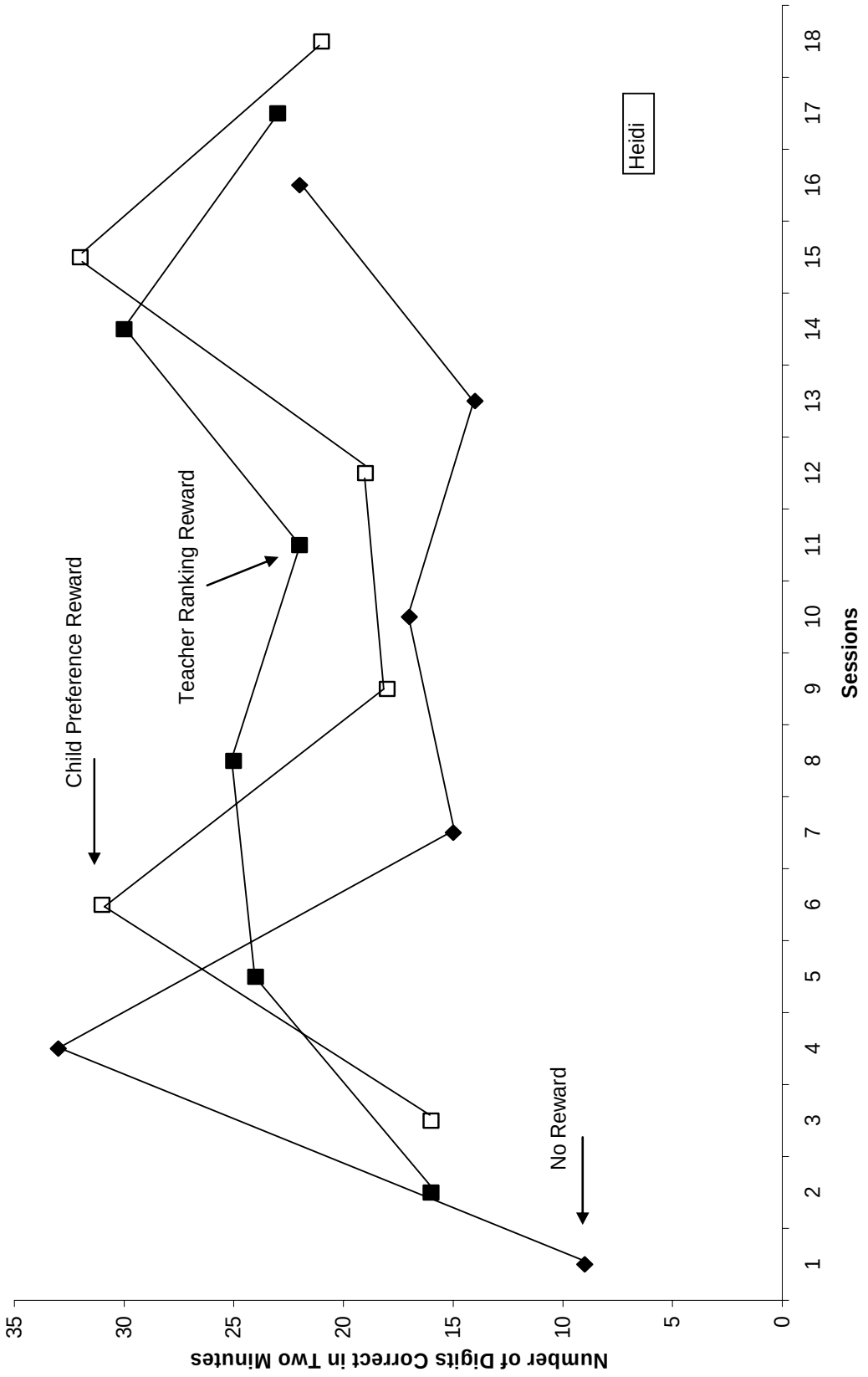

Figure 2. Number of Digits Correct in Two Minutes for Heidi.

condition. The means show that there was only a small difference between the preference assessment and teacher ranking reward conditions that is likely to reflect chance variations. The greatest difference was that between the no reward and teacher ranking reward conditions (5 digits).

Caleb's digits correct in two minutes per condition are presented in Figure 3. It can be seen that Caleb differentiated among the experimental conditions and the no reward condition clearly and early on. The experimental conditions produced significantly greater amounts of digits correct. The teacher ranking reward and preference assessment reward conditions remain undifferentiated.

When examining the mean digits correct across the study for Caleb, the no reward condition produced .3, the teacher ranking reward condition produced 14.3, and the preference assessment reward condition produced 15.3. There is only a 1 digit difference in the mean digits correct for the experimental conditions. Meanwhile, there is a large 14 digit difference between the teacher ranking reward condition and the no reward condition, and, similarly, a 15 digit difference between the preference assessment reward condition and the no reward condition.

Emma's data for all three conditions are presented in Figure 4. Like Kailey and Heidi, Emma's initial responding was undifferentiated. Like the other 3 participants, the teacher ranking reward condition and preference assessment reward condition remain undifferentiated throughout. However, differentiated responding did emerge for Emma, with very low levels of responding over the last three sessions of the no reward condition. The mean digits correct across the study for Emma were: 15.9 for the no reward condition, 23.3 for the teacher ranking reward condition, and 29.9 for the preference assessment reward.

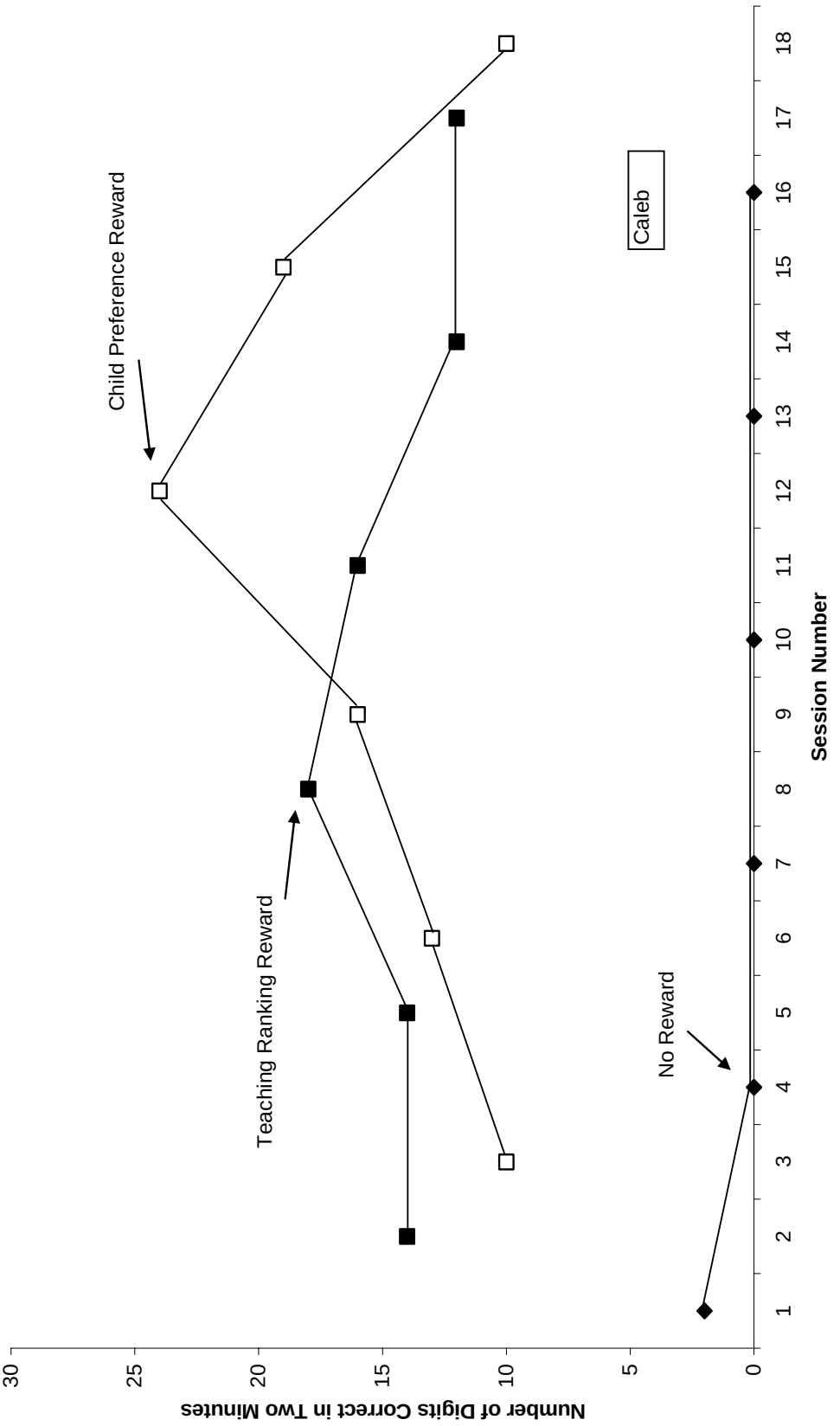

Figure 3. Number of Digits Correct in Two Minutes for Caleb.

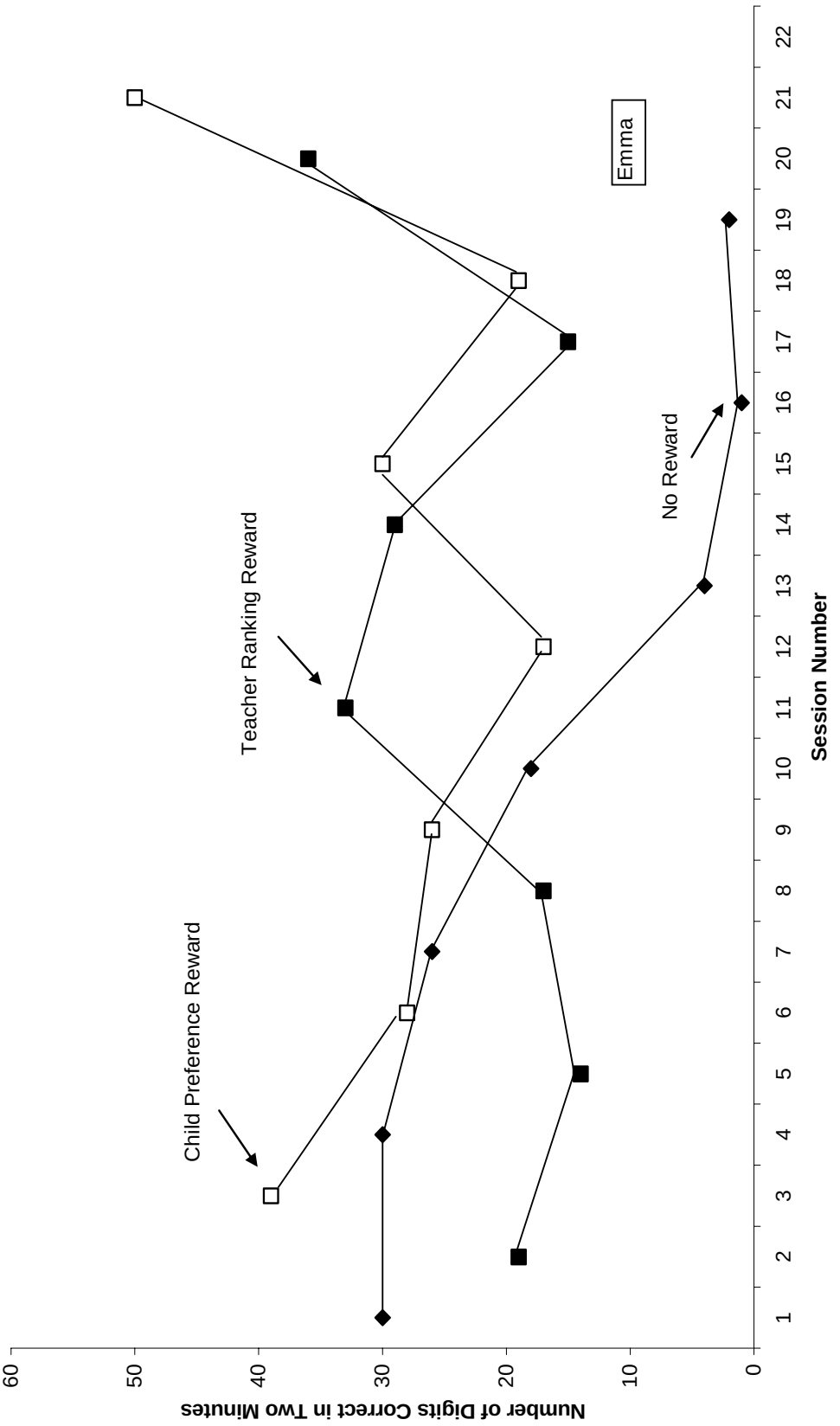

Figure 4. Number of Digits Correct in Two Minutes for Emma.

Discussion

This study examined the reinforcing effectiveness on digits answered correctly in two minutes for three conditions: preference assessment reward, teacher ranking reward, and no reward. It was found that both the preference assessment reward and teacher ranking reward produced more digits correct in two minutes on average than the no reward condition for all four participants. Interestingly, there was no clear differentiation between the preference assessment reward and teacher ranking reward for these participants.

The findings of this study differed greatly from those found by Green et al. (1988) and Green et al. (1991). In these two studies, the authors found that items identified as highly preferred through a single stimulus preference assessment reinforced behavior but items identified as preferred by an institutional staff caregiver survey did not reinforce behavior. This study found no difference in the reinforcing effectiveness of highly preferred rewards chosen through a MSWO preference assessment and highly preferred rewards chosen by teacher ranking. There are many possible explanations for this difference, one of which may be that different populations were examined. It may be that typically developing children are more likely to find diverse stimuli reinforcing than severely or profoundly mentally disabled individuals. If this is true, it may not always be necessary to perform preference assessments before selecting items for use in a reinforcement based intervention with this population.

Interpretations of these results should be tempered by consideration of the study's weaknesses. First, it is possible that reactivity to the experimenter affected child performance. Even though the participants all fit the study's selection criteria for performance deficits, two participants continued to solve problems under the no reward condition. Future studies might consider having someone the child is familiar with to conduct

the reinforcer assessment, such as a teacher or parent. This may better represent child behavior in the environment where the intervention would take place. Second, the children were all told their score to beat and what they scored in the no reward condition. It is possible that the two children who continued to solve problems during the no reward condition were reinforced enough by the act of beating their math score. This does seem unlikely, though, since all participants were identified as exhibiting a performance deficit in math. Nevertheless, this question would have been answered if an additional condition were explored in which the child was not told the score to beat, was not offered a reward, and was not told his or her score upon finishing. This is also a possibility to examine in further studies. Finally, it is possible that all of the items considered in the survey acted as reinforcers for these children. If this was the case, one would not expect any differentiation between teacher identified rewards and child identified rewards. Each participant would be just as likely to choose a reinforcer. Future studies may wish to first identify items as reinforcing or not for a particular child and then include an equal mixture of both in the surveys. This way one could determine whether teachers and children were more likely to choose items that were reinforcing over those that were not.

Future studies should also consider the importance identifying preferred items in developing interventions for typically developing children. A concurrent operant reinforcement assessment could be carried out to examine the effectiveness of a treasure chest full of items identified as highly preferred versus a treasure chest full of teacher or parent selected items. Another possible study could evaluate the effectiveness of receiving an item identified as highly preferred versus the choice between a variety of items that were ranked as less preferred.

In summary, for these four participants, there was little difference in the reinforcing effectiveness of a MSWO preference assessment selected highly preferred reward and a teacher ranking selected highly preferred reward for digits correctly completed in two minutes. The children did differentiate between the experimental conditions and the no reward condition either initially or as the analysis progressed, with all participants completing more digits correct in two minutes on average. This study implies that, for this population, who chooses the item or how the item is chosen may not be as important as it has been found to be for some more vulnerable populations (Green et al., 1988; Green et al., 1991). What may be important is that a reward is offered for improving performance.

Experiment III: Comparison of Preference Assessment Results across Sessions and Time

Methods

This experiment compared the outcomes of preference assessments consisting of varying numbers of sessions using a group design. A Spearman rho correlation was used to determine how well the results of varying numbers of sessions correlated. The dependent measure was item preference rankings. The major purpose of this experiment was to determine how well the outcomes of brief and mini MSWO preference assessments correlate with outcomes of the standard MSWO preference assessment.

The secondary purpose of this experiment was to determine how stable individual preferences are over time. The dependent measure was item preference rankings. A Spearman rho correlation was used to compare rankings of items across sessions.

Participants, Setting, and Materials

The participants were the sixteen children from experiment one who did not participate in experiment two, nine females and seven males. Six children were in kindergarten, four children were in first grade, and six children were in second grade. None of the children had any diagnoses or were receiving any special services. All participants were Caucasian.

The setting was identical to that in experiment one. The items used for the MSWO preference assessment were items identified as highly preferred based on the child survey in experiment one. Up to seven items were considered. If the child indicated that more than seven items were highly preferred during his or her survey, seven of those items were randomly chosen for use in this experiment.

Five separate sessions were carried out with at least a week separating each session. Due to school holidays and school-wide testing, some sessions were separated by two weeks. All assessments were completed within seven weeks.

Data Collection and Dependent Variables

Preference Assessment Scores. Seven different items were assessed for each child, with a maximum possibility of seven item choices. Items were scored in the order they were chosen, with the first item chosen being scored as a 1, the second item as a 2, and so on. If an item was not chosen, it was scored as an 8. See Appendix G for the scoring sheet.

Preference Assessment Reliability. Two scorers simultaneously but independently recorded the order of items chosen and the associated points received for 60% of the preference assessments. Interscorer agreement was calculated by dividing all point agreements by the sum of point agreements and disagreements multiplied by 100. Overall interscorer agreement was 100%.

Procedure

Preference Assessment. The preference assessment utilized a multiple-stimulus without replacement (MSWO) method. Before beginning the assessment, the high preference ranked items according to the child survey in experiment one were brought in and laid out in a random array on the table about five centimeters apart in a semi-circle. The same procedure as that used in experiment one was carried out using these items. The major difference was that the assessment considered fewer items and was conducted for the standard five sessions.

Data Analysis

A Spearman rho correlation was used to determine the relationship between mean rankings received with the standard five session assessment, brief three session assessment,

and mini one session assessment. The one session assessment contained rankings received on the first session carried out. The three session assessment rankings were averaged across the first three sessions. The five session assessment rankings were averaged across all five sessions. Next, a test-retest reliability of rankings for the group as a whole across sessions was computed using a Spearman rho correlation. For each child participant, rankings from each session for each item were compared to one another.

Results

A Spearman rho correlation was run to determine the relationship between a mini one session MSWO, the mean of a brief three session MSWO, and the mean of a standard five session MSWO. Results showed a significant relationship between the mini one session MSWO and the brief three session MSWO ($\rho=.81$, $p<.01$). A significant relationship was also found between the mini one session MSWO and standard five session MSWO ($\rho=.68$, $p<.01$), although it was not as strong. The strongest significant relationship found was between the brief three session MSWO and the standard five session MSWO ($\rho=.86$, $p<.01$).

A Spearman rho correlation was run to determine the reliability of rank order selections over time for the entire group. The overall reliability was significant for 7 out of 10 comparisons: session one and session two ($\rho=.41$, $p<.01$), session one and session three ($\rho=.30$, $p<.01$), session one and session five ($\rho=.29$, $p<.01$), session two and session four ($\rho=.22$, $p<.05$), session three and session four ($\rho=.28$, $p<.01$), session three and session five ($\rho=.36$, $p<.01$), and session four and session five ($\rho=.44$, $p<.01$).

Discussion

This study was conducted to determine two things. First, the author wished to compare a mini one session MSWO assessment, a brief three session MSWO assessment,

and a standard five session MSWO assessment in order to determine if a shorter assessment would be about as reliable as the standard five session MSWO assessment. Second, the authors examined whether each child's rankings across sessions were stable over time.

It was found that the brief three session MSWO assessment had the strongest relationship with the standard five session MSWO assessment at .86. This falls in line with the Carr et al. (2000) finding that a brief MSWO assessment was able to predict items that would act as reinforcers in a reinforcer assessment. This finding contradicts the Graff and Ciccone (2002) conclusion that professionals would be best served by the standard five session MSWO assessment with fewer items.

This study also found that the mini one session assessment correlated highly (ρ= .81) with the brief three session assessment. This replicated the Carr et al. (2000) findings that the one session and three session MSWO assessments had a strong relationship. In addition to examining this relationship, this study extended the past literature by determining the correlation between the mini one session assessment and standard five session assessment, which was found to be moderate to high at .68.

It should be noted that some caution must be taken when interpreting these correlations. It was expected that these assessments would have some relationship due to the overlap found between assessments from averaging sessions. The mini one session MSWO assessment overlaps the three session MSWO assessment by 33%, and the brief three session MSWO assessment overlaps the standard five session assessment by 60%.

The second correlational analysis for this experiment found that rankings over time were moderately related. The highest relations found were between the first and second (ρ=.41) and fourth and five (ρ=.44) sessions, which could be attributed to the fact that these

sessions were closer together in time. However, these relationships account for less than half the variance. It appears that this population's preferences changed somewhat over time rather than remaining stable as the participants in the Green et al. (1991) study. These results are more in line with the Mason et al. (1989) finding that participant preferences changed over a month.

This study has several limitations. First, the sample size was rather small at 16 participants. Further studies should examine these questions using a larger sample size. Second, all sessions included seven items for each child. This number was chosen based on the average amount used in previous studies (DeLeon & Iwata, 2006). Graff and Ciccone (2002) found that the number of items examined, or trials, affected how well the assessments correlated. Future studies may want to examine how related the one session, three session, and five session assessments are when the number of items examined vary. Third, this experiment only considered the stability of rankings across a time span of about two months with at least one week between each session. Other researchers may wish to examine the stability of items across shorter or longer periods of time.

This experiment found that the brief three session MSWO assessment was strongly correlated with the standard five session MSWO assessment. With this strong relationship, it seems that it may be more efficient for professionals to use a brief three session MSWO assessment when attempting to identify preferred items. In addition, this study found that preferences were not entirely stable over time for this population. If a professional intends to use MSWO preference assessments with this population, he or she would be advised to run the assessments more frequently (perhaps daily based on the Mason et al. (1989) study) to determine changes in preference.

General Discussion

These three experiments examined ways to identify preferred items that may be effective and time efficient. It is important to identify preferred items for use in multiple reinforcement based interventions. If the task of identifying preferred items can be made easier and less time consuming, it may allow the professional to develop and put in place an effective intervention sooner.

Several reinforcer surveys have been developed for the purpose of identifying reinforcers for use in interventions either through questioning individuals or their caregivers (Cautela & Brion-Meisels, 1979; Fantuzzo et al., 1991; Keat, 1974; Matson et al., 1999), but none of these surveys have been compared to preference or reinforcer assessments. Northup and colleagues (1995) were the first to test the validity of child surveys. Northup (2000) continued this line of research and found that a reinforcement survey was able to predict items that would act as reinforcers in a reinforcer assessment 57% of the time. Green et al. (1991) and Green et al. (1988) studied the validity of caregiver surveys and found that caregiver ratings of items did not predict their effectiveness as reinforcers as well as a single stimulus preference assessment.

To further the research on the validity of child and caregiver surveys, experiment one was conducted in order to determine how well child surveys, teacher surveys, and teacher rankings correlate with a brief MSWO preference assessment. Results showed that child surveys, teacher surveys, and teacher rankings did not correlate highly with preference assessments. This fits with Northup's (2000) finding that child surveys could only predict items identified by a reinforcer assessment a little over half the time. This experiment also replicated the Green et al. (1988) study that found caregiver surveys did not highly correlate

with preference assessment ratings. This study also extends on the Green et al. (1988) study by showing that teachers of typically developing children were also unable to choose items either via survey or ranking that would be rated as high preference during a preference assessment.

Past studies have examined whether caregivers were able to identify items that would act as reinforcers for their clients (Green et al., 1991; Green et al., 1988). In both studies, Green and colleagues compared the reinforcing effectiveness of items ranked as highly preferred by a staff survey and a single stimulus preference assessment. Green et al. (1988) found that stimuli which were ranked low by a preference assessment but highly preferred by a staff survey did not act as reinforcers for any of the five participants, but stimuli that were ranked as highly preferred by both the preference assessment and staff survey were associated with the highest levels of responding for four of the five participants. Green et al. (1991) replicated these findings.

Experiment two tested the reinforcing effectiveness of four conditions: no reward (control), the item identified as most preferred by the teacher ranking, and the item identified as most preferred by the brief MSWO preference assessment. The number of digits correct in two minutes was the dependent variable. Results showed that, although preference assessments did not correlate with teacher rankings, items identified as highly preferred through teacher rankings produced performance similar to items identified as highly preferred through preference assessments. This finding contradicts that of Green et al. (1988) and Green et al. (1991), who found that items only acted as reinforcers if they were selected as highly preferred via the preference assessment.

There are a few possible reasons why this study's findings differed from past studies. First, the population in the Green et al. (1998) and Green et al. (1991) studies were institutionalized individuals of varying ages who were severely mentally disabled while the population in this study was typically developing children. It is possible that these two populations differ in the range of items they find reinforcing. Second, the previous studies compared staff rankings while the present study compared teacher rankings. It is possible that these two groups differ in their ability to identify reinforcing items. Third, earlier studies used a single stimulus preference assessment, but this study utilized a brief MSWO preference assessment. The MSWO preference assessment has been shown to more readily identify reinforcers and produce less false positives than the single stimulus preference assessment (DeLeon & Iwata, 1996; Fisher et al., 1992). Fourth, definitions of "highly preferred" items for each study differed. In the Green et al. (1998) and Green et al. (1991) studies it was defined as any stimulus that was approached 80% or more of the time in the single stimulus preference assessment or ranked in the top 80% when averaged across staff rankings. In this study, the item ranked as first among all twenty items assessed was labeled as highly preferred for both the teacher ranking and MSWO preference assessment. This could lead to different items being labeled as highly preferred depending on the definition used.

A few studies have examined the benefits and risks of conducting preference assessments containing fewer sessions. Carr et al. (2000) found that a brief three session MSWO assessment was able to identify items that worked as effective reinforcers. They also found a strong relationship between the brief three session MSWO and a mini one session MSWO assessment. In 2002, Graff and Ciccone examined whether the same items would be

identified as most preferred if less sessions and trials were used. When a three session assessment with seven trials was utilized, they found that the same items were identified 70% of the time, but the five session assessment with three trials resulted in the same items being identified 81% of the time.

Experiment three set out to determine how well a mini one session MSWO assessment and brief three session MSWO assessment would correlate with the standard five session MSWO assessment. It found that time can be saved by using a brief three session MSWO assessment, which highly correlated (.86) with a standard five session MSWO assessment. This showed the brief three session MSWO assessment to be much more related to the standard five session MSWO than reported by Graff and Ciccone (2002). It was also found that the mini one session MSWO assessment correlated with the brief three session MSWO assessment at .81, which replicates the earlier finding by Carr et al. (2000) that these two assessments were highly related.

Two past studies have considered the stability of preference over time (Green et al., 1991; Mason et al., 1989). Mason et al. (1989) showed that a daily preference assessment identified different reinforcers that increased adaptive behavior for three children with autism. This study showed that child preferences did change over time. Green et al. (1991) assessed the preferences of 12 institutionalized individuals with severe mental retardation over a time period of four to twenty eight months using a single operant method. They found significant correlations for 11 of the 12 participants, revealing that preferences were pretty consistent over time for this population.

Experiment three examined whether typically developing children's preferences changed over time. It was found that preference stability over five sessions spaced over two

months ranged from .22 to .44. This shows that, for the most part, preferences did change over time. This finding is more in line with the Mason et al. (1989) finding than the Green et al. (1991) results. Again, this could be due to the population differences. Also, Green et al. (1991) surveyed preferences over a longer period of time. It is possible that if preferences had been measured over a longer period of time in this study, a larger sample of rankings would have produced a greater correlation.

Taken together, these experiments imply that preference assessments may not correlate well with more time efficient indirect assessments, but this may *not* be a critical issue. It may be that typically developing children perform similarly when an item is chosen for them as a reinforcer as when they choose it through a preference assessment. This is a possibility that needs to be studied further in order to be confirmed or rejected. If this is the case, the fact that children's preferences change may not mean that we need to constantly perform assessments, but that we need to change reinforcers used over time. Future studies should compare the reinforcing effectiveness of items selected by daily reinforcer assessments versus various items selected daily by teachers, psychologists, or parents. Finally, if future studies do show that children's performance is significantly better when they select their own reinforcers through preference assessments, preference assessments can be made shorter by utilizing a brief three session MSWO assessment or one session MSWO assessment, depending on the level of reliability required.

References

Atkeson, B. M., & Forehand, R. (1979). Home-based reinforcement programs designed to modify classroom behavior: A review and methodological evaluation. Psychological Bulletin, 86(6), 1298-1308.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1(1), 91-97.

Barth, R. (1979). Home-based reinforcement of school behavior: A review and analysis. Review of Educational Research, 49(3), 436-458.

Caffyn, R. E. (1987). Rewards and Punishments in Schools: A study of their effectiveness as perceived by secondary school pupils and their teachers. School Psychology International, 8, 85-94.

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(3), 353-357.

Cautela, J. R., & Brion-Meisels, L. B. (1979). Children's reinforcement survey schedule. Psychological Reports, 44, 327-338.

Delquadri, J., Greenwood, C.R., Whorton, D., Carta, J. J., & Hall, R. V. (1986). Classwide peer tutoring. Exceptional Children, 52(6), 535-542.

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcers preferences. Journal of Applied Behavior Analysis, 29(4), 519-533.

Drasgow, E., & Yell, M. L. (2001). Functional behavioral assessments: Legal requirements and challenges. School Psychology Review, 30(2), 239-251.

Fantuzzo, J. W., Rohrbeck, C. A., Hightower, A. D., & Work, W. C. (1991). Teachers' use and children's preferences of rewards in elementary school. Psychology in the Schools, 28, 175-181.

Fisher, W. W., Piazza, C. C., Bowman, L. G., & Amari, A. (1996). Integrating caregiver report with a systematic choice assessment to enhance reinforcer identification. American Journal on Mental Retardation, 101(1), 15-25.

Fisher, W. W., Piazza, C. C., Bowman. L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. Journal of Applied Behavior Analysis, 25(2), 491-498.

Fisher, W. W., & Mazur, J. E. (1997). Basic and applied research on choice responding. Journal of Applied Behavior Analysis, 30(3), 387-410.

Freeland, J. T., & Noell, G. H. (1999). Maintaining accurate math responses in elementary school students: The effects of delayed intermittent reinforcement and programming common stimuli. Journal of Applied Behavior Analysis, 32(2), 211-215.

Graff, R. B., & Ciccone, F. J. (2002). A post hoc analysis of multiple-stimulus preference assessment results. Behavioral Interventions, 17, 85-92.

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.

Green, C. W., Reid, D. H., White, L. K., Halford, R. C., Brittain, D. P., & Gardner, S. M. (1998). Identifying reinforcers for persons with profound handicaps: Staff opinion versus systematic assessment of preferences. Journal of Applied Behavior Analysis, 21(1), 31-43.

Greenwood, C. R., Arreaga-Mayer, C., Utley, C. A., Gavin, K. M., & Terry, B. J. (2001). Classwide peer tutoring learning management system: Applications with elementary level English language learners. Remedial and Special Education, 22(1), 34-47.

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

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.

Ivancic, M. T. (2000). Stimulus preference and reinforcer assessment applications. In J. Austin & J. E. Carr (Eds.), Handbook of applied behavior analysis (pp.19-38), Reno, NV; Context Press.

Jacob, S., Daly, M. D., King, D. W., & Cheramie, G. (1984). The accuracy of teacher predictions of student reward preferences. Psychology in the Schools, 21, 520-524.

Keat, D. B. (1974). Survey schedule of rewards for children. Psychological Reports, 35, 287-293.

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(2), 171-179.

Matson, J. L., Bielecki, J., Mayville, E. A., Smalls, Y., Bamburg, J. W., & Baglio, C. (1999). The development of a reinforcer choice assessment scale for persons with severe and profound mental retardation. Research in Developmental Disabilities, 20(5), 379-384.

Noell, G. H., Freeland, J. T., Witt, J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students, Journal of School Psychology, 39(4), 335-355.

Noell, G. H., Gansle, K. A., Witt, J. C., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. H., Gilbertson, D. N., & Northup, J. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. Journal of Applied Behavior Analysis, 31(4), 659-663.

Northup, J. (2000). Further evaluation of the accuracy of reinforcer surveys: A systematic replication. Journal of Applied Behavior Analysis, 33(3), 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.

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(1), 99-100.

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(3), 249-255.

Paclawskyj, T. R., & Vollmer, T. R. (1995). Reinforcer assessment for children with developmental disabilities and visual impairments. Journal of Applied Behavior Analysis, 28(2), 219-224.

Pavchinski, P., Evans, J. H., & Bostow, D. E. (1989). Increasing word recognition and math ability in a severe learning-disabled student with token reinforcers. Psychology in the Schools, 26, 397-411.

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), 1-9.

Sulzer-Azaroff, B., & Mayer, R. G. (1986). Achieving education excellence: Using behavioral strategies. New York: Holt, Rinehart, & Winston.

Windsor, J., Piche, L. M., & Locke, P. A. (1994). Preference testing: A comparison of two presentation methods. Research in Developmental Disabilities, 15(6), 439-455.

Witt, J. C. (2002). STEEP Administration Manual. Retrieved January 9, 2003 from http://www.joewitt.org/instruction.

Appendix A

Parent Consent Form A

Dear Parent(s),

We are writing to request your permission for your child to participate in a study that is being conducted at A. Elementary by a team from the Department of Psychology at Louisiana State University. Your child's participation in this study would require that a master's level graduate student assess his or her preferences for rewards. Your son or daughter has been nominated by his or her teacher to represent children in his or her classroom. If you agree, an assessment of your child's reward preferences will be performed. This will consist of a survey and a direct assessment in which the child will be asked to identify rewards he or she likes. The assessment time will be determined by your son or daughter's teacher to ensure that no important school activities will be missed. The assessment will take approximately 10 to 15 minutes a day and will be conducted on several separate days at the school during the time period of February 14 through March 18. Your child's principal has approved this assessment with your consent.

The information obtained from this assessment will be used as part of research that is being conducted by graduate students who are doctoral candidates at Louisiana State University. The name of the study is "Evaluation of Cost Effective Preference Assessments for use in Regular Education Settings." This research is being conducted in hopes of developing a more efficient way to assess children's reward preferences. This assessment will not affect your son or daughter's school grade or standing. There are no known risks associated with this study. Any data collected concerning your child will be remain confidential and your child's name will not be included in any research reports. The LSU Institutional Review Board (which oversees research) may inspect the study's records. Your child's records will not be released to anyone outside the research team without your permission. You may choose not to participate in the study if you prefer. You may withdraw your child from this activity at any time with no penalty to yourself or your child.

A list of the rewards being assessed is attached to this note. If your son or daughter is allergic to any of the candy being used, please do not allow him or her to participate. If your son or daughter is allergic to peanuts, chocolate, or red dyes, he or she should not participate.

If you have any questions about this assessment, please feel free to contact us at your earliest convenience.

Sincerely,

Jennifer Resetar, M.A.	M.S.	George H. Noell, Ph.D.
Graduate Consultant	School Principal	Associate Professor
	A. Elementary	Graduate Supervisor

Please Keep This Portion For Your Records

List of Rewards That Will be Assessed

1. Stickers
2. Hershey Kisses™ chocolates
3. Bead necklaces
4. Book marks
5. Colorful pencils
6. Erasers
7. Colorful notepads
8. Mini Reeses™ peanut butter cups
9. Sweet Tart™ candies
10. Goldfish™ cheese crackers
11. Toy rings
12. Mini Milky Way™ candy bars
13. Gummy bear candies
14. Bear shaped graham crackers
15. Colored markers
16. Mini Snickers™ candy bars
17. Animal crackers
18. Small toy cars
19. Small toy dinosaurs
20. Small bouncing balls

Please Check One and Return This Portion to School

_____ Yes, I give my permission for my child to be assessed. I have looked over the list of rewards that will be used and I approve of their use with my child.

_____ No, I do not give my permission for my child to be assessed.

Print Name: _____

Signature: _____

If you have additional questions about participant's rights or other concerns regarding the research component of this activity you can contact: Robert C. Mathews, Institutional Review Board, Louisiana State University, (225) 578-8692.

Appendix B

Teacher Consent Form

Dear Teacher,

We are writing to ask for your participation in a research study being conducted at A. Elementary by a team from the Department of Psychology at Louisiana State University. We are asking that you nominate one student in your classroom to be assessed. We are interested in assessing how effective particular rewards are for individual children, so we are asking that you nominate one student whom you would like to identify an effective reward or reinforcer for. We will assess the students' reward preferences with several measures. These measures will consist of a survey and direct measurement. The direct assessment will be conducted on several separate days throughout the time period of February 14 through April 8. Each assessment will take approximately 10 to 15 minutes. In addition, we will also ask you to fill out a brief teacher survey based on your experience with the child. Each survey should take about 5 minutes to complete. The survey will ask you to rate how much you believe the participating child would prefer certain rewards. Appropriate times for assessment will be determined according to your schedule.

In addition, first and second grade teachers should nominate a child who they believe may have the skills to complete work but are currently not motivated to do so. These students may be screened for participation in a second portion of the study that would provide math practice. It will also be determined if offering rewards for correct work completion will increase work performance.

The information obtained from this assessment will be used as part of research that is being conducted by graduate students who are doctoral candidates at Louisiana State University. The name of the study is "Evaluation of Cost Effective Preference Assessments for use in Regular Education Settings." This research is being conducted in hopes of developing a more efficient way to assess children's preference. Your principal has approved your participation with your consent. There are no known risks associated with this study. Any data collected will remain confidential and your name will not be included in any research reports. You may choose not to participate in the study if you prefer. You may withdraw from this activity at any time with no penalty to yourself or your students.

If you have any questions about this assessment, please feel free to contact us at your earliest convenience.

Sincerely,

Jennifer Resetar, M. A.	M. S.	George H. Noell, Ph.D.
Graduate Consultant	School Principal	Associate Professor
	A. Elementary	Graduate Supervisor

Please Keep This Portion For Your Records

Please Check One and Return This Portion

_____ Yes, I wish to participate in this study.

_____ No, I do not wish to participate in this study.

Print Name: _____

Signature: _____

If you have additional questions about participant's rights or other concerns regarding the research component of this activity you can contact: Robert C. Mathews, Institutional Review Board, Louisiana State University, (225) 578-8692.

Appendix C

Child Assent Form

I, _____, agree to be in a study to find out what kind of rewards kids like to work for. I will have to tell the psychologist that comes to work with me which rewards I like best. I can point to the rewards or say their names. I have to follow all the classroom rules, even when I am working with the psychologist. I can decide to stop being in the study at any time without getting in trouble.

Child's Signature _____

Age _____ Date _____

Witness _____ Date _____

Appendix D

Preference Survey - Teacher

Student Information

Student's name _____ Date _____

School _____ Grade_____

Gender: Male _____ Female _____

Ethnic Group: American Indian _____ Asian _____ Black _____ Hispanic _____

White_____ Other_____

Teacher Information

Teacher's name_____

Grade taught _____ Years of Experience_____

Certifications held _____

Gender: Male _____ Female _____

Ethnic Group: American Indian _____ Asian _____ Black _____ Hispanic _____

White_____ Other_____

Part 1 - Instructions

Following are some items that might be used to reward a child for appropriate behavior. Please circle the number corresponding to how much you believe this child would like to receive the stated reward on a scale of 1 to 5: 1 representing very much like, 2 representing like, 3 representing indifferent, 4 representing dislike, 5 representing very much dislike.

	Very Much	Like	Don't Like or Dislike	Dislike	Very Much Dislike
1. How much would the child like a bead necklace?	1	2	3	4	5
2. How much would the child like a bookmark?	1	2	3	4	5
3. How much would the child like a colorful pencil?	1	2	3	4	5
4. How much would the child like an eraser?	1	2	3	4	5
5. How much would the child like a colorful notepad?	1	2	3	4	5
6. How much would the child like a toy ring?	1	2	3	4	5
7. How much would the child like a mini Reeses™ peanut butter cup?	1	2	3	4	5
8. How much would the child like a gummy bear?	1	2	3	4	5
9. How much would the child like a Hershey Kisses™ chocolate?	1	2	3	4	5
10. How much would the child like a colored marker?	1	2	3	4	5
11. How much would the child like a mini Milky Way™ candy bar?	1	2	3	4	5
12. How much would the child like an animal cracker?	1	2	3	4	5
13. How much would the child like a mini Snickers™ candy bar?	1	2	3	4	5
14. How much would the child like a small toy car?	1	2	3	4	5
15. How much would the child like a small toy dinosaur?	1	2	3	4	5
16. How much would the child like a small bouncing ball?	1	2	3	4	5
17. How much would the child like a bear shaped graham cracker?	1	2	3	4	5

	Very Much	Like	Don't Like or Dislike	Dislike	Very Much Dislike
18. How much would the child like a sticker?	1	2	3	4	5
19. How much would the child like a Sweet Tart™ candy?	1	2	3	4	5
20. How much would the child like a Goldfish™ cheese cracker?	1	2	3	4	5

Appendix E

Preference Ranking – Teacher

Instructions

Following are some items that might be used to reward a child for appropriate behavior. Please rank these items 1-20, with 1 representing the reward that you believe the child would like to receive the most and 20 representing the reward that you believe the child would like to receive the least.

Bead necklace _____

Book mark _____

Colorful pencil _____

Eraser _____

Colorful Notepad _____

Toy Ring _____

Hershey Kisses™ chocolate _____

Gummy bear _____

Mini Reeses™ peanut butter cup _____

Colored marker _____

Mini Milky Way™ candy bar _____

Animal crackers _____

Mini Snickers™ candy bar _____

Small toy car _____

Small toy dinosaur _____

Small bouncing ball _____

Bear shaped graham cracker_____

Sticker _____

Sweet Tart™ candy _____

Goldfish™ cheese cracker _____

Appendix F

Preference Survey - Child

Student Information

Student's name _____ Date _____

School _____ Grade_____

Gender: Male _____ Female _____

Ethnic Group: American Indian _____ Asian _____ Black _____ Hispanic _____

White_____ Other_____

Teacher's name_____

Instructions

To be read aloud to child:
I'm going to name some things that kids sometimes get in school. I want to know how much you like each of these
things. After I name each thing, I'm going to ask you to show me how much you like each item. I want you to point to the face that shows how you feel about the item. (Point to each item as explaining) You might like the item very much, and that's this very happy smiley face with a grin. You might just like the item, and that's this smiley face with a smile. You might not really like it but you don't dislike it, and that's the straight face. You might dislike it, and that's this frown face. You might dislike it very much, and that's the angry frown face.

To experimenter:
Read each item to the child aloud. After reading each item, ask the child which face shows how they would feel about receiving that item. If the child does not automatically point to a smiley face, you should prompt again.

	Very Much Like	Like	Don't Like or Dislike	Dislike	Very Much Dislike
					
1. How much would you like a bead necklace?					
2. How much would you like a bookmark?					
3. How much would you like a colorful pencil?					
4. How much would you like an eraser?					
5. How much would you like a colorful notepad?					
6. How much would you like a toy ring?					
7. How much would you like a Hershey Kisses™ chocolate?					
8. How much would you like a gummy bear?					
9. How much would you like a mini Reeses™ peanut butter cup?					
10. How much would you like a colored marker?					
11. How much would you like a mini Milky Way™ candy bar?					
12. How much would you like an animal cracker?					
13. How much would you like a mini Snickers™ candy bar?					
14. How much would you like a small toy car?					
15. How much would you like a small toy dinosaur?					
16. How much would you like a small bouncing ball?					
17. How much would you like a bear shaped graham cracker?					
18. How much would you like a sticker?					
19. How much would you like a Sweet Tart™ candy?					
20. How much would you like a Goldfish ™ cheese cracker?					

Appendix G

Preference Assessment Data Sheet

Student name _____ Date _____ Time _____

Experimenter _____ Observer _____

Preference assessment experiment number and session number _____

Rank the below items in the order they are chosen and put the corresponding points received in parentheses:

Bead necklace _____

Bear shaped graham cracker _____

Book mark _____

Sticker _____

Colorful pencil _____

Sweet Tart™ candy _____

Eraser _____

Goldfish™ cheese cracker _____

Colorful notepad _____

Toy ring _____

Gummy Bear _____

Mini Reeses™ peanut butter cup _____

Hershey Kisses™ chocolate _____

Colored marker _____

Mini Milky Way™ candy bar _____

Animal cracker _____

Mini Snickers™ candy bar _____

Small toy car _____

Small toy dinosaur _____

Small bouncing ball _____

Appendix H

Parent Consent Form B

Dear Parent(s),

We are writing to request your permission for your child to participate in a study that is being conducted at A. Elementary by a team from the Department of Psychology at Louisiana State University. Your child's participation in this study would require that a master's level graduate student assess his or her preferences for rewards. Your son or daughter has been nominated by his or her teacher to represent children in his or her classroom. If you agree, an assessment of your child's reward preferences will be performed. This will consist of a survey and a direct assessment in which the child will be asked to identify rewards he or she likes. The assessment time will be determined by your son or daughter's teacher to ensure that no important school activities will be missed. The assessment will take approximately 10 to 15 minutes a day and will be conducted on several separate days at the school during the time period of February 14 through March 18. Your child's principal has approved this assessment with your consent.

If your child qualifies, he or she will also receive practice in grade level math skills. A graduate student will work with your child to help him or her become more fluent in math. Occasionally, your child will be given rewards for beating his or her score for most problems correctly completed in two minutes.

A list of the rewards being assessed is attached to this note. If your son or daughter is allergic to any of the candy being used, please do not allow him or her to participate. If your son or daughter is allergic to peanuts, chocolate, or red dyes, he or she should not participate.

The information obtained from the assessment and math tutoring will be used as part of research that is being conducted by graduate students who are doctoral candidates at Louisiana State University. The name of the study is "Evaluation of Cost Effective Preference Assessments for use in Regular Education Settings." This research is being conducted in hopes of developing a more efficient way to assess children's reward preferences. This assessment will not affect your son or daughter's school grade or standing. There are no known risks associated with this study. Any data collected concerning your child will be remain confidential and your child's name will not be included in any research reports. The LSU Institutional Review Board (which oversees research) may inspect the study's records. Your child's records will not be released to anyone outside the research team without your permission. You may choose not to participate in the study if you prefer. You may withdraw your child from this activity at any time with no penalty to yourself or your child.

If you have any questions about the assessment or math tutoring, please feel free to contact us at your earliest convenience.

Sincerely,

Jennifer Resetar, M.A.	Marsha Sherburne	George H. Noell, Ph.D.
Graduate Consultant	School Principal	Associate Professor
	A. Elementary	Graduate Supervisor

Please Keep This Portion For Your Records

List of Rewards That Will be Assessed

21. Stickers
22. Hershey Kisses™ chocolates
23. Bead necklaces
24. Book marks
25. Colorful pencils
26. Erasers
27. Colorful notepads
28. Gummy Bears candies
29. Sweet Tart™ candies
30. Goldfish™ cheese crackers
31. Toy rings
32. Mini Reeses™ peanut butter cups
33. Mini Milky Way™ candy bars
34. Colored markers
35. Mini Snickers™ candy bars
36. Animal crackers
37. Bear shaped graham crackers
38. Small toy cars
39. Small toy dinosaurs
40. Small bouncing balls

Please Check One and Return This Portion to School

_____ Yes, I give my permission for my child to be assessed and participate in math tutoring. I have looked over the list of rewards that will be used, and I approve of their use with my child.

_____ No, I do not give my permission for my child to be assessed.

Print Name: _____

Signature: _____

If you have additional questions about participant's rights or other concerns regarding the research component of this activity you can contact: Robert C. Mathews, Institutional Review Board, Louisiana State University, (225) 578-8692.

Vita

Jennifer L. Resetar grew up in Albany, Louisiana. She received her Bachelor of Arts degree in psychology from the University of Louisiana at Monroe in May 2001 and her Master of Arts degree in school psychology from Louisiana State University in December 2003. She is currently a candidate for the degree of Doctor of Philosophy in school psychology at Louisiana State University, where she works under Dr. George H. Noell. In addition, she is completing her internship at Girls and Boys Town in Omaha, Nebraska, under the supervision of Dr. Patrick C. Friman. After receiving her doctorate, she plans to pursue postdoctoral work with children and adolescents in an outpatient clinic setting.