

EFFECTS OF PRESENTATIONS OF ASSESSMENT ROUNDS ON PREFERENCE  
STABILITY

By

Alexandria Thomas

A THESIS

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Applied Behavior Analysis – Master of Arts

2021

## ABSTRACT

### EFFECTS OF PRESENTATIONS OF ASSESSMENT ROUNDS ON PREFERENCE STABILITY

By

Alexandria Thomas

Behavior interventions have been found to be the most effective treatments for behaviors associated with Autism Spectrum Disorder (ASD). Children diagnosed with ASD tend to experience barriers in terms of communication, thus communicating wants and needs during treatment may be difficult. As a result, clinicians have used preference assessments to identify potentially reinforcing stimuli to use during behavior interventions to increase the likelihood of desired behaviors occurring again in the future. Previous research on preference assessments has looked at evaluating brief preference assessments and stability of responding across time and assessments. The purpose of this study is to evaluate the extent to which responding is stable across rounds within a single MSWO for children aged 3-5 with a diagnosis of ASD. Results showed that overall, stability in responding across rounds of a single MSWO varied across participants regardless the type of stimuli used during the assessment (all edible or all tangible stimuli).

*Keywords:* autism, preference assessments, stability

## ACKNOWLEDGEMENTS

I would like to thank Michigan State University for providing the resources needed to conduct this thesis. I would also like to give a special thanks to my advisor, Dr. Matthew Brodhead, for continued patience, guidance, and wisdom through this process.

## TABLE OF CONTENTS

LIST OF TABLES .....	v
LIST OF FIGURES .....	vi
INTRODUCTION .....	1
METHOD .....	5
Secondary Data Analysis .....	5
Participants and Setting .....	5
MSWO Preference Assessments .....	6
Dependent Measures .....	7
Interobserver Agreement .....	8
Procedure .....	8
RESULTS.....	10
General Analysis of Stability .....	19
DISCUSSION.....	20
Potential Limitations.....	23
APPENDIX.....	27
REFERENCES .....	30

## LIST OF TABLES

Table 1: r-values calculated for participants in this study .....	28
---	----

## LIST OF FIGURES

Figure 1: A flowchart depicting participant recruitment and inclusion this study.....29

## INTRODUCTION

Autism Spectrum Disorder (ASD) is a developmental disorder that affects social interactions and communication, and is often accompanied by repetitive behaviors and restricted interests (American Psychiatric Association, 2013). In effort to support people with ASD, researchers have evaluated several methods to address and improve these core deficits. Behavioral interventions are one such example; with outcome studies suggesting behavioral interventions can improve upon core deficits and result in increased independence in people with ASD (e.g., Eldevik et al., 2009).

A cornerstone of an effective behavior intervention is the use of reinforcers to encourage positive behavior and social interactions. A reinforcer is defined as a stimulus added or removed, contingent on behavior, and as a result, increases the future frequency of that behavior (Catania, 2013). However, due to impairments associated with communication, individuals with ASD may not effectively express preference for stimuli, which may, in turn, function as a reinforcer (Hanley et al., 1999). While a typically developing 3-5 year old child may express their preference for one toy over another by simply saying the name of the toy they want, many children with ASD, especially during early intervention, may reach for items they want as a means to express preference for the item. Thus, many clinicians use preference assessments in order to identify potential reinforcers to use during behavior interventions, when these communication deficits are present.

Preference assessments are one strategy to overcome barriers in communication and identify putative reinforcers. Preference assessments typically involve an observation of a client's interaction with stimuli, data collection of responses, and a comparison of responses (Mangum et al., 2011). Results are then used to identify potentially reinforcing stimuli to

reinforce desired behaviors of the client during behavior interventions. Some preference assessments that have been created to address these unique concerns are paired choice preference assessments, free operant preference assessments, multiple stimulus without replacement preference assessments (MSWO) and multiple stimulus with replacement preference assessments (MSW).

Because of their strong predictive validity in identifying putative reinforcers for people with ASD (see Tullis et al., 2011), preference assessments are commonly used in practice. However, still an open question is to what extent participant selections vary, or remain consistent, within preference assessment administrations. The degree to which responding is stable or consistent during a single preference assessment may have several implications, especially in applied settings (stability in this study simply refers to the pattern of responding). This study will focus on multiple stimulus without replacement (MSWO) preference assessment.

MSWOs are preference assessments in which selections are made without replacing them back in the array (DeLeon & Iwata, 1996). An MSWO contains 5 rounds of assessments, each of which may begin with 8 stimuli. Stable responding across rounds during a single MSWO may indicate that preference is fixed. For example, if a student participated in a full MSWO (which includes 5 rounds of preference assessments of the same 8 items), and selected items in a similar order across all five rounds, results may indicate that student preference is unlikely to vary, giving clinicians a shortcut in determining which reinforcers to use for the student during treatment. However, if another student participated in a full MSWO and chose items in different orders across all five rounds, this pattern of responding may suggest their responding was unstable. Unstable responding may suggest that there is a lack in preference for the items presented during the MSWO or failure to discriminate between stimuli, for example. In the case



of unstable responding, clinicians may conduct more recurring assessments to identify potential reinforcers to use during treatment, introduce different stimuli into the preference assessment array, or improve upon item discrimination skills before administering another MSWO.

Though no researchers have evaluated stability within assessment administrations, some have evaluated stability of responding across preference assessment administrations. For example, Carr et al. (2000) found that 3 participants displayed relatively stable responding across 9 MSWO assessment administrations. Participant responding yielded high agreement across assessment administrations, suggesting stable responding across administrations and across time (following the initial assessment, 8 ongoing assessments took place across 4 weeks). In a study by Hanley et al. (2006), preference stability of leisure activities was evaluated in 10 adults with developmental disabilities. Researchers found stable responding of preference for leisure activities for 8 of 10 participants. However, stability in this study may have been linked to the way in which preference was assessed, as authors noted that there was variability in assessment methods.

Research on preference assessments has also looked at the stability of responding with a large group (Kelley et al., 2016). In this study, researchers evaluated the stability of responding for 21 participants across several paired-choice preference assessments, but focused on the last 10 preference assessments administered for each participant. The study found that 16 of the 21 participants demonstrated stable responding across assessments.

Though previous research has evaluated the extent to which responding may be stable across preference assessments, previous research has not yet evaluated the extent to which responding may be stable within an assessment. Given the potential importance of understanding the extent of stability of participant responding across rounds of assessment, the current study

looks to extend this research by evaluating the extent to which responding is stable within a single MSWO. This study evaluates stability within a single MSWO by comparing responding across rounds within the assessment. The purpose of this study is to extend the research on preference assessments by conducting a secondary data analysis of an original preference assessment study on preference displacement, by determining the extent to which responding is stable for children aged 3-5 during an MSWO preference assessment (Sipila-Thomas et al., 2021). Thus the research questions for this study are what are the effects of presentations of assessment rounds on preference stability of leisure items on subsequent rounds of assessments; and what are the effects of presentations of assessment rounds on preference stability of edible items on subsequent rounds of assessments?

## **METHOD**

### **Secondary Data Analysis**

This study involved a secondary data analysis of an original study on preference assessments. The original study involved video recordings of 25 participants that completed three MSWOs: one all tangible MSWO, one all edible MSWO, and one combined edible and tangible MSWO (Sipila-Thomas et al., 2021) and MSWO rounds all began with the presentation of 8 stimuli. This current study focused on videos that included edible only and tangible only MSWOs administered in Sipila-Thomas et al.

### **Participants and Setting**

This study involved 27 participants (see Figure 1) each of which participated in two MSWOs each (one tangible only and one edible only). Participants included both male and female children ranging in ages 2-5 years old. All participants had a diagnosis of ASD. All participants were also enrolled in early intensive behavioral intervention (EIBI) services in which they had received 6-18 months of services prior to the start of the study (see Plavnick et al., 2021, for a description). Participants also had previous experience completing preference assessments at their EIBI center (e.g., MSWOs and paired-stimulus preference assessments). Sessions for the study were conducted in a separate room outside of the child's typical treatment room.

Inclusion criterion for this study involved requiring all participants to be able to visually scan an array of eight stimuli and point at or touch the stimuli as a selection response. This study originally involved 27 participants with 54 different MSWOs (one all edible MSWO and one all tangible MSWO for each participant). However, 7 participants were excluded from the study due to non-selection of stimuli, lost access to their session videos, problem behavior that interrupted

sessions, or not enough stimuli were selected during the session to accurately evaluate preference stability. Of the 7 participants eliminated, neither their edible or tangible graphs met inclusion criterion.

### **MSWO Preference Assessments**

This study evaluated selection responses during two types of MSWOs (all edible and all tangible MSWOs) following similar procedures described by DeLeon and Iwata (1996). Sessions were conducted in a quiet room in which a camera, a table and chairs, stimuli, and paperwork for data collection were present. The researcher pre-exposed each participant to the stimuli that would be used during the assessment, by allowing the participant to engage in tangible items for 30 seconds each, and engage in edible items until the item was consumed. In the event the participant pushed the item away or did not consume the item, the researcher consulted the child's clinical director in order to identify other items to use during assessments. The participant was pre-exposed to the new stimulus to assess engagement with the item as well.

Following pre-exposure, researchers conducted the MSWO. The researcher began by sitting across from the participant at a table and placed the eight stimuli on the table in a horizontal row equidistant from each other. The researcher then told the participant to "pick one" or "choose one", in which the participant would respond by grabbing the item to consume or engage with while the researcher removed all other items from the table. During edible MSWOs, the participant was allowed to consume the food until it was gone. During tangible MSWOs, the researcher allotted 30s of play per selected item. After a selection of a stimulus and engagement with the item, the researcher rotated the array of remaining items, without replacing the item that was selected. In the event that the participant stopped responding or refused to select items 30 seconds after the presentation of items, the session was terminated and the remaining items were

scored as “not selected”. After all items were selected, the researcher repeated this process four more times (total of five rounds of assessment for each MSWO).

### **Dependent Measures**

Researchers in this study evaluated the extent to which responding was stable across rounds and compared rounds within a single MSWO. Thus, the primary dependent variable was *selection*, which was used to then derive stimulus rankings and then calculate measures of patterns of selection of edible items (for the all edible MSWO) and patterns of selection of tangible items (for the all tangible MSWO).

**Selection.** A selection occurred when a participant used their hand or finger to touch an item after the experimenter said, “pick one” or “choose one” (Sipila-Thomas et al., 2021).

**Ranking.** Researchers ranked items in the order in which they were selected for each round within the MSWO. Each round within the MSWO started with 8 stimuli. Researchers first randomly assigned stimuli to code numbers (1-8). After assigning each stimulus to a code number, the researcher ranked items in the order in which they were selected. For example, code numbers were assigned to stimuli as follows, 1) potato chip, 2) m&m, 3) fruit snack, 4) veggie straw, 5) cheese puff, 6) oreo, 7) goldfish 8) sour patch. The researcher then ranked items in the order in which the participant selected them (for example, a ranking of 5, 8, 7, 2, 4, 1, 3, 6, would indicate that cheese puffs were chosen first and oreos were chosen last). Researchers ranked items based on the order the participant selected the items; thus, the first item (along with its corresponding code number) selected was ranked first and so on until all items were ranked.

**Stability.** Stability in this study refers to patterns in responding. Correlation coefficients (or *r*-values) were calculated based on the rank in which items were selected. This study used a stability criterion of 0.6 based on previous research on preference stability (Hanley et al., 2006).

*R*-values at or exceeding 0.6 indicated relatively stable responding across rounds within an MSWO. *R*-values lower than 0.6 indicated relatively unstable responding across rounds within an MSWO. *R*-values were calculated by using the Pearson's correlation coefficient formula.

### **Interobserver Agreement**

Interobserver Agreement (IOA) was calculated on a trial by trial bases for 40% of rounds across all participants. IOA was calculated by dividing the total number of agreements with the total number of trials to yield a percentage (Cooper et al., 2020). Overall IOA was 99.82%; IOA was 100% across all participants for all administrations, except for Participant 16's edible tangible assessment, which was 94%.

### **Procedure**

Researchers first randomly assigned numbers to each stimulus used during that session (1-8). Researchers then recorded the order in which items were selected using the assigned numbers given to each stimulus. These steps were completed for the entire 5 rounds within the assessment. After the numbers were ranked in the order in which they were selected, the researcher then compared rounds of assessments. *r*-values were calculated by comparing rank orders of rounds 1 to 2, rounds 2 to 3, rounds 3 to 4, and rounds 4 to 5. These comparisons generated 4 different *r*-values to determine the extent to which responding was stable.

Participant MSWOs were individually evaluated by categorizing *r*-values. Of the four *r*-values generated for each MSWO, if 0-1 round *R*-Values met or exceeded the inclusion criterion of 0.6, it was categorized as demonstrating relatively low stability in responding across rounds. If 2 of the 4 *r*-values met or exceeded the inclusion criterion of 0.6, it was categorized as demonstrating relatively moderate stability in responding across rounds. If 3-4 of the 4 *r*-values met or exceeded the inclusion criterion of 0.6, it was categorized as demonstrating relatively high

stability in responding across rounds. Thus, for example, if the 4  $r$ -values generated from an edible MSWO were 0.44, 0.39, 0.62, and 0.73, the participant demonstrated relatively moderate stability in responding across rounds (2  $r$ -values met or exceeded the 0.6 stability criterion).

## RESULTS

Below we describe the results from participants. See Table 1.

### Participant 1

Participant 1 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.95. The  $r$ -value generated by comparing rounds 2 to 3 was 0.88. The  $r$ -value generated by comparing rounds 3 to 4 was 0.76. The  $r$ -value generated by comparing rounds 4 to 5 was 1. Thus, 4 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively stable responding across rounds.

Participant 1 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.52. The  $r$ -value generated by comparing rounds 2 to 3 was 0.88. The  $r$ -value generated by comparing rounds 3 to 4 was 0.88. The  $r$ -value generated by comparing rounds 4 to 5 was 0.83. Thus, 3 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating moderately stable responding across rounds.

### Participant 2

Participant 2 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.38. The  $r$ -value generated by comparing rounds 2 to 3 was -0.4. The  $r$ -value generated by comparing rounds 3 to 4 was 0.31. The  $r$ -value generated by comparing rounds 4 to 5 was -0.14. Thus, 0 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating low stability in responding across rounds.

Participant 2 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.79. The  $r$ -value generated by comparing rounds 2 to 3 was 0.81. The  $r$ -value generated by comparing rounds 3 to 4 was 0.88. The  $r$ -value generated by comparing



rounds 4 to 5 was 0.83. Thus, 3 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating moderately stable responding across rounds.

### **Participant 3**

Participant 3 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.24. The  $r$ -value generated by comparing rounds 2 to 3 was 0.24. The  $r$ -value generated by comparing rounds 3 to 4 was -0.43. The  $r$ -value generated by comparing rounds 4 to 5 was 0.26. Thus, 0 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively low stability in responding across rounds.

Participant 3 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.07. The  $r$ -value generated by comparing rounds 2 to 3 was 0.95. The  $r$ -value generated by comparing rounds 3 to 4 was 0.19. The  $r$ -value generated by comparing rounds 4 to 5 was 0.26. Thus, 1 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively low stability in responding across rounds.

### **Participant 4**

Participant 4 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.31. The  $r$ -value generated by comparing rounds 2 to 3 was -0.29. The  $r$ -value generated by comparing rounds 3 to 4 was 0.05. The  $r$ -value generated by comparing rounds 4 to 5 was 0.02. Thus, 0 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively low stability in responding across rounds.

### **Participant 6**

Participant 6 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.69. The  $r$ -value generated by comparing rounds 2 to 3 was -0.33. The  $r$ -value generated by comparing rounds 3 to 4 was 0.29. The  $r$ -value generated by comparing rounds 4 to

5 was 0.69. Thus, 2 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively moderate stability in responding across rounds.

Participant 6 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.05. The  $r$ -value generated by comparing rounds 2 to 3 was 0. The  $r$ -value generated by comparing rounds 3 to 4 was 0.07. The  $r$ -value generated by comparing rounds 4 to 5 was 0.40. Thus, 0 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively low stability in responding across rounds.

### **Participant 8**

Participant 8 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.81. The  $r$ -value generated by comparing rounds 2 to 3 was -0.33. The  $r$ -value generated by comparing rounds 3 to 4 was 0.36. The  $r$ -value generated by comparing rounds 4 to 5 was 0.07. Thus, 1 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively low stability in responding across rounds.

### **Participant 9**

Participant 9 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.88. The  $r$ -value generated by comparing rounds 2 to 3 was 0.77. The  $r$ -value generated by comparing rounds 3 to 4 was 0.77. The  $r$ -value generated by comparing rounds 4 to 5 was 0.95. Thus, 4 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

Participant 9 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.19. The  $r$ -value generated by comparing rounds 2 to 3 was 0.43. The  $r$ -value generated by comparing rounds 3 to 4 was 0.76. The  $r$ -value generated by comparing rounds 4 to

5 was 0.95. Thus, 2 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively moderate stability in responding across rounds.

### **Participant 10**

Participant 10 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.38. The  $r$ -value generated by comparing rounds 2 to 3 was 0.55. The  $r$ -value generated by comparing rounds 3 to 4 was 0.26. The  $r$ -value generated by comparing rounds 4 to 5 was 0.24. Thus, 0 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively low stability in responding across rounds.

Participant 10 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.81. The  $r$ -value generated by comparing rounds 2 to 3 was 0.31. The  $r$ -value generated by comparing rounds 3 to 4 was 0.86. The  $r$ -value generated by comparing rounds 4 to 5 was 0.71. Thus, 3 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

### **Participant 11**

Participant 11 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.71. The  $r$ -value generated by comparing rounds 2 to 3 was 0.83. The  $r$ -value generated by comparing rounds 3 to 4 was 0.83. The  $r$ -value generated by comparing rounds 4 to 5 was 0.48. Thus, 3 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

Participant 11 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.95. The  $r$ -value generated by comparing rounds 2 to 3 was 0.40. The  $r$ -value generated by comparing rounds 3 to 4 was 0.83. The  $r$ -value generated by comparing rounds 4 to 5 was 0.48. Thus, 3 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

5 was 0.55. Thus, 2 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively moderate stability in responding across rounds.

### **Participant 12**

Participant 12 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was -0.21. The  $r$ -value generated by comparing rounds 2 to 3 was 0.24. The  $r$ -value generated by comparing rounds 3 to 4 was 0.65. The  $r$ -value generated by comparing rounds 4 to 5 was 0.67. Thus, 2 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively moderate stability in responding across rounds.

Participant 12 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was -0.02. The  $r$ -value generated by comparing rounds 2 to 3 was -0.36. The  $r$ -value generated by comparing rounds 3 to 4 was 0.62. The  $r$ -value generated by comparing rounds 4 to 5 was 0.69. Thus, 2 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively moderate stability in responding across rounds.

### **Participant 13**

Participant 13 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.64. The  $r$ -value generated by comparing rounds 2 to 3 was 0.43. The  $r$ -value generated by comparing rounds 3 to 4 was 0.72. The  $r$ -value generated by comparing rounds 4 to 5 was 0.54. Thus, 2 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively moderate stability in responding across rounds.

### **Participant 14**

Participant 14 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.76. The  $r$ -value generated by comparing rounds 2 to 3 was 0.33. The  $r$ -value generated by comparing rounds 3 to 4 was 0.76. The  $r$ -value generated by comparing rounds 4 to

5 was 0.60. Thus, 3 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

Participant 14 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.88. The  $r$ -value generated by comparing rounds 2 to 3 was 0.90. The  $r$ -value generated by comparing rounds 3 to 4 was 0.81. The  $r$ -value generated by comparing rounds 4 to 5 was 0.88. Thus, 4 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

### **Participant 15**

Participant 15 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.90. The  $r$ -value generated by comparing rounds 2 to 3 was 0.88. The  $r$ -value generated by comparing rounds 3 to 4 was 0.76. The  $r$ -value generated by comparing rounds 4 to 5 was 0.81. Thus, 4 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

Participant 15 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.83. The  $r$ -value generated by comparing rounds 2 to 3 was 0.64. The  $r$ -value generated by comparing rounds 3 to 4 was 0.83. The  $r$ -value generated by comparing rounds 4 to 5 was 0.60. Thus, 4 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

### **Participant 16**

Participant 16 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.66. The  $r$ -value generated by comparing rounds 2 to 3 was 0.37. The  $r$ -value generated by comparing rounds 3 to 4 was 0.80. The  $r$ -value generated by comparing rounds 4 to

5 was 1.00. Thus, 3 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

Participant 16 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.52. The  $r$ -value generated by comparing rounds 2 to 3 was 0.57. The  $r$ -value generated by comparing rounds 3 to 4 was 0.29. The  $r$ -value generated by comparing rounds 4 to 5 was 0.67. Thus, 1 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively low stability in responding across rounds.

### **Participant 17**

Participant 17 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.69. The  $r$ -value generated by comparing rounds 2 to 3 was 0.76. The  $r$ -value generated by comparing rounds 3 to 4 was 0.76. The  $r$ -value generated by comparing rounds 4 to 5 was 0.86. Thus, 4 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

Participant 17 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.62. The  $r$ -value generated by comparing rounds 2 to 3 was 0.76. The  $r$ -value generated by comparing rounds 3 to 4 was 0.90. The  $r$ -value generated by comparing rounds 4 to 5 was 0.88. Thus, 4 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

### **Participant 18**

Participant 18 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.79. The  $r$ -value generated by comparing rounds 2 to 3 was -0.05. The  $r$ -value generated by comparing rounds 3 to 4 was -0.26. The  $r$ -value generated by comparing rounds 4

to 5 was 0.48. Thus, 1 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively low stability in responding across rounds.

Participant 18 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.36. The  $r$ -value generated by comparing rounds 2 to 3 was 0.69. The  $r$ -value generated by comparing rounds 3 to 4 was 0.62. The  $r$ -value generated by comparing rounds 4 to 5 was 0.42. Thus, 2 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively moderate stability in responding across rounds.

### **Participant 19**

Participant 19 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.52. The  $r$ -value generated by comparing rounds 2 to 3 was 0.86. The  $r$ -value generated by comparing rounds 3 to 4 was 0.36. The  $r$ -value generated by comparing rounds 4 to 5 was 0.60. Thus, 2 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively moderate stability in responding across rounds.

Participant 19 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.57. The  $r$ -value generated by comparing rounds 2 to 3 was 0.71. The  $r$ -value generated by comparing rounds 3 to 4 was 0.86. The  $r$ -value generated by comparing rounds 4 to 5 was 1.00. Thus, 3 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

### **Participant 20**

Participant 20 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.73. The  $r$ -value generated by comparing rounds 2 to 3 was 0.12. The  $r$ -value generated by comparing rounds 3 to 4 was 0.76. The  $r$ -value generated by comparing rounds 4 to

5 was 0.55. Thus, 2 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively moderate stability in responding across rounds.

### **Participant 26**

Participant 26 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was -0.02. The  $r$ -value generated by comparing rounds 2 to 3 was 0.45. The  $r$ -value generated by comparing rounds 3 to 4 was -0.17. The  $r$ -value generated by comparing rounds 4 to 5 was 0.12. Thus, 0 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively low stability in responding across rounds.

Participant 26 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.74. The  $r$ -value generated by comparing rounds 2 to 3 was 0.31. The  $r$ -value generated by comparing rounds 3 to 4 was 0.69. The  $r$ -value generated by comparing rounds 4 to 5 was 0.95. Thus, 3 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively high stability in responding across rounds.

### **Participant 27**

Participant 27 completed an all edible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.31. The  $r$ -value generated by comparing rounds 2 to 3 was 0.12. The  $r$ -value generated by comparing rounds 3 to 4 was 0.57. The  $r$ -value generated by comparing rounds 4 to 5 was 0.45. Thus, 0 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively low stability in responding across rounds.

Participant 27 completed an all tangible MSWO. The  $r$ -value generated by comparing rounds 1 to 2 was 0.55. The  $r$ -value generated by comparing rounds 2 to 3 was 0.43. The  $r$ -value generated by comparing rounds 3 to 4 was 0.69. The  $r$ -value generated by comparing rounds 4 to



5 was -0.17. Thus, 1 of 4 round comparisons met or exceeded the 0.6 criterion, demonstrating relatively low stability in responding across rounds.

### **General Analysis of Stability**

Seventeen edible MSWOs met the criteria of inclusion for this study. Of the 17, 7 MSWOs (41%) demonstrated relatively high stability in responding across rounds, 3 MSWOs (18%) demonstrated relatively moderate stability in responding across rounds, and 7 MSWOs (41%) demonstrated relatively low stability in responding across rounds.

Eighteen tangible MSWOs met the criteria of inclusion for this study. Of the 18, 7 MSWOs (39%) demonstrated relatively high stability in responding across rounds, 6 MSWOs (33%) demonstrated relatively moderate stability in responding across rounds, and 5 MSWOs (28%) demonstrated relatively low stability in responding across rounds.

Thirty-five MSWOs in total (both edible and tangible MSWOs) met the criteria of inclusion for this study. Of the 35, 14 MSWOs (40%) demonstrated relatively high stability in responding across rounds, 9 MSWOs (26%) demonstrated relatively moderate stability in responding across rounds, and 12 MSWOs (34%) demonstrated relatively low stability in responding across rounds.

## DISCUSSION

The purpose of the current study was to evaluate the extent to which responding was stable across rounds of administration within a single MSWO for both edible and tangible preference assessments. The stability of responding, in general, varied across participants. Overall, for 40% of the MSWOs administered across participants, participant demonstrated relatively high stability across rounds, 26% demonstrated relatively moderate stability in responding across rounds, and 34% demonstrated relatively low stability in responding across rounds, regardless of the MSWO type.

Of the 17 participants who met inclusion criteria for the edible MSWO preference assessments, 7 participants (41%) demonstrated patterns of relatively high stability in responding across rounds, 3 participants (18%) demonstrated patterns of relatively moderate stability in responding across rounds, and 7 participants (41%) demonstrated patterns of relatively low stability in responding across rounds. Thus, the degree to which participant responding was stable during MSWOs with edible-only stimuli tended to vary across participants. Our findings suggest, then, that participant responding is likely to be stable across rounds, or have low stability across rounds, and that moderate levels of stability are not likely to be found, at least in our participants completing the edible preference assessment.

Of the 18 participants who met inclusion criteria for the tangible MSWOs preference assessments, 7 participants (39%) demonstrated relatively high stability in responding across rounds, 6 participants (33%) demonstrated relatively moderate stability in responding across rounds, and 5 participants (28%) demonstrated relatively low stability in responding across rounds. Given the results, the degree to which responding was stable across participants during tangible-only MSWOs varied. In fact, there was no notable difference between categorical

groupings (high, moderate, low stability). This may suggest that other variables such as environmental factors, motivating operation, and selected stimuli to use during assessments may influence stability in responding rather than the stimulus type alone.

For this study, of all the participants, 15 participants' edible and tangible MSWOs met the inclusion criteria. Of those 15 participants, seven participants' responding (47%) during the edible and tangible MSWOs stayed within the same categorical group. For example, participant 1's responding during the edible and tangible MSWO were categorized as demonstrating high levels of stability. For both their edible and tangible MSWO, 3 of 4 of the round comparisons met or exceeded the 0.6 criterion. The implications of these results (participants demonstrating similar levels of stability across their edible and tangible MSWO) may have an applicable use during treatment and future preference assessments. Participants with MSWOs that stayed within the same categorical group may suggest that the participant may have the tendency to have more consistent or variable responding, based on their own history associated. For example, a student with a strong reinforcement history with mesh balls may be more likely to choose the mesh ball prior to less familiar stimuli. Contrarily, a student that has had little to no exposure to mesh balls may instead select a more familiar item prior to selecting the mesh ball. With nearly half of the participant's (with both MSWOs meeting the inclusion criterion) MSWOs being categorically identical, participants themselves may influence the likelihood of preference stability, regardless of the MSWO type.

If participant responding during both MSWOs are within the same category in terms of relative levels of stability (for example, both their edible and tangible MSWO demonstrated low stability across rounds), this may indicate to clinicians that stability in responding for this participant tends to vary, regardless of the MSWO type. As a result, clinicians may consider

changing stimuli or run more recurring preference assessments in order to determine potential reinforcers to use during treatment.

High stability across rounds may suggest that participant responding is more predictable. This may have an applicable utility in the clinical settings for people administering MSWOs routinely. Determining that a participant's categorical grouping for their MSWOs were high may suggest to clinicians that their preference for the items presented in the MSWOs are stable. Thus, clinicians may not run full MSWOs, but a brief MSWO as a result (Carr et al., 2000). Brief MSWOs consist of three rounds of assessment instead of five. Brief MSWOs would shorten time allotted for treatment and save resources for clinics as fewer stimuli would be used for assessments.

Low stability across rounds may suggest that participant responding is variable. Implications that can be drawn from participant MSWOs categorized as having low stability, involve the significance of frequent administrations of preference assessments. Variable responding may be influenced by the type of stimuli used during assessments, immediate changes in preference during assessments, satiation, or low exposure to stimuli (little history associated with particular stimuli; McSweeney, 2004). As a result, variable responding (or low levels of stability), may indicate to clinicians to change stimuli used during assessments, administer more frequent preference assessments, and consider environmental factors that may influence responding. Environmental factors that could have an influence on stability of responding is the time of day in which the assessment was conducted, the motivating operation at the time of the assessment, and potential stimulus control associated with stimuli used during the assessment. Clinicians may administer these brief MSWOs more frequently (before each

program begins for the treatment day) to determine which stimulus to use immediately as reinforcement.

Previous research took individual  $r$ -values across time and have averaged the numbers to generate a single number to compare against the study's criterion (typically 0.58-0.6). This study instead, looked at raw numbers without averaging  $r$ -values, which is a much more accurate representative of participant responding. If this study averaged out  $r$ -values instead of evaluating each  $r$ -value comparison, readers would have lost out on information relevant to the stability of responding. This method of analyzing  $r$ -values offers a much more clear and accurate representation of stability of responding.  $r$ -values that are averaged has the potential to sway the averaged number especially when they're an outlier in the group of  $r$ -values (an  $r$ -value significantly higher or lower than the rest of the  $r$ -values).

Looking at individual  $r$ -values as opposed to averaged  $r$ -values gives a complete story of stability across rounds. For example, participant 27's tangible MSWO  $r$ -values were 0.55, 0.43, 0.69, and -0.17. If these  $r$ -values were averaged, the number generated would be 0.38, suggesting relatively moderate stability across rounds. However, this averaged number does not convey elements of the round comparisons that may be useful when interpreting results such as the inverse relation between the round 4 to 5 round comparison (-0.17). Reviewing each round comparison separately gives a more complete account of the data sets.

### **Potential Limitations**

This study presents several limitations. First, during preference assessments, at times the participant reached toward/grabbed an item while the assessor placed stimuli on the table or immediately after stimuli were placed on the table. Since a selection was made prior to the assessor saying the discriminative stimulus (SD) "Pick one", the assessor blocked the response or

took the item away then presented the SD. As a result, some participants then chose a different item instead. The response blocking/removal of item (not allowing the participant to select an item prior to saying "Pick one" or taking an item a participant grabs before presenting the SD) may have punished the response, which in turn resulted in a different selection in some cases and may have affected patterns of stability; which is a limitation of the original study.

Lack of participant demographics is also a limitation of the original study. The clinic the participants all attended did not keep record of this information, thus it was not included in the methods section of this study. Including demographics in this study would have provided more information on individual characteristics and implications may have been drawn based on similarities between demographics and patterns in responding (Brodhead et al., 2014).

Another potential limitation of the original study involves selection of stimuli used during MSWOs. Assessors interviewed participants' BCBA's in order to identify stimuli to use during MSWOs. The same set of stimuli was used for all participants, apart from one participant that "only ate white food". As an accommodation one food item was substituted for a snack he typically eats. Using the same stimuli, especially in relation to food, may have limited researchers' ability to actually assess preference rather than inferring preference based on the stability of responding. For the purpose of this study, stimuli were consistent across rounds to promote consistent procedural methods across participants, however the lack of variation does impact the overall external validity of this study. Variation of stimuli used during MSWOs likely would have also aided in interventions as personalized stimuli used as reinforcers may have more of a reinforcing effect.

Lastly, a limitation in this study was the lack of a social validity measure. Researchers in this study did not interview clients or their parents in order to determine which stimuli to use

during MSWOs. Participants in this study come from an array of different cultural backgrounds, which influence preference of food and play items. Interviewing clients and their parents would have likely given researchers a better inclination of preference for edible and tangible items. Stability in responding does not guarantee preference for a particular stimulus but instead suggests relative preference of a stimulus over another, through its ranking system. Including a social validity measure to interview stakeholders in the participants' lives would have given researchers the opportunity to understand how they viewed procedures and their overall thoughts on the significance of the study.

This study contributes to the literature in several ways. Previous research on stability of responding during preference assessments has evaluated stability across time and across preference assessments (Hanley et al., 2006). This study is the first preference assessment study to evaluate stability across rounds within the same MSWO. Results of the study add to the literature as it indicates that half of the participants with both edible and tangible MSWOs meeting inclusion criterion, had both MSWOs fall within the same categorical group based on relative stability (Ex: Participant 3's edible and tangible MSWOs both demonstrated low levels of stability). The findings of this study differ from other studies on stability of responding during preference assessments. Contrary to previous research, the results of this study indicated variable patterns of responding across participants, MSWO type, and rounds of assessment. Previous research has reported more stable responding across participants and across assessments (Kelley et al., 2016).

The results of this study ensue potential practical considerations for the clinical setting. For example, procedural errors such as inconsistent stimulus exposure during assessments may have had an influence on responding. Inconsistencies involving clinician interactions with

stimuli used during assessments to expose items to participants may have swayed participant responding inadvertently. Overall history associated with stimuli used during assessments may contribute to participant responding. Failure to properly ensure that participants knew how to use all stimuli for tangible MSWOs may also have contributed to participant responding. Overall, this study provides a way to interpret the relative stability of participant responding across rounds of a single assessment. Several practical implications must be considered when administering these assessments in order to get the most accurate representation of stability of participant responding during preference assessment.



## **APPENDIX**

**Table 1**

Participant R-Values						
Participant	MSWO Type	Round 1 to Round 2	Round 2 to Round 3	Round 3 to Round 4	Round 4 to Round 5	R-Values meeting or exceeding 0.6 criterion
P1	Edible	0.95	0.88	0.76	1.00	4
	Tangible	0.52	0.88	0.88	0.83	3
P2	Edible	0.38	-0.40	0.31	-0.14	0
	Tangible	0.79	0.81			2
P3	Edible	0.24	0.24	-0.43	0.26	0
	Tangible	0.07	0.95	0.19	0.26	1
P4	Edible	0.84				1
	Tangible	0.31	-0.29	0.05	0.02	0
P6	Edible	0.69	-0.33	0.29	0.69	2
	Tangible	0.05	0.00	0.07	0.40	0
P7	Edible	Only Chose One Item				0
	Tangible					0
P8	Edible	0.81	-0.33	0.36	0.07	1
	Tangible	0.64				0
P9	Edible	0.88	0.77	0.77	0.95	4 (N/S for some responses)
	Tangible	0.19	0.43	0.76	0.95	2
P10	Edible	0.38	0.55	0.26	0.24	0
	Tangible	0.81	0.31	0.86	0.71	3
P11	Edible	0.71	0.83	0.83	0.48	3
	Tangible	0.95	0.40	0.83	0.55	2
P12	Edible	-0.21	0.24	0.65	0.67	2
	Tangible	-0.02	-0.36	0.62	0.69	2
P13	Edible	Only Chose One Item				
	Tangible	0.64	0.43	0.72	0.54	2 (N/S for some responses)
P14	Edible	0.76	0.33	0.76	0.60	3
	Tangible	0.88	0.90	0.81	0.88	4
P15	Edible	0.90	0.88	0.76	0.81	4
	Tangible	0.83	0.64	0.83	0.60	4
P16	Edible	0.66	0.37	0.80	1.00	3 (N/S for some responses)
	Tangible	0.52	0.57	0.29	0.67	1
P17	Edible	0.69	0.76	0.76	0.86	4
	Tangible	0.62	0.76	0.90	0.88	4
P18	Edible	0.79	-0.05	-0.26	0.48	1
	Tangible	0.36	0.69	0.62	0.42	2
P19	Edible	0.52	0.86	0.36	0.60	2
	Tangible	0.57	0.71	0.86	1.00	3 (N/S for some responses)
P20	Edible					
	Tangible	0.73	0.12	0.76	0.55	2
P26 (6A)	Edible	-0.02	0.45	-0.17	0.12	0
	Tangible	0.74	0.31	0.69	0.95	3
P27 (7A)	Edible	0.31	0.12	0.57	0.45	0
	Tangible	0.55	0.43	0.69	-0.17	1

Table 1: r-values calculated for participants in this study

**Figure 1**

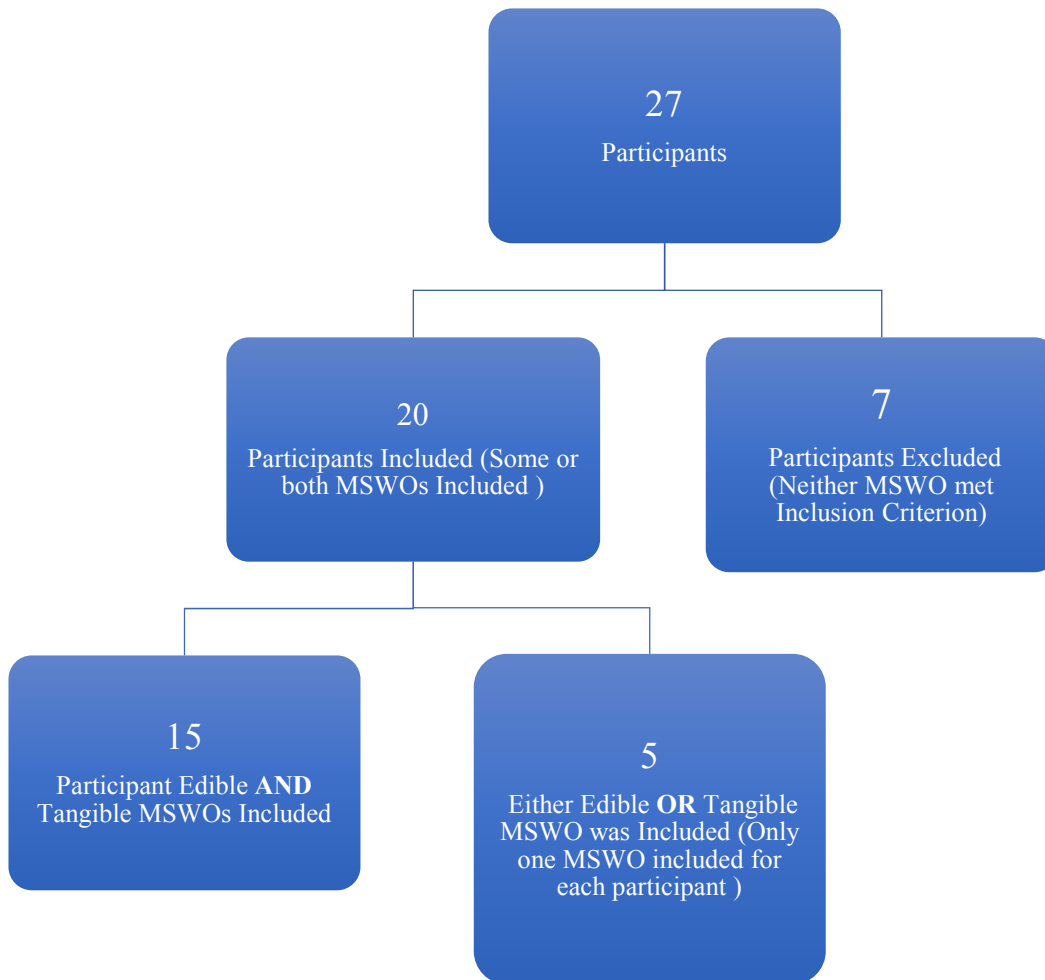


Figure 1: A flowchart depicting participant recruitment and inclusion this study

## REFERENCES

## REFERENCES

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5<sup>th</sup> ed.). American Psychiatric Publishing.
- Carr, J. E., Nicolson, A. C., & Higbee, T. S. (2000). Evaluation of a brief multiple-stimulus preference assessment in a naturalistic context. *Journal of Applied Behavior Analysis, 33*, 353–357. <https://doi.org/10.1901/jaba.2000.33-353>
- Catania, A. C. (2013). *Learning*. Cornwall-on-Hudson, NY: Sloan Pub.
- Copper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis (3<sup>rd</sup> ed.)*. Pearson.
- Brodhead, M. T., Durán, L., & Bloom, S. E. (2014). Cultural and linguistic diversity in recent verbal behavior research on individuals with disabilities: a review and implications for research and practice. *The Analysis of Verbal Behavior, 30*, 75–86. <https://doi.org/10.1007/s40616-014-0009-8>
- DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*, 519–533. <https://doi.org/10.1901/jaba.1996.29-519>
- Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Clinical Child and Adolescent Psychology, 38*, 439–450.
- Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis, 25*, 491–498. <https://doi.org/10.1901/jaba.1992.25-491>
- Hanley, G. P., Iwata, B. A., & Lindberg, J. S. (1999). Analysis of activity preferences as a function of differential consequences. *Journal of Applied Behavior Analysis, 32*, 419–435.
- Hanley, G. P., Piazza, C. C., Fisher, W. W., Contrucci, S. A., & Maglieri, K. A. (1997). Evaluation of client preference for function-based treatment packages. *Journal of Applied Behavior Analysis, 30*, 459–473.
- Hanley, G. P., Piazza, C. C., Fisher, W. W., & Maglieri, K. A. (2005). On the effectiveness of and preference for punishment and extinction components of function-based interventions. *Journal of applied behavior analysis, 38*, 51–65. <https://doi.org/10.1901/jaba.2005.6-04>

- Kelley, M. E. (2016). *Stability of daily preference across multiple individuals* Society for the Experimental Analysis of Behavior. doi:10.1002/jaba.288
- Hanley, G. P., Iwata, B. A., & Roscoe, E. M. (2006). Some determinants of changes in preference over time. *Journal of Applied Behavior Analysis, 39*, 189–202. doi: 10.1901/jaba.2006.163-04
- Mangum, A., Roane, H., Fredrick, L., & Pabico, R. (2012). The Role of Context in the Evaluation of Reinforcer Efficacy: Implications for the Preference Assessment Outcomes. *Research in autism spectrum disorders, 6*, 158–167.  
<https://doi.org/10.1016/j.rasd.2011.04.001>
- McSweeney F. K. (2004). Dynamic changes in reinforcer effectiveness: satiation and habituation have different implications for theory and practice. *The Behavior Analyst, 27*, 171–188.  
<https://doi.org/10.1007/BF03393178>
- Plavnick, J. B., Bak, M. Y. S., Avendaño, S. M., Dueñas, A. D., Brodhead, M. T., & Sipila-Thomas, E. S. (2020). Implementing early intensive behavioral intervention in community settings. *Autism: International Journal of Research and Practice, 24*, 1913-1916.
- Sipila-Thomas E. S., Foote A. J., White A. N., Melanson I. J., & Brodhead M. T. (2021). A replication of preference displacement research in children with autism spectrum disorder. *Journal of Applied Behavior Analysis, 54*, 403-416. doi: 10.1002/jaba.775
- Taylor, R. (1990). Interpretation of the Correlation Coefficient: A Basic Review. *Journal of Diagnostic Medical Sonography, 6*, 35–39.  
<https://doi.org/10.1177/875647939000600106>
- Tullis, C. A., Cannella-Malone, H. I., Basbigill, A. R., Yeager, A., Fleming, C. V., Payne, D., & Wu, P. (2011). Review of the choice and preference assessment literature for individuals with severe to profound disabilities. *Education and Training in Autism and Developmental Disabilities, 46*, 576- 595