AN EVALUATION OF PREFERENCE STABILITY WITHIN MSWO PREFERENCE ASSESSMENTS IN CHILDREN WITH AUTISM

By

ISAAC JOSEPH MELANSON

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

AN EVALUATION OF PREFERENCE STABILITY WIHTIN MSWO PREFERENCE ASSESSMENTS IN CHILDREN WITH AUTISM

By

ISAAC JOSEPH MELANSON

The purpose of this paper was to analyze the effects of presentations of assessment rounds on preference stability during subsequent rounds of a multiple stimulus without replacement (MSWO) preference assessment in preschool aged children with autism. We conducted a secondary data analysis based on videos recorded during the data collection phase of Sipila-Thomas et al. (2021) and calculated preference stability across consecutive rounds using Spearman rank-order correlation coefficients (Spearman's ρ) for 13 participants with autism. We defined preference stability as the Spearman's ρ critical r value meeting or exceeding .6 for consecutive round comparisons. Additionally, we present a new definition for patterns of stability and variability across rounds of a MSWO preference assessment. We observed patterns of preference stability for 10 participants and patterns of preference variability for 3 participants. The implications of these results are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Reinforcement is the cornerstone of effective behavioral interventions (Cooper et al., 2019). Identifying effective reinforcers, therefore, is an important component to skill-acquisition and behavior-management programs (Verriden & Roscoe, 2016). As a result, there is robust support for the use of preference assessments as a means to identify potential reinforcers to use in behavioral programs (Kang et al., 2013; Cannella et al., 2005).

Behavior analysts can infer preference for stimuli through observing selection of those stimuli. If one stimulus is consistently selected sooner than another stimulus, that stimulus is said to be more preferred and is often implemented as a putative reinforcer. Although previous research has found preference assessments to be a useful way of selecting putative reinforcers, little is known about how preference for stimuli maintains over time (Butler & Graff, 2021). If selection remains constant over time, individuals may require less frequent preference assessments which can save time and resources. Reducing the number of preference assessments administered may also reduce the potential abolishing effects of repeated exposure to stimuli (Morris & Vollmer, 2020).

If selection is variable across stimuli, individuals may require more frequent preference assessments to identify potential reinforcers, requiring more time and effort. Variable responding may result from little discrimination between stimuli. In this situation, practitioners may consider modifying the array of the preference assessment with stimuli from different stimulus classes (edible items, leisure items, social attention) to increase the likelihood that a preferred stimulus functions as a reinforcer (see Karsten et al., 2011).

Two ways to measure preference stability are to evaluate shifts in preference over time with Spearman rank-order correlation coefficients (Spearman's ρ) or Pearson correlation coefficients (Pearson's ρ), which range on a scale from the critical *r* value of -1 (perfect negative correlation) to the critical *r* value of 1 (perfect positive correlation). Previous research on preference stability has defined stability for both the Spearman's ρ and Pearson's ρ meeting or exceeding the critical *r* value of .58 (Hanley et al., 2006; Kelley et al., 2016) or meeting or exceeding the critical *r* value of .6 (Verriden & Roscoe, 2016; Morris & Vollmer, 2020; Butler & Graff, 2021). While researchers have reported that variability in preference is often observed when repeated preference assessments are conducted (Hanley et al., 2006; Zhou et al. 2001), more recent published studies show these patterns of stability are idiosyncratic across participants.

For example, ten adults diagnosed with a developmental delay from Hanley et al. (2006) completed a range of six to 16 paired choice preference assessments over 25 days. Preference assessments for seven of ten (70%) participants produced Pearson's ρ correlation coefficients greater than the .58 criterion. Similarly, 21 children enrolled in an early intensive behavioral intervention (EIBI) program from Kelley et al. (2016) completed paired choice preference assessments on ten consecutive days of treatment. Preference assessments for 16 of the 21 participants (76%) produced mean Pearson's ρ correlation coefficients greater than the .58 criterion.

Results from Butler & Graff (2021) showed similar patterns of stability across short term stability assessments (month-by-month), and less stability across long term assessments (greater than 2 months). Four students aged 16-21 from a residential program serving individuals with Autism Spectrum Disorder (ASD) participated in three separate paired choice preference assessments (edible, leisure, and social attention) over 12 months. The researchers compared the average preference rankings from the first preference assessment to each subsequent preference assessment and calculated a mean *r* value to analyze long term stability while also comparing each preference assessment to the preference assessment from the subsequent month to analyze for short term stability. Results indicated that preference was more stable when evaluated over short periods (75% of comparisons) than over long periods (41% of comparisons).

Still, no study to date has assessed stability within rounds of the same preference assessment (rounds 1-5 in a single MSWO preference assessment). Without a direct analysis of the effects of presentations of assessment rounds on preference stability during subsequent rounds of an MSWO preference assessment, patterns of stability within a single MSWO preference assessment is unknown. This is problematic for a few reasons, the first is that stability comparisons between many preference assessments have often reported mean *r* values (Verriden & Roscoe, 2016; Morris & Vollmer, 2020). Evaluating results of stability by comparing the mean *r* value may reduce its validity as sudden short-term changes, round by round, are eliminated. The second reason is that calculating a mean selection percentage across rounds may also miss potential satiation effects from repeated exposure to stimuli within the preference assessment. For example, if a participant selects a stimulus early at the start of a preference assessment, but repeated exposure to that stimulus leads to progressively later selection over the course of the assessment, outlying scores can artificially inflate or deflate the mean. Direct comparisons of selection, round-by-round, may make it easier to identify which participants may be more sensitive to repeated exposure of stimuli.

Finally, round-by-round comparisons may highlight individual behavioral interventions to pursue. For instance, if responding is variable for a specific participant, it is possible that they have yet to acquire the prerequisite skills necessary (discrimination, scanning an array) to accurately complete the assessment. In this case, practitioners may need to first assess if these skills are in the individual's current repertoire, if not, they should be acquired before continuing with the assessment. If prerequisite skills are currently in the individual's repertoire, new stimuli from a range of stimulus classes may need to be added to get a better understanding of individual preference.

Due to the importance of preference assessments to identify putative reinforcers for behavioral interventions, analyzing the effects of assessment rounds on preference stability is needed to inform individualization of assessments for recipients of behavioral services and to further understand how preference assessment results are implemented as putative reinforcers. Therefore, the purpose of this study was to analyze the effects of presentations of assessment rounds on preference stability during subsequent rounds of an MSWO preference assessment in preschool aged children with autism.

METHOD

Secondary Data Analysis

We conducted a secondary data analysis (Heaton, 2003) based on videos recorded during the data collection phase of Sipila-Thomas et al. (2021).

Participants and Setting

A total of 13 participants (11 males and 2 females) of the original 25 participants from Sipila-Thomas et al. (2021) were included in the present analysis. Twelve participants from the original study were excluded from the current analysis because their videos were no longer available, they did not select a stimulus from the preference assessment array within 30 s, or because they did not make the ample number of selections each round to accurately assess preference stability (see selection definition). The remaining 13 participants were between the ages of 2-5 years old and had a confirmed medical diagnosis of ASD.

Participants received services from a community-based EIBI program for children with ASD (Plavnick et al., 2020) for approximately 6-18 months and had previous exposure to MSWO preference assessments. The Mullen Scales of Early Learning (Mullen, 1995) was administered for all participants upon enrollment in the EIBI program. Raw scores from this assessment were calculated to determine each participant's developmental quotient (Eapen et al., 2013). See Table 1 for participant demographics and overall developmental quotient results. Participants were included in the original study if they visually scanned an array of eight and made a selection response upon request. Participants were excluded from the original study if they engaged in problematic behavior that interfered with their ability to make a selection. In an effort to control for distractions, all research sessions were conducted in a separate room from the typical treatment area.

MSWO Preference Assessments

The following description provides experiment-specific details from Sipila-Thomas et al. (2021) that pertain to the current study. Please refer to Sipila-Thomas et al. for a complete description of the methodological details from that study.

Participants from Sipila-Thomas et al. (2021) completed a 5-round, combined (leisure and edible) MSWO preference assessment similarly to that described by DeLeon and Iwata (1996). Participants entered a separate room from their typical treatment area and sat across from the experimenter at a table. Before beginning the assessment, the experimenter presented each leisure stimulus to the participant and provided the participant with an opportunity to interact with the stimulus for 30 s. Each stimulus was only presented one time during the presession exposure period before it was included in the preference assessment.

Following presession exposure, the experimenters conducted two MSWO preference assessments to identify highly preferred edible and leisure stimuli, with the first preference assessment including eight edible stimuli and the second preference assessment including eight leisure stimuli. Each MSWO preference assessment included five rounds. The experimenters calculated a selection percentage (operational definitions are described in more detail below) for each stimulus by dividing the total number of times each stimulus was selected by the total number of times that same stimulus was presented in the array, and multiplied by 100 to yield a percentage. The four leisure stimuli and four edible stimuli with the highest selection percentage were then included in a final, combined MSWO preference assessment. In an effort to control for potential satiation effects, the combined MSWO preference assessment was conducted on the following day. In the present analysis, we only observed videos from the combined MSWO preference assessment.

Sipila-Thomas et al. conducted the combined assessment (the assessment evaluated in this secondary data analysis) in a manner identical to that of the edible and leisure assessments, except it consisted of four leisure stimuli and four edible stimuli, as noted above. Specifically, eight stimuli were presented in an array on the table and the participant was instructed to "pick one" or "choose one". Following each selection, the stimulus was removed from the array and the participant was given 30 s to engage with the leisure stimulus, or an opportunity to consume the edible stimulus. The remaining stimuli in the array were then shifted one spot to the left with the stimulus that was placed furthest left in the previous trial rotated to the far-right end of the array. This process continued until the final stimulus for the round was selected and each stimulus was given a ranking for that round (defined below). After the final selection, the eight stimuli were presented in an array again and a new round of the assessment was conducted. The assessment continued until five full rounds were complete.

Dependent Measures

Selection. We collected data on the stimulus selected in each trial of the combined MSWO preference assessment. Selection was defined as a participant touching a stimulus with their hand or finger within 30 s of the instruction to "pick one" or "choose one". If the participant did not engage with the leisure stimulus or consume the edible stimulus after each selection, we still scored the stimulus as selected.

Ranking. Stimuli were ranked highest (1) to lowest (8) in a preference hierarchy (Morris & Vollmer, 2020) for each round of the assessment. Since each stimulus was removed after it was selected during each assessment round, no additional calculation was needed. Rankings in each round were independent of one another. For example, if the mesh ball was selected first in round 1, it received a ranking of 1 for the first round and that round continued until all eight stimuli were selected and each stimulus received a ranking. During round 2, if the mesh ball was selected third, it received a ranking of 3 for the second round. During round 3, if the mesh ball was once again selected first, it received a ranking of 1 for the third round. This process continued until all five assessment rounds were completed and five independent preference hierarchy rankings of stimuli 1 through 8 were created. If a participant did not make a selection within 30 s of the instruction, or if they did not make eight selections in each round of the MSWO preference assessment, they were excluded from the analysis.

Preference Stability and Variability. Following the ranking of selection for each round of the preference assessment, we used the Spearman's ρ to compare stability in selection between consecutive rounds (e.g., MSWO assessment round 1 vs. MSWO assessment round 2). Preference stability was defined as a Spearman's ρ equal to or exceeding the critical *r* value of .6, consistent with previous research on preference stability (Verriden & Roscoe, 2016; Morris & Vollmer, 2020), and based on methods described by Hanley et al. (2006). Conversely, we defined preference variability as a Spearman's ρ equal to or less than the critical *r* value of .59 (see data analysis for more detail).

Patterns of Stability and Variability. Patterns of stability were defined as at least three out of four critical *r* values between consecutive MSWO preference assessment rounds meeting or exceeding the .6 criterion for preference stability. Patterns of variability were defined as zero, one, or two out of four critical *r* values between consecutive MSWO preference assessment rounds meeting or exceeding the .6 criterion for preference stability.

Interobserver Agreement. Interobserver agreement (IOA) for stimulus selection was calculated for 40% of MSWO preference assessment rounds for each participant. For each trial of that specific round, a second, independent observer recorded stimulus selection. An agreement was scored if both observers recorded the same stimulus selected. A disagreement was scored if the second observer recorded a different stimulus selected than the primary observer. The experimenter then calculated IOA for each selection by dividing the total number of agreements by the total number of agreements plus disagreements and multiplying by 100 to yield a percentage (Cooper et al., 2019). IOA was calculated at 100% across all participants.

Data Analysis

As mentioned above, to obtain stimulus rankings needed to calculate correlation coefficients, we observed videos for each participant during their combined MSWO preference assessment and ranked stimulus selection in each of the five rounds of the assessment. After the preference hierarchy rankings for each participant were collected, we measured for preference stability in selection between consecutive rounds. In addition, we compared selection between the final round and first round.

For each participant, rankings from assessment round 1 were compared to rankings from assessment round 2, rankings from assessment round 2 were compared to rankings from assessment round 3, rankings from assessment round 3 were compared to rankings from assessment round 4, and rankings from assessment round 4 were compared to rankings from assessment round 5. Finally, we measured for preference stability across the entirety of the assessment by comparing rankings from assessment round 5 to assessment round 1. Each of the five comparisons generated an independent *r* value used to determine its level of stability (for an example of this procedure, see Figure 1).

RESULTS

Table 2 depicts the participant results. Patterns of stability between consecutive MSWO preference assessment rounds were obtained for 10 participants (77%). Specifically, patterns of stability in which all four *r* values met or exceeded the .6 criterion were obtained for six participants (46%), Tim, Larry, Linda, Luke, Robert, and Henry (see figure 2); and patterns of stability in which three out of four *r* values met or exceeded the .6 criterion were obtained for four participants (31%), Ryan, Sam, Ethan, and Andrew (see figure 3). Patterns of variability between consecutive MSWO preference assessment rounds were obtained for three participants (23%). Specifically, patterns of variability in which two out of four *r* values met or exceeded the .6 criterion for stability were obtained for two participants (15%), Stanley and Oliver; and patterns of variability in which zero *r* values met or exceeded the .6 criterion for stability was obtained for one participant (8%), Megan (see figure 4).

Results for comparisons between round 5 and round 1 for each participant were similar to those obtained from comparisons between consecutive rounds. Specifically, preference stability, in which the *r* value met or exceeded the .6 criterion were obtained for nine participants (69%), Larry, Linda, Sam, Luke, Ethan, Andrew, Robert, Oliver, and Henry; and preference variability, in which the *r* value was equal to or less than .59 were obtained for four participants (31%), Stanley, Tim, Ryan, and Megan.

For three participants, *r* values obtained from comparisons between round 5 and round 1 differed from patterns observed from consecutive rounds. Specifically, patterns of stability were obtained for Tim and Ryan during consecutive round comparisons, but comparisons from round 5 and round 1 were variable. In contrast, patterns of variability were obtained for Oliver during consecutive round comparisons, but the comparison from round 5 and round 1 was stable.

DISCUSSION

Results indicate that pre-school aged children with autism in this study were more likely to engage in patterns of preference stability than patterns of preference variability within the same preference assessment. Ten of the 13 participants (77%) engaged in patterns of preference stability when consecutive rounds of the same preference assessment were compared. These results are similar to those reported in recent research on preference stability across separate preference assessments. For example, Hanley et al. (2006) reported stable preference for seven out of 10 participants (70%) across up to 16 preference assessments, and Kelley et al. (2016) reported stable preference for 16 out of 21 participants (76%) across 10 preference assessments. Results from Butler and Graff (2021) were similar during what they defined as “short-term” month-by-month preference assessment comparisons, reporting stable preference for all four participants (100%) during edible assessments, three out of four participants (75%) during leisure assessments, and two out of four participants (50%) during social attention assessments. Still, there remains much difference between relative preference across a 30 min length of time and 1 month. Therefore, future research on preference stability should better define and analyze length of time between preference assessments to see how increasing time length may affect stability of preference.

We also compared stability from round 5 and round 1 to observe the effects of repeated exposure to stimuli on preference. Of the 13 participants, only two (Tim and Ryan) who had an *r* value of .6 or better across three out of four round comparisons had an *r* value below .6 when round 5 and round 1 were compared. For these participants, repeated exposure to certain stimuli may have decreased its momentary evocative effectiveness (Michael, 1993), resulting in a reduction in *r* value from round 1 to round 5. The participant's responding also may have been more sensitive to consumption of stimuli over time, resulting in highly preferred stimuli at the beginning of the assessment ranking lower across rounds. It is also possible that over the entirety of the assessment, participants may have become more familiar with each stimulus. As the novelty of each stimulus was reduced, the value of each stimulus may have also reduced. Future research should analyze the effects of repeated exposure of stimuli on patterns of preference stability by comparing the change in ranking of each stimulus round-by-round to evaluate how establishing operations may increase or decrease the value of stimuli across individual rounds of MSWO preference assessments.

Although our study was the first to compare preference stability across rounds of a preference assessment, we only compared consecutive rounds to determine whether or not participants engaged in patterns of preference stability. For example, we did not compare round 1 vs. round 3 for any participant. The rationale for only comparing consecutive rounds was to highlight consecutive round changes and how they may lead to a better understanding of patterns of selection to guide practice. It is possible that comparing all rounds with each and every round would have resulted in different results than were presented in the study as consecutive round comparisons may have led to false positives for patterns of preference stability. If participant's consecutive rounds comparisons resulted in small and consistent changes across rounds, those *r* values would be greater than the .6 criterion for at least 3 out of 4 comparisons. Conducting a complete analysis across rounds may show greater variability in preference. Future research may consider calculating each round comparison by calculating additional *r* values across all rounds and comparing those to consecutive rounds to measure the validity of our definition for patterns of preference stability.

In our study, we presented and defined a new way to measure patterns of stability and variability across rounds of the same preference assessment. Previous research has consistently measured mean *r* values to determine stability. Instead, we defined patterns of stability as at least three out of four critical *r* values between consecutive MSWO preference assessment rounds meeting or exceeding the .6 criterion for preference stability. Measuring stability like this is beneficial because non-selected stimuli are accounted for and not washed out as they would have been if we calculated mean *r* values. Upon further analysis, defining patterns of stability in this way appears to be a more conservative measure of stability. For example, if we had measured stability as the mean *r* value of comparisons across all rounds of the preference assessment as meeting or exceeding .6, 12 out of 13 participants would have met or exceeded the .6 criterion. Specifically, Stanley and Oliver showed patterns of variability in our analysis, but mean *r* values across all consecutive comparisons would have indicated that their preferences were stable with *r* values of .63 and .64 respectively. Recent studies on choice have begun assessing more conservative or refined measures to analyze results on choice (see Carter & Zonneveld, 2019; Sipila-Thomas et al., 2021). Future research on preference stability, and patterns of choice in general, should consider similar methodologies when analyzing results to ensure variable responding is not removed from analysis when averages are calculated. Also, further defining behavioral patterns may better allow researchers to categorize behavior and further individualize assessment and subsequent treatment.

Although our findings were similar to previous studies that report preference stability across multiple preference assessments (e.g., Hanley et al. 2006; Kelly et al., 2016; Butler & Graff, 2021), it is still unclear if the participants in previous research who engaged in patterns of stability across separate preference assessments would have also engaged in patterns of stability across rounds within the same preference assessment. It is also unclear how time may affect stability results. Comparisons between consecutive rounds are brief and an assessment typically lasts less than 30 min. If value for certain stimuli changes over longer periods of time, the validity of round-by-round comparisons to measure preference stability may not be accurate. Future research is needed within and across preference assessments to evaluate if participants who engage in patterns of stability across rounds of a single preference assessment also engage in stable responding across multiple preference assessments.

We defined preference stability in a manner similar to previously published research (meeting or exceeding the critical *r* value of .58 or .6). However, it is important to remember that stability is a construct and defining stability numerically may not indicate that highly preferred stimuli that are stable across rounds will function as a reinforcer, as higher *r* values may not directly relate to reinforcer effectiveness. Future researchers should conduct reinforcer assessments with highly ranked stimuli from participants whose round-by-round comparisons resulted in different ranges of *r* values to evaluate if highly ranked stimuli from assessments with *r* values closer to 1 (perfect correlation) are more likely to function as reinforcers in comparison to participants with highly ranked stimuli from assessments with *r* values closer to 0 (no correlation).

The findings from this study provide a variety of practical implications. First, if selection is stable across the first three rounds of a MSWO preference assessment, a brief MSWO preference assessment as described by Carr et al. (2000) may be sufficient in determining preference. All seven participants (Tim, Larry, Linda, Luke, Ethan, Robert, and Henry) whose two *r* values from comparisons between the first three rounds (round 1-2 & round 2-3) met or exceeded .6, also engaged in patterns of stability across the full five round assessment. Practitioners may consider terminating the assessment after three rounds to save time and lessen the likelihood of the potential abolishing effects of frequent presentations of stimuli (Langthorne & McGill, 2009).

Patterns of stable responding may also indicate that less frequent preference assessments are needed. In a survey of psychologists and their use preference assessments by Graff and Karsten (2012), 81.4% of participants identified lack of time as a barrier to conducting regular preference assessments. If responding is stable across rounds of a single preference assessment, and results from previous studies show selection is likely to remain stable over many preference assessments, practitioners may be more willing to assess preference with confidence that a single preference assessment is sufficient in identifying potential reinforcers.

Variable selection across rounds of a single preference assessment can also present important information for practitioners. If selection is variable, certain prerequisite skills, such as scanning an array, discrimination of similar stimuli, or necessary motor skills may need to be further assessed to identify if the individual is ready to complete a MSWO preference assessment. In our study, we did not assess for prerequisite discrimination and motor skills before administering the preference assessment. It is possible that variable responding resulted from an absence of untested prerequisite skills. Practitioners should review the individual's current programming and ensure that the necessary skills to complete a MSWO preference assessment are in the individual's repertoire. If these skills are not present, they may consider assessing preference in different ways, such as a paired-stimulus (Fisher et al., 1992) or free-operant (Roane et al., 1998) preference assessment. If these skills are present, stimuli from a wider range of classes may be necessary to better capture preference and identify effective reinforcers. To determine which preference assessment may be best suited for each individual's current performance level, see Karsten et al. (2011).

The findings from our study are not without limitations. The first limitation is the use of a secondary data analysis to measure patterns of stability across rounds. Collecting data on pre-recorded videos led to a few obstacles that could have been avoided if data were collected in-person. The first main obstacle was the exclusion of six participants from the original study due to missing videos. Without these additional videos, patterns of stability and variability from these participants are unknown. Furthermore, the use of a secondary data analysis relies on existing data that attempts to answer a different research question, possibly resulting in an absence of important variables necessary to answer the new research question (Cheng & Phillips, 2014). A second limitation refers to the exclusion criteria in our study. Participants that did not make a selection on every opportunity of their assessment were excluded from the study. It is common for individuals completing a preference assessment to not select a stimulus as the array shortens and lesser preferred stimuli remain. Future research should consider methodological strategies to measure stability across rounds of a single preference assessment when participants do not select a stimulus on all selection opportunities. Finally, it is unclear how measures of stability and variability were affected by the physical location of each stimulus presented. Individuals with autism may engage in position-biased responding during preference assessments and select lesser preferred stimuli over more preferred stimuli as a result of their location in the array (Bourret et al., 2012). It is unknown if any of the participants engaged in position-biased responding and if so, to what extent that bias affected the obtained results.

The use of preference assessments to identify potential reinforcers for individuals with autism has received robust support. However, less is known regarding patterns of choice within those assessments. Our results show preliminary data that preference across rounds of a single MSWO preference assessment are more stable than variable, consistent with results from previous studies assessing stability across multiple preference assessments. Still, continued research on patterns of choice within preference assessments is needed.

APPENDICES

APPENDIX A

Tables

Table 1:

Participant Descriptions

Participant	Age	Gender	Overall DQ
Stanley	3 years, 6 months	Male	65.25
Tim	4 years, 10 months	Male	66.03
Larry	5 years, 1 month	Male	80.26
Ryan	5 years, 2 months	Male	53.07
Linda	5 years, 0 months	Female	39.73
Sam	5 years, 4 months	Male	77.16
Luke	4 years, 2 months	Male	60.91
Ethan	4 years, 7 months	Male	72.45
Andrew	4 years, 2 months	Male	99.43
Robert	5 years, 3 months	Male	48.22
Megan	5 years, 3 months	Female	65.18
Oliver	3 years, 6 months	Male	67.22
Henry	3 years, 5 months	Male	-

Note. DQ = developmental quotient. When the original study was conducted, the overall developmental quotient score was not available for Henry.

Table 2:

Spearman's ρ critical r values

Participant	Round 1 & 2	Round 2 & 3	Round 3 & 4	Round 4 & 5	Round 5 & 1
Stanley	0.71*	0.52	0.81*	0.48	.07
Tim	0.69*	0.9*	0.81*	0.88*	.57
Larry	0.74*	0.83*	0.98*	0.83*	.69*
Ryan	0.38	0.83*	0.64*	0.71*	.45
Linda	0.83*	0.95*	0.88*	0.93*	.81*
Sam	0.79*	0.55	0.71*	0.79*	.83*
Luke	0.81*	0.79*	0.62*	0.9*	.61*
Ethan	0.71*	0.69*	0.24	0.79*	.71*
Andrew	0.33	0.6*	0.98*	0.83*	.64*
Robert	0.81*	0.98*	0.95*	0.9*	.81*
Megan	-0.14	-0.26	0.02	0.26	.07
Oliver	0.47	0.52	0.71*	0.86*	.97*
Henry	0.81*	0.83*	0.95*	0.93*	.95*

Note. * Values represent patterns of stability

APPENDIX B

Figures

Figure 1:

Example of procedure.

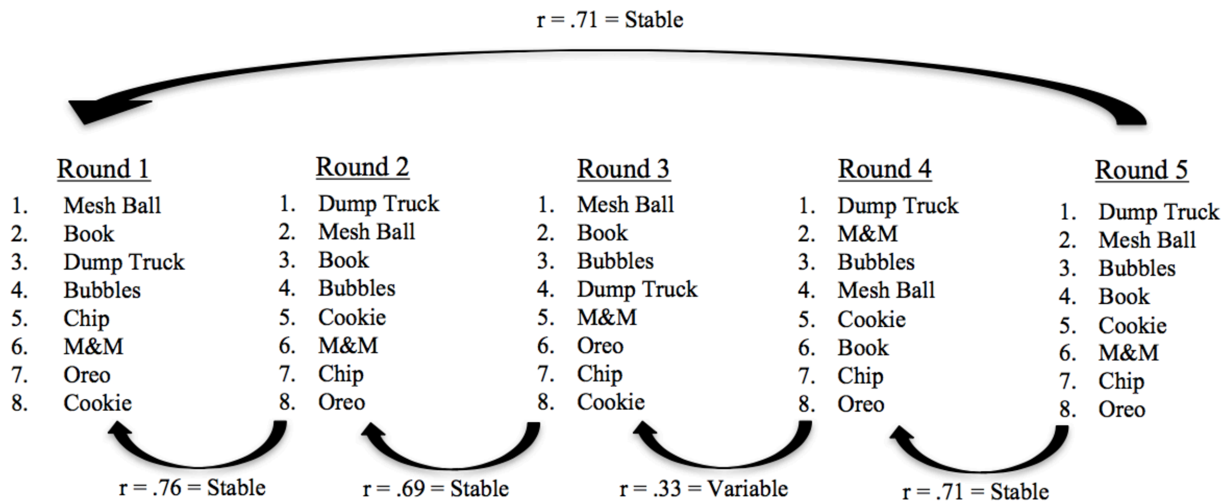


Figure 2:

Results for participants who engaged in patterns of preference stability in which all four critical *r* values met or exceeded the .6 criterion.

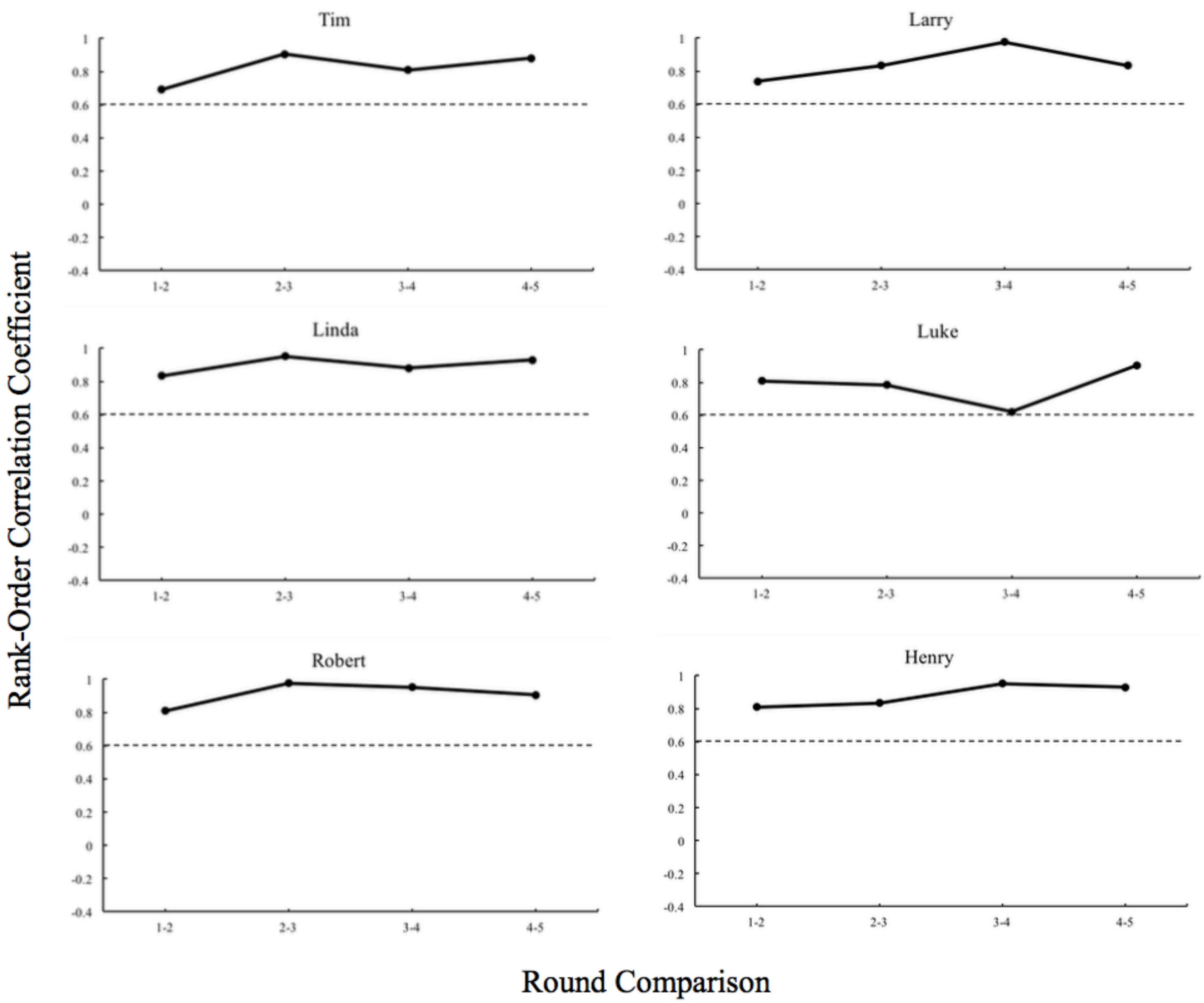


Figure 3:

Results for participants who engaged in patterns of preference stability in which three out of four critical *r* values met or exceeded the .6 criterion.

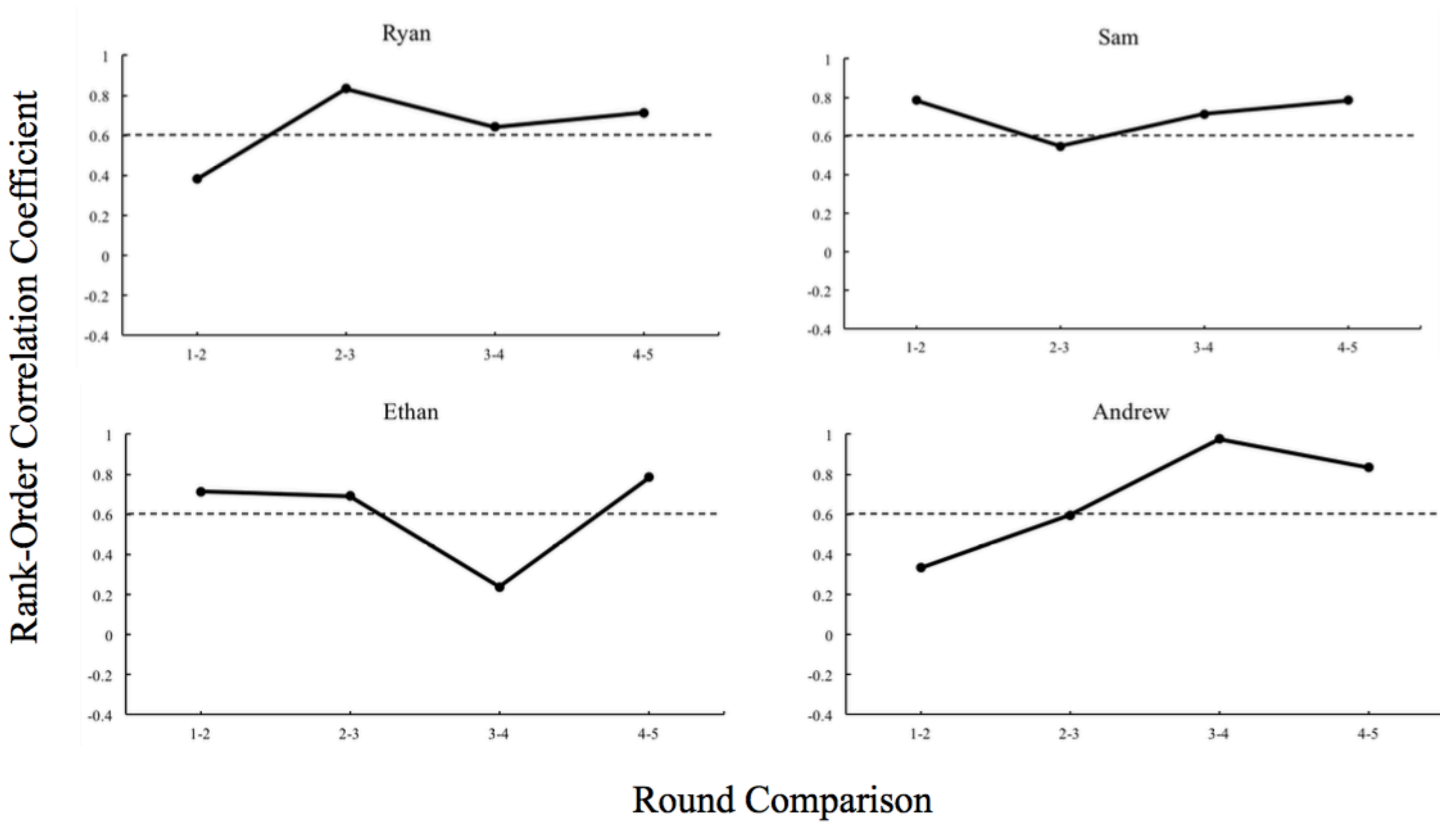
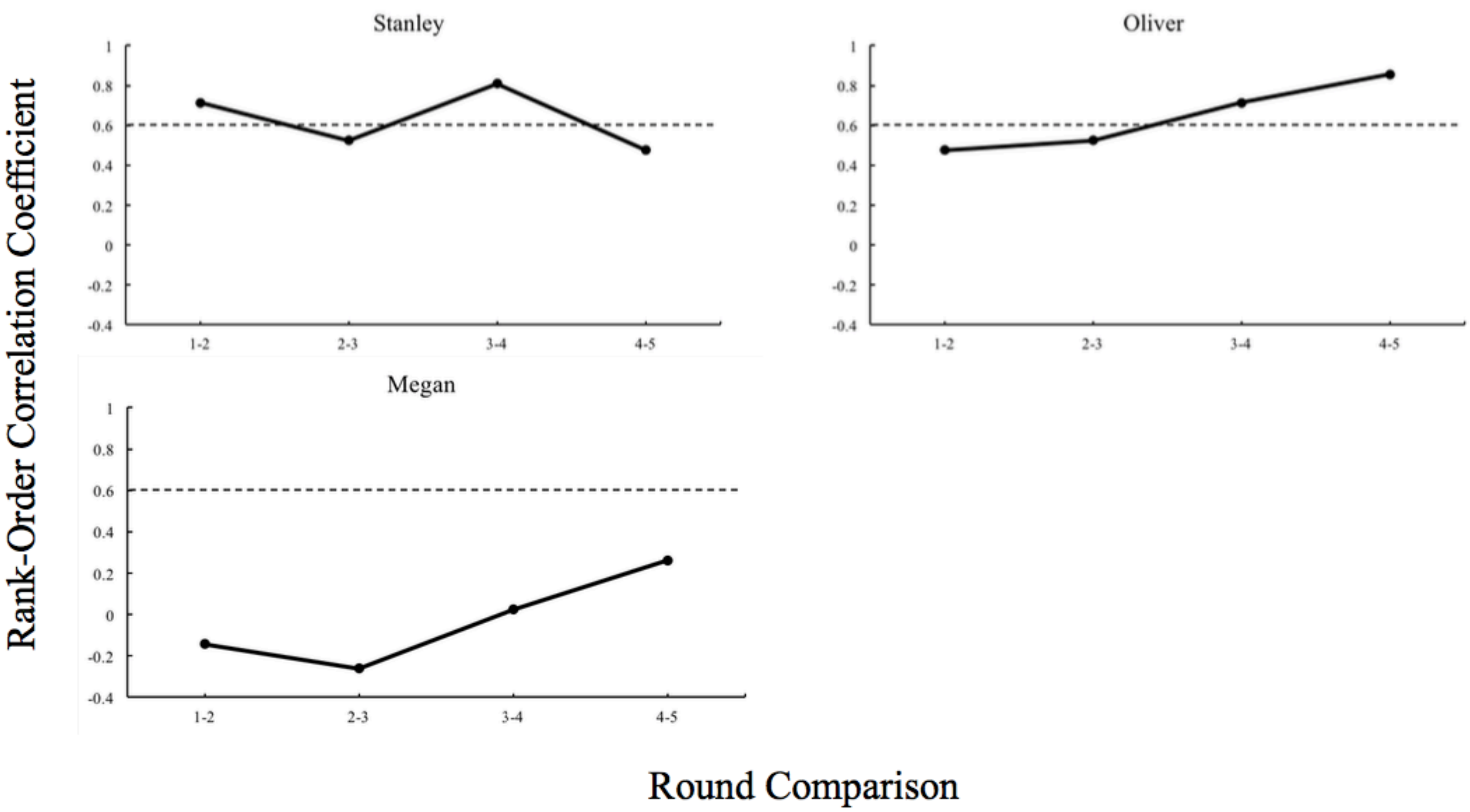


Figure 4:

Results for participants who engaged in patterns of preference variability in which zero, one, or two out of four critical *r* values met or exceeded the .6 criterion.



REFERENCES

REFERENCES

Bourret, J. C., Iwata, B. A., Harper, J. M., & North, S. T. (2012). Elimination of position-biased responding in individuals with autism and intellectual disabilities. *Journal of Applied Behavior Analysis, 45*(2), 241-250. https://doi.org/10.1901/jaba.2012.45-241

Butler, C., & Graff, R. B. (2021). Stability of preference and reinforcing efficacy of edible, leisure, and social attention stimuli. *Journal of Applied Behavior Analysis.* Advance online publication. https://doi.org/10.1002/jaba.807

Canella, H. I., O'Reilly, M. F., & Lancioni, G. E. (2005). Choice and preference assessment research with people with severe to profound developmental disabilities: a review of the literature. *Research in Developmental Disabilities, 26*(1), 1-15. https://doi.org/10.1016/j.ridd.2004.01.006

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. https://doi.org/10.1901/jaba.2000.33-353

Carter, A. B., & Zonneveld, K. L. M. (2019). A comparison of displacement and reinforcer potency for typically developing children. *Journal of Applied Behavior Analysis, 53*(2), 1130-1144. https://doi.org/10.1002/jaba.636

Cheng, H. G., & Phillips, M. R. (2014). Secondary analysis of existing data: opportunities and implementation. *Shanghai Archives of Psychiatry, 26*(6), 371-375. https://doi.org/10.11919/j.issn.1002-0829.214171

Cooper, J. O., Heron, T., E., & Heward, W. L. (2019). *Applied Behavior Analysis* (3rd edition). Hoboken, NJ: Pearson Education.

DeLeon, I. G., & Iwata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behavior Analysis, 29*(4), 519-533. https://doi.org/10/1901/jaba.1996.29-519

Eapen, V., Crncec, R., & Walter, A. (2013). Clinical outcomes of an early intervention program for preschool children with autism spectrum disorder in a community group setting. *BMC Pediatrics, 13*(3), 1-9. https://www.biomedcentral.com/1471-2431/13/3

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. https://doi.org/10.1901/jaba.1992.25-491

Graff, R. B., & Karsten, A. M. (2012). Assessing preferences of individuals with developmental disabilities: A survey of current practices. *Behavior Analysis in Practice, 5*(2), 37-48. https://doi.org/10.1007/BF03391822

Hanley, G. P., Iwata, B. A., & Roscoe, E. M. (2006). Some determinants of changes in preference over time. *Journal of Applied Behavior Analysis, 39*(2), 189-202. https://doi.org/10.1901/jaba.2006.163-04

Heaton, J. (2003). Secondary data analysis. In R. Miller & J. Brewer (Eds.), *The A-Z of Social Research* (pp. 285-288). Sage, London.

Kang, S., O'Reilly, M., Lancioni, G., Falcomata, T. S., Sigafoos, J., & Xu, Z. (2013). Comparison of the predictive validity and consistency among preference assessment procedures: A review of the literature. *Research in Developmental Disabilities, 34*(4), 1125-1133. https://doi.org/10.1016/j.ridd.2012.12.021

Karsten, A. M., Carr, J. E., & Lepper, T. L. (2011). Description of a practitioner model for identifying preferred stimuli with individuals with autism spectrum disorders. *Behavior Modification, 35*(4), 347-369. https://doi.org/10.1177/0145445511405184

Kelley, M. E., Shillingsburg, M. A., & Bowen, C. N. (2016). Stability of daily preference across multiple individuals. *Journal of Applied Behavior Analysis, 49*(2), 394-398. https://doi.org/10.1002/jaba.288

Langthorne, P., & McGill, P. (2009). A tutorial on the concept of the motivating operation and its importance to application. *Behavior Analysis in Practice, 2*(2), 22-31. https://doi.org/10.1007/BF03391745

Michael, J. (1993). Establishing operations. *The Behavior Analyst, 16*(2), 191-206. https://doi.org/10.1007/BF03392623

Morris, S. L., & Vollmer, T. R. (2020). Evaluating the stability, validity, and utility of hierarchies produced by the social interaction preference assessment. *Journal of Applied Behavior Analysis, 53*(1), 522-535. https://doi.org/10.1002/jaba.610

Mullen, E. M. (1995). *Mullen scales of early learning.* American Guidance Service.

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*(7), 1913-1916. https://doi.org/10.1177/1362361320919243

Roane, H. S., Vollmer, T. R., Ringdahl, J. E., & Marcus, B. A. (1998). Evaluation of a brief stimulus preference assessment. *Journal of Applied Behavior Analysis, 31*(4), 605-620. https://doi.org/10.1901/jaba.1998.31-605

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*(1), 403-416. https://doi.org/10.1002/jaba.775

Verriden, A. L., & Roscoe, E. M. (2016). A comparison of preference-assessment methods. *Journal of Applied Behavior Analysis, 49*(2), 265-285. https://doi.org/10.1002/jaba.302

Zhou, L., Iwata, B. A., Goff, G. A., & Shore, B. A. (2001). Longitudinal analysis of leisure-item preferences. *Journal of Applied Behavior Analysis, 34*(2), 179-184. https://doi.org/10.1901/jaba.2001.34-179