Spring 5-1-2013

Methods for Aggregating Descriptive Assessment Data Prior to Conducting a Matching Analysis

Mary-Jo Robinson
Syracuse University

Methods for Aggregating Descriptive Assessment Data Prior to Conducting a Matching Analysis

A Capstone Project Submitted in Partial Fulfillment of the
Requirements of the Renée Crown University Honors Program at
Syracuse University

Mary-Jo Robinson
Candidate for B.A. Degree
and Renée Crown University Honors
May 2013

Honors Capstone Project in Psychology

Capstone Project Advisor: ________________________
Brian Martens, Ph.D.

Capstone Project Reader: ________________________
Tanya Eckert, Ph.D.

Honors Director: ________________________
Stephen Kuusisto, Director

Date: April 23, 2013

Abstract

Data collected from descriptive assessments of children's behavior and caregiver responses can be summarized according to matching theory, which states that relative rates of responding match relative rates of reinforcement. The extent to which matching applies to descriptive assessment data can be evaluated by application of the generalized matching equation (GME). However, three limitations exist in previous applications of the GME: (a) the most appropriate method of aggregating the data is unclear, (b) consequences must be manipulated in order to determine reinforcement, and (c) individual differences in sensitivity can influence the results. This Capstone project addressed those three issues by comparing the results of a descriptive assessment and an experimental analysis for two children. The results showed that aggregating data into 2.5 min bins prior to applying the GME provided the closest approximation to matching under experimental conditions. Furthermore, the descriptive data were more variable than the experimental data. Lastly, the results of this study support use of the GME to detect individual differences in matching.

Table of Contents

Acknowledgements

It is with immense gratitude that I acknowledge the support and help of Professor Brian Martens, who served as my Capstone advisor. This Capstone was only made possible by his expertise, encouragement, support, and patience. I would also like to extend my appreciation to Tonya Lambert. As a Ph.D. candidate, Tonya was responsible for leading the experimental and descriptive phases of data collected for the data used for this Capstone. Furthermore, she was a great support to me throughout this entire process. In addition, I wish to thank Professor Eckert, who acted as my Capstone reader. Her flexibility and positive feedback have been immensely appreciated. Lastly, I would like to thank the Renée Crown University Honors Program staff for the endless support that has been made available. Thank you all for your support and encouragement. It has certainly made a difference.

Advice to Future Honors Students

I have found the process of completing a Capstone project to be very rewarding. It has challenged my work ethic, my brain, and my passion for field of work more than anything else in my college career. Although exhausting and often frusterating, these challenges resulted in great satisfaction and pride.

My first piece of advice is to find an advisor that will support you, challenge you, and keep you on schedule. These are things that I was lucky enough to find in my advisor, and it made all the difference in this process. If you don't feel that you are getting these things, my advice is to seek help from the honors department and find a new advisor. That is how crucial having an excellent advisor was for the completion of this Capstone.

My second piece of advice is to start early. My data was collected in its entirety in the Spring semester of my Junior year. Without having that done so early, I feel as though my experience would not have been nearly as positive. Furthermore, continue to work at a regular pace throughout the year. That way, you feel as prepared as possible when turn-in day arrives. Trust me, it arrives much faster than you would expect!

Lastly, be proud of the work that you are doing, and don't sell yourself short. Be confident that you will master the material that you are working on, because you eventually will. It may take time to feel like you understand your project at all! Don't let this shake your confidence in your abilities. It all comes together in the end.

Introduction

The principles of operant conditioning describe how learning occurs through the reinforcement and punishment of behavior (Cataldo et al., 2012). These principles also occur in various natural settings, including classrooms, when teachers reinforce children's behavior with consequences such as attention, tangible rewards, or escape from task demands (Martens, Gertz, Werder, & Rymanowski, 2010). Behavior is said to have an operant function when it allows children to consistently obtain such consequences. One way to identify potential operant functions of children's behavior in classrooms is to conduct a descriptive assessment of behavior and its consequences (Martens, DiGennaro, Reed, Szczech, & Rosenthal, 2008). According to Martens and colleagues, a descriptive assessment is a type of functional behavior assessment (FBA) that seeks to describe antecedent-behavior-consequence relationships that exist in the natural environment.

During an FBA, observations of child behavior are conducted at the time and place of their natural occurrence. Neither the setting nor the data are manipulated in a descriptive assessment. Therefore, descriptive assessment provides an accurate account of the events surrounding the occurrence of observed child behaviors. As an example, Anderson and Long (2002) observed three boys and one girl between the ages of 6 and 13, all of whom were being treated for severe behavior problems. Observation sessions were conducted throughout the day and each child's caregiver (either a teacher or parent) was instructed to respond to problem behavior as they typically would. In their descriptive assessment, Anderson and Long sequentially recorded child behaviors and categories of caregiver responses that were defined prior to conducting the study. The resulting output provided not only the frequencies of child behaviors and caregiver responses, but also the sequence in which they occurred. Recording these naturally occurring sequences provided meaningful information about which caregiver responses were acting as potential reinforcers for certain child behaviors. In order to identify such relationships, Anderson and Long calculated and graphed various conditional probabilities from the descriptive assessments to develop interventions to reduce the children's severe problem behaviors.

Applications of Matching Theory

The data collected from descriptive assessments can provide insight into the interaction of behavior and reinforcers within natural environments, and have been summarized according to matching theory (St. Peter et al., 2005). Matching theory states that relative rates of responding will match relative rates of reinforcement provided for two alternative behaviors (Herrnstein, 1961). More recently, matching has been used to identify which specific events may be acting as reinforcers for on- and off-task child behavior. For example, in the work conducted by St. Peters and colleagues, the relationship between adult attention and child behavior was explored. Three students between the ages of 14 and 19, all with diagnosed developmental disabilities, participated in the study. A descriptive assessment was conducted in a playground setting and in a classroom setting. Observations were collected for at least 3.5 hours for all three students. A matching analysis was then applied to the descriptive data using the generalized matching equation. Based on the results of the matching analysis, St. Peter et al. found a correlation between the relative amounts of adult attention and child problem behavior, providing support for matching and suggesting that as the rate of adult attention increases, so does the rate of child problem behavior.

As illustrated by St. Peter et al. (2005), the generalized matching equation (GME) is commonly used to evaluate the extent to which matching applies to descriptive assessment data. The GME is a linear equation in which the logarithmic (log) transformation of behavior ratios for responses one and two (R1/R2) are regressed on the log of reinforcement ratios for the two responses (r1/r2) to predict relative rates of behavior. For strict matching, the slope of the line will equal 1.0, and the intercept will equal 0, indicating a unit change in behavior ratios with a change in reinforcement ratios (slope = 1.0) and no bias favoring either behavior (intercept = 0).

Although matching has been a useful tool for evaluating the operant function of problem behavior, there are limitations to fitting the GME to descriptive data. First, without actually manipulating consequences one cannot be sure which adult responses are functioning as reinforcers for child behavior. Second, the most accurate method of aggregating descriptive assessment data prior to conducting the analyses has not been determined. Because researchers have used various methods of aggregation, the accuracy of matching for descriptive data is still in question. For example, some researchers have measured adult responses to child behaviors based on closeness in time, determining how many seconds have passed between a specific behavior and a response before considering that response as a potential reinforcer (e.g., Borrero & Vollmer, 2002). However, others, such as St. Peter et al. (2005), have partitioned observations into specific blocks of time (e.g., 10 min), and counted the number of responses and behaviors that occurred within each block or “bin”.

Variations in how the data are binned may result in different matching parameters, calling into question the accuracy of matching analyses. For example, Borrero et al. (2007) conducted a study in which they binned experimental data in multiple ways, and the resulting matching parameters were different for each method of binning. A total of 25 undergraduate students served as participants. The experimental sessions were conducted two-on-one with one undergraduate student and two confederates in a laboratory room. The confederates followed strict schedules of administering statements of agreement in response to answers given by the participant in an effort to reinforce verbal responses and attending to each of the confederates. The GME was used to analyze the experimental data in an attempt to determine whether or not the rate of agreeable statements influenced the rate or duration of student responses. In order to conduct this analysis, Borrero et al. separated the data in three ways: (a) they applied the GME to data from only the first 5 min of the session, (b) to data from only the last 5 min of the session, and (c) to data from the entire session (20 or 30 min). The results of this study showed that the data from the first 5 min of the session provided a closer approximation to perfect matching than the data from the last 5 min of the session. In other words, this finding implies that participants were immediately sensitive to reinforcers (i.e., agreeable statements given by the experimenter), but that this sensitivity decreased near the session end. The researchers also found that less variance was accounted for in the 30 min sessions that in the 20 min sessions, also implying differences in participant matching.

A third limitation to using the GME may result from individual differences in subjects' sensitivity to reinforcement across different bins or time windows. According to Baum and Aparicio (1999), the payoff ratio of one reinforcer compared to that of an alternative reinforcer plays a role in determining choice behavior. For example, if a child receives some form of reinforcement after every 60 s of on-task behavior, and some form of reinforcement after every 90 s of off-task behavior, then the payoff ratio is greater for on-task behavior. According to matching theory, the child should choose on-task behavior more frequently to ensure a higher payoff of reinforcers. However, a child that is highly sensitive to reinforcement schedules may recognize that at around 90 s, they can achieve a higher payoff ratio on the leaner alternative if they temporarily switch to off-task behavior to receive reinforcement. Therefore, individual sensitivity to reinforcement may influence behavior allocation, as the child may not behave in such a way that optimizes the payoff ratio for each alternative. Baum and Aparicio found this to be true in their study of non-humans, in which the payoff ratio achieved by the subjects for two individual reinforcers was not equal. Instead, the subjects showed a strong preference toward one reinforcer over the other. As concluded by Prelec (1982), it is possible for some individuals to display insensitivity to relative reinforcement during brief time windows which would appear as a failure in matching (Prelec, 1982). As a result, the existence of individual differences in reinforcement sensitivity may interact with the length of bins used to aggregate data resulting in different degrees of matching across participants.

Purpose of the Present Study

The goals of this capstone project were (a) to determine the most accurate method of aggregating descriptive assessment data into bins prior to conducting matching analyses, (b) to compare the results of matching from descriptive data on naturally occurring rates of teacher attention to experimental data produced by programmed schedules of experimenter attention, and (c) to examine individual differences in children's sensitivity across different bins of time, which may then impact the accuracy of the aggregated data in terms of matching. In order to achieve these goals, predefined schedules of adult attention for both on-task and off-task child behavior were manipulated, and experimental data were collected. These data were then analyzed at three levels of aggregation or bin lengths (5 min, 2.5 min, and 1 min). The number of teacher responses and duration of child on- and off-task behaviors in each bin were considered. Data from each aggregation level were compared to the GME parameters obtained from analyzing the data from all sessions based on the means of each condition (i.e., the highest and most reliable level of aggregation). Once the most accurate method of aggregation was determined from the experimental data, the descriptive data were analyzed using the same method. Results from analysis of the descriptive data and the experimental data were then compared. This was done to determine how accurately matching defines the operant function of behavior within the children's natural environment using results from the experimental data as criterion.

Method

Participants and Setting

The participants in this study were two, 4-year-old boys. Both boys, Mark and Jack (pseudonyms), were enrolled in an inclusive preschool program in Central New York. Mark and Jack were both labeled as a preschooler with a disability due to a functional delay in one or more areas (e.g., cognitive, language, motor, socio-emotional, adaptive development). Each child was identified by the head teacher in their classroom as a student who displayed disruptive behaviors during center activities in order to gain attention. For the collection of descriptive data, the children were observed in the classroom setting in which five adults and 7-10 students were typically present. During the collection of descriptive data, the children were engaged in center-based activities. Examples of center activities included puzzles, finger painting, building blocks, and computer programs. At each center, both Mark and Jack worked one-on-one with an adult. This allowed for careful observation and recording of both student and teacher behaviors and social reinforcers delivered by teachers during each session.

Materials

To collect both the descriptive and experimental data, a laptop containing DataPal, a computer software program was utilized. DataPal allows for the recording of both duration (second by second) and frequency of behavior. To record duration, a key was pressed to start recording the time, and then a separate key was pressed to stop recording the time for a defined behavior. To record frequency, the assigned key was pressed each time that the assigned behavior was observed. Therefore, each category of teacher and child behavior was assigned to a specific key to measure either duration or frequency.

For the collection of experimental data, an individualized schedule of adult attention was formulated for each child. These schedules were created based on the patterns of behavior and teacher attention that were observed during the collection of descriptive data. To ensure that these schedules were followed accurately during the experimental sessions, a vibrating timer was used to signal the experimenter when to provide a given form of attention.

Response Definitions and Recording

Five categories of teacher behavior and seven categories of student behavior were defined before collecting data. With respect to teacher behavior, a *demand* was defined as any “do” or “choice” command, instruction, or prompt given by the teacher (e.g., “Now color the box blue, Mark.”). A *reprimand* was defined as any “don’t” command or statement negatively evaluating student behavior (e.g., “We don’t hit in this classroom”). *Physical attention* was defined as any physical contact between the child and teacher. Physical attention could be positive, such as a “high five”, or negative, such as holding the child’s hand to block hitting. Every 3 s of physical contact was counted as an additional response. *Praise* was defined as any statement positively evaluating or praising student behavior (e.g., “Good job” or “nice asking”). Lastly, a *neutral response* was defined as any statement directed toward the student related to the task that was not a direct demand, praise, or reprimand (e.g., "Did you hear that?"). These five teacher responses were recorded throughout each session of both experimental and descriptive data collection.

Seven categories of student behavior were also defined and recorded throughout the collection of both descriptive and experimental data. *Aggression* was defined as any grasping of the teacher or her clothing, or any forceful contact with the teacher (e.g., hitting). *Destruction* was defined as any occurrence of banging objects, throwing items, pushing over furniture, swiping items off the table, or destroying materials. *Compliance* was defined as the student's independent completion of a command or an approximation of the command that was issued by the teacher. Compliance was not recorded if physical guidance from the teacher was required. *Inappropriate vocalization* was defined as non-word vocalizations above conversational level, separated by a breath. Examples of inappropriate vocalizations included crying, screaming, and whining. After every 3 s, another inappropriate vocalization was recorded. *Flopping* was defined as the student dropping from a seated or standing position to the floor without being instructed to do so. Lastly, *on-task* behavior was defined as engagement with task materials while facing the table or teacher. Behavior was not recorded as on-task if the student was destroying, hitting with, or throwing materials, even though the student was directly engaged with the materials.

Whereas some of these student and teacher responses were recorded as frequencies, others were recorded based on duration. The student behaviors of on-and off-task were recorded as duration in seconds. Therefore, for each session, the total amount of time that a child was on-task verses off-task was recorded. Additionally, a second-by-second thread of on-task and off-task behavior was also noted in the program output. All other student responses (aggression, destruction, compliance, inappropriate vocalization, and flopping) were recorded as frequencies. Therefore, a total count of each of these defined behaviors was also recorded. Additionally, the second-by- second output also identified at which points throughout the session each of these behaviors occurred.

All teacher responses were recorded as frequencies. Therefore, the count for each defined teacher behavior was recorded for each session. Additionally, just as for student responses, the second-by-second output identified at which specific points throughout the session each of the teacher behaviors occurred.

Observations throughout the collection of experimental and descriptive data were conducted by undergraduate research assistants including the Capstone author. Prior to conducting observations, each research assistant was trained to use DataPal correctly, as well as on how to appropriately identify and record teacher and student behaviors. This was done by having each student watch a number of recordings of mock classroom scenarios. Several variations of scenarios were observed to ensure that research assistants could adequately identify all necessary behaviors and record these behaviors at the exact time they occurred to ensure accuracy of the data.

While conducting observations, the research assistants sat near the perimeter of the classroom, away from the activity center that the child was working at, but close enough to accurately observe all behaviors. This was to ensure that the child was not altering his behavior due to awareness of being observed.

Experimental Design and Procedures

Descriptive phase. The first phase of this study involved the collection of descriptive data. During these sessions of observation, approximately 5 min of one-on-one student/teacher interaction was observed within the classroom environment. Observations were collected during individual center-based activities. Throughout these sessions, there was no experimenter involvement or experimental manipulation. The duration of child time on-task, duration of time off-task, and the frequency of the other student and teacher responses were recorded. A total of 12 observation sessions with two different teachers were conducted for each child.

Experimental phase. The second phase of data collection for the study involved experimental sessions. Based on the pattern and frequency of teacher responses to child behavior from the descriptive observation sessions, three individualized, concurrent variable interval (CONC VI) schedules of adult attention were created for each student. Each schedule was designed to reinforce both on- and off-task child behavior within the same session (i.e., concurrently) but at different relative rates (i.e., 90%/10%, 50%/50%, 10%/90%). During each experimental session, an experimenter would follow the given CONC VI schedule for that day. Specific teacher responses would be given at specific points in time throughout the session contingent on either child on- or off-task behavior. Just as for the descriptive sessions, the duration of each child's on- and off-task behavior were recorded, as well as the frequency of disruptive student responses. For each student, the first schedule employed was the 90%/10% schedule, which was designed to mimic the descriptive data that were previously collected. Eight 90%/10% schedule sessions were conducted for Jack, and nine were conducted for Mark. The second schedule was the 50%/50% schedule. Ten 50%/50% schedule sessions were conducted for Jack, and nine were conducted for Mark. The third schedule implemented was the 10%/90% schedule. Six 10%/90% schedule sessions were conducted for Jack and eight were conducted for Mark. To create a reversal design, the experimental sessions were concluded with a second implementation of the 90%/10% schedule.

Interobserver Agreement and Procedural Integrity

During observational sessions, interobserver agreement (IOA) was assessed during 58% of sessions across both students. IOA for student responses ranged from 70% to 100%, with a mean IOA of 88.5%. The mean IOA for time on-task was 96.95% (range = 67 to 100%). The mean IOA for teacher responses was 82.31%, with a range of 58.33% to 100%.

During the experimental phase of this study, IOA was assessed during 30% of sessions across all conditions for both students. However, computer program issues on one of the computers resulted in usable IOA data for only 27% of the experimental sessions. The mean IOA for teacher responses was 90%

(range = 64-100%). The mean IOA for student responses during experimental sessions was 81% with a range of 54 to 100%. Lastly, the mean IOA for time on-task was 92%, with a range of 71% to 99.6%.

A step-by-step protocol was used to verify that the experimental conditions were implemented correctly (i.e., procedural integrity). During the experimental phase, integrity was assessed for 36% of all sessions across all conditions and both children with a mean of 98.7% (range = 88-100%).

Data Preparation

After all descriptive and experimental data were collected, the Capstone author recoded the data into bins by hand, transformed the ratios logarithmically, and then conducted all GME analyses. To begin, for each individual session (both descriptive and experimental and across both students), the data were separated into three time windows or bins (i.e., 1 min, 2.5 min, and 5 min). For the 1 min bins, for example, all data observed from zero to 60 s of a session were separated from the data observed from 60.1 s to 120 s. For each bin, the duration of child on-task behavior, child off-task behavior, number of teacher or experimenter responses while the child was on-task, and number of teacher or experimenter responses while the child was off-task were calculated. Using this information, data for each bin size for both descriptive and experimental sessions were compiled into a spreadsheet and further analysis using the GME. To do so, first the ratio of on-task time verses off-task time for all sessions was calculated and labeled as behavior ratios. The log10 transformation was then calculated for each of these ratios excluding those sessions with values of zero for which log transformations are undefined. Also for all sessions, the ratio of reinforcement received while on-task verses reinforcement received while off-task was calculated and labeled as reinforcement ratios. The log10 transformation for each reinforcement ratio was also calculated again excluding zero values.

Using the log10 values for each behavior and reinforcement ratio, the slope, intercept, and variance accounted for were calculated for each bin size using linear regression. This was done separately for each student and for both the experimental and descriptive sessions. The experimental condition means for each child were also analyzed according to the GME, and these values were used as an overall index because they provided the most accurate estimates of matching by encompassing all of the experimental data. The results were compared across students as well as across conditions.

Results

Figures 1 and 2 depict the matching results for Mark and Jack, respectively, and from both the experimental conditions and the descriptive assessment. It can be seen by looking at the experimental results that Mark showed little variability in either slope or intercept across all bin lengths. In other words, within the experimental condition, Mark showed nearly perfect matching with a mean slope of 1.04 and a mean intercept of -0.04. Across all bin lengths, including the condition means, Mark's slope values ranged from 0.71 (at the 2.5 min bin length) to 1.04 (the mean across conditions). Mark's intercept values ranged from -0.04 (the mean across conditions) to 0.09 (at the 5 min bin length). Variance accounted for (VAF) in Mark's data ranged from 41% (at the 1 min bin length) to 92% (the mean across conditions). By looking at the top panel of Figure 1, it can be observed that under the experimental condition, as the size of the bin increased from 1 min to 5 min, Mark's approximation to matching steadily improved.

By looking at the top panel of Figure 2, it can be observed that Jack also showed a steady improvement of approximation to matching as bin length increased from 1 min to 5 min. However, Jack's experimental data were more variable than Mark's, particularly at the 1 min bin length, with a slope of 0.25 (compared to Mark's 1 min slope of 0.78), an intercept of .12 (compared to Mark's 1 min intercept of 0), and only 2% VAF (compared to Mark's 41% VAF). Furthermore, as can be seen in Table 1, Jack consistently under-matched (slope values ranging from .25 at the 1 min bin length to 0.83 for the mean across all conditions) and showed a biased toward off-task behavior (intercept values ranging from -0.02 at the 2.5 min bin length to 0.23 for the mean across all conditions), with VAF ranging from 2% (at the 1 min bin length) to 87% (at the 5 min bin length).

Figure 2 shows that greater variability existed in the descriptive data than in the experimental data for both students. For Mark, the range in slope increased from 0.33 in the experimental condition to 0.63 in the descriptive assessment data, and the range in intercept increased from 0.13 to 0.44, respectively. The mean VAF across the three bin lengths for Mark decreased from 57% for the experimental data to 38% for the descriptive assessment data. For Jack, the range in slope increased from 0.57 in the experimental condition to 0.70 in the descriptive assessment data, and the range in intercept increased from 0.25 to 0.38, respectively. The mean VAF across the three bin lengths for Jack decreased from 44% for the experimental data to 25% for the descriptive assessment data. Despite this increased variability in the descriptive data results, for both Jack and Mark, the matching results at the 2.5 minute bin level most closely approximated matching as defined by the condition means for each child. This can be observed in Figure 1 and 2.

Discussion

There were three main goals of this Capstone project. The first of these goals was to determine the most accurate method of aggregating descriptive assessment data into bins prior to conducting matching analyses. The second was to compare the results of the descriptive and experimental matching results. The last goal was to examine individual differences in children's sensitivity across different bins of time, which may then impact the accuracy of the aggregated data in terms of matching.

As can be seen in Figures 1 and 2, dividing the data into 2.5 min bin lengths appears to be the most accurate method of aggregating descriptive assessment data. At this level of aggregation, fit parameters from the GME most accurately reflected those of the condition means for each child. Also, the most variance was accounted for at the 2.5 min bin length for both children. One possible explanation for this finding lies in the fact that data were collected in sessions of approximately 5 min. Therefore, at the 2.5 minute level, the data provided information on approximately half of each session. It may take half of a session for the child's behavior to become sensitive to adult attention, or for the experimenter (or teacher) to settle into a routine. In that case, the 2.5 min data would capture the portion of interaction between the child and adult that most closely approximated matching.

By examining the VAF values in Table 1, it becomes clear that, with the exception of the 2.5 min bin length, the GME provided a poorer description of the descriptive assessment data than the experimental data. This suggests that descriptive data collected in the natural environment may be more variable than data collected under conditions when reinforcers are manipulated. This variability may exist for a variety of reasons. First, there were several differences between the experimenter that conducted the experimental sessions and the teachers who were observed during the descriptive assessment. Individual differences in the how attention was delivered may have caused discrepancies that were not reflective of the child's sensitivity to reinforcement.

Second, the data used from the descriptive assessment to create the schedules of reinforcement that were employed during the experimental sessions may not have accurately reflected the teachers' response patterns. For example, during the descriptive assessments, more than one teacher was observed interacting with each child. Although patterns of reinforcement differed among these teachers, but data from both teachers were combined to program the experimental conditions.

Third, although the experimental sessions were also conducted within the classroom setting, these sessions were conducted with the child and experimenter separated from the rest of the classroom, unlike the descriptive assessment. Therefore, the child was much more available to distractions during the descriptive assessment sessions, which may have resulted in more off-task behavior, confounding the effects of the reinforcers alone.

When examining the data, it is clear that individual differences in matching existed between the two children, providing support for the generality of the GME as a description of choice behavior. Figures 1 and 2 show that Mark's behavior approximated matching more closely than Jack's. In other words, Mark's behavior was more sensitive to the schedule of reinforcement that was being employed by the teacher or experimenter. However, this could also be a reflection, in part, of the differences in teaching styles, as these children were observed with different teachers.

Limitations

This study provides important insight into the most accurate length of time (i.e., bin) in which to aggregate data in a matching analysis, the comparability of experimental and descriptive matching analysis results, as well as the effect of individual differences in children's sensitivity to relative reinforcement rates. However, several limitations may have implications for the accuracy of the study's findings. First and foremost, the use of only two participants in this study limits the generality of these findings to other children. Future research involving a larger number of participants is needed to replicate the current findings, which in turn would provide a clearer picture of, and more support for, the findings of the study.

Second, multiple sessions of experimental and descriptive data were eliminated from the study due to zero values. For example, if a child engaged in only on-task behavior for an entire bin (e.g., a 60 s bin), then zero would be recorded for "time spent off-task". As a result, this session would be eliminated from the GME analyses as zero values result in an undefined log10, and the GME cannot be applied. Therefore, conducting a larger number of descriptive and experimental sessions would have provided more data, compensating for those sessions that were eliminated.

Third, because sessions were conducted within the children's classroom, environmental factors may have influenced both child and teacher (or experimenter) behaviors. For example, the children at times were distracted by the activities and events occurring around them. Not only would this affect child behavior, but it would also affect teacher responses toward the child. Furthermore, the children often were aware that they were being observed which also may have influenced their behavior. Also, differences between the children's regular teacher (observed during the descriptive sessions) and the experimenter (observed during the experimental sessions) may have further influenced the child's behavior, which would skew the actual impact of the reinforcement schedule being implemented by the experimenter. These influences on behavior may have implications for the accuracy of the data collected.

Fourth, only three reinforcement schedules were employed in this study (90%/10%, 50%/50%, and 10%/50%). This leaves a large gap in relative reinforcement percentages between the 90%/10% and 50%/50% schedules.

Therefore, including more schedules of reinforcement in future research may provide a more accurate portrayal of matching as it exists in the natural environment.

Implications and Directions for Future Research

One possible direction for future research is to replicate the study with more participants and more schedule variations. Another possible direction for future research is to replicate this study with a group of typically developing children and a group of children with developmental or behavior disabilities (e.g., Autism Spectrum Disorders). This direction could shed light on whether or not children's sensitivity to reinforcement could be used as a tool in the classroom to more effectively engage students with disabilities. For example, children with specific disabilities may show lower sensitivity to reinforcement. This information could possibly assist in the diagnosis of disabilities, as well as implementing more effective reinforcement-based programs in inclusive and special education settings. These more effective programs may include more individualized and specific reinforcement schedules within the classroom (i.e., more explicit and/or frequent reinforcement for specific students). A third possible direction for future research would be to conduct a longitudinal study, in which a matching analysis is conducted at the pre-school age and then again in the primary grades. This may shed light on how children's sensitivity to reinforcement changes as they get older and how schedules of reinforcement may be adjusted in order to maximize on-task and positive behaviors.

Conclusion

Although limitations exist, the results of this study support that the GME is a useful tool in determining not only the extent to which matching applies to descriptive assessment data, but also in describing children's choice behavior. If data are aggregated appropriately (i.e., binned in lengths of 2.5 mins), applying the GME can provide an accurate description of an individual's pattern of matching. Therefore, future use of this method may have positive implications in numerous research areas such as the study of developmental disabilities.

References

Anderson, C.M., & Long, E.S. (2002). Use of a structured descriptive assessment methodology to identify variables affecting problem behavior. *Journal of Applied Behavior Analysis, 35,* 137-154.

Baum, W.M., & Aparicio, C.F. (1999). Optimality and concurrent variable-interval variable-ratio schedules. *Journal of the Experimental Analysis of Behavior, 71*, 75-89.

Borerro, J.C., Crisolo, S.S., Tu, Q., Rieland, W.A., Ross, N.A., Francisco, M.T., Yamamoto, K.Y. (2007). An application of the matching law to social dynamics. *Journal of Applied Behavioral Analysis, 40,* 589-601.

Borerro, J.C., & Vollmer, T.R. (2002). An application of the matching law to severe problem behavior. *Journal of Applied Behavior Analysis, 35,* 13-27.

Cataldo, M. F., Kahng, S., DeLeon, I. G., Martens, B. K., Friman, P. C., Cataldo, M., & Matthews, K. (in press). Behavioral principles, assessment, and therapy. In M. L. Batshaw, N. J., Roizen, & L. Pelligrino (Eds.), *Children with Disabilities* (7th[th] Edition). Baltimore, MD: Paul H. Brookes Publishing Co.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior, 4,* 267-272.

Martens, B.K., DiGennaro, F.D., Reed, D.D., Szczech, F.M., & Rosenthal, B.D. (2008). Contingency space analysis: An alternative method for identifying contingent relations from observational data. *Journal of Applied Behavior Analysis, 41,* 69-81.

Martens, B.K., Gertz, L.E., Werder, C.S., & Rymanowski, J.L. (2010). Agreement between descriptive and experimental analyses of behavior under naturalistic test conditions. *Journal of Behavioral Education, 19,* 205-221.

Prelec, Drazen. (1982). Matching, maximizing, and the hyperbolic reinforcement feedback function. *Psychological Review, 89,* 189-230.

St. Peters, C.C., Vollmer, T.R., Bourret, J.C., Borrero, C.S., & Sloman, K.N. (2005). On the role of attention in naturally occurring matching relations. *Journal of Applied Behavioral Analysis, 38,* 429-443.

Tables and Figures

Table 1

Matching Results

Participant	Condition	Number of Sessions	Bin Length (min)	Slope	Intercept	VAF
Jack	Experimental	9	5	0.82	0.10	87%
		13	2.5	0.68	-0.02	42%
		15	1	0.25	0.12	2%
		24	Condition Means	0.83	0.23	77%
	Descriptive	5	5	0.21	0.55	17%
		7	2.5	0.91	0.23	41%
		8	1	0.53	0.17	18%
		24	Condition Means	0.83	0.23	77%
Mark	Experimental	7	5	0.81	0.09	80%
		11	2.5	0.71	0.06	49%
		9	1	0.78	0	41%
		25	Condition Means	1.04	-0.04	92%
	Descriptive	5	5	0.41	0.37	9%
		6	2.5	1.37	-0.40	75%
		7	1	1.38	-0.23	29%
		25	Condition Means	1.04	-0.04	92%

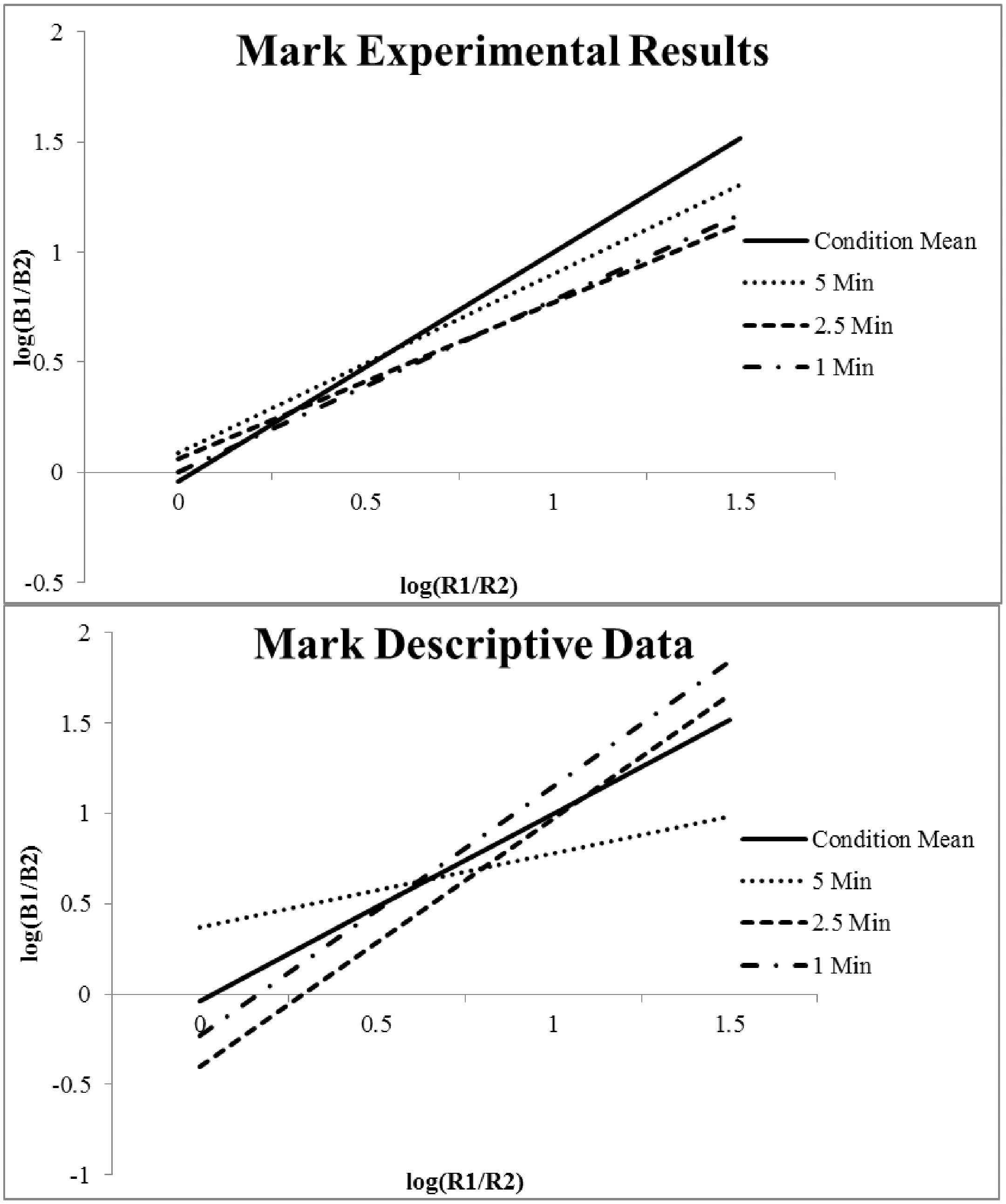
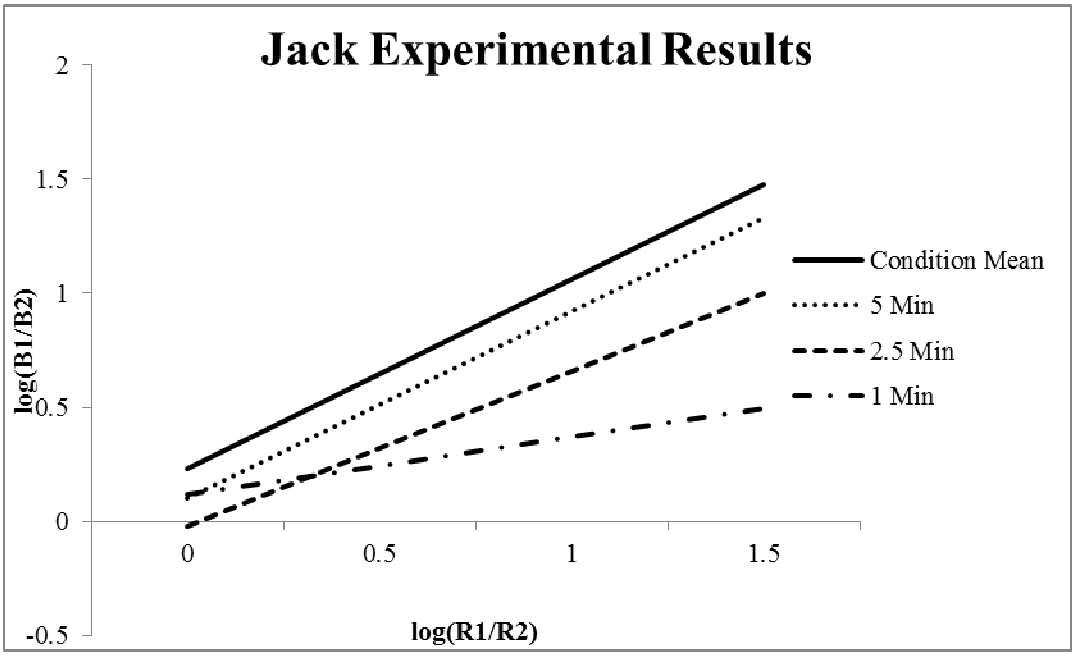


Figure 1. Linear graphs of Mark's experimental (top) and descriptive (bottom) data for each bin length and for the condition mean. Data were analyzed using the general matching equation (GME).



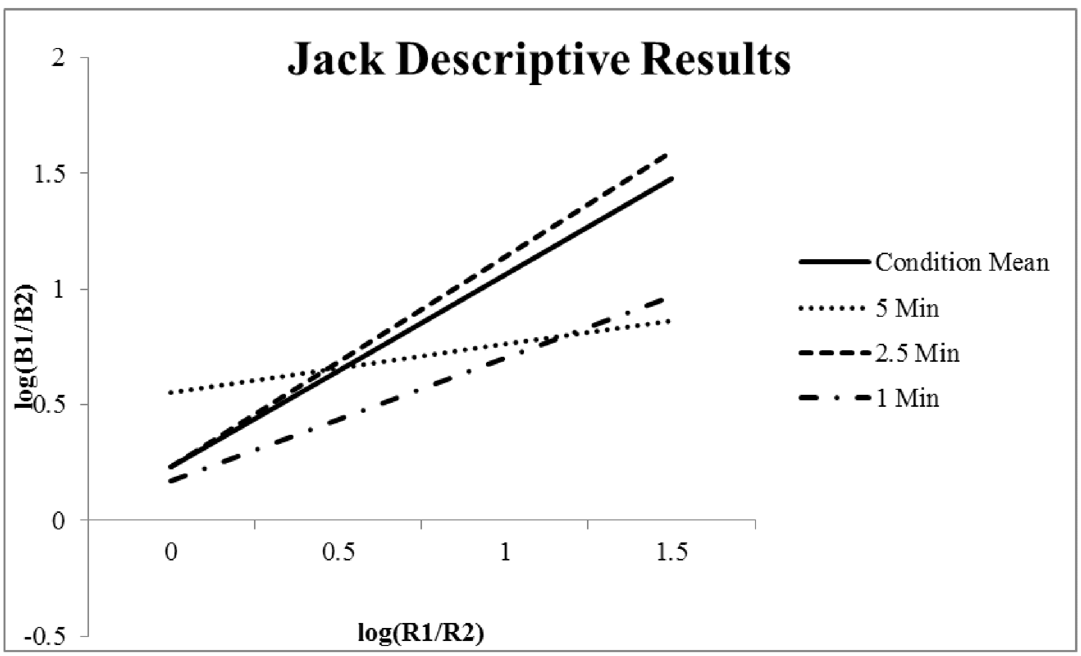


Figure 2. Linear graphs of Jack's experimental (top) and descriptive (bottom) data for each bin length and for the condition mean. Data were analyzed using the general matching equation (GME).

Summary of Capstone Project

The relationship between behavior and reinforcement is one that has been of great interest in the field of psychology. Reinforcers influence not only what behaviors are portrayed, but also for how long those behaviors persist, and how frequent those behaviors occur. Behaviors are influenced by a number of reinforcers, depending on setting, context, and the relationship between the individuals involved. In a classroom setting, a child's behavior may be influenced by reinforcers such as statements of praise (i.e. – "What a great drawing!"), statements of reprimand (i.e. – "We don't hit others in this classroom."), tangible rewards (i.e. – presenting achild with a sticker), or providing relief from a task (i.e. – allowing the child to stop practicing writing simply because he or she does not want to continue). Therefore, the relationship between behaviors and reinforcers has a significant impact on learning within a classroom. As a result, understanding this relationship, and how reinforcement may be used to encourage positive and on-task behavior a useful tool, not only in classrooms, but other settings as well.

In order to understand how a child's behavior is influenced by reinforcers in the classroom, one must first identify which adult behaviors are acting as reinforcers. One way to identify reinforcers is to observe the interactions that occur between a child and a teacher or caregiver in the natural classroom environment. More specifically, without engaging in or manipulating the interactions, one can observe and record the duration and frequency of the child's behaviors, as well as the duration and frequency of the teacher's responses to the child's behaviors. This observation of the natural setting provides an accurate account of the events surrounding the occurrence of certain child behaviors, therefore making it possible to identify potential reinforcers.

Once data is collected from the natural setting, shedding light on the teacher responses that are acting as reinforcers of behavior, it can be summarized according to the matching theory to explain how behavior was influenced by reinforcement. The matching theory states that for two alternative behaviors, relative rates of responding will match relative rates of reinforcement. For example, if the two alternatives being examined are time spent off-task and time spent on-task, and the reinforcer is teacher attention, according to the matching theory, as the relative rate of teacher attention increases, the relative rate of on-task behavior and off-task behavior will change at the same rate (i.e.-on-task behavior will increase at the same rate, and off-task behavior will decrease at the same rate).

In order to apply the matching theory to data collected regarding ehaviors and reinforcements, a linear equation known as the general matching equation (GME) is used. This equation creates a line displaying the relative rate of behavior, as they change in response to a change in the relative rate of reinforcement. Therefore, perfect matching would produce a line with a slope of 1 and an intercept of 0. In other words, in perfect matching, each unit change in reinforcement would result in a unit change of behavior.

Although the matching theory is a useful tool for evaluating how specific reinforcers may be influencing problem behaviors, there are also some problems in using the GME to explain this relationship. Three of the main problems are that (a) without manipulating the provision of reinforcers, one cannot be sure which adult responses are actually acting as reinforcers, (b) the most accurate method for separating the data prior to analyzing the data by using the GME has not been determined, and (c) individual differences in sensitivity to reinforcers may affect the matching results as provided by use of the GME. Therefore, this Capstone project had three goals, each of which attempted to address each of these problems. These goals were (a) to determine the most accurate method of separating the data into bins prior to conducting matching analyses, (b) to compare the results of matching from observational data to experimental data produced by manipulating experimenter attention, and (c) to examine individual differences in children's sensitivity across different bins of time, which may then impact the accuracy of the aggregated data in terms of matching. In order to achieve these goals, predefined schedules of adult attention for both on-task and off-task child behavior were manipulated, and experimental data were collected. The data were then broken into bins three times (first, into 1 min bins, then 2.5 min bins, and lastly, 5 min bins). At each level of separation, the data were then analyzed according to matching theory by using the GME. This method was used to determine which bin length most accurately reflected matching, to determine how accurately matching defines the relationship between reinforcers and behavior in the natural environment by comparing the observational and experimental data, and to explore the role of individual sensitivity to reinforcers.

Results from this study showed that breaking the data into bins of 2.5 mins in length most accurately reflected matching in the natural environment. One possible explanation for this finding lies in the fact that data were collected in sessions of approximately 5 min. Therefore, at the 2.5 minute level, the data provided information on approximately half of each session. It may take half of a session for the child's behavior to become sensitive to adult attention, or for the experimenter (or teacher) to settle into a routine. In that case, the 2.5 min data would capture the portion of interaction between the child and adult that most closely approximated matching.

The results also showed that the GME provided a poorer description of the descriptive assessment data than the experimental data. This suggests that descriptive data collected in the natural environment may be more variable than data collected under conditions when reinforcers are manipulated. This variability may exist for a variety of reasons. First, there were several differences between the experimenter that conducted the experimental sessions and the teachers who were observed during the descriptive assessment. Individual differences in the how attention was delivered may have caused discrepancies that were not reflective of the child's sensitivity to reinforcement.

Lastly, the results made it clear that individual differences in matching existed between the two children, providing support for the generality of the GME as a description of choice behavior. Use of the GME was able to show that one child's behavior approximated matching more closely than another child's. However, this could also be a reflection, in part, of the differences in teaching styles, as these children were observed with different teachers.

This study provides important insight into the use of matching theory to understand the relationship between adult reinforcement and child behavior. Furthermore, this study provides insight to ideas for future research. One possible direction for future research is to replicate the study with more participants and more schedule variations. Another possible direction for future research is to replicate this study with a group of typically developing children and a group of children with developmental or behavior disabilities (e.g., Autism Spectrum Disorders). This direction could shed light on whether or not children's sensitivity to reinforcement could be used as a tool in the classroom to more effectively engage students with disabilities. For example, children with specific disabilities may show lower sensitivity to reinforcement. This information could possibly assist in the diagnosis of disabilities, as well as implementing more effective reinforcement-based programs in inclusive and special education settings. These more effective programs may include more individualized and specific reinforcement schedules within the classroom (i.e., more explicit and/or frequent reinforcement for specific students). A third possible direction for future research would be to conduct a longitudinal study, in which a matching analysis is conducted at the pre-school age and then again in the primary grades. This may shed light on how children's sensitivity to reinforcement changes as they get older and how schedules of reinforcement may be adjusted in order to maximize on-task and positive behaviors. With further research, matching theory may become not only a more useful tool for research in the field of behavior analysis, but also a tool for diagnosis and treatment of children showing problem behaviors.