AN EVALUATION OF DPPXYR PARAMETERS ON MULTIELEMENT DESIGN FUNCTIONAL ANALYSIS GRAPHS

A Thesis Submitted to the

Temple University

Graduate Board

In Partial Fulfillment

of the Requirements for the Degree

MASTER OF SCIENCE IN

EDUCATION

By

Kasey Prime

August 2022

Thesis Approvals:

Saul Axelrod, Ph.D., BCBA-D, Art Dowdy, Ph.D., BCBA-D, Matt Tincani, Ph.D., BCBA-D, Jason Travers, Ph.D., BCBA-D – Department of Teaching and Learning

Philip Hineline, Ph.D. – Department of Psychology

Donald A. Hantula, Ph.D. – Department of Psychology and Neuroscience

Corey Peltier, Ph.D. – Department of Educational Psychology: University of Oklahoma

ABSTRACT

Single-case experimental design (SCED) allows for the identification of a functional relationship between a dependent and independent variable. For individuals who work as Board Certified Behavior Analysts (BCBAs), it is crucial and part of the ethics code that they can interpret SCED graphed outcomes to make data-based decisions. Visual analysis is the primary method to detect outcomes of SCED; however, research suggests graph characteristics may impact the reliability and estimation of intervention effects, specifically the ratio of the x-to-y axes. This study examined how the alteration of the data-points-per-x-to-y-ratio (DPPXYR) influences BCBA's ability to detect a function in functional analysis multielement graphs. A primary finding was the overall low agreement in BCBAs selecting function compared to the modified visual inspection (MVI). Descriptive results showed participants had highest accuracy in detecting the function when the DPPXYR was 0.09, although the difference across manipulation was minimal.

DEDICATION

This paper is dedicated to my parents, family, and friends. All of whom acted as my pillars of strength, guidance, and inspiration through this journey. Thank you to my parents; you both have been my guidepost, always offering your sound advice and encouragement.

ACKNOWLEDGEMENTS

I would like to thank Dr. Art Dowdy for his supervision, feedback, and direction throughout this entire thesis process. I am incredibly grateful that he took the time to help guide me. In addition, I would like to thank Dr. Corey Peltier for his guidance and feedback during graph and study creations. Finally, I would like to thank all of the Temple committee members who helped me get to the point I am today and guiding me thought these last few years at Temple.

TABLE OF CONTENTS

CHAPTER 1

REVIEW OF LITERATURE

Single-Case Experimental Design

Single-case experimental design (SCED) methodology includes rigorous experimental procedures that evaluate an intervention's effects on socially significant outcomes for individuals (Kazdin, 2020). The essence of SCED design (Ledford & Gast, 2018; SCED) allows researchers and clinicians to evaluate the impact of the intervention on a single participant by repeatedly measuring the dependent variable in a time-series fashion across baseline and intervention phases. SCED allows for the systematic adaption or modification of an intervention if little or no change occurs based on the data path properties (e.g., level, trend, variability) that are observed (Lane & Gast, 2018). A host of different graphic displays can be used to present data, such as cumulative recording, semi-logarithmic charts, bar graphs, and line graphs (Lane & Gast, 2014), with time-series graphs having evidence of having highest user acceptability (Kinney et al., 2020). A vital benefit of SCED is its ability to allow researchers and clinicians to develop the tool set to evaluate an individual's response in an applied setting in a flexible and systematic manner.

The effectiveness of an intervention in SCED is determined when a functional relationship is present; specifically, this refers to an independent variable's effect on a dependent variable. In most cases, experimental control is demonstrated when three demonstrations of the experimental effect occur at three different points in time with a single participant (within-subject replication) or across various participants (inter-subject replication; Horner et al., 2005). Randomization is another way to control for threats against internal validity and can be experimentally controlled (Jacobs, 2019). SCED allows for the empirical demonstration of a functional relationship and, through repeated replication of an effect (Hantula, 2019), SCED research may result in evidence-based practices in behavior analysis and related fields.

Randomized controlled trials (RCT) or group-comparison designs also allow for the investigation of effects on interventions using standardized and validated instruments with a greater number of participants compared to SCED (Vlaeyen et al., 2020). Depending upon the research question, RCT generally is restricted to only mean outcomes of participants who are masked of individualized characteristics (Onghena et al., 2017). This limiting property of RCT has historically motivated the use of SCED within the behavioral science community. Sidman (1960) stated, "it appears, then, that when different groups of subjects are used to obtain the points determining a functional relation, the mean curve does not provide the information necessary to make statements concerning the function for the individual" (p. 268).

Due to the lack of inherited individualization in RCT, SCED has played a critically important role in behavior science. The systematic and detailed analysis of individual responses to SCED is a sustainable option for researchers across disciplines, with over 45 professional journals reported to include SCED (Freeman & Sugai, 2013). SCEDs represent an innovation in clinical trial methodology by balancing individual needs with methodological rigor while providing an individualized approach to assessing treatment effects (Vlaeyen et al., 2020). Sidman (1960) discussed, "Whenever we are forced to use groups of subjects or large behavior samples from individual subjects in order to smooth the data, we are demonstrating a lack of experimental control over our subject matter" (pp. 16-17). With that said, SCED has allowed the field of behavior analysis and related fields to provide data-driven interventions for heterogenous populations of varied needs (Kazdin, 2020).

Visual Analysis of SCED

In SCED, visual analysis serves as the primary method to evaluate the effectiveness of interventions (Dowdy et al., 2022; Ledford et al., 2018; Tanious & Onghena, 2021). Properties of visually analyzing SCED data include analyzing both within phase and between phase comparisons of the level, trend, variability, consistency, overlap, and immediacy of data paths (Kratochwill et al., 2021). Level refers to the change in the number of behavior that occurs between baseline and intervention phases. Trend refers to the movement in data over time, with specific attention given to the direction of a data path within and between conditions. Accelerating, decelerating, or zero-celerating are commonly used trends along the ordinate scale. At the same time, variability refers to how data points are similar regarding value (ordinate scale). Consistency refers to how data patterns are the same within like conditions. Overlap refers to the extent to which data from one condition are at the same level as data from an adjacent condition. Finally, immediacy is how data change simultaneously with a condition change (Ledford et al., 2018).

Visual analysis is the practice of interpreting graphs by visually assessing different characteristics of the data to determine the effects. A visual analyst can draw a reasonable conclusion or make a plausible hypothesis about any relationship or the lack of a relationship (Parsonson & Baer, 1992). In SCED research, participants serve as their control when evaluating change. When graphed, a visual analyst may determine if an intervention was effective based on the comparison of two adjacent conditions, also known as the A-B comparison (Lane & Gast, 2014). When collecting baseline data, the visual analyst should make sure the data are stable before introducing the independent variable.

In SCED, data are collected and transferred to a graphing software (Kranak et al., 2019), and data are graphed to then be visually analyzed. Visual analysts tend to select the best graphic display to represent their data. Once graphed, the visual analyst assesses the effect using visual analysis and observes the data's trend, level, and stability across baseline and interventions phases. If additional conditions are added, the visual analysts should introduce them systematically with a phase change line (Lane & Gast, 2014). This allows visual analysts to determine what aspect of the intervention is most effective in changing the target behavior. Visual analysis is an individualized approach that allows for the systematic adaptation or modification of behaviors based on characteristics observed during assessment or intervention sessions (Manolov et al., 2014).

SCED and EBPs

Furthermore, SCED plays an integral role in determining evidence-based practices in behavior analysis and related fields. Evidence-based practices (EBP) are a professional decision-making model in which practitioners incorporate the best available evidence with client values, context, and clinical expertise to provide services for clients (Slocum et al., 2014). EBPs allow clinicians a framework of the best available evidence to work within their scope of practice on complex cases, settings and, more broadly, an effort to improve decision making in applied settings and improve outcomes. EBPs are a core value of applied behavior analysis, and the field is committed to utilizing evidence-based decision-making (Smith, 2013). Recognizing the importance and need of developing quality EBPs, requires rigorous standards and high exceptions to offer solutions to problems.

One way to detect the evidence base of an intervention is to evaluate the intervention using SCED and engage in replication studies across various contexts (e.g., settings, participant characteristics) and evaluate consistency and comparability across studies (Manolov et al., 2022; Smith, 2013). Visual analysis may serve an integral role when determining the evidence base of an intervention. To illustrate, suppose a collection of SCED studies on a single intervention are not conducted using high procedural fidelity, may not have robust outcomes, and/or graph dimensions that vary when conducting visual analysis. These studies might be more susceptible to a Type I or Type II error. A Type I error refers to the incorrect rejection of a true null hypothesis (a false positive), in other words saying an intervention was effective when in fact it is not.. In contrast, a Type II error means not rejecting the null hypothesis when it is actually false (a false negative), in other words determining an intervention is ineffective when in fact it was effective. Incorrect evaluation of data could result in a Type I or Type II error that, in turn, results in an incorrect classification of an evidence-based practice. For practitioners, this would result in an effective treatment selection for clients. The identification of EBPs must be grounded in rigorous methodology to reduce the likelihood of Type I and Type II errors. Overall, with varying information on graph ratio, it is important to understand these effects on visual analysis and how they can contribute to Type I and II errors before establishing and disseminating EBPs to clinicians.

Disagreements Surrounding Visual Analysis

Although visual analysis is most widely used, there is mixed research regarding the reliability of this primary method in detecting an effect in SCED (Ninci et al., 2015; Wolfe & McCammon, 2022). In applied behavior analysis, the BACB task list emphasizes data-based decision making; however, there are a limited number of items related to visual analysis, and those that do are relatively general (e.g., C-11: Interpret graphed data; D-5: Use single-subject experimental design; BACB 2017). Limited specificity of the task list may relate to possible inconsistencies of detecting a function by BCBAs when analyzing SCED graphs (Wolfe & McCammon, 2022). Despite the inconsistencies, visual analysis continues to be the most used method, perhaps due to its flexibility and relatively low technical demand. One potential decision-altering characteristic that might impact BCBAs when visually analyzing SCED graphs may involve graph construction, and more specifically ratios related to x- and y-axis lengths. Despite the potential reasons, evidence of inconsistent interpretation of outcomes of SCED has been shown across both researchers and clinicians.

Researchers

Wolfe et al. (2016) reviewed interrater agreement on the visual analysis of individual tiers and functional relations in multiple baseline designs. Fifty-two experts reviewed 31 multiple baseline design graphs to judge the changes in the dependent variable. The results showed that the interrater agreement was just at or below minimally adequate levels for both decisions. In addition, the agreement at the individual tier level often resulted in agreement about the overall functional relation. All 52 raters agreed on the functional relation decision for 11 of the 31 graphs (34%). The pairwise percent agreement for tier-level decisions ranged from 60% to 97%, with a mean of 83%. This study shows the inconsistency in visual analyst rating on determining a functional relation with SCED graphs.

While research has resulted in mixed findings when detecting the reliability in visual analysis among researchers (Wolf et al., 2016), advancements have been made that include a structured approach to help reduce the likelihood of incorrect detection of functional relationships. Dowdy et al. (2022) conducted a brief review of the literature between the years 2015 to 2020 in the *Journal of Applied Behavior Analysis* to assess the extent to which structured visual analysis was used to guide or supplement researchers' interpretations of SCED graphs.

Although recent structured visual analysis advancements have been developed to help address the inconsistent interpretations of SCED data, research shows that, to date, these methods appear to be rarely used by researchers to analyze SCED data.

Clinicians

Inconsistencies when interpreting SCED graphs also appear to impact clinicians' decision-making. Wolfe & McCammon (2022) examined the content and methods that professors use to teach visual and statistical analysis of SCED data in a verified course sequence. A total of 37 course instructors completed the survey. Findings showed that there was inconsistency across instructors in some fundamental aspects of data analysis and consistency in others. For example, the number of effects required for a functional relationship showed much variability across instructors. Specifically, 19% *(n=7)* of respondents selected "at least three," 16% *(n=6)* selected "at least five," and the majority of respondents *(65%; n = 24)* selected "it depends on the characteristics of the data." Despite the inconsistencies reported, almost all instructors emphasized visual analyses over statistical analyses. Wolfe and McCammon's (2022) findings highlight how visual analysis training inconsistencies predict disparate conclusions about functional relationships.

Rader et al. (2021) conducted a quantitative analysis of accuracy, reliability, and bias in judgments of functional analyses. A total of 121 doctoral level Board Certified Behavior Analysts (BCBA-D) experienced in visual analysis responded to ten FA graphs to examine their consistency in judgment. The researchers first looked at identifying the function and, secondly examined which action they would take next. On average, 108 attentive participants produced a modest *d*' of 1.59 analogous to 63% accuracy. In addition, when the data was slightly elevated, visual analysts were not able to detect the difference. Overall, the data suggested that reliability is at best modest for visual analysis, even among experts.

Lieberman et al. (2010) investigated participants' visual analysis of multiple baseline designs across participants graphs when change was delayed. A total of 36 expert reviewers from the field analyzed multiple baseline design graphs. The expert reviewers were asked to examine 16 graph sets and share their confidence that the graph demonstrated a functional relationship. Researchers found that graphs with steep slopes after treatment began were far more likely to be judged as showing a functional relation. The intraclass correlation coefficient for the steep versus shallow distinction was only 0.12, indicating 88% of the variance was due to factors other than graph sets (e.g., expert raters). The intraclass correlation coefficient for the distinction between each of the 16 graph sets was only 0.13. Overall, reviewers disagreed on which graphs showed a functional relationship.

Graph Construction Elements that Appear to Impact Visual Analysis

While research on visual analysis has illustrated inconsistencies in the reviewer's detection of a functional relationship, less research has been conducted on how graph construction may impact the visual analyst's evaluation of data. Dart and Radley (2018) shared that published SCED data may be collected with less experimental rigor and evaluated using visual analysis of a linear graph without quantitative effect sizes. This may be of concern as an emerging body of literature suggests that simple elements of the graphical display (e.g., ordinate axis scaling, ratio of X to Y-axis length [DPPXYR]) can impact judgments made by visual analysts. They suggested that a DPPXYR be between 0.14 and 0.16 to reduce Type I and Type II error rates. Preliminary evidence showed that altering the vertical axis scaling and the DPPXYR influences visual analysis impacting the Type I and Type II error rates and subsequently could affect the identification of EBPs.

Kubina et al. (2017) conducted a critical review of line graphs in applied behavior analytic journals. They examined 4,313 graphs from 11 journals as the sample for this study, in which researchers surveyed essential quality features of line graphs in behavior analytic journals. The quality features of a line graph convey representativeness and continuity of graphically displayed time-oriented data. Kubina et al. identified that a critical interpretation variable is the physical proportion of the vertical to the horizontal axis of a graph. They suggested a vertical to horizontal axis ratio ranging from 5:8 to 2:3, with a maximum of 3:4. The results of their search showed a high degree of deviation from standards of graph construction. Only 15% of the 4,313 reviewed graphs had a proportional construction ratio falling between 63 and 75%. In addition, 44% of vertical axes and 39% of horizontal axes had different physical lengths.

Dart and Radley (2017) evaluated the impact of ordinate scaling on the visual analysis of SCED data. Participants included 32 experts in visual analysis who were asked to examine graphs that were presented with different maximum ordinate scaling metrics (i.e. 100%, 80%, 60%, 40%). Eight ABAB design data sets were presented, and two graphs had null, small, moderate, and large effects. In addition, three iterations of each graph with only the ordinate scale manipulated were observed for a total of 32 graphs. The results showed that as the ordinate scale was truncated, visual analysts were more likely to overestimate the size of the treatment effect (i.e., 100% ordinate had 0% commission, 80% ordinate had 4.7% commission, 60% ordinate had 6.3% commission, and 40% ordinate had 21.9% commission).

Data Points Per X to Y Ratio (DPPXYR)

Although it has been suggested that graphs should have an x- to y- ratio between 5:8 to 2:3, with a maximum of 3:4 for proper visual analysis this may be an oversimplification of a more complex element of graph construction (Kubina et al., 2017). This recommendation emphasized the importance of considering the horizontal and vertical axis lengths when constructing time-series graphs. However, the density of the data points plotted along the x-axis may also interact with visual analysis (Ledford et al., 2019).

This focus on x-axis to y-axis ratio while also considered the density of the data points plotted along the x-axis led to the creation of the data points per x- to y-axis ratio (DPPXYR). Radley et al. (2018) investigated the effects of DPPXYR on visual analysts' evaluation of multiple-baseline design graphs. Participants included 29 school psychologists with experience in SCED and visual analysis. The participants reported an average of (a) 8.8 years of experience using visual analysis to evaluate the effects in SCED, (b) 27.8 hours of instruction in SCED, and (c) 21.2 hours of training in visual analysis as part of their graduate training. The study was conducted with an anonymous survey distributed using the Qualtrics platform via email. Participants were asked to estimate the magnitude of intervention effects for five different variation of the DPPXYR on eight graphs for a total of 40 SCED graphs. The results showed that participants scored effects as larger when DPPXYR was smaller and effects as smaller when the DPPXYR was larger. In sum, Radley’s recommended that a DPPXYR of 0.14 or 0.16 should be used for multiple baseline designs, to maintain in the interest of a more conservative effect size judgment.

More recently, empirical investigations have taken place to determine optimal graph ratios for visual analysis (Peltier et al., 2021b). Data points per x- to y- ratio (DPPXYR) may be a more comprehensive metric to determine the effects of the axis ratio on the data rather than a simple evaluation of the ratio. DPPXYR is calculated by dividing the x-axis length by the y-axis length and dividing that quotient by the number of data points that could fit along the x-axis.

Peltier et al. (2021a) examined the ordinate scaling and axis proportions of SCED graphs from 2010 to 2019. The study included 40 SCEDs, including 258 graphs (i.e., ABAB, alternating treatment design, or an AB as part of a multiple-baseline design) that were published in the last 10 years in *Behavioral Disorders* and *Journal of Emotional and Behavioral Disorders*. The results showed a large variation in the axis portions as measured using standardized x:y (i.e., x-axis length divided by y-axis height) and the DPPXYR. Across all SCED graphs, the mean value of DPPXYR was 0.12 (SD = 0.09), and the median was 0.10, with a range of 0.02-0.41. Peltier et al. identified 101 (62.3%) graphs included in multiple-baseline designs below the suggested DPPXYR metric (0.14). Furthermore, due to published graphs presented at a less-than-recommended DPPXYR, there may be an increased likelihood of Type I errors during visual analysis.

Similarly, Peltier et al. (2021b) conducted a decade review of two potential analysis altering variables in graph construction for individuals with autism spectrum disorder. They reviewed SCED graphs published from 2010 to 2020 in the four journals with a mission statement focused on individuals identified with autism (Journal of Autism and Developmental Disorders, Autism, Research in Autism Spectrum Disorders, and Focus on Autism and Other Developmental Disabilities). Researchers included 348 articles and 2675 graphs. The results showed large variation across and within types of SCEDs when evaluating the vertical and horizontal axis lengths using the x:y ratio and the DPPXYR, with few adhering to current recommendations. The mean DPPXYR for studies using AB design (0.25), withdrawal design

(0.08), alternating treatments design (0.13), multiple-baseline or multiple-probe design (0.11), and changing criterion design (0.06) all resulted in metrics below the suggested DPPXYR range. Peltier et al. (2021b) noted that additional empirical research was necessary to identify optimal DPPXYR boundaries on multielement/alternating treatment designs.

Multielement Design Functional Analysis Graphs

Functional Analysis (FA) is most often used to identify primary variables or the function of problem behavior to develop effective function-based treatment subsequently (Tincani et al., 2018). Accurate identification of a behavior function is essential for decreasing the problem behavior, by replacing the problem behavior with a functionally equivalent replacement behavior. When interventions are not function-based, behaviors may not change because the replacement behavior may not be functionally equivalent. By identifying contingencies that currently maintain problem behavior, BCBAs use functional analysis technology to alter relevant consequences, and associated discriminative stimuli and establishing operations to reduce problem behavior (Hanley et al., 2012).

Functional analysis was created to improve the assessment of problem behavior by first demonstrating control over problem behavior by the suspected maintaining environmental variables. During a functional analysis, the problem behavior is evoked, and contingent reinforcement is delivered; also, problem behavior is eliminated when those same reinforcers are provided noncontingently in the control condition (Jessel et al., 2020). These methods allow for the identification of variables that maintain problem behavior (i.e., "function" of problem behavior). These behaviors serve a specific purpose or function for an individual. The results of the FA can be used to develop interventions that target identified functions (Iwata et al. 1982/1994).

The standard FA was the first protocol developed to help determine the function of an individual's behavior. This procedure involves manipulating the antecedents and consequences to help identify the maintaining variables based on four functions of behavior (e.g., escape/demand, attention, tangible, automatic) (Iwata et al.,1982;1994). A standard FA involves testing the behaviors for four common functions of problem behavior. Each condition differs along with one or more of the following dimensions: (1) play materials (present vs. absent), (2) experimenter demands (high vs. low), and (3) social attention (absent vs. noncontingent vs. contingent) (Iwata et al., 1982;1994). Beavers et al. (2013) provided a description of common conditions in a standard functional analysis. During the attention condition, the client is provided social deprivation, and attention is delivered when the problem behavior is emitted. During the escape condition, the client is presented with an identified unpreferred activity; if the participant emits the target behavior, social negative reinforcement in the form of escape is provided. During the tangible condition, a preferred item is placed out of reach and if the problem behavior is emitted the item is provided to the individual. In an alone condition, the individual is in a room alone without any materials that may compete with problem behavior and the clinician collects data on the target behavior if it occurs. Finally, toys and activities are presented in the play condition, and the practitioner interacts socially with the participant. The play condition serves as a comparison or control condition to rule out confounding variables.

Since the inception of the standard functional analysis, other functional analysis formats (e.g., brief, latency-based, trial-based, IISCA) have been developed to improve analysis efficiency and control, with most formats retaining fundamental components of the standard and some omitting all (e.g., interview-informed, synthesized contingency analysis; IISCA; Jessel et al., 2016). Brief functional analyses (BFA; Northup et al. 1991) is an abbreviated assessment methodology derived from traditional extended functional analysis methods. BFAs are most used during situations in which time constraints in a school, clinic, or home are of concern. Similarly, the trial-based functional analysis (TBFA; Rispoli et al. 2018) is a method to identify the function of target behaviors that are often presented in a discrete trial format. The TBFA format is often used in schools because it allows for trials to be embedded into regularly scheduled classroom activities throughout the day (Lambert et al., 2012).

Another type of functional analysis format is latency-based functional analysis (Sunde et al., 2022). This format is a viable option when safety due to dangerous problem behavior is of concern. Latency-based FAs allow for experimental control while evoking a fraction of the problem behaviors commonly observed during traditional FAs (Lambert et al., 2017). Additionally, this format allows an analyst to examine the absence of treatment-specific establishing operations while discriminative stimulus evokes behavior soon after the initial presentation (Caruthers et al., 2015).

A final type of functional analysis format is named the interview-informed synthesized contingency analysis (IISCA; Jessel, 2021; Coffey et al., 2019). The IISCA begins with an open-ended interview followed by the presentation of synthesized antecedents and consequences that emulate the context of natural contingencies of problem behavior reported during the interview. Coffey et al. (2019) research found that IISCA produced effective treatment gains (i.e., 90% reduction in problem behavior) across all 102 IISCA applications in 17 studies. IISCA benefits are (a) outcomes result in improved analytic efficiency while maintaining demonstrations of control, (b) the format can be used to develop effective language and skill-based treatment among participants, and (c) it has been found to be highly acceptable among caregivers and teachers.

While the development of functional analysis has allowed for the identification of an individual's function of behavior, there remain questions about the methods used to visually analyze FAs. Roane et al. (2013) created a modified visual inspection (MVI) criterion based upon Hagopian et al. (1997) to support accurate function identification when assessing functional analysis outcomes. To apply the MVI criterion, an upper and lower criteria line is placed one standard deviation above and below the mean of all data points in the toy play condition. MVI allows visual analysts to assess data point percentages, trends, rates of behavior, and multiple maintaining variables, all in the perspective of set criterion lines. Although structured visual analysis advancements, such as MVI, have been developed to help address the inconsistent interpretations of SCED data, research showed that these methods are rarely used to analyze published SCED data in the *Journal of Applied Behavior Analysis* (Dowdy et al., 2022).

The functional analysis compares the data collected using SCED. The most commonly used SCED for functional analysis are multielement designs (Desrochers & Fallon, 2014). Multielement design is a SCED that consists of a rapid alternation of two or more conditions or treatments (e.g., independent variable) from session to session or within sessions. The conditions are generally alternated across days to reduce the confounding effects of order (e.g., Day 1: demand, alone, play, attention; Day 2: attention, demands, alone, play).

Multielement designs play a significant role in functional analysis, and practitioners use them to make function-based decisions for selecting interventions to decrease target behavior and increase alternative adaptive functionally equivalent behavior. Graph construction may result in differing evaluations of times-series data when engaging in visual analysis of graphed data (e.g., Dart & Radley, 2017). Radley et al. (2018) presented data suggesting the manipulation of DPPXYR in multiple baseline design graphs resulted in manipulated decisions of visual analysis in the magnitude of treatment effect. Thus, it is essential to evaluate if graph construction of multielement design graphs has the same effect on multielement designs due to the nature of them being to develop function-based interventions.

Purpose

Research has shown that visual analysis may result in the incorrect detection of a functional relation of SCED graphs based on several elements (e.g., graphical characteristics, clinician background, data patterns); however, less research has been conducted to identify how varied DPPXYR indices impact visual analysis. To the author's knowledge, no research has investigated the effects of manipulating x:y axis ratios, or the DPPXYR, on BCBAs interpretations of multielement functional analysis design SCED graphs. This study examines how varying DPPXYR metrics may influence BCBAs' detection of a function of behavior when presented using multielement design graphs. Similar to Radley et al. (2018), it was hypothesized that BCBAs would be more likely to identify a functional relation when data are presented on a graph with a smaller DPPXYR. Further, it was also hypothesized that visual analysts would be more likely to identify an effect as large when data are presented on a graph with a smaller DPPXYR on multielement graphs.

Research Question 1: What percent of BCBAs detect the correct function of behavior when presented graphs with different manipulated DPPXYR?

Research Question 2: How are SCED parameters (level, trend, variability, consistency, etc.) reported when the DPPXYR is manipulated?

Research Question 3: Are their common moderating variables that impact BCBA's interpretations of MED graphs (geography, experience, etc.)?

CHAPTER 2

METHOD

Participants

Participants were recruited by distributing a survey through the Behavior Analyst Certification Board, Inc. (BACB) to Board-Certified Behavior Analysts (BCBAs) with experience in SCED and visual analysis. The BACB is the certification board that provides practice requirements, an ethics code, and a disciplinary system designed to provide a base minimum requirement to be a practicing BCBA. The BCBA, BCaBA, and RBT certification programs require applicants to obtain a high standard of supervision and knowledge before passing the requirements. The BACB task list states that BCBAs should have experience with SCED and use the methodology to make data-based decisions (e.g., C-11: Interpret graphed data; D-5: Use single-subject experimental design; BACB 2017). For BCBAs', visual analysis is the primary method to analyze multielement design (MeD) graphs to determine a functional relationship or function of problem behavior which can help identify interventions to decrease a target behavior.

An email was sent to solicit BCBAs and BCBA-Ds, which requested their participation in a research study. Specifically, participants were informed that the intent was to observe the decisions that BCBAs make when evaluating SCED data, particularly functional analysis MeD graphs. The email included a link to a Qualtrics survey and stated that the survey would take between 20 to 30 minutes to complete if they should decide to participate. Other behavior-oriented related personnel, who generally do not make graph-based decisions and do not hold a BCBA, were excluded from this research. These certifications which were excluded from the study were Board Certified Assistant Behavior Analysts (BCaBAs), State Certified Behavior

Specialist, and Registered Behavior Technician (RBT). These applicants were excluded from the study due the less stringent level of expertise (cf. BCBA certification) required to practice.

The email was sent to 22,353 BCBA/BCBA-D candidates. A total of 9,348 candidates opened the email, and 571 people opened the study by clicking on the URL. Data collected showed that 126 individuals began the survey, and a total of 59 (47%) individuals completed all the survey elements. Of the sample who did not entirely complete the survey, 27 (21%) of these individuals did not finish completing the demographic questionnaire, and 40 (32%) of individuals discontinued after starting the visual analysis portion of the survey. No individuals were excluded from the survey due to a lack of meeting study qualifications. Out of the participants who completed the full study, the majority of the participants had their BCBA 46 (78%), in addition to 13 individuals (22%) who had their BCBA-D. The mean age of participants was 39 years old (range = 24-69). Forty-four participants were female (75%), with 14 (24%) male participants and 1 participant who preferred not to say. Forty-nine of participants identified as white (83%), with 3 (5%) identified as black, 2 identified as (3.4%) Latino/a, 2 (3.4%) identified as Asian, 1 (1.7%) identified as Native American, and 2 (3.4%) preferred not to say. Twenty-two of participants were from the Northeast (37.2%), in addition to 13 (22%) participants from the West Coast, 9 (15%) from the Midwest, 8 (13.5%) from the Southeast, and 7 (11.8%) from the Southwest.

Materials

All data were collected on an anonymous survey on the Qualtrics® platform. Qualtrics is a web-based software that allows users to create and distribute surveys. In addition, it allows researchers to generate reports without having any previous programming knowledge. This study was accessed by participants using a link that was disseminated via email through the BACB.

Visual Analysis Task

Once demographic information was collected, participants visually analyzed multielement functional analysis graphs. Participants were asked to analyze 32 SCED graphs that were presented using a multielement design that represented a functional analysis. Thirty-two graphs were selected for this study based on Radley et al. 2018 research on the effect of DPPXYR on multiple baseline design graphs. For each graph that was presented, participants were asked to identify one, multiple, or no function of the behavior in the MeD graph. One graph was shown per page in the survey. The conditions that were included on each MeD graph were play, attention, demand, tangible, and alone.

A primary purpose of this study was to examine how the alteration of the DPPXYR (Radley et al., 2018) impacts BCBAs' interpretation of detecting a function of behavior for 32 graphs. Participants were shown four variations of eight graphs (0.06, 0.09, 0.13, & 0.14). For all eight graphs, only the DPPXYR was altered, and the level, trend, and variability of the data paths remained constant. Pseudo functional analysis data were generated, and each graph consisted of a distinct DPPXYR metric created explicitly for this study. Appendix A shows a screenshot of the graphs that were used in the Qualtrics survey.

Graph Creation

Prior to creating graphs, the researcher, with the support of her academic advisor, reviewed Radley et al. (2018) and selected a graph format that hadn't yet been empirically evaluated. Radley et al. reviewed 105 studies and 295 SCED graphs for the *Journal of Applied School Psychology, Journal of School Psychology, Psychology in Schools, School Psychology Quarterly, and School Psychology Review* between 2010 and 2015. The second most used SCED graph was alternating treatment designs with 60 graphs, after multiple baseline designs with 153 graphs. In applied behavior analysis, a functional analysis of problem behavior is often depicted in a multielement format (Ledford & Gast, 2018). Thus, the multielement design graph was selected for this project.

As alternating treatment design (ATD) was the second most frequently published design the next logical step after Radley et al. (2018) research on manipulating the DPPXYR of multiple baseline designs was to examine if ATD graphs had the same results. Graphs were constructed to ensure they did not violate one of the What Works Clearinghouse criteria (2020) to eliminate potential concerns regarding design that may impact a rater's ability to evaluate the graphs besides manipulated graph construction. As such, each graph included 4 data points per condition (i.e., the WWC specifies a minimum of 3 per condition for alternating treatments designs). GraphPad Prism version 9.3.1 was used to create graphs based on the simulated data. The multielement graphs included 20 sessions, in which play, tangible, demand, attention, and alone conditions were presented in a quasi-random order across sessions.

After the eight multielement graphs were created, the DPPXYR of each panel was scaled to the average DPPXYR of multielement design graphs based on Peltier et al. (2021). Eight graphs were chosen to stay consistent with Radley et al.'s (2018) research on the effects of DPPXYR. Peltier et al. (2021) reviewed a total of 425 graphs that included a multielement design and found the mean DPPXYR was 0.13 (SD = 0.18). DPPXYR was calculated by using the screenshot tool on Mac (i.e., Shift-Command-4). This tool reported the height and width of a selected area in the number of pixels (see Appendix B). To measure the axes, the measurement tool was selected and the extreme upper point of the y-axis was clicked. Next, the researcher dragged the tool to the extreme right point of the x-axis, which yielded a height and width measurement for the panel. As the length of axes was affected by the display size of the graph, no descriptive statistics were calculated for the average length of axes. Instead, units of measurement were standardized by dividing the measured length of the x-axis by the measured length of the y-axis. Finally, the number of possible data points for each panel was determined by consulting the maximum value of the corresponding x-axis. The value of the x-axis length divided by the y-axis length was then divided by the number of possible data points to yield the DPPXYR. Below is the formula for computing the DPPXYR.

$$\frac{[x - axis\ length/\ y - axis\ length]}{\#\ of\ possible\ data\ points} =\ DPPXYR$$

Modified Visual Inspection

A structured approach named modified visual inspection (MVI; Roane et al., 2013) was used to detect the function of multielement design graphs. To detect a function MVI criterion requires placing an upper and lower criterion line one standard deviation above and below the mean of all points in the play condition. The guide allows visual analysts to assess data point percentages, trends, rates of behavior, and multiple maintaining variables, all in perspective of the criteria lines.

Interrater Reliability

Interrater reliability (IRR) of MVI was collected on 50% of the graphs developed using the Roane et al. (2013) criteria. The IRR observer held her BCBA and had 11 years of experience developing and analyzing SCED graphs which included MeD functional analysis graphs. To begin computing MVI, the BCBA read the criteria to calculate the Roane et al. (2013) criteria that aimed to detect a function of the behavior. The researcher modeled computing the MVI outload on a separate graph and then provided a different MeD graph for the observer to compute MVI alone. When IRR obtained 100% accuracy during training, they then analyzed a randomly selected sample of 50% of the MeD graphs included in the Qualtrics survey.

The researcher calculated IRR for the agreement of functions based on the MVI criteria. An agreement was scored if both observers recorded the same function, functions, or lack of function for each graph. Next, the graphs that resulted in agreements were divided by the total number of graphs and multiplied by 100 ([# of interval of 100% agreement]/ [total # of intervals] × 100). The IRR for the MVI criteria was 100%.

Next, two professors reviewed all eight graphs to verify the functions based on the MVI criteria. Both professors held a doctoral degree in either behavior analysis/special education or educational psychology, had a minimum of 10 years of experience in interpreting SCED graphs, and had a minimum of 15 peer-reviewed publications that involved SCED. Both professors completed the finished study and provided feedback before the study was sent to the BACB.

DPPXYR Manipulation

After identifying a function of the included MeD graphs using MVI, the data were replotted on three additional yet exact graphs that only varied by DPPXYR. Existing graphs were plotted using the DPPXYR of 0.06, 0.09, and 0.14. These DPPXYR estimates were identified by computing the first quartile, median, and third quartile of the alternating treatments design/multielement design graphs included and analyzed in Peltier et al. (2021). The minimum (0.001) and maximum (2.26) DPPXYR identified in Peltier et al. (2021b) where not included due to the difficulty in the legibility of the graphs when manipulated to a minimum and maximum estimate. During manipulation of the DPPXYR, only the x-axis length was adjusted, with the y-axis length held constant across all variations of DPPXYR. This procedure resulted in four variations of each of the eight graphs, yielding a total of 32 graphs. Similarly, Radley et al. (2018) included 40 graphs, however, due to the legibility of the graphs, only 32 graphs were included in this study.

Procedure

Before data collection, all study procedures were approved by the affiliated university's Institutional Review Board #28829. Hyperlinks were distributed through the BACB listserv. Individuals who followed the link to participate were initially presented with a consent document explaining the purpose of the study. Potential participants were informed that the study was intended to better understand how BCBAs detect functional relationships. However, due to the purpose of the study, the consent document did not mention manipulation of the DPPXYR. Once participants agreed to consent to the study, they proceeded to the demographic questionnaire. The demographic questionnaire asked questions pertaining to the participants' experience, monthly use of SCED, and education level. After completing the demographic section, participants were given instructions for the visual analysis task. All graphs were displayed with fixed heights and widths of graphs that varied because of differences in x-axis length.

For each of the 32 graphs, participants were asked to identify all of the functions presented, or none, if no function was present. After answering all questions for 32 graphs, the visual analysis task ended, and participants were thanked for their participation. The average survey completion time was 108 minutes (range = 4.2-4175 min.). It was possible for participants to start and stop the survey at their convenience and some participants completed if over the course of several days. Excluding two participants who required more than 22 hrs., the average completion time was 14.4 minutes (4.2-85 min.).

CHAPTER 3

RESULTS

Research Question 1

Research question one evaluated the effects of a BCBA detecting the function of behavior when the DPPXYR was manipulated. *Table 1* shows participants' percent correct in selecting the correct function based on the MVI criteria for determining the function of the behavior. The data shows that participants were slightly more likely to select the correct function when the DPPXYR was at 0.09 for four out of the eight graphs. Graph 1 showed that the DPPXYR of 0.13 was the most accurate with, 98.3% of participants selecting the correct function based on the MVI criteria. For graphs 6, 7, and 8, all participants did not select the correct function that the MVI criteria did not consider a function for the behavior.

Table 1
Percent Correct for Participants Based Upon Roane's Criteria

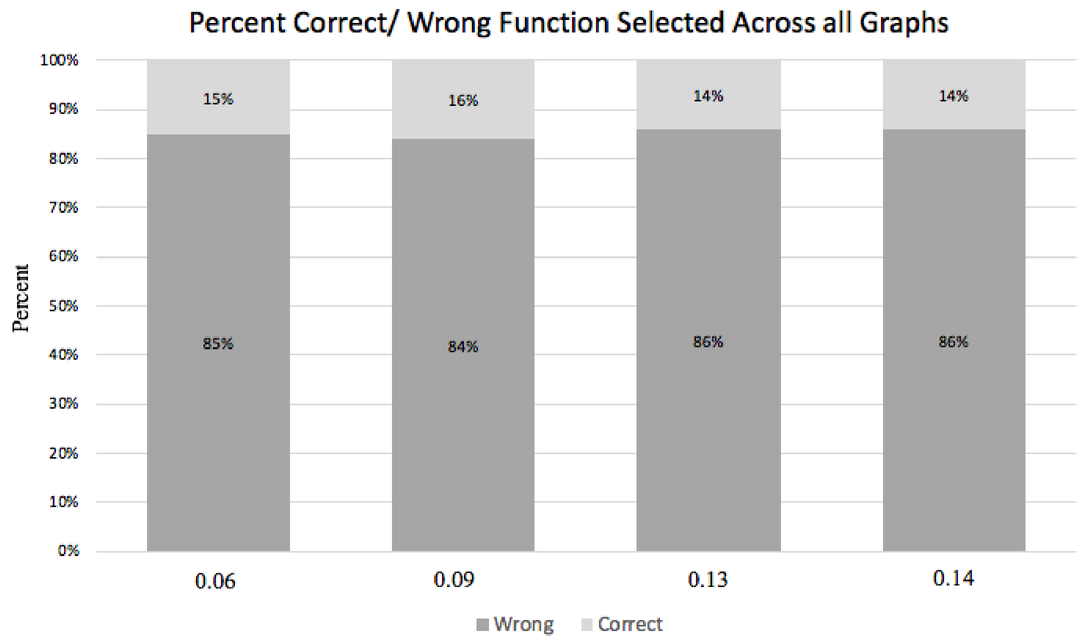
DPPXYR	**Graph 1**	**Graph 2**	**Graph 3**	**Graph 4**	**Graph 5**	**Graph 6**	**Graph 7**	**Graph 8**
0.06	96.6%	0%	1.7%	22%	0%	0%	0%	0%
0.09	93.2%	3.4%	3.4%	25.4%	1.7%	0%	0%	0%
0.13	98.3%	0%	1.7%	15.3%	0%	0%	0%	0%
0.14	94.9%	1.7%	1.7%	15.3%	0%	0%	0%	0%

Note: DPPXYR = data points per XY ratio

Figure 1 shows participants' percent of correct and incorrect function selection based on the MVI criteria across all graphs for each DPPXYR. Data show that 80% of participants selected the incorrect function based upon MVI when outcomes were combined across DPPXYR indices. While the DPPXYR of 0.09 had the highest percentage (16%) of participants selecting the correct function across all eight graphs, graphs 2, 5, and 7, all had 70% or more agreement on a different function of the graph that did not agree with the MVI criteria.

Figure 1

Percent of correct and wrong function selection from participants across all manipulated DPPXYR and graphs



Research Question 2

Research question two examined how SCED parameters (level, trend, variability, consistency, etc.) were reported when the DPPXYR was manipulated. The MVI criteria for interpreting functional relations, included all of these SCED parameters to help determine the function of the behavior. Level refers to the number of behaviors that occurs, as indicated by the ordinate scale value (Ledford & Gast, 2018). The MVI criteria examined the number of data points above and below the upper and lower criterion line. Namely, the level of behavior that determines if the condition is differentiated when 50% or more of the data point fall above the test conditions' upper criteria line. Graph 1 (see Appendix A) shows how participants assessed the level of the data path that occurred at a higher rate compared to all other conditions tested.

Trend. Trend is the slope and direction of a data series, or the direction data are moving over time (increase, decreasing, or remaining the same). In the MVI criteria, lines of each conditions are used to assess if the condition is on a downward or upward trend. Depending on the trend line, this can help a visual analyst determine if a condition is the function of the graph. In Graph 6 (see Appendix A), several participants selected attention or an automatic function without considering the trend lines of the graph. The attention condition presented a consistent downward trend, while the tangible condition's trend was on a consistent upward trend, with 50% of the points falling above the upper criteria line in the last half of the assessment. Trend may have been overlooked by some participants given that Graph 6's function was a tangible function due to the trend of data.

Variability. Variability is a fluctuation of one data point to the next; in data with no trend (i.e. zero celerating), variability can be summarized as the range of data values within a condition or as the percentage of data points falling within a given stability envelope (Ledford & Gast, 2018). In graph 2 (see Appendix A), many participants selected tangible as the function of the behavior; however, it appeared that the variability of the data might have been overlooked when detecting a function. Specifically, participants appeared to overlook the consistent upward trend of the demand condition. On the other hand, stability is the predictability and consistency of data values within a condition or lack of fluctuations in adjacent data points (i.e., lack of variability) (Ledford & Gast, 2018). In graph 3 (see Appendix A), the demand condition was stable and consistent during every test condition; however, it appeared that some participants examined the inconsistencies of the other functions and chose tangible/automatic or no function as the function.

Immediacy of change across adjacent conditions is the degree to which behavior change occurs as soon as the intervention is introduced. When a large change in level occurs immediately after the introduction of a new condition, it is referred to as an abrupt change in level, which show visual analyst the immediate effect on the intervention (Ledford & Gast, 2018). In Graph 7 (see Appendix A), the attention condition immediately presents higher rates of behavior when the condition is tested and stays consistent. **Consistency** refers to the extent to which data patterns in one condition are similar to data patterns in other conditions (Ledford & Gast, 2018). For the attention condition, the rate of behavior stays at the same level each time the condition is tested.

Overlap refers to the value of data in one condition that is in the same range of value of data in the subsequent, adjacent condition (Ledford & Gast, 2018). In Graph 5 (see Appendix A), both the tangible and demand conditions stay at the same range throughout the test conditions. In addition, they show the same magnitude of change when each test condition is trialed. **Magnitude** of the effect is assessed by comparing the amount and consistency of change across conditions and cases within a study that is directly attributed to the intervention (Ledford & Gast, 2018). The behavior immediately increased when both the tangible and demand conditions were being tested.

Research Question 3

Descriptive information was gathered from participants regarding demographics, experience with SCED, and visual analysis. Most participants were from the Northeast (*n* = 22: 37.2%), in addition the West Coast (*n* = 13, 22%), Midwest (*n* = 9, 15%), Southeast (*n* = 8, 13.5%), and Southwest (*n* = 7, 11.8%). Participants reported an average of 11 years (range =1-45) of experience in SCED and an average of two courses (range = 0-15) of training in visual analysis as part of their graduate training. Participants reported an average of 12 hours per month (range = 0-150), visually analyzing graphs to detect intervention and treatment effects using SCED data. Additionally, participants reported nine years (range = 0-45) of experience interpreting FA graphs presented in a multielement design format based upon Iwata et al. (1982; 1994). Participants also reported an average of 2 publications (range = 0-42) that incorporated SCED.

Figure 2

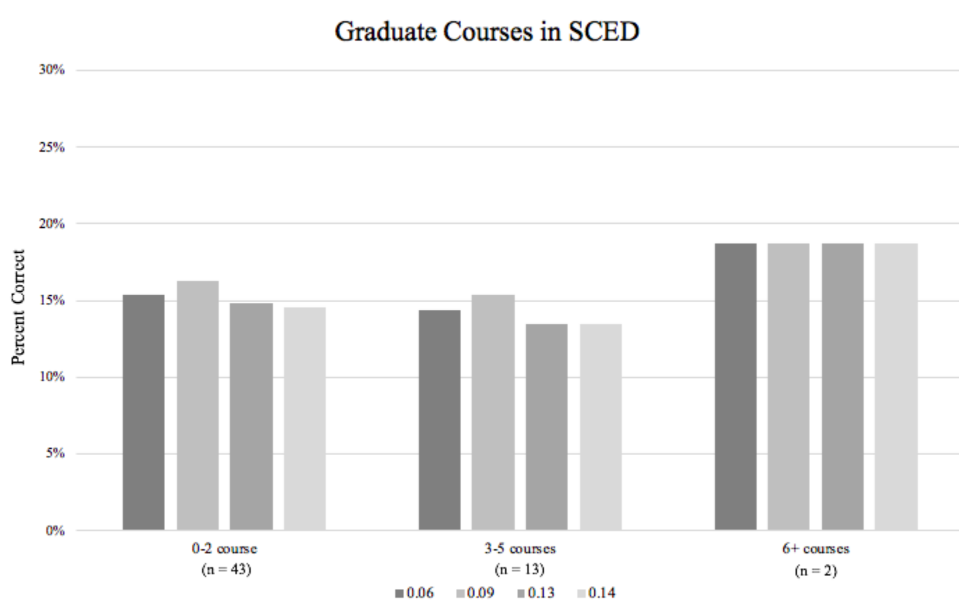
Years of experience with visual analysis and the percentage of selecting the correct function

Note. *N = the number of participants in each field*

Research question three asked what common moderating variables might impact BCBA's interpretations of MeD graphs (i.e., geography, experience). Figure 2 examined the relationship between the number of years that a participant had used visual analysis to detect effects of SCED graphs and how experience interacted with their accuracy of selecting the correct function of each graph based upon the MVI criteria. Participants reported an average of 11 years (range =1-45) of experience using SCED. With 19 participants having 0 to 5 years' experience, 21 participants having 6 to 10 years, nine participants having 11 to 15, two participants having 16 to 20, and eight participants having over 20 years' experience. Figure 1 shows, participants with 0 to 5 years' experience (*n = 19*) selected the 0.09 DPPXYR metric most often. Similar outcomes occurred with participants that had 16 to 20 years' experience (*n* = 2), the 0.09 DPPXYR metric was most often selected when determining the correct function. However, participants with 6 to 10 years' experience (*n* = 21), performed at the same rate when the DPPXYR was 0.06, 0.09, and 0.13. While participants with 11 to 15 years' experience (*n* = 9) and participants with over 20 years' experience (*n* = 8) both selected the correct function aligned with the MVI criteria when the DPPXYR was 0.06.

Figure 3

Number of graduate courses in single case experimental design and the percentage of selecting the correct function

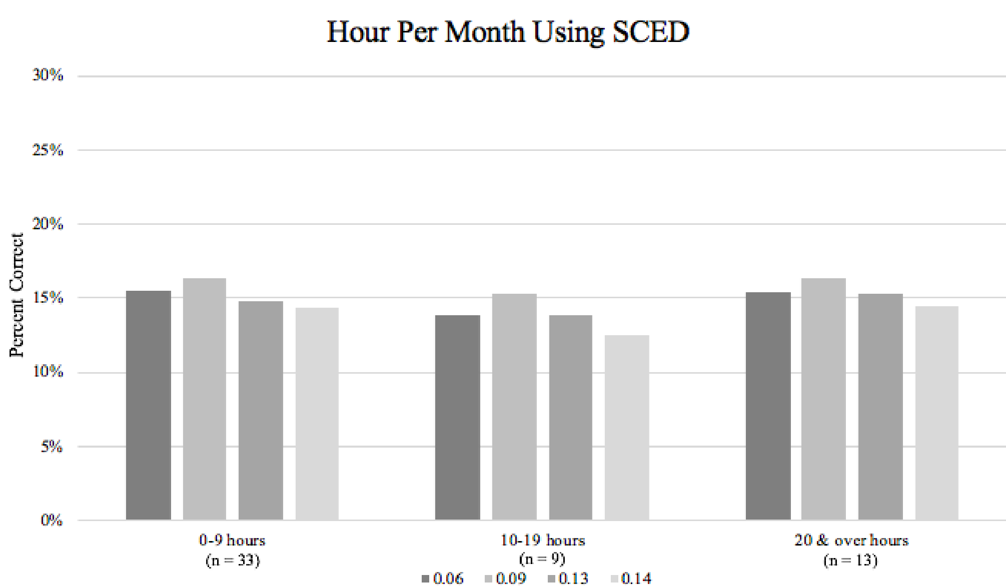


Note. *N = the number of participants in each field*

Figure 3 shows the number of graduate courses in SCED that were completed and how training interacts with the correct function selection based upon Roane's criteria. Participants had an average of two courses (range = 0 - 15) of training in visual analysis as part of their graduate training. Participants who completed 0-2 graduate courses (*n* = 43) in SCED selected the 0.09 DPPXYR metric most often. Similarly, participants who completed 3-5 graduate courses (*n* = 13) also selected the 0.09 DPPXYR metric most often. For individuals who completed 6 or more graduate courses (*n* = 2) in SCED, they appeared to have high consistent detection of the function for all four DPPXYR graph manipulations. In addition, participants who completed 6 or more graduate courses preformed with a higher percentage of accuracy across all DPPXYR manipulated graphs.

Figure 4

Number of hours per month a participant uses visual analysis to interpret multi-element design graphs and the percentage of selecting the correct function



Note. *N = the number of participants in each field*

Figure 4 examines the interaction between the number of hours per month a participant used visual analysis to interpret MeD graphs and their accuracy with selecting the correct function of each graph based upon the MVI criteria. Mean visual analysis of interpreting MeD graphs for participants was 12 hours per month (range = 0-150). For each hour category, participants preformed best when the DPPXYR was presented as 0.09. Participants who had 20 hours and over using visual analysis per month to interpret MeD graphs preformed at a slightly higher rate for all DPPXYR manipulated graphs.

CHAPTER 4

DISCUSSION

While our research showed that accuracy was only slightly higher, the results generally support that participants selected the correct function of behavior based upon the MVI criteria as the x-axis was shortened to become closer in length to the y-axis. Participants selected the correct function based on the MVI criteria most frequently when the DPPXYR was presented as 0.09. In addition, when the DPPXYR was presented as 0.06 participants had the second highest frequency of selecting the correct function based upon the MVI criteria. While our differences were small, similar to other research, Radley et al. (2018), found that when the DPPXYR became smaller as more participants' confidence in the presence of a functional relation and their judgement of effect size magnitude increased when the x-axis was shortened to become closer in length to the y-axis. These results supported that the DPPXYR can impact visual analysts' judgment of determining a function of behavior as well as the treatment effects.

Data indicated that the more graduate courses in SCED the participant had taken the more accurately they were at determining the function of behavior based on the Roane criteria across all manipulated DPPXYR graphs. Similar, Lane et al. (2019) showed that multicomponent visual analysis training resulted in participants increasing their accuracy of visually analyzing SCED graphs. This data suggests that with increased training in SCED, participants may be able to more accurately detect the function of a behavior when functional analyses are presented in a multielement design format.

Rader et al. (2021) found that visual analysts' accuracy, reliability and bias in judgement of functional analyses were questionable and further exploration of decision aids is warranted. These findings supported that the participants' accuracy of determining the function of behavior varied. The number of hours the participant utilized in SCED per month, as well as, the years of experience they had did not increase the accuracy of determining the function. Therefore, using visual analysis to determine the function of a graph is still variable across BCBA/BCBA-D certifications (Dowdy et al., 2020).

Implications

These results hold several implications for practitioners and researchers. First, both should be aware of the impact that the DPPXYR has on visual analysis and single-case data. Visual analysis may serve an integral role when determining the evidence base of an intervention. Visual analysts engaging in identifying in the function of behavior should ensure that the effects of graphical display characteristics like DPPXYR are limited during graph construction. They should also be aware that implementing these characteristics could help minimize the likelihood of their data being misleading. Applying these graphical display characteristics like DPPXYR could help reduce the threats of Type I and Type II errors.

Furthermore, additional researchers should attempt to investigate whether DPPXYR manipulation impacts visual analysis decisions across all graph types. Data may show that certain designs are more robust against variations in certain graphical display characteristics. Radley et al. (2018) suggests multiple baseline designs have a DPPXYR of 0.14 or larger in the interest of more conservative effect size judgement. On the other hand, these findings conservatively suggest that MeD graphs which have a DPPXYR of 0.09 appear to result in the most accurate selection the correct function across BCBAs and BCBA-D certifications.

Limitations

The results in this study can be contextualized in light of several limitations. First, although unmanipulated DPPXYR data were acquired from Peltier et al. (2021) research, and given the recency of this work there was no additional data to support its use as the comparison value.

Secondly, the same size for this research was relatively small. The survey was distributed to 22,353 BCBA/BCBA-D candidates. A total of 9,348 candidates opened the email, and 571 people opened the study by clicking on the URL. Data collected showed 126 individuals began the survey, a total of 59 individuals completed all elements of the survey. Overall, the number of completed responds was relatively low for the number of candidates that it was sent to.

Finally, the eight MeD graphs were not directly visually analyzed and utilized the MVI criteria as the basis of determining the function of behavior as being correct or incorrect. Cox et al. (2021) found that accuracy was marked higher when participants used the MVI strategies compared to traditional visual-inspection strategies, while they observed more modest increases in reliability coefficients. However, if the MVI was incorrect in detecting a function then the participants correspondence could be incorrect.

Conclusion

Characteristics of graphic display of SCED data have received minimal attention and very little empirical investigation for years, however these findings show that it appears to be a critical variable to take into consideration when analyzing graphs, particular MeD functional analysis graphs. Outcomes of the DPPXYR metrics that were analyzed in this study supported that different DPPXYRs affected BCBAs BCBA-Ds determinations when asked to identify the function of behavior in multielement design graphs and should warrant consideration in graph construction.

REFERENCES

Beavers, G. A., Iwata, B. A., & Lerman, D. C. (2013). Thirty years of research on the functional analysis of problem behavior. *J Appl Behav Anal. 2013 Spring*; 46(1):1-21. https://doi.org/10.1002/jaba.30

Caruthers, C.E., Lambert, J.M., Chazin, K.M. *Harbin, E.R., & Houchins-Juarez, N.J. (2015).* Latency-based FA as baseline for subsequent treatment evaluation. *Behav Analysis Practice* 8, 48–51 (2015). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5048249/

Coffey, A. L., Shawler, L. A., Jessel, J., Nye, M. L., Bain, T. A., & Dorsey, M. F. (2019). Interview-informed synthesized contingency analysis (IISCA): Novel interpretations and future directions. *Behavior analysis in practice*, *13*(1), 217–225. https://doi.org/10.1007/s40617-019-00348-3

Cox, A. D., Zonneveld, K. L. M., & Tardi, L. D. (2021). Further evaluating interobserver reliability and accuracy with and without structured visual-inspection criteria. *Behavioral Interventions*, *36*(3), 652–666. https://doi.org/10.1002/bin.1793

Dart, E. H., & Radley, K. C. (2017). The impact of ordinate scaling on the visual analysis of single-case data. Journal of School Psychology. 105-118. https://doi.org/10.1016/j.jsp.2017.03.008

Dart, E. H., & Radley, K. C. (2018). Toward a standard assembly of linear graphs. *School Psychology Quarterly*, *33*(3), 350–355. https://doi.org/10.1037/spq0000269

Desrochers, M. & Fallon, M. (2014). Instruction in functional assessment. *Open SUNY Textbooks*. 978-0-9897226-7-4

Dowdy, A., Jessel, J., Saini, V., & Peltier, C. (2022). Structured visual analysis of single-case experimental design data: Developments and technological advancements. *Journal of Applied Behavioral Analysis*. 55(2):451-462. https://pubmed.ncbi.nlm.nih.gov/34962646/

Dowdy, A., Tincani, M. & Fisher, A.G. (2020). Interrater Reliability and Convergent Validity of the Trial-Based Functional Analysis. *Adv Neurodev Disord* **5,** 61–70 (2021). https://doi.org/10.1007/s41252-020-00174-7

Freeman, J., & Sugai, G. (2013). Identifying Evidence-Based Special Education Interventions from Single-Subject Research. *TEACHING Exceptional Children*, *45*(5), 6–12. https://doi.org/10.1177/004005991304500501

Hanley, G. P. (2012). Functional assessment of problem behavior: Dispelling myths, overcoming implementation obstacles, and developing new lore. *Behavior Analysis and Practice*, 5, 54-72. https://doi.org/10.1007/BF03391818

Hantula, D.A. (2019). Editorial: replication and reliability in behavior science and behavior analysis: A call for a conversation. *Perspect Behav Sci* **42,** 1–11 (2019). https://doi.org/10.1007/s40614-019-00194-2

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*(2), 165 179. https://doi.org/10.1177/001440290507100203

Iwata, B. A., Pace, G. M., Dorsey, M. F., Zarcone, J. R., Vollmer, T. R., Smith, R. G., . . . Willis, K. D. (1994). The functions of self-injurious behavior: An experimental-epidemiological analysis. *Journal of Applied Behavior Analysis*, 27, 215 240 https://doi.org/10.1901/jaba.1994.27-215

Jacobs, K. W. (2019). Replicability and randomization test logic in behavior analysis. *Journal of* *the Experimental Analysis of Behavior*. https://doi.org/10.1002/jeab.501

Jessel, J. (2022). Practical functional assessment. In: Leaf, J.B., Cihon, J.H., Ferguson, J.L., Weiss, M/J/ (eds) Handbook of Applied Behavior Analysis Interventions for Autism. Autism and Child Psychopathology Series. Springer, Cham. https://doi.org/10.1007/978-3-030-96478-8_23

Jessel, J., Hanley, G. P., & Ghaemmaghami, M. (2016). Interview-informer synthesized contingency analysis: Thirty replications and reanalysis. *Journal of Applied Behavior Analysis*. https://doi.org/10.1002/jaba.316

Jessel, J., Metras, R., Hanley, G. P., Jessel, C., & Ingvarsson, E. T. (2020). Evaluating the boundaries of analytic efficiency and control: A consecutive controlled case series of 26 functional analyses. *J Appl Behav Anal*; 53(1):25-43 https://pubmed.ncbi.nlm.nih.gov/30734285/

Kazdin, A. E. (2020). Single-case experimental designs: characteristics, change, and challenges. *Journal of the Experimental Analysis of Behavior*. https://doi.org/10.1002/jaba.926

Kinney, C. E., Begeny, J. C., Stage, S. A., Patterson, S., & Johnson, A. (2022). Three alternatives for graphing behavioral data: A comparison of usability and acceptability. *Behavior Modification*, *46*(1), 3-35. https://doi.org/10.1177%2F0145445520946321

Kranak, M. P., Shapiro, M. N., Sawyer, M. R., Deochand, N., & Neef, N. A. (2019). Using behavioral skills training to improve graduate students' graphing skills. *Behavior Analysis: Research and Practice*, *19*(3), 247–260. https://doi.org/10.1037/bar0000131

Kratochwill, T. R., Horner, R. H., Levin, J.R., Machalicek, W., Ferron, J., Johnson, A. (2021). Single-case design standards: an update and proposed upgrades. *Journal of School Psychology*. https://doi.org/10.1016/j.jsp.2021.10.006

Kubina, R. M., Kostewicz, D. E., Brennan, K. M., & King, S. A. (2017). A critical review of line graphs in behavior analytic journals. Educational Psychology Review, 29, 583-598. https://doi.org/10.1007/s10648-015-9339-x

Lambert, J. M., Bloom, S. E., & Irvin, J. (2012). Trial-based functional analysis and functional communication training in an early childhood setting. *Journal of Applied Behavior Analysis*, *45*(3), 579-84. https://pubmed.ncbi.nlm.nih.gov/23060670/

Lambert, J. M., Staubitz, J. E., Roane, J. T., Houchins-Juárez, N. J. Juárez, A. P., Sanders, K. B., & Warren, Z. E. (2017). Outcome summaries of latency-based functional analyses conducted in hospital inpatient units. *Journal of Applied Behavioral Analysis*. https://pubmed.ncbi.nlm.nih.gov/28556903/

Lane, J. D., Gast, D. L. (2014). Visual analysis in single case experimental design studies: brief review and guidelines. *Neuropsychol Rehabil*. 2014;24(3-4):445-63. https://pubmed.ncbi.nlm.nih.gov/23883189/

Lane, J. D., Shepley, C., & Spriggs, A. D. (2021). Issues and Improvements in the Visual Analysis of A-B Single-Case Graphs by Pre-Service Professionals. *Remedial and Special Education*, *42*(4), 235–247. https://doi.org/10.1177/0741932519873120

Ledford, J. R, & Gast, D. L. (2018). Single case research methodology: Applications in special education and behavioral science (3rd ed.). *London: Routhledge*. https://doi.org/10.4324/9781315150666

Ledford, J. R., Lane, J., & Severini, K. (2018). Systematic use of visual analysis for assessing outcomes in single case design studies. *Brain Impairment*, 19(1), 4-17. https://doi.org/10.1017/BrImp.2017.16

Lieberman, Rebecca & Yoder, Paul & Reichow, Brian & Wolery, Mark. (2010). Visual analysis of multiple baseline across participants graphs when change is delayed. *School Psychology Quarterly*. https://doi.org/10.1037/a0018600

Manolov, R., Gast, D. L., Perdices, M., & Evans, J. J. (2014). Single-case experimental designs: Reflections on conduct and analysis. *Neuropsychological Rehabilitation*, *24*(3/4), 634–660. https://pubmed.ncbi.nlm.nih.gov/24779416/

Manolov, R., Tanious, R., & Fernández-Castilla, B. (2021). A proposal for the assessment of replication of effects in single-case experimental designs. *Journal of Applied Behavior Analysis*. Advanced online publication. https://doi.org/10.1002/jaba.923

Ninci, J., Vannest, K. J., Willson, V., & Zhang, N. (2015). Interrater agreement between visual analysts of single-case data: A meta-analysis. *Behavior Modification*, *39*(4), 510-541. https://doi.org/10.1177%2F0145445515581327

Northup, J., Wacker, D., Sasso, G., Steege, M., Cigrand, K., Cook, J., & DeRaad, A. (1991). A brief functional analysis of aggressive and alternative behavior in an outclinic setting. *J Appl Behav Anal. 1991 Fall;24(3):509-22.* https://doi.org/10.1901/jaba.1991.24-509

Onghena, P., Michiels, B., Jamshidi, L., Moeyaert, M., & Van den Noortgate, W. (2018). One by one: Accumulating evidence by using meta-analytical procedures for single-case experiments. *Brain Impairment*, https://doi.org/10.1017/BrImp.2017.25

Parsonson, B. S., & Baer, D. M. (1992). The visual analysis of data, and current research into the stimuli controlling it. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 15–40). Lawrence Erlbaum Associates, Inc.

Peltier, C., McKenna, J. W., Sinclair, T. E., Garwood, J., & Vannest, K. J. (2021a). Brief report:

ordinate scaling and axis proportions of single-case graphs in two prominent EBD journals from 2010 to 2019. *Behavioral Disorders*, *47*(2), 134–148. https://doi.org/10.1177/0198742920982587

Peltier, C., Muharib, R., Haas, A., & Dowdy, A. (2021b). A decade review of two potential analysis altering variables in graph construction. *Journal of Autism and Developmental Disorders.* 52, 714–724 (2022). https://doi.org/10.1007/s10803-021-04959-0

Rader, A.E., Young, M.E., Leaf, J.B. (2021). A quantitative analysis of accuracy, reliability and bias in judgements of functional analyses. *Journal of the Experimental Analysis of Behavior*. https://doi.org/10.1002/jeab.711

Radley, K. C., Dart, E. H., & Wright, S. J. (2018). The effect of data points per x- to y-axis ratio on visual analysts evaluation of single-case graphs. *School Psychology Quarterly*, 33(2), 314–322. https://doi.org/10.1037/spq0000243

Rispoli, M., Brodhead, M., Wolfe, K., & Gregori, E. (2018). Trail-based functional analysis informs treatment for vocal scripting. *Behavior Modification*. 42(3) 441-465 https://doi.org/10.1177/0145445517742882

Roane, H. S., Fisher, W. W., Kelley, M. E., Mevers, J. L., & Bouxsein, K. J. (2013). Using modified visual-inspection criteria to interpret functional analysis outcomes. *Journal of Applied Behavior Analysis*, 46, 130-146. https://doi.org/10.1002/jaba.13

Sidman, M. (1960). *The tactics of scientific research*. New York, NY: Basic Books.

Slocum, T. A., Detrich, R., Wilczynski, S. M., Spencer, T. D., Lewis, T., & Wolfe, K. (2014). The evidence-based practice of applied behavior analysis. *Behav Anal*;37(1):41-56. https://doi.org/10.1007/s40614-014-0005-2

Smith, T. (2013). What is evidence-based behavior analysis? *Association for Behavior Analysis*

International; 36(1):7-33. https://doi.org/10.1007/BF03392290

Sunde, E., Briggs, A. M., & Mitteer, D. R. (2022). Reliability and validity of using structured visual-inspection criteria to interpret latency-based functional analysis outcomes. *Journal of Applied Behavior Analysis*. https://doi.org/10.1002/jaba.926

Tanious, R., & Onghena, P. (2021). A systematic review of applied single-case research published between 2016 and 2018: Study designs, randomization, data aspects, and data analysis. *Behavior Research Methods*, *53*(4), 1371-1384. https://doi.org/10.3758/s13428-020-01502-4

Tincani, M., Lorah, E. R., & Dowdy, A. (2018). Functional behavioral assessment. *PRO-ED*.

Vlaeyen, J. W. S., Wicksell, R. K., Simons, L. E., Gentili, C., De, T. K., Tate, R. L., Vohra, S., Punja, S., Linton, S. J., Sniehotta, F. F., & Onghena, P. (2020). From boulder to Stockholm in 70 years: Single case experimental designs in clinical research. *The Psychological Record*, *70*(4), 659–670. https://doi.org/10.1007/s40732-020-00402-5

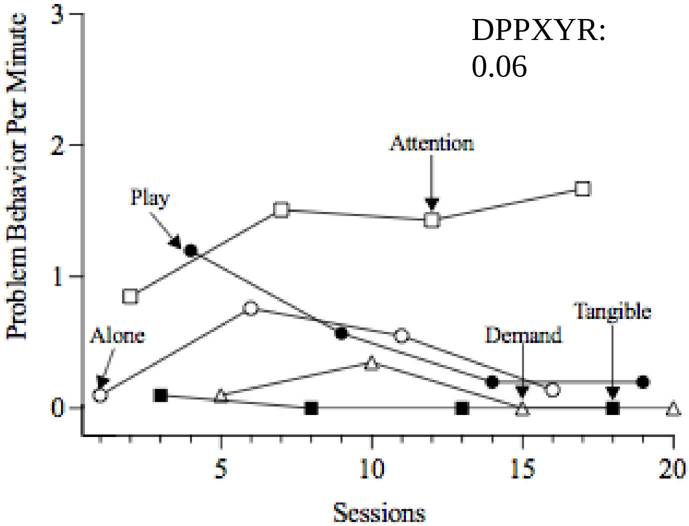
Wolfe, K., Seaman, M. A., Drasgow, E. (2016). Interrater agreement on the visual analysis of individual tiers and functional relations in multiple baseline designs. *Behav Modif*; 40(6):852-873. https://doi.org/10.1177/0145445516644699

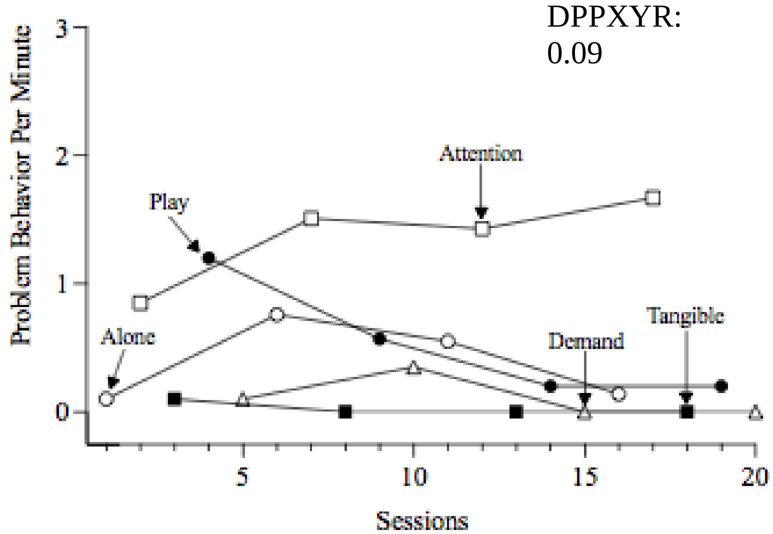
Wolfe, K., & McCammon, M. N. (2022). The analysis of single-case research data: Current instructional practices. *Journal of Behavioral Education*, *31*(1),2842. https://doi.org/10.1007/s10864-020-09403-4

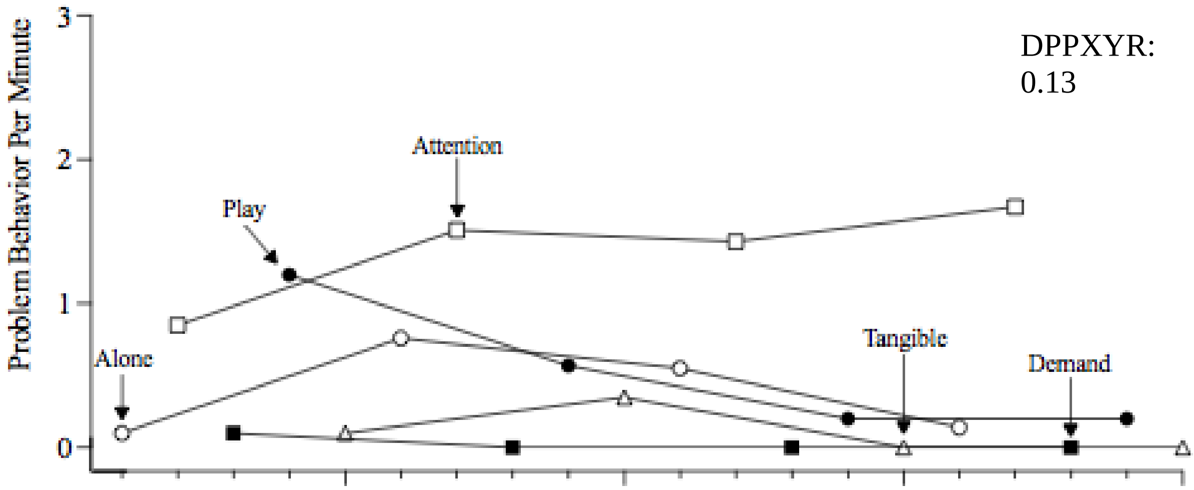
What Works Clearinghouse. (2020). What Works Clearinghouse Standards Handbook, Version 4.1. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

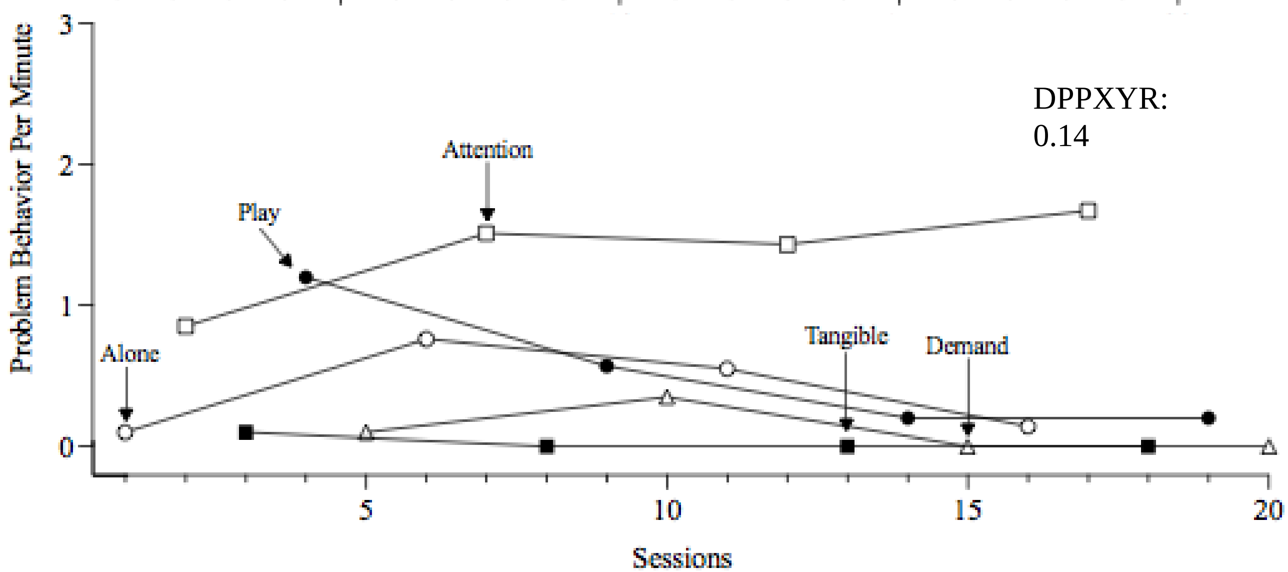
APPENDIX A
GRAPH MANIPULATIONS

Graph 1

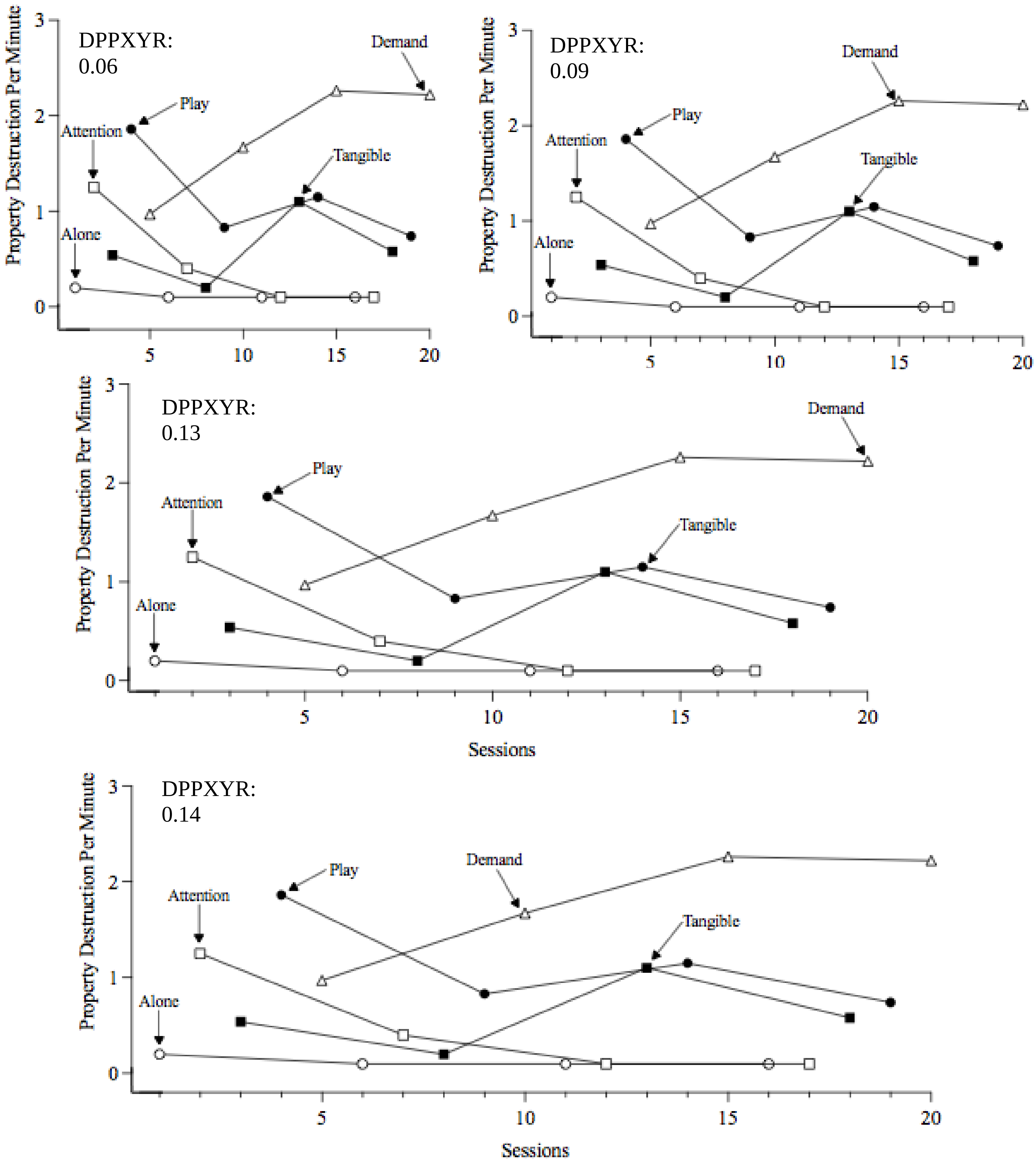






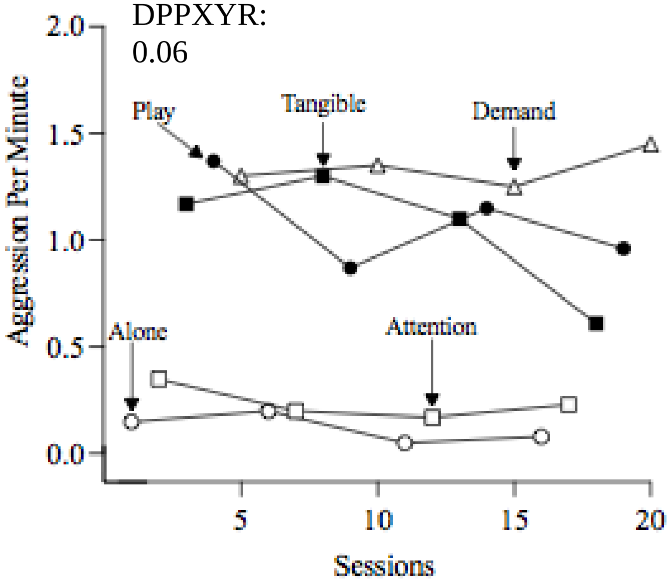


Graph 2

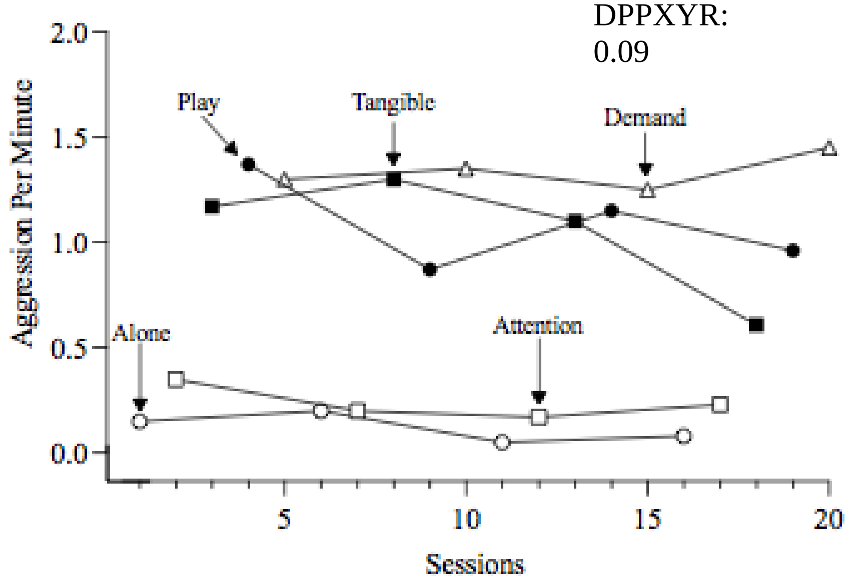

DPPXYR:
0.06
DPPXYR:
0.09
DPPXYR:
0.13
DPPXYR:
0.14
Property Destruction Per Minute
Sessions
Demand
Play
Attention
Tangible
Alone
0
1
2
3
5
10
15
20

Graph 3

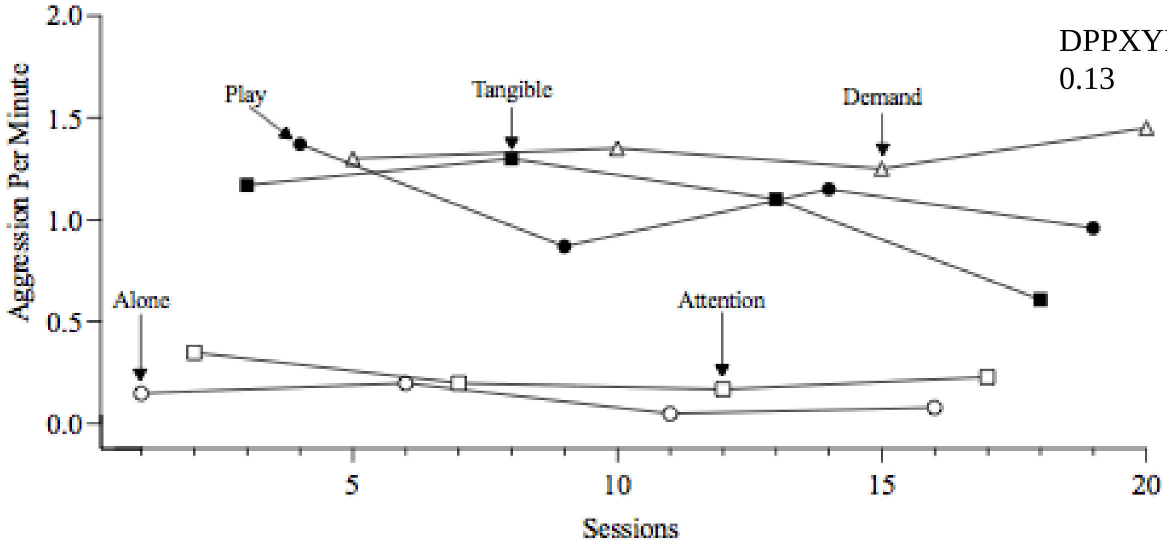
DPPXYR: 0.06



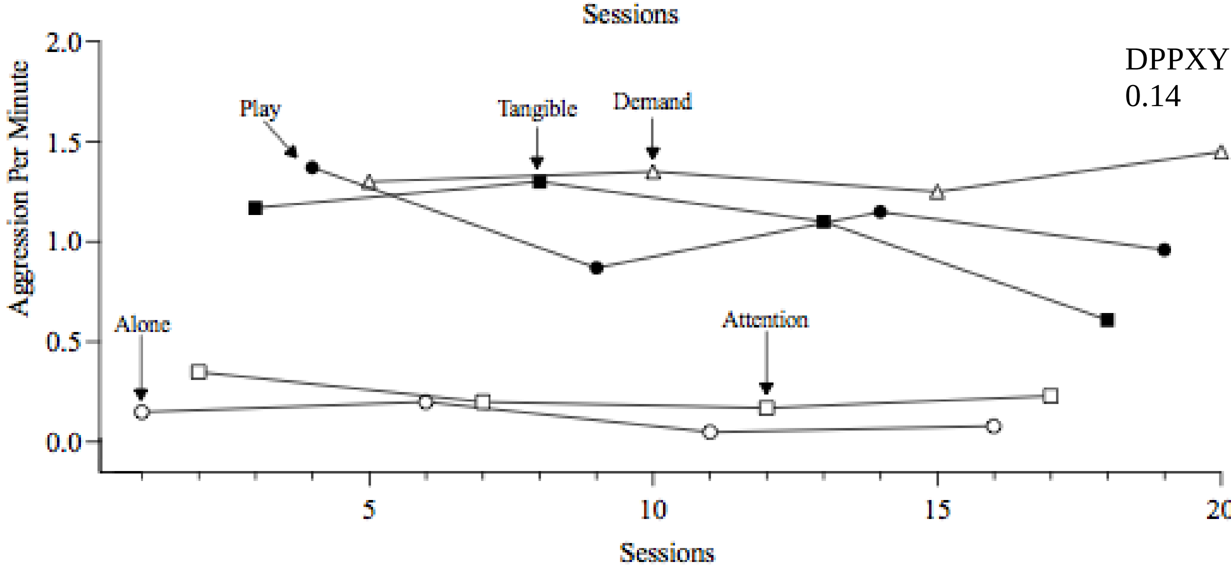
DPPXYR: 0.09



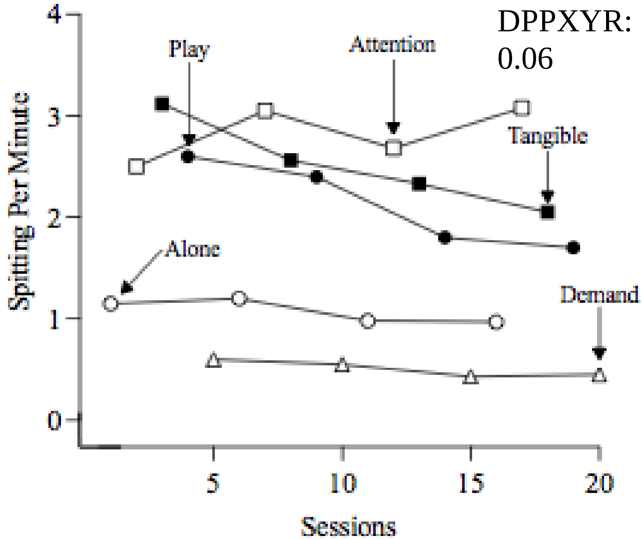
DPPXYR: 0.13

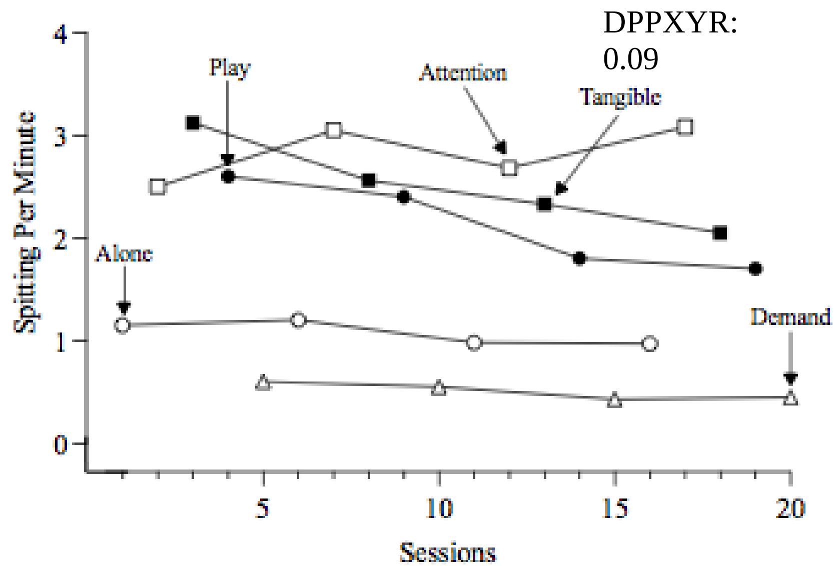


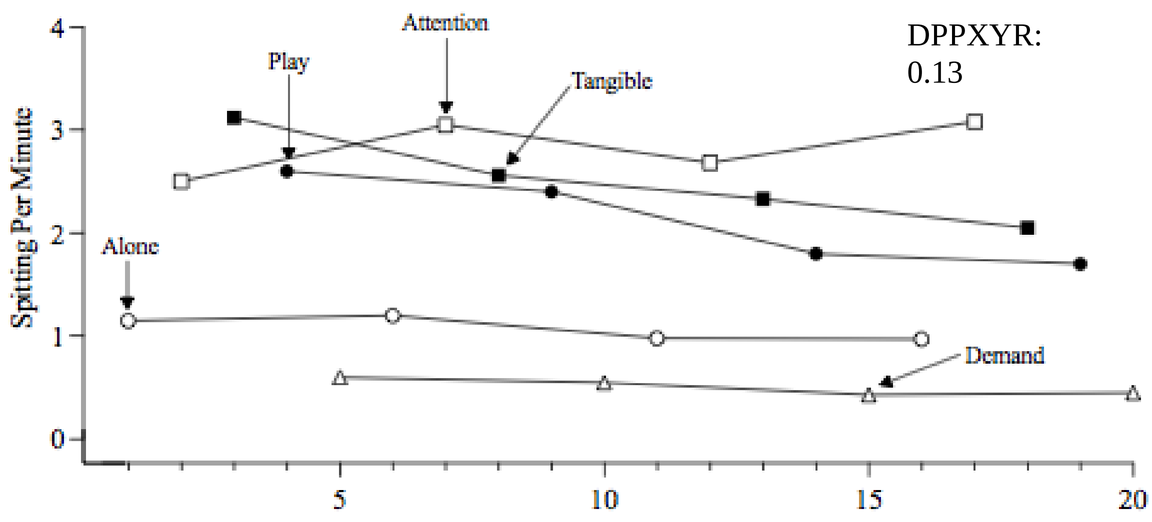
DPPXYR: 0.14

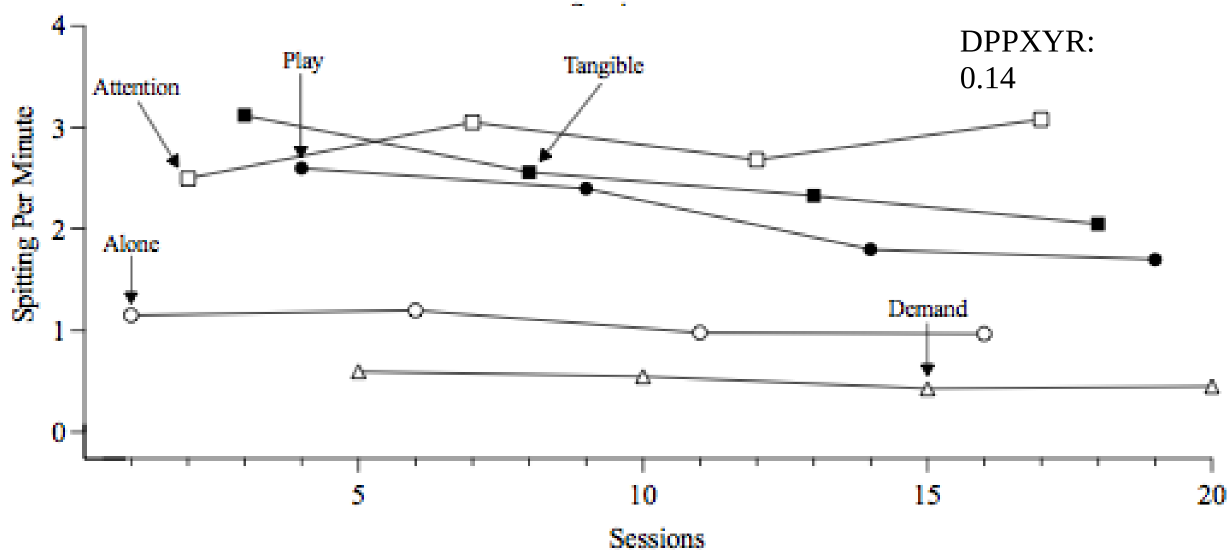


Graph 4

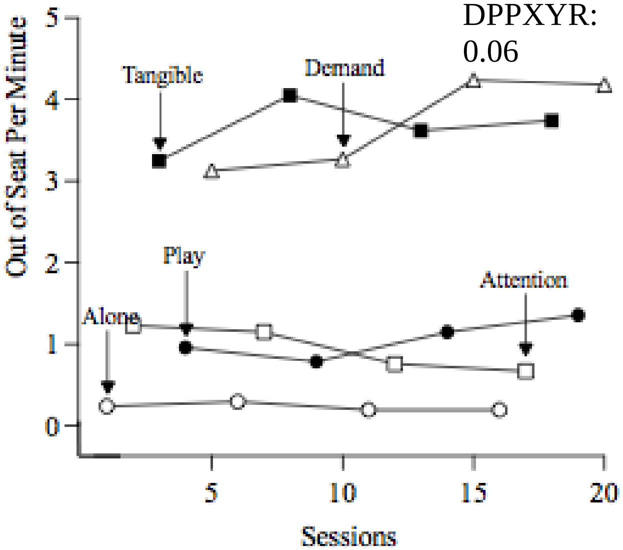

DPPXYR:
0.06
Play
Attention
Tangible
Alone
Demand
Spitting Per Minute
Sessions


DPPXYR:
0.09
Play
Attention
Tangible
Alone
Demand
Spitting Per Minute
Sessions

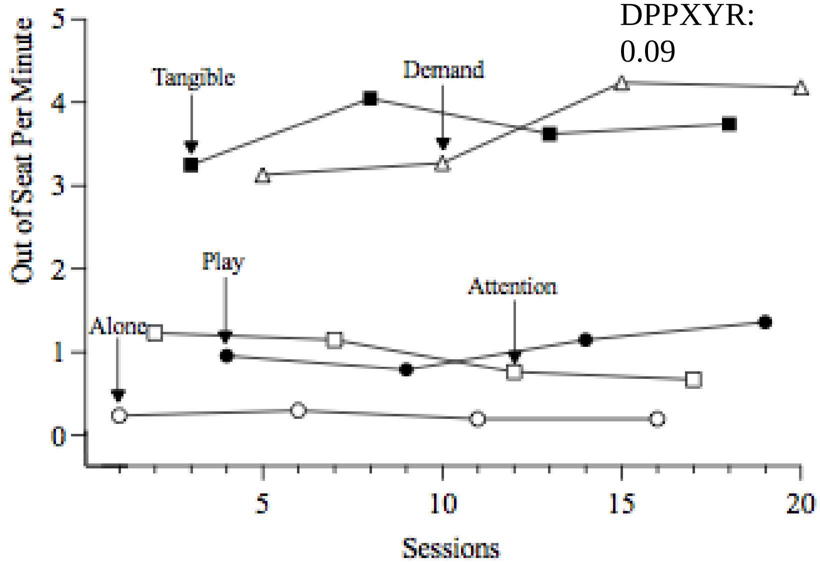

Attention
DPPXYR:
0.13
Play
Tangible
Alone
Demand
Spitting Per Minute


DPPXYR:
0.14
Play
Tangible
Attention
Alone
Demand
Spitting Per Minute
Sessions

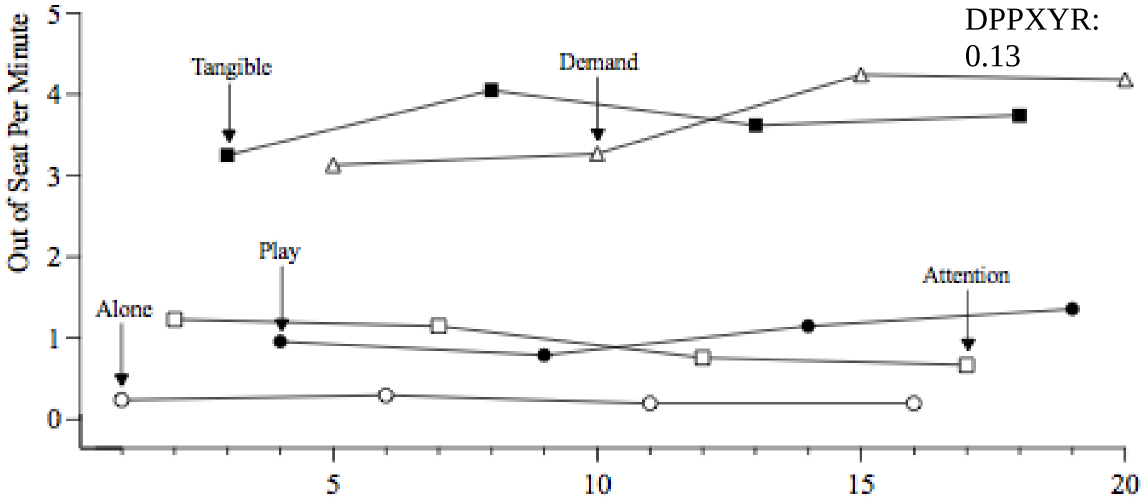
Graph 5



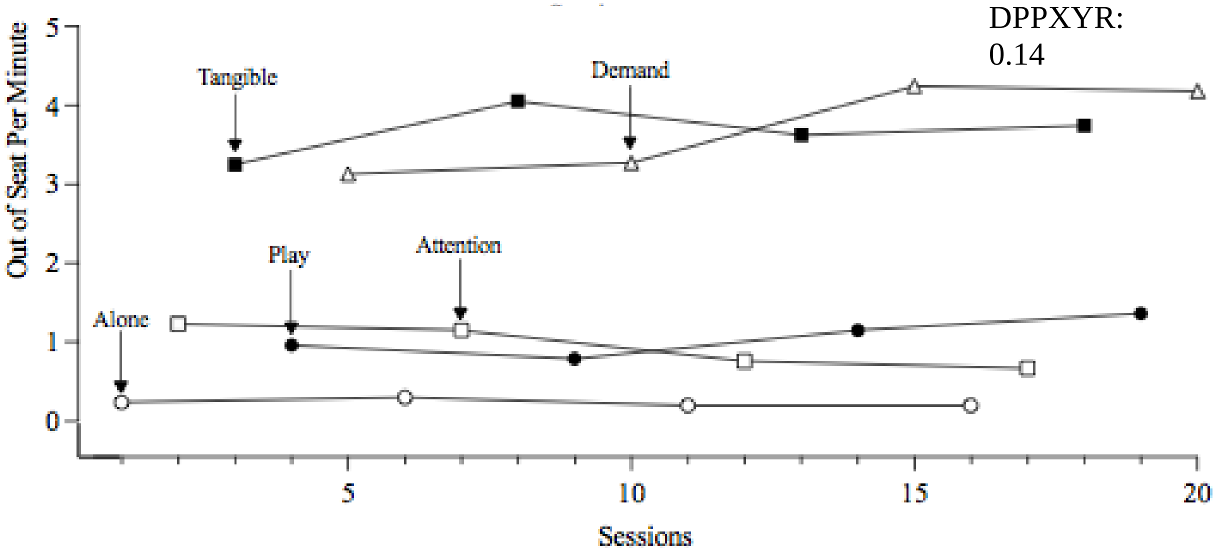
DPPXYR:
0.06



DPPXYR:
0.09

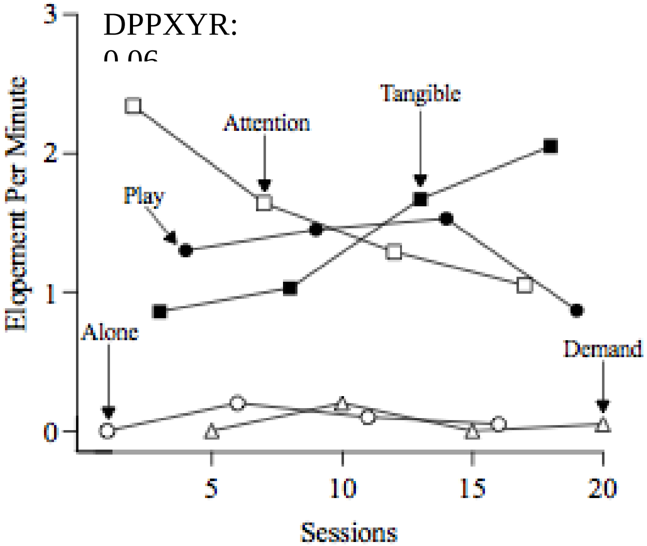


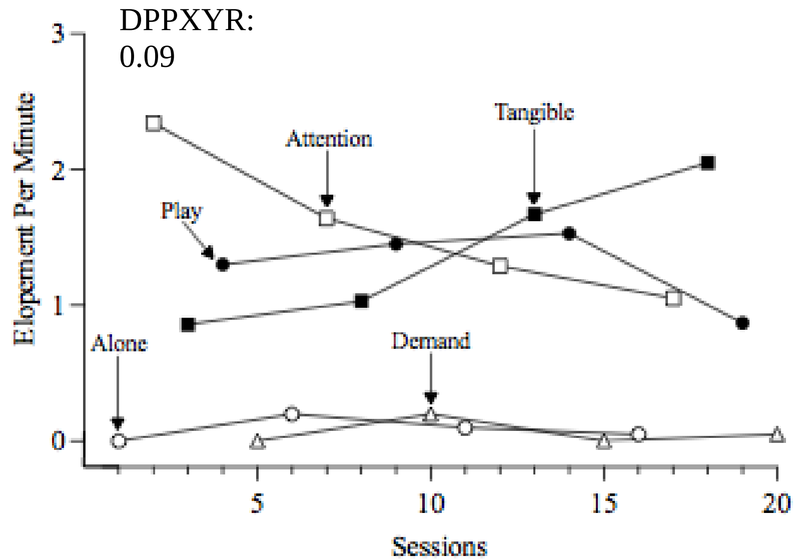
DPPXYR:
0.13

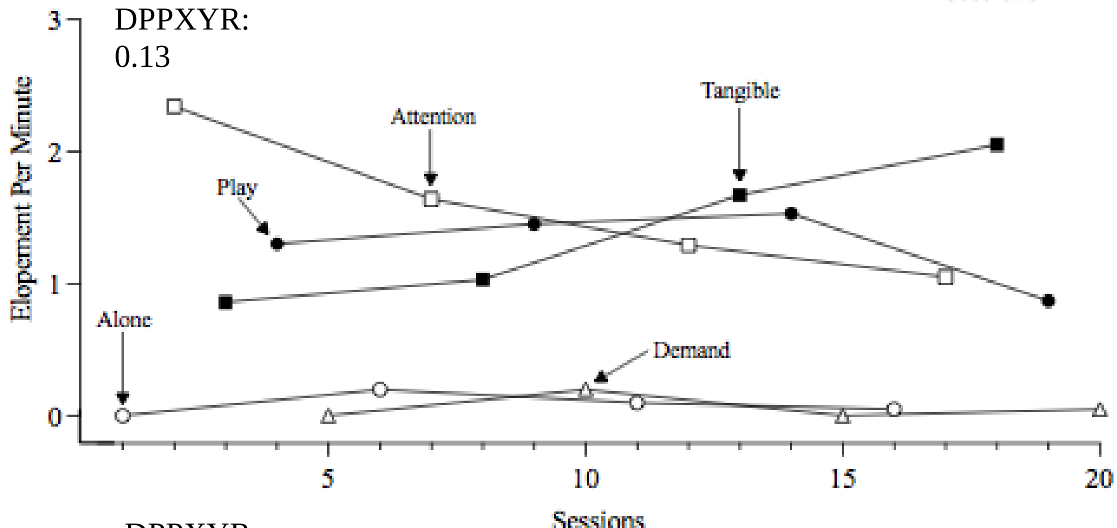


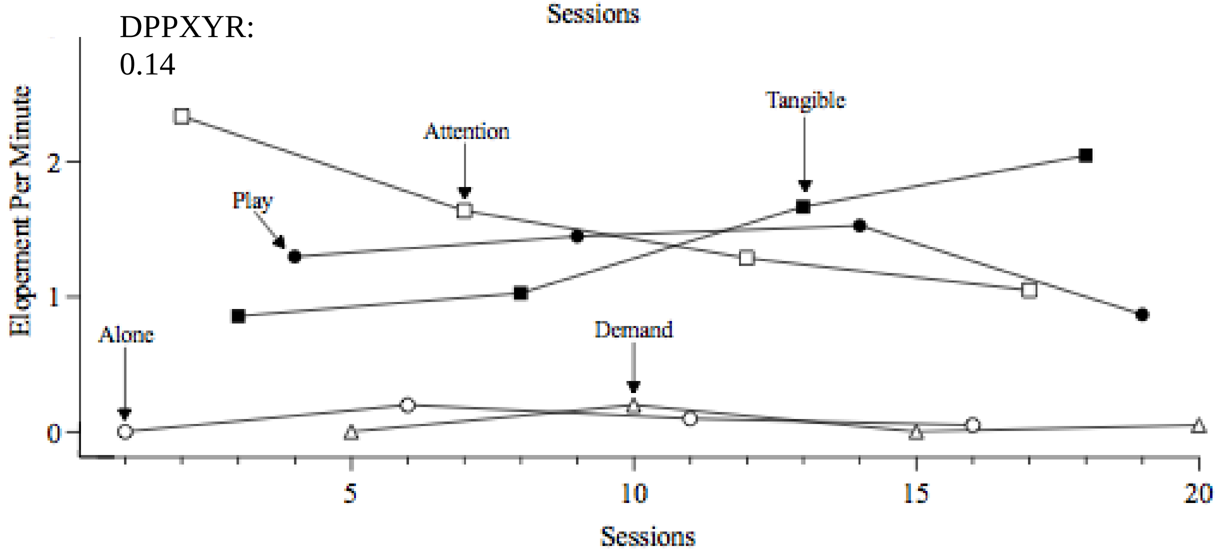
DPPXYR:
0.14

Graph 6

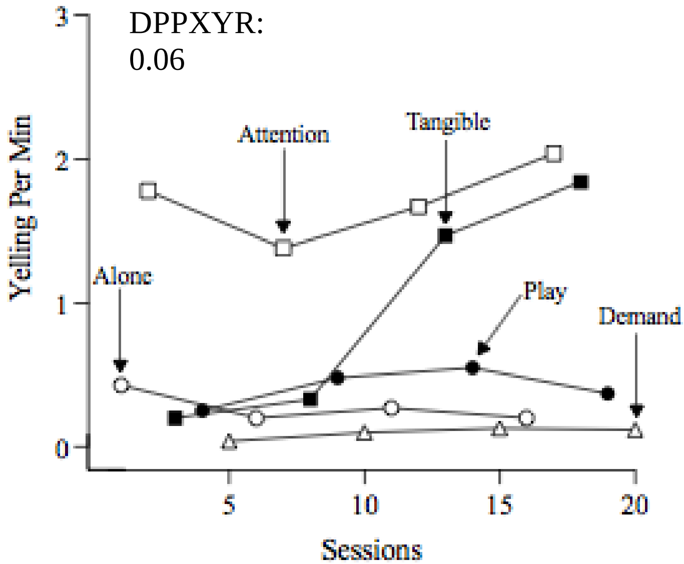

DPPXYR:
Tangible
Attention
Play
Alone
Demand
Elopement Per Minute
Sessions
0
1
2
3
5
10
15
20

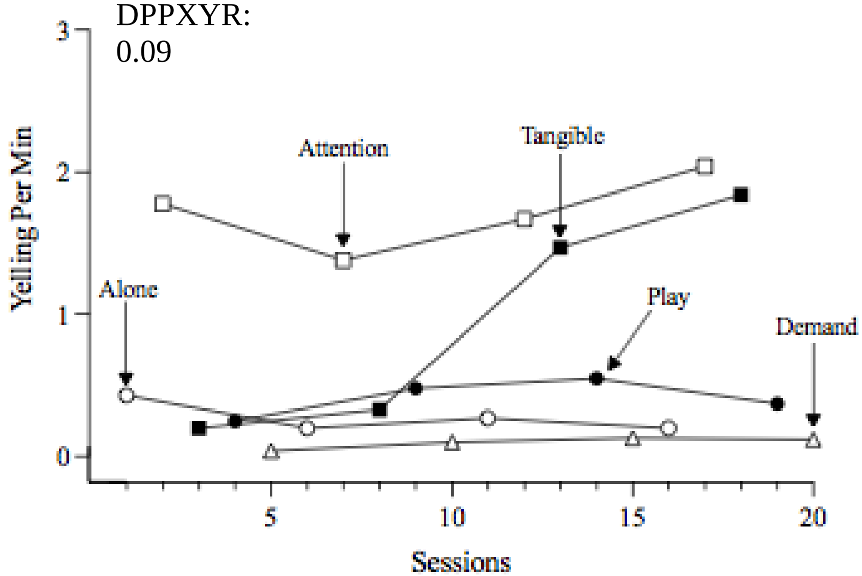

DPPXYR:
0.09
Tangible
Attention
Play
Alone
Demand
Elopement Per Minute
Sessions
0
1
2
3
5
10
15
20

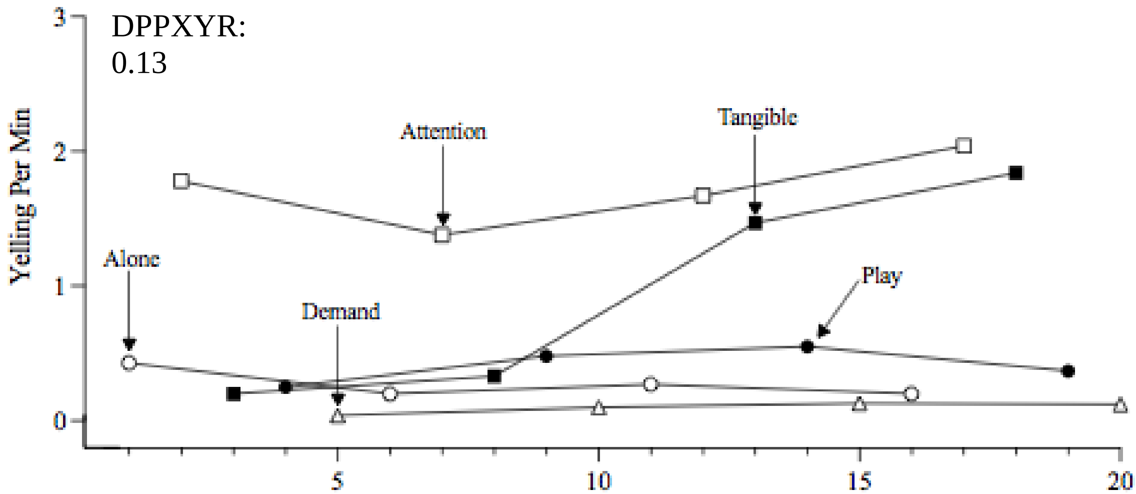

DPPXYR:
0.13
Tangible
Attention
Play
Alone
Demand
Elopement Per Minute
Sessions
0
1
2
3
5
10
15
20

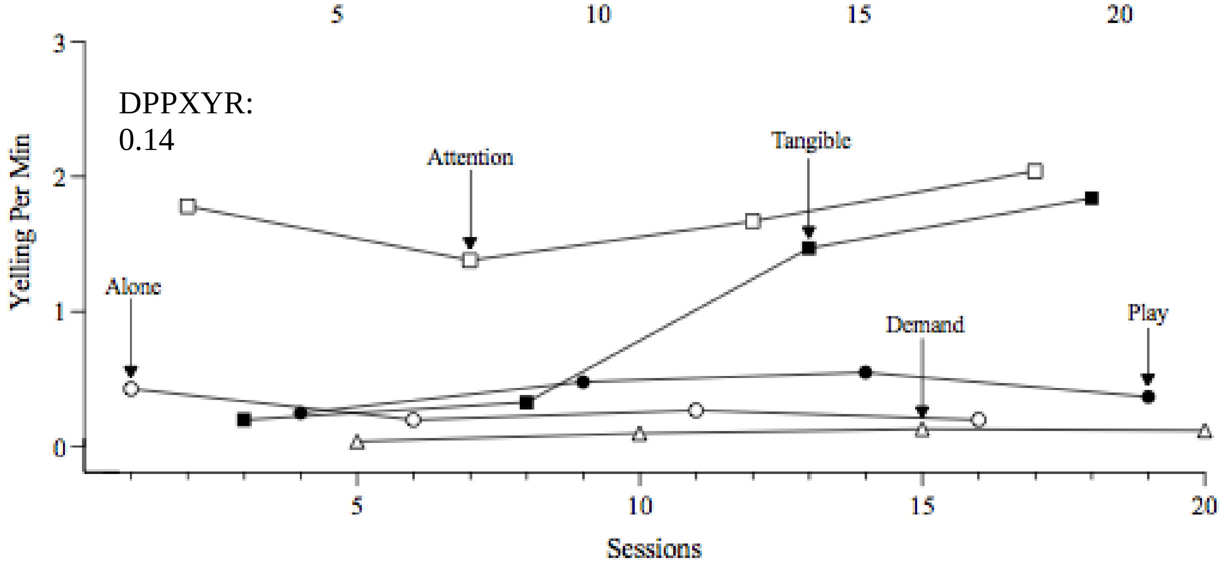

DPPXYR:
0.14
Tangible
Attention
Play
Alone
Demand
Elopement Per Minute
Sessions
0
1
2
5
10
15
20

Graph 7

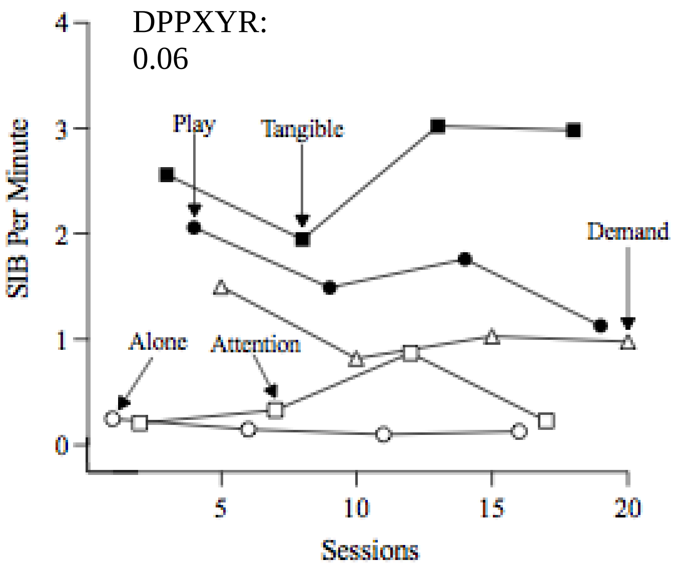

DPPXYR:
0.06
Attention
Tangible
Alone
Play
Demand
Yelling Per Min
0
1
2
3
5
10
15
20
Sessions

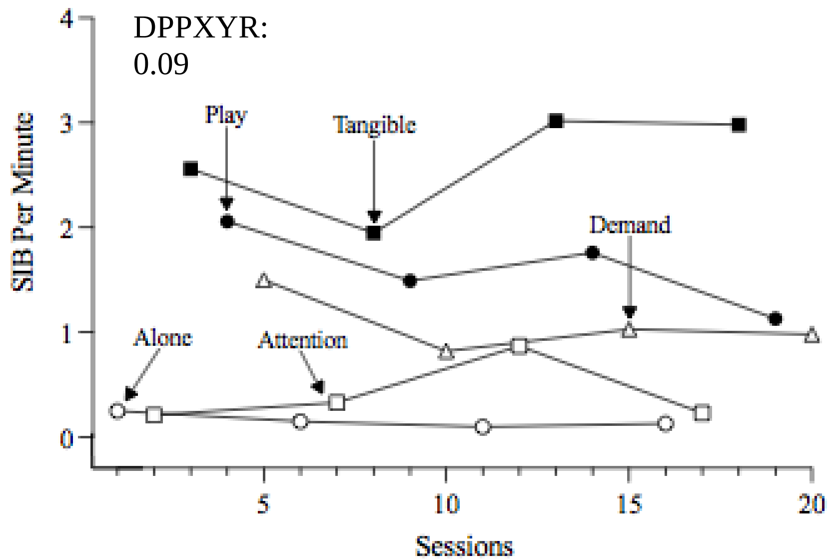

DPPXYR:
0.09
Attention
Tangible
Alone
Play
Demand
Yelling Per Min
0
1
2
3
5
10
15
20
Sessions

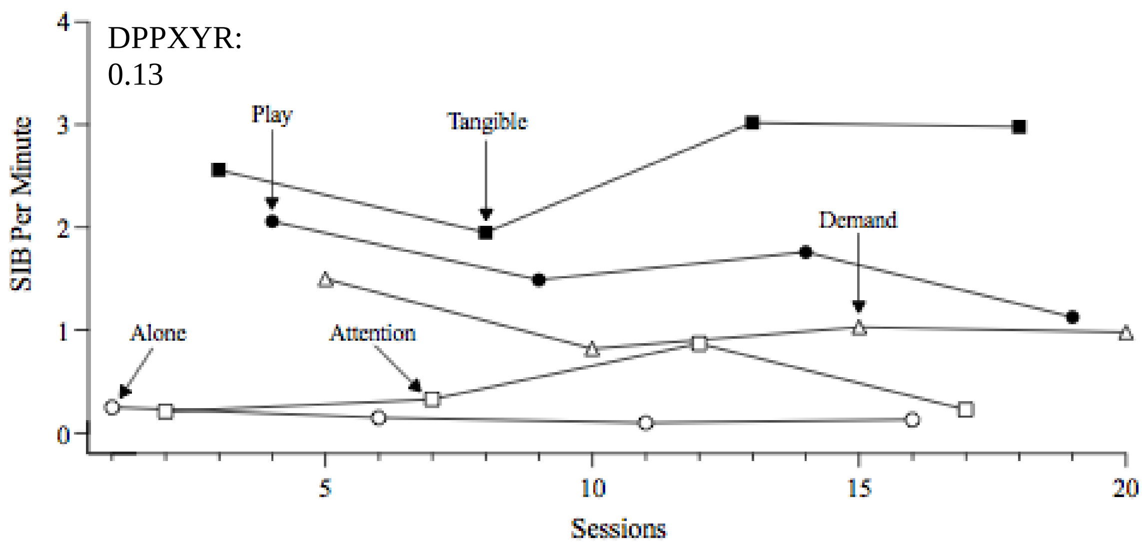

DPPXYR:
0.13
Attention
Tangible
Alone
Play
Demand
Yelling Per Min
0
1
2
3
5
10
15
20

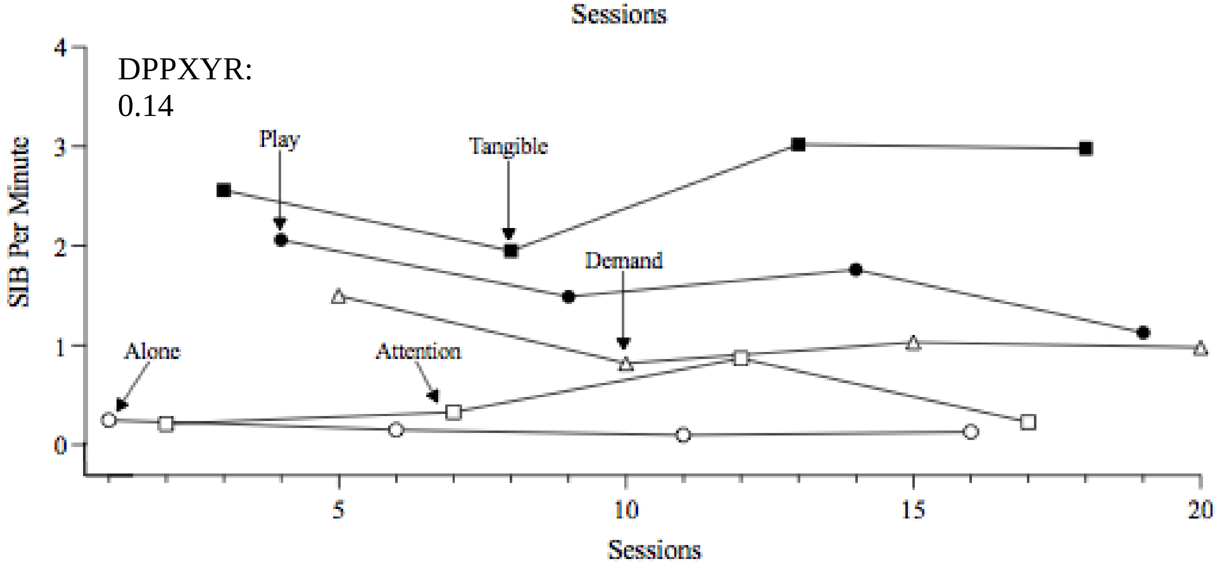

DPPXYR:
0.14
Attention
Tangible
Alone
Demand
Play
Yelling Per Min
0
1
2
3
5
10
15
20
Sessions

Graph 8


DPPXYR:
0.06
Play
Tangible
Demand
Alone
Attention
SIB Per Minute
Sessions
0
1
2
3
4
5
10
15
20


DPPXYR:
0.09
Play
Tangible
Demand
Alone
Attention
SIB Per Minute
Sessions
0
1
2
3
4
5
10
15
20


DPPXYR:
0.13
Play
Tangible
Demand
Alone
Attention
SIB Per Minute
Sessions
0
1
2
3
4
5
10
15
20


DPPXYR:
0.14
Play
Tangible
Demand
Alone
Attention
SIB Per Minute
Sessions
0
1
2
3
4
5
10
15
20

APPENDIX B
CALCULATING DPPXYR