EFFECTS OF THE FUNCTIONAL ANALYSIS CELEBRATION CHART ON DECISION MAKING FOR BOARD CERTIFIED BEHAVIOR ANALYSTS

A Dissertation in

Special Education

by

Salvador Ruiz

© 2018 Salvador Ruiz

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

May 2018
The dissertation of Salvador Ruiz was reviewed and approved* by the following:

Richard M. Kubina Jr.  
Professor of Special Education  
Dissertation Adviser  
Chair of Committee

Charles A. Hughes  
Professor of Special Education

Jonte C. Taylor  
Assistant Professor of Special Education

David Guthrie  
Associate Professor of Education

Mary C. Scheeler  
Associate Professor of Special Education  
Professor in Charge of Special Education

*Signatures are on file in the Graduate School
ABSTRACT

Functional Analysis (FA) has contributed significantly to the field of Applied Behavior Analysis. The use of visual analysis to determine function has largely proven effective. However, challenges still exist, namely how can behavior analysts unify analysis and interpretation of results. The present study will examine the extent to which Board Certified Behavior Analysts (BCBA) could determine function of challenging behavior by using a standardized graphical display called the Functional Analysis Celeration Chart (FACC). Participants with BCBA certification received an invitation to participate in the survey. Participant demographic questions, FA data recharted on the FACC, and decision making questions will appear on visual displays and Likert scale questions.

Keywords: functional analysis, visual analysis, and functional analysis chart
# TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................ vi

LIST OF FIGURES .................................................................................................... vii

ACKNOWLEDGEMENTS ............................................................................................ viii

Chapter 1  INTRODUCTION ......................................................................................... 1
  Visual Analysis ........................................................................................................ 2
  Supplements to Visual Analysis ............................................................................. 3
  Present Study .......................................................................................................... 4

Chapter 2  METHOD .................................................................................................... 6
  Participants .............................................................................................................. 6
  Survey ..................................................................................................................... 6
  Functional Analysis Celeration Chart .................................................................... 7
  Determining Function ............................................................................................ 7
  Dependent Measures ............................................................................................. 8
  Procedure ............................................................................................................... 8
  Tutorial .................................................................................................................. 9
  Accuracy ................................................................................................................ 9
  Social Validity ........................................................................................................ 9

Chapter 3  RESULTS .................................................................................................. 11
  Determining Function with the FACC .................................................................... 11
  Quantification of Level .......................................................................................... 12
  Preference on Graphical Display .......................................................................... 13

Chapter 4  DISCUSSION ............................................................................................. 14
  Determining Function with the FACC .................................................................... 14
  Quantification, Technology, and the FACC .......................................................... 18
  Preference on Graphical Display .......................................................................... 18
  Limitations ............................................................................................................ 19
  Future Directions ................................................................................................. 20

References ............................................................................................................... 21

Appendix A  Survey .................................................................................................... 25

Appendix B  Tables .................................................................................................... 50

Appendix C  Figures .................................................................................................... 52
Appendix D  Literature Review ........................................................................................................55

Appendix E Recruitment Materials ................................................................................................95

Recruitment Email sent to the Behavior Analysis Certification Board .........................95

Informed Consent .........................................................................................................................96
LIST OF TABLES

Table 1: Preference on the Sequential View FACC..........................................................50
Table 2: Preference on the Nonsequential View FACC..................................................50
Table 3: Percentage of displays when Quantification was used in Decision Making ........51
Table 4: Participants Preference on Graphical Displays..................................................51
LIST OF FIGURES

Figure 1: Participant Years of Experience Practicing Behavior Analysis. ..........................52
Figure 2: Participant Years of Experience using Visual Analysis........................................53
Figure 3: Participant Years of Experience Conducting/Analyzing Functional Analysis.......53
Figure 4: Correct Responding by type of Result.................................................................54
Figure 5: Correct Responding by Type of FACC Display....................................................54
ACKNOWLEDGEMENTS

Completing a dissertation is a contribution to science. I can say that without the support of my committee, family, and friends I would not have had the ability to accomplish my goals. I wish to offer a wholehearted thank you to the following people.

First, I would like to thank my mentor, Dr. Richard Kubina. The lessons you have taught me will serve me well moving forward and I feel extremely prepared, confident, and happy in all of my choices. I can’t express my appreciation for all you have done. Next, I would like to thank Dr. Charles Hughes. We have had many tough conversations and your guidance became invaluable as I navigated through the program. Dr. Taylor, thank you for your support, feedback, and guidance. I learned a lot from you. Dr. Guthrie, learning from you both in and out of the classroom was a great experience that I will never forget.

Throughout my time at Penn State I received a lot of support from Dr. Wolfe and Dr. Lee. I thank both of you for your time and kind words when they were needed. I consider myself fortunate to receive so much encouragement from the Chartlytics team as well. Thank you for that.

Last, I would like to thank my mother and father for all their support. I appreciate all that you have done and love both of you. To Tony and Megan, thank you so much for your great friendship when things haven’t been the best. I consider both of you family and your support helped throughout this process. Thank you to my friends who have provided words of encouragement and celebrations!
Chapter 1

Introduction

Functional Analysis (FA) provides guidance for the treatment of challenging behavior by determining what type of reinforcement will maintain a target behavior. FA has several defining features: (1) The assessment provides a series of test conditions and a control condition; (2) a condition specific motivating operation and reinforcement for each test condition creates deprivation and increases the likelihood of the occurrence of the target behavior; and (3) a control condition that allows for a comparison against the test conditions (Neef & Peterson, 2007). Recorded instances of the target behavior appear on a visual display for analysis. The level of each data set determines the condition(s) that produced the most occurrences of the target behavior when compared to the control.

FA has benefitted from widespread literature demonstrating the ability of the assessment to determine function of challenging behavior, effectiveness of variations, training professionals on implementation, development of sub-types of functions, the use of structured criteria to determine agreement of function, and the use of statistical methods to augment visual analysis (e.g. Brossart, Parker, Olson, Mahadevan, 2006; Journal of Applied Behavior Analysis, 1994; JABA, 2003; JABA, 2013). Comprehensive reviews of functional analysis findings have further demonstrated the effectiveness and importance of the issue (Hanley, Iwata, & McCord, 2003; Beavers, Iwata, & Lerman, 2013). The use of FA in the treatment of problem behavior has led to the development of function-based interventions and decreases in the use of punishment-based interventions (Pelios, Morren, Tesch, & Axelrod, 1999).

Variations of FA have included slight modification to procedures and measurement tactics. For example, Trial Based Functional Analysis modifies procedures by conducting the assessment in the natural environment and will typically use a bar graph (See Rispoli, Ninci, Neely, & Zaini, 2014 for a comprehensive review). The use of bar graph provides a direct view
of level, but discounts sequencing effects. Sequencing effects demonstrate if a participant has learned to access reinforcement in an environment, which may lead to false positive results (Johnston & Pennypacker, 2009). Additionally, if two or more conditions have a similar level of behavior, the ability to correctly determine function may become difficult due to the inability to see variability and trend on a bar graph.

**Visual Analysis**

The use of visual analysis has contradictory perspectives in the literature (e.g. DeProspero & Cohen, 1979; Khang, Chung, Gutshall, Pitts, Kao, & Girolami, 2010). Research has shown inconsistent analytical conclusions across analysts (Jones, Weinrott, & Vaugt, 1978; DeProspero & Cohen, 1979; Ninci, Vannest, Wilson, & Zhang, 2015). Experimenter’s have shown low agreement between statistical methods and visual analysis (DeProspero & Cohen, 1979; Jones et al., 1978). The findings suggest visual analysis may not provide enough evidence to support effective interventions. For example, inaccurate analysis may occur due to an error in the visual display. Further, the analyst may lack the ability to appropriately analyze the data. The use of statistical methods provides additional checks such as confidence intervals and effect sizes (Brossart, Parker, Olson, & Mahadevan, 2006; Michaels, 1974).

Additionally, research has shown that in Trial Based Functional Analysis conflicting results may occur using different graphical displays with the same analysts (Ruiz & Kubina, 2017). The issue of reliability of analysis becomes imperative with regards to FA because of the severity of the challenging behavior targeted. Notably, Bloom and colleagues did not match three Trial Based FA’s to traditional FA and had one partial match (Bloom, Iwata, Fritz, Roscoe, & Carreau, 2011). Similarly two other manuscripts either had a partial match or no match when comparing FA and Trial Based FA (LaRue, Lenard, Weiss, Bamond, Palmieri, & Kelly, 2010; Rispoli, Davis, Godwyn, & Camargo, 2013). The previously described findings indicate using different graphical displays may produce different conclusions in regards to function.
Researchers have attempted to address visual analysis inconsistencies through two methods. First, the use of structured criteria to determine the function of challenging behavior with ten data points per condition and varying data points per condition (Hagopian, Fisher, Thompson, Owen-DeSchryver, Iwata, & Wacker, 1997; Roane, Fisher, Kelley, Mevers, & Bouxsein, 2013). Second, statistical methods that enhance the use of visual analysis to determine if a functional relation exist (Davis, Gangnè, Fredrick, Alberto, Waugh, & Haardörfer, 2013). Both methods have demonstrated effectiveness.

**Supplements to Visual Analysis**

The use of structured criteria provides standardized decision rules for professionals. Using a criterion that provides rules facilitates the consistent examination of data for determining a function. Development of structured criteria relies on an appropriately constructed display. However, line graphs appear with variable construction features and with widely discrepant proportions and scaling. For example, the axes often do not follow the proportional construction rule, which governs the ratio of size between the vertical (5/8 to 3/4) to the horizontal axis (Kubina, Kostewicz, Brennan, & King, 2017). An ill-formed graph visually affects the placement of trend, variability, and level.

To mitigate the many subjective properties of nonstandard linear graphs (e.g. scaling, labeling, placement of tick marks), researchers have experimented with statistical methods and begun pairing statistical analysis with visual analysis (Brossart et al., 2006; Davis et al., 2013; Parker, Cryer, & Byrns, 2006). The use of quantification can provide a level of certainty supporting visual analysis. For example, FA data sets with multiply maintained results can have elevated levels of behavior in several test conditions. The test conditions may not have an equal level to each other and outliers may cloud judgment.

Providing a numerical value for the occurrence of behavior in each test condition can facilitate the identification of interventions most effectively leading to a decrease in targeted behavior. While visual analysis will show elevated levels of behavior in a test condition, the
variability of the data path may cause concern in determining if a test condition maintains a behavior. Quantification augments visual analysis by using a measure of central tendency across a condition to determine the value of the level. Also, quantification of variability reflects the degree of control or influence of the present condition on the target behavior.

Another possibility aside from supplementing visual analysis with statistical analysis resides in having a chart that directly offers statistical information - the Standard Celeration Chart or SCC (Pennypacker, Gutierrez, & Lindsley, 2003). The SCC, a standardized ratio chart, has graphic properties allowing the chart reader to not only visualize the data in a standard capacity, but also quantify the change features (e.g., trend – celeration, variability – bounce). The numerical values improve visual analysis, and direct statistical information on the occurrence of a target behavior during a test condition.

**Present Study**

A modified version of the SCC called the Functional Analysis Celeration Chart (FACC) can quantify level and visually analyze FA data sets (See Figure 1 and 2). The FACC displays data in a sequential and condition-grouped view, offering a numerical comparison of level across conditions, and uses different symbols to represent each condition (Chartlytics, 2016). The different data displays each provide a unique advantage. The sequential display allows for detection of sequencing effects, which assist in determining a potential problem with procedures. For example, a participant learning how to access reinforcement in the session would have an increase in behavior across time. The condition-grouped display affords an easier visual inspection of the data by grouping each condition on the graph (See Figure 2). When an undifferentiated result occurs, experimenter’s will typically conduct additional sessions testing one function at a time called an extended analysis, or pairwise analysis (Betz & Fisher, 2011). The additional sessions expose the participant to continued reinforcement, potentially strengthening the response. Additionally, the FACC delivers numerical values that eliminate the need for pairwise analysis assuming behavior occurs in any of the conditions.
The FACC may further assist with undifferentiated results. Undifferentiated results occur when the levels of problem behavior are variable across conditions, without any of the test conditions displaying a higher pattern of responding than the control (Betz & Fisher, 2011). Current practice suggests extending the analysis to conduct additional sessions if the behavior remains undifferentiated a retest of socially mediated conditions may provide conclusive results (Betz & Fisher, 2011). However, the FACC quantifies existing data contributing guidance for which condition yielded the most behavior. While the pattern may appear undifferentiated with visual analysis alone, the use of quantification as an analytical tactic allows for the development of an intervention.

The proposed study seeks to examine the effectiveness of the FACC by measuring participants’ ability to detect the function of challenging behavior in sequential and conditioned grouped views. Further, determining the role of quantification and preference may predict use of the display moving forward. Specifically, the present study asks the following questions:

1. To what extent will Board Certified Behavior Analysts determine the function of challenging behavior using the FACC?

2. How frequently will BCBA’s use the quantification of level to determine the function of challenging behavior?

3. Will participants find a standardized graphical display preferable for analysis in comparison to equal interval line graphs?
Chapter 2

Method

Participants

The participants had a BCBA certification and received a request for participation via the Behavior Analyst Certification Board’s website. All participants completed coursework in Applied Behavior Analysis, a minimum of 1,500 experience points, and passed a certification exam. Additional questions inquired about years of experience practicing behavior analysis, time spent interacting with graphical displays of behavior, and experience conducting FA assessments. Further, questions pertained to type of employment held to clarify roles of participants. Participants had the option of not taking part in the survey. Anyone who began the survey but did not complete had their answers excluded from the results.

Participant Characteristics. A total of 20,956 people received a request to take the survey. All participants have a BCBA credential and have their contact information on the Behavior Analysis Certification Board Registry (BACB.com). Behavior Analysts that begin the survey can enter their job responsibilities in a provided text box. Job responsibilities may include direct service implementation, consulting with families whose children receive behavior analytic services (in school, or home), and provide consultation to school districts for children that qualify for behavior analytic services. All participants have a minimum of 15 credit hours in behavior analytic content and completed a minimum of a master’s degree.

Survey

A survey administered via Qualtrics® asked participant demographic questions, as well as, display charts with functional analysis data (See Appendix A). All FA data selected has appeared in published manuscripts and recharted on a Functional Analysis Celeration Chart by the experimenter. The selected FA data will come from a highly referenced study (Iwata, Dorsey, Slifer, Bauman, & Richman 1994). The function for each target behavior appears in the original
manuscript. Additional FA data came from an article that implemented visual criteria for
decision making rules in-regards to determining the function of a challenging behavior (Roane,
Fisher, Kelley, Mevers, & Bouxsein, 2013). The manuscripts demonstrated functional relations
of behavior in the selected graphs. Manuscripts selected had used traditional FA procedures.
After the participants determine the function of each graph, they would answer additional
questions regarding the method they used to determine function and preferences of analytical
tactics.

**Functional Analysis Celeration Chart (FACC)**

The online Chartlytics graphing platform allows for quantification of data and a standard
visual display. The FACC displays functional analysis data using a standard, ratio line chart
(Chartlytics, 2016). The data display can appear in a sequential, or non-sequential (i.e. grouped
by condition) order. Standardized symbols represent each main condition and the user sets what
the symbols represents. For example, an open circle can symbolize an alone condition and a star
can depict a demand condition. The FACC also quantifies data, (e.g. level). A user who hovers a
cursor over the horizontal level line sees the numerical value of the level. Each condition has
level values appearing in a table facilitating a quantifiable comparison for all conditions
conducted in an FA. Software users can select different options for level calculation: geometric
mean, arithmetic mean, or median. FACCs included in the survey will have a static display
generated by screen shots of data. The displays contained data with level lines and a table with
the level values (See Figures 1 & 2).

**Determining Function**

The survey asked participants to select from condition specific answers, a combination of
conditions (e.g. attention, escape, tangible), or undifferentiated results. Each data set had the
possibility of behavior maintained by one function, multiple functions, or undifferentiated (i.e.,
unable to determine a function). In all, 5-6 choices will appear under each data set. If a data set
demonstrated behavior maintained by multiple functions, participants had to select each function.
If a participant does not select all of the functions in a multiply maintained result, the selection scores as an incorrect response. Participant responses, recorded electronically, receive a score of correct (response matched the function in the manuscript) or incorrect (responses did not match the function in the manuscript).

**Dependent Measures**

The dependent variable consisted of FA data from published manuscripts, recharted on an FACC. Each FACC survey question has the ability to select functions that could maintain the behavior (See Figure 3). Of the 18 graphs included in the survey, nine charts have a one-function result (n=9), two charts had an undifferentiated result, and six (n=6) have a multiply maintained result. Demographic and analysis questions rely on a Likert scale to quantify participant responses (See Figure 4 & 5).

**Procedure**

A recruitment email with a link to the survey sent to participants requested participation. In addition, the email provided a rationale for the study. Prior to accessing the survey, participants received a tutorial on visual analysis with the FACC. The tutorial included the components of an FACC in both views (Sequential and Condition-Grouped), FA data, quantification table, and how to analyze data using the FACC. Participants selected the survey via an embedded link. The email contained information regarding identification of selection, a link to the survey, and information regarding a chance to win a $50.00 Visa Gift Card. The first page included demographic questions. Each subsequent page contained one FACC per page and a table with the numerical values with the option of selecting if the data show one function (with each function listed as an option), multiply maintained, or undifferentiated results. Participants could select more than one option per question. The last page included participants reacting to the FACC and decision making tactics.
Tutorial

A pre-recorded tutorial explained the use of visual analysis (i.e. level as a decision making tactic) and how to read the quantification of the level (i.e., calculated with the geometric mean) on the FACC lasted approximately 10 minutes. Data appeared in both sequential and non-sequential views. The tutorial included two types of visual displays. A sequential view FACC shows data visually and with a table that compare each condition to the control. Participants received instruction on how the sequential view FACC displays data and how to read the table to analyze the data. A condition-grouped FACC displayed the same data as the sequential view FACC with instruction on the differences between the two types and examining the level of each condition.

Accuracy

Minimizing error in measurement allows experimenters the ability to make better decisions. A true value represents a standard value arrived at through extra efforts involved in minimizing measurement error. (Kostewicz, King, Datchuk, Brennan, & Casey, 2016). The use of accuracy portrays how closely observed values approximate what actually happened (Johnston & Pennypacker, 2009). The present study established a true value by examining the participant responses (i.e., the observed values) on each survey in relation to the responses provided by the participants through Qualtrics®. Participants could perform the behavior by selecting a response. An instance of a response demonstrates an instance of behavior (i.e. participants selected a response from the options available). The experimenter viewed selected responses to each survey question and compare the responses to the results from Qualtrics® (i.e., the true value).

Social Validity

Social Validity guides program planning and evaluation (Schwartz & Baer, 1991). Upon completion of participant responses to how analysis and function judgments occurred, additional questions asking about the preference of using the FACC followed. The Likert scale included a range from strongly disagree to strongly agree. Strongly disagree would indicate that the FACC
would not have preference as a possible graphical display for measuring FA data and strongly agree would indicate that the display had preference.
Chapter 3

Results

A total of 503 participants responded to the survey and answered a minimum of one question. Not all participants responded to every question. Further, a link for a tutorial included in the participation request allowed the experimenter to track number of views. A total of 53% participants watched the tutorial. All participants had obtained a BCBA™ prior to participation in the study. Demographic questions indicated 35.17% of the sample population had 4-7 years of experience practicing behavior analysis (See Figure 1). Additionally, nearly 40% had 4-7 years practicing visual analysis (See Figure 2) indicating a disconnect between year’s experience and examining graphical displays. The majority of respondents had 0-3 year’s experience conducting and/or analyzing FA data (See Figure 3). Demographic results suggest participants had more experience with practicing behavior analysis and visual analysis, with less exposure to FA.

Determining Function using the FACC

A total of nine sets of data appearing in sequential and condition grouped views yielded a total of 18 graphs. Of the visual displays included in the current study, 10 had a single function result, four showed multiply maintained functions, and four had no differentiation in responding. The FA literature demonstrates a large portion of single function results, followed by multiply maintained (Beavers et al., 2013). Across the three types of results, the sample responded to single function results most accurately (See Figure 4).

Single Function Results

All of the common functions of behavior had multiple displays (e.g. Attention, Demand, Tangible, and Automatic). The automatic function choice on each question had a label of alone to depict the test condition label in the graphical display. Each participant had the option of selecting multiple response for every question regarding graphical displays. Correct responding for single function results occurred 62.47% of the total responses.
**Multiply maintained results.** Multiply maintained results occur when the control condition remains stable with a low level and more than one test has an elevated level of behavior (Betz & Fischer, 2011). Two graphs had two functions maintaining the target response. An additional two graphs had three functions maintaining the behavior. These survey questions required participants to select all the test conditions that had an elevated level of behavior in comparison to the control. The sample of participants responded correctly 31.7% of the total responses (See Figure 4).

**Undifferentiated results.** Undifferentiated patterns of responding occur when no discernable difference in level between test and control take place. Further, the level of any of the conditions should place at the bottom of the graphical display (Betz & Fischer, 2011). A raised level of behavior in each session would likely indicate an automatic function. A total of four questions in the two views had a correct response of undifferentiated. The response choice appeared in all (n= 18) display questions. Participants correctly responded to 49.8% of the total number of opportunities (See Figure 4).

**Response by type of display.** Each display had the same data set presented in both views. For example, if one data set appeared in a sequential view it similarly appeared in a condition grouped view. Participants responded correctly to approximately 53% of the visual displays in either view (See Figure 5), indicating that neither display evoked significantly more correct responding. However, in the condition grouped view an outlier appeared. One FACC had three test conditions maintaining the behavior. Participants largely selected undifferentiated pattern of responding (38.4%). If calculating the average correct responding with the median, condition grouped correct responses occurred for 58% of the total questions.

**Quantification of Level**

Level lines on visual displays use measures of central tendency to determine line placement. The FACC used the Geometric Mean to determine the level line. The value of each conditions level can then receive comparison to any other condition. For example, if the control
condition shows no change (i.e. no change in responding) the level would then show x1, any condition compared to the control with an elevated level of responding would display at a higher value indicating more responding occurred. Participants had access to the level values for each of the 18 displays. When asked how frequently participants used the table for decision making 34.8% of respondents reported examining the table for 0-5 displays reflecting a heavy reliance on visual analysis (See Table 3). However, 28% of the sample reported using the table for most of the FACC charts.

**Preference on Graphical Display**

Participants reported on preference of graphical display and satisfaction with each display (i.e. sequential vs. condition grouped). When asked to report on preference of graphical display a visual image of a line graph received inclusion although no survey question included the image. Most respondents reported their greatest preference for the line graph (See Table 4). Additionally, 3.2% of the sample reported no preference for any of the three options. A Likert scale ranging from strongly satisfied to strongly dissatisfied appeared to determine preference of the two FACC displays. Nearly half of the sample reported some level of satisfaction with the sequential FACC (See Table 1). Similar results occurred for the condition grouped view. Preference on display did not vary based on display.
Chapter 4

Discussion

A review of the relevant literature indicates that visual analysis requires support to maximize effectiveness (See Appendix D). As an assessment, functional analysis or FA uses the graphical element “level” as the primary analytic tactic. Previous studies have shown a low level of agreement when using only visual analysis sans intervention (e.g. structured criteria; visual aids). The use of structured criteria has demonstrated an increase in agreement across raters, increasing the likelihood of correct identification of function (Hagopian et al., 1997; Roane et al., 2013). However, the criteria do have limitations. Primarily, the criteria rely on a truth-by-agreement standard. While agreement may provide more confidence in analysis, consensus among observers does not provide any information on the accuracy of the observed events (Johnston & Pennypacker, 2009; Kostewicz et al., 2016). The Functional Analysis Celeration Chart (FACC) can help behavior analysts with function detection through objective quantification. The present study asked the following questions:

1. To what extent will Board Certified Behavior Analysts determine the function of challenging behavior using the FACC?
2. How frequently will BCBA’s use the quantification of level to determine the function of challenging behavior?
3. What preference will participants have for a standardized graphical display?

Determining Function using the FACC

Participants viewed data on the FACC and had an opportunity to select the correct function(s) maintaining the behavior in each visual display. Further, participants could choose multiple responses for each display question. The results showed correct responding occurred most frequently with single function results, followed by undifferentiated, and then multiply maintained.
Single Function Results. Most FA results that appear in the literature have a single function maintaining the response (Beavers et al., 2013). That makes the ability to identify single function results important for practicing behavior analysts because of the likelihood of encountering one during analysis. Determining a single test condition maintaining a behavior relies on low levels of the control and the other conditions. Visually a clear separation in level must occur.

The current study found single function results as the most commonly identified correct function suggesting that the FACC allows analysts to clearly determine one function of a challenging behavior. A similar study found 46% interrater agreement across analysts (Hagopian et al., 1997). However, the authors did not specify agreement across types of results. Comparatively, participants in the current study performed at higher levels (i.e., 62% function identification). Further, the data used had come from well referenced articles that identified a function. The participants determined the same function as the previously referenced studies on the FACC.

The implications for higher identification scores in the present study suggest behavior analysts can use a standard ratio graph to determine function. The FACC has a ratio scale which offers ratio comparisons of data sets with different magnitudes, show growth or rate of change better than linear graphs, and normalize variance (Few, 2009; Harris, 1999; Kubina & Yurich, 2012). The view of the data on the ratio scale may have led to cleaner comparisons between different test conditions. Additionally, the variability may have also flattened thereby showing more stability and thus a clear function.

Multiply Maintained Results. Multiply maintained results can occur when multiple test conditions have an elevated level in comparison to the control (Betz & Fischer, 2011). Participants had varying degrees of correct responding. Specifically, the FACC charts with two functions maintaining the challenging behavior resulted in participants answering correctly more often than when a behavior had three conditions maintaining the response. When multiple test
conditions maintain a behavior, detecting function proves challenging. For example, the level of the control condition would dictate if a behavior is maintained by other tests or automatic reinforcement. Visual analytic tactics (e.g. trend, stability, variability) have shown to create issues with FA results (Diller, Barry, & Gelino, 2016). It remains unclear if the ratio scale of the FACC helped or hindered the process of function identification.

The FACC does advance the use of quantification of the levels, a seldom-used approach for determining function. Quantification has proven effective with other types of research designs (e.g. ABAB) and using statistical testing (Brossart et al., 2006; Davis et al., 2013). The current study differs in that it focused on multielement designs. The effects of quantification did not seem apparent as participants did not respond correctly frequently. A tutorial provided explained how to use quantification. However, half of the sample did not view the tutorial that reviewed the quantification of level. The tutorial would have demonstrated how quantification could assist when determining multiply maintained results. With only half of the sample watching the tutorial, determining the effect of quantification of level with the table requires future study.

**Undifferentiated Results.** Undifferentiated results occur in approximately 5% of the FA literature (Beavers et al., 2013; Hanley et al., 2003) indicating that behavior analysts cannot typically identify functions of behavior. The current study had a small sample of undifferentiated FA results as well. An undifferentiated result occurs when all conditions (e.g. tests, control) provide no visually discernable difference in level (Betz & Fischer, 2011). Essentially the analyst looks for a pattern in responding and when one does not seem apparent an FA has undifferentiated result. It remains unclear how quantification or visual analysis of the FACC impacted undifferentiated results.

Participants detected an undifferentiated result using the FACC for half of the graphs. Over half the sample correctly identified undifferentiated responses in two separate questions. Each question had a different view (i.e., sequential and condition grouped). Quantification may
have impacted decision making as the level values did not have numerical separation. Visually, the non-sequential view demonstrates similar level lines across conditions and sequentially the data do not separate and remain at low levels. While research in the reliability of detecting functions on multi-element designs appears limited, one study found low levels of reliability across analysts (Diller et al., 2016). The researchers discovered level distances created issues with analysis. The issue appeared most prominently for variability within conditions. Raters performed poorly when data paths have variability and did not reach agreement. The participants in the current study may have experienced a similar issue with undifferentiated and multiply maintained responding.

Response by type of display. The FACC has the unique ability to examine data in sequential and condition grouped views. To date, no other study has examined a condition grouped view of data on a ratio scaled visual display. The condition grouped view takes the data out of time, but allows for easy viewing of the level of each condition. The sequential view places the data in the order the conditions transpired, allowing for a more traditional view of data. Participants performed slightly better in the condition grouped view (See Figure 5) suggesting that grouping by condition may provide better analysis. The grouped data may more distinctly present differences between the conditions when compared the data laid on top of each.

The condition grouped FACC collates data by condition which also differs from a bar graph in that analysts can see all sessions. The ability to scan and see where the level lines lay in each condition make decision making efficient. Trial Based Function Analysis (TBFA) uses bar graphs in many instances, however, the data do not appear individually (See Ruiz & Kubina, 2017 for a comprehensive review). Condition grouping while still observing individual data points appears exclusively on the FACC.

Despite overall higher rates of correct responding, an outlier appeared in the condition grouped FACC. Many participants responded to one condition grouped view incorrectly leading to a 5% decrease in overall correct responses. When using the median, the correct response rate
increased when compared to the sequential view. The outlier had three functions maintaining the behavior. If participants did not select all three function responses, the question scored incorrect. While the outlier only appeared in one question, more research examining multiply maintained FACCs in the condition grouped view would provide more clarity.

**Quantification, Technology, and the FACC**

As stated earlier, quantification can assist with decision making (Hagopian, Rooker, & Zarcone, 2015). The FACC calculates level line values and allows comparison with every condition. The level offers an individual performance statistic for each condition stating the average rate of responding immediately available to individual analysts. However, as stated earlier it seems participants largely did not use the table as nearly half of the sample did not even view the tutorial. The explanation of how to read the table appeared in the tutorial as a formal decision making tactic for determining function. Therefore, the present results do not reflect the potential for behavior analysts to use quantification for analysis and function detection.

The present survey questions used static displays, live technology and an understanding of the data table and subsequent statistics could prove useful in the future. A recent study demonstrated how telehealth technology allowed behavior analysts 200 miles away from clients engage in successful functional analyses (Lee et al., 2015). The technology of conducting functional analyses remotely still has many issues that need addressed, but in the present study behavior analysts did better than average with single function results. Therefore, the FACC as technology tool holds promise for displaying data for FAs.

**Preferences for Graphical Displays**

Within contemporary behavior analysis BCBAs have the option to choose whatever type of visual display they like. Currently, the field of behavior analysis exclusively uses line graphs to display, analyze, and communicate data (Cooper et al., 2007). A question asked participants preferences for visual displays. While most participants preferred the line graph, the sample did report satisfaction with the FACC. Over half the sample population reported satisfaction with the
sequential FACC and half with the non-sequential view (See Table 1). Willingness to accept and use the FACC may help in future studies.

The benefits of the FACC have common features with other standardized ratio graphs. For example, the Standard Celeration Chart (SCC) allows for quick and reliable interpretation of data (Datchuk & Kubina, 2011). Further, the SCC quantifies data by using celeration, bounce, and level (Datchuk & Kubina, 2011). The FACC shares standardized construction and quantification features with the SCC. Exposing participants to the display will allow for consistency in analysis as each display follows similar features. Further, previous research that examines agreement with multi element designs uses a standardized construction and quantification feature for linear graphs (Hagopian et al., 2015). The use of both has increased agreement. Similarly, both tactics appear on the FACC.

History and experience with linear graphs may have contributed to the strong preference towards line graphs. Further, participants may not have had previous exposure to the FACC. A recently published meta-analysis that examined multiple factors in-regards to agreement determined that experience may play a role in visual analysis (Ninci et al., 2015). More exposure and practice using the FACC may impact its overall preference among behavior analysts.

**Limitations**

The present study has several limitations. First, one of the survey questions resulted in a large amount of incorrect responding indicating a potential issue with the item. An outlier in either the test or control conditions may have caused an error in responding. Outliers have impacted decision making of analysts (Nelson, Van Norman, & Christ, 2016) making the need for quantifiable values of level across a data set in FA important. The response had to include three functions maintaining the behavior. However, each survey question had the same wording. Second, although quantification of the level of each condition appeared on each graph a decision making rule did not. For example, a difference in level of x2 would indicate function. A rule may have increased correct responding. Third, more participant responses could provide more
information on the usefulness of the FACC as a tool. Fourth, participants did not have explicit knowledge that multiple choices could receive selection per question. Regardless of the amount of conditions that could maintain the behavior, every question could have all options selected.

**Future Directions**

Future research should provide increased response rates, as well as, the development of a numerical value that would indicate function to support visual analysis. The value should receive calculation by relying on published studies that included a determined function. Additionally, the FACC needs more research to better establish its utility. From basic components such as examining levels in each condition to viewing trends and variability in the data, future research would show the technical merits of the FACC. Using the chart with variations of traditional FA will provide additional insight and analysis.

**Conclusion**

The FACC has the potential to assist in determining functions of challenging behavior. Its standardized construction features and quantification of level can befit users. Participants in the current study correctly identified function most frequently with single function results followed by undifferentiated results. With the development of a rule, quantification can assist in decision making and future research should seek to increase response rates.

Correct responses exceeded previous research baseline levels (Hagopian et al., 1997). Further, providing another graphical display to assist analysts in circumstances where other displays may no demonstrate a functional relation. Visual Analysis with supports has been proven to assist those seeking to determine relations with intervention and behavior (Davis et al., 2013; Nelson et al., 2016; Ninci et al., 2015). The FACC provides those supports.
References


Appendix A

Survey

Figure 1: Sample FACC in Sequential View
Figure 2: Sample FACC in Condition Grouped View
The following behavior is maintained by?

☐ Alone
☐ Attention
☐ Demand
☐ Undifferentiated

*Figure 3: Sample Question FACC Condition Grouped View*

---

How many years of experience do you have practicing ABA?

☐ 0-3 Years
☐ 4-7 Years
☐ 8-10 Years
☐ 11+ Years

*Figure 4: Sample Demographic Question*
Did you like the FACC data presented in sequential view?

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

*Figure 5: Sample Preference Question*
Default Question Block

How many years of experience do you have practicing ABA?

- 0-3 Years
- 4-7 Years
- 8-10 Years
- 11+ Years

How many years of experience do you have using visual analysis?

- 0-3 Years
- 4-7 Years
- 8-10 Years
- 11+ Years

How many years of experience do you have conducting and/or analyzing Functional Analysis data?

- 0-3 Years
- 4-7 Years
- 8-10 Years
- 11+ Years

What type of job do you currently hold? (e.g. School Based Consultant, Private Practice, Teacher, etc.)
The following behavior is maintained by?

- Alone
- Attention
- Demand
- Undifferentiated
The following behavior is maintained by?

- Demand
- Free Play
- Attention
- Alone
- Undifferentiated
The following behavior is maintained by?

- Alone
- Demand
- Attention
- Undifferentiated
The following behavior is maintained by?

- [ ] Attention
- [ ] Demand
- [ ] Toy Play
- [ ] Tangible
- [ ] Alone
- [ ] Undifferentiated
The following behavior is maintained by?

- Free Play
- Tangible
- Demand
- Attention
- Undifferentiated
The following behavior is maintained by?

☐ Attention
☐ Demand
☐ Toy Play
☐ Tangible
☐ Alone
☐ Undifferentiated
The following behavior is maintained by?

- Ignore
- Attention
- Free Play
- Tangible
- Demand
- Undifferentiated
The following behavior is maintained by?

- Alone
- Attention
- Demand
- Free Play
- Undifferentiated
The following behavior is maintained by?

- Free Play
- Attention
- Demand
- Tangible
- Undifferentiated
The following behavior is maintained by?

- Free Play
- Ignore
- Attention
- Tangible
- Demand
- Undifferentiated
The following behavior is maintained by?

- Attention
- Demand
- Free Play
- Tangible
- Ignore
- Undifferentiated
The following behavior is maintained by?

- Ignore
- Tangible
- Free Play
- Demand
- Attention
- Undifferentiated
The following behavior is maintained by?

- Attention
- Demand
- Free Play
- Ignore
- Tangible
- Undifferentiated
The following behavior is maintained by?

- Ignore
- Attention
- Demand
- Tangible
- Free Play
- Undifferentiated
The following behavior is maintained by?

- Demand
- Free Play
- Attention
- Alone
- Undifferentiated
The following behavior is maintained by?

- Alone
- Attention
- Demand
- Free Play
- Undifferentiated
The following behavior is maintained by?

- Alone
- Attention
- Demand
- Free Play
- Undifferentiated
The following behavior is maintained by?

- Free Play
- Demand
- Attention
- Alone
- Undifferentiated
Did you like the FACC data presented in sequential view?

- Strongly Agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Did you like the FACC data presented in a non-sequential view?

- Strongly Agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

To what extent did you use the table to make decision in-regards to function?

- 0-24% of the displays
- 25-49% of the displays
- 50-74% of the displays
- 75-100% of the displays
Which type of Graphical Display do you most prefer?

- Sequential View FACC
- Non Sequential View FACC
- Equal Interval Line Graph
- None

If you would like a chance to win a $100 Visa gift card please enter your email below. A total of 2 winners will be selected and contacted.
## Appendix B

### Tables

**Table 1**: Preference on the Sequential View FACC.

<table>
<thead>
<tr>
<th>Satisfaction with Sequential View FACC</th>
<th>Percentage Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Satisfied</td>
<td>8.87%</td>
</tr>
<tr>
<td>Satisfied</td>
<td>25.81%</td>
</tr>
<tr>
<td>Somewhat Satisfied</td>
<td>17.74%</td>
</tr>
<tr>
<td>Neither Satisfied or Dissatisfied</td>
<td>20.97%</td>
</tr>
<tr>
<td>Somewhat Dissatisfied</td>
<td>12.1%</td>
</tr>
<tr>
<td>Dissatisfied</td>
<td>10.89%</td>
</tr>
<tr>
<td>Strongly Dissatisfied</td>
<td>3.63%</td>
</tr>
</tbody>
</table>

**Table 2**: Preference on the Nonsequential View FACC.

<table>
<thead>
<tr>
<th>Satisfaction with the Nonsequential View FACC</th>
<th>Percentage Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Satisfied</td>
<td>7.63%</td>
</tr>
<tr>
<td>Satisfied</td>
<td>22.49%</td>
</tr>
<tr>
<td>Somewhat Satisfied</td>
<td>19.28%</td>
</tr>
<tr>
<td>Neither Satisfied or Dissatisfied</td>
<td>23.29%</td>
</tr>
<tr>
<td>Somewhat Dissatisfied</td>
<td>12.05%</td>
</tr>
<tr>
<td>Dissatisfied</td>
<td>11.24%</td>
</tr>
<tr>
<td>Strongly Dissatisfied</td>
<td>4.02%</td>
</tr>
</tbody>
</table>
Table 3: Percentage of displays where quantification was used in decision making.

<table>
<thead>
<tr>
<th>Decision Making Using Quantification</th>
<th>Percentage Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-24% of Displays</td>
<td>34.8%</td>
</tr>
<tr>
<td>25-49% of Displays</td>
<td>16%</td>
</tr>
<tr>
<td>50-74% of Displays</td>
<td>21.2%</td>
</tr>
<tr>
<td>75-100% of Displays</td>
<td>28%</td>
</tr>
</tbody>
</table>

Table 4: Participant preferences on graphical displays.

<table>
<thead>
<tr>
<th>Preference of Graphical Display</th>
<th>Percentage Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential View FACC</td>
<td>12.4%</td>
</tr>
<tr>
<td>Nonsequential View FACC</td>
<td>10%</td>
</tr>
<tr>
<td>Equal Interval Line Graph</td>
<td>74.4%</td>
</tr>
<tr>
<td>None</td>
<td>3.2%</td>
</tr>
</tbody>
</table>
Appendix C

Figures

Figure 1: Participant years of experience.
Figure 2: Participant years of experience using visual analysis.

Figure 3: Participant years of experience conducting/analyzing FA assessments and data.
Figure 4: Correct responding by type of result on the FACC.

Figure 5: Correct responding by type of FACC display.
Appendix D

Review of the Relevant Literature Outline

Functional Analyses (FA) have a successful history in the field of behavior analysis. Current procedures and practices closely mirror work originally published in 1982 and reprinted in 1994 (Iwata, Dorsey, Slifer, Bauman, & Richman, 1994). Essentially, experimenters developed an assessment that superimposed conditions with a specific reinforcer and collected information on the occurrence of the target behavior. The visual data displayed would suggest a function of behavior when a test condition occurred at a discernable difference in level higher than a control condition (Iwata et al., 1994). The success of the assessment has led to the publication of three special issues in the Journal of Applied Behavior Analysis, numerous variations of the procedure, and wide spread adoption by behavior analysts (JABA 1994; 2003; 2013).

FA relies on visual analysis to determine function. Primarily, using a multielement design, a variation of an alternating treatment design, FA compares each test condition to a control condition (Cooper, Heron, & Heward, 2007). The reliability of visual analysis requires raters to determine a functional relation between a dependent (e.g. behavior) and independent variable (e.g. intervention/assessment). If an analyst has difficulties determining the effects of an independent variable a need exists to improve it. In order to ensure raters can agree on what data show, graphical displays must follow appropriate construction guidelines. In concert with such rules, visual aids also help guide analysis of the data.

Implementing visual aids on graphical displays supports determination of the degree of change in the data, the course of the data, and the level of the overall data set. Visual aids include examples such as trend and level lines (e.g., Nelson, Van Norman, & Christ, 2016). Visual aids should provide additional guidance to analysts. Nevertheless, construction issues may skew the aids making data appear to have more dramatic effects. Pairing visual aids with statistical tests, a
suggestion by some researchers, may then create disagreement between visual analysis and quantification.

Manuscripts examining the literature on FA have analyzed publication rates across journals, participant characteristics, topographies of behavior, and methodological characteristics (Beavers, Iwata, & Lerman, 2013; Hanely, Iwata, & McCord, 2003). The current review differs in that it examines the reliability of visual analysis, how experimenter’s have made visual analysis more reliable, and investigates the occurrence of errors in the construction of FA graphs. More specifically, the current review asks the following questions:

1. Does visual analysis offer a reliable method to analyze behavior data?
2. What suggestions have experimenters made to improve visual analysis?
3. Do construction errors appear in the FA literature?

**Reliability of Visual Analysis**

Applied Behavior Analysis uses visual analytic tactics to determine if a functional relation exists between interventions and targets behaviors (Cooper et al., 2007). FA differs in that the assessment does not function as an intervention but instead reinforces instances of the challenging behavior. After the assessment, the level of a test condition(s) will occur at higher rates than the control (Betz & Fischer, 2011). However, the need to determine the differences in level between test and control conditions relies on the visual display of data. Issues regarding the reliability of visual analysis have received attention in the behavior analytic literature since the 1970’s (e.g., Baer, 1977; DeProspero & Cohen, 1979; Jones, Weinrott, & Vaughn, 1978).

Two viewpoints have appeared in the discussion of reliability, but experimental literature examining visual analysis reliability has not frequently appeared. One perspective states that visual analysis offers a reliable and conservative estimate of the data, while another argues that improvements can occur. A total of seven studies have directly examined the reliability of visual analysis (n=7). Five of the articles found visual analysis as an unreasonably method to analyze data (DeProspero & Cohen, 1979; Fisch, 2001; Jones et al., 1978; Matyas & Greenwood, 1990). The
remaining two studies suggest visual analysis can reliably analyze data (Bobrovitz & Ottenbacher, 1998; Khang, Chung, Gutshall, Pitts, Kao, & Girolami, 2010).

Opposing Viewpoints

When considering the reliability of visual analysis, assumptions of the data should assist in clarification of statistical testing. For example, inferential statistics relies on one datum predicting the next. Single case data each operate independently from one another (Johnson & Pennypacker, 2009). Thus, leaving other statistical tests that can determine effect size post intervention (e.g. Hierarchical Linear Modeling) and descriptive statistics to determine a numerical value for level, which provides an added support for analysts. Using visual analysis alone has produced poor to adequate agreement (DeProspero & Cohen, 1979; Diller, Barry, & Gelino, 2016; Fisch, 2001; Jones et al., 1978; Matayas & Greenwood, 1990; Ninci, Vannest, Wilson, & Zhang, 2015).

Jones and colleagues asked raters to determine if a meaningful change occurred between baseline and intervention(s) phases (1978). The researchers selected journal editors as participants. Mean agreement across graphs came to .39. However, the term meaningful has subjectivity and may not have adequately captured the type of change needed to make a decision. A similar study published a year later calculated agreement among analysts at .61 (DeProsepero & Cohen, 1979). Both results fall below a range of acceptable agreement and indicate that a display without a clear demonstration of a functional relation fosters disagreement among raters. Both experimenters used editors of behavior analytic journals (DeProspero & Cohen, 1979; Jones et al, 1978).

A literature review examined a series of experiments that included the use of inferential statistics and discovered raters may miss treatment effects using visual analysis alone (Fisch, 2001). Five manuscripts included for review found several important points to consider. First, using visual analysis by itself creates a likelihood of treatment effects and trends in data. For example, if a treatment effect and trend occur at the same time only one will likely receive
attention. Second, statistical tests exist that do not rely on parametric assumptions. If a statistical analysis allows for independent observations, the quantification of effects aids assessment of the impact of a given treatment. Third, statistical tests should not replace visual analysis (Fisch, 2001). While a given statistical test may prove effective in determining an effect, other problems may arise. For example, determining sequencing effects becomes difficult when data appear numerically and not visually. The author of the review published all the included articles (Fisch, 2001).

Another issue in regard to reliability comes in the form of Type 1 and 2 errors. A Type 1 error occurs when an analyst determines a functional relation when one does not exist. A Type 2 error takes place when a functional relation appears but goes undetected (Matyas & Greenwood, 1990). If visual analysis provides a conservative method to analyze data, one would conclude a low rate of Type 1 error. To examine Type 1 and 2 errors, experimenters recruited 37 graduate school students that had minimal experience outside of one single case design course at the university level. Then, they asked participants to rate 27 graphs. Results showed the participants had a high Type 1 error rate. The article demonstrated low reliability of visual analysis generally and suggested a problem with agreement and increased Type 1 error (Matyas & Greenwood, 1990). Although the graphs included used an AB design, a Type 1 error in a FA could create problems during intervention.

Interrater agreement has become a common way to determine the effects of an independent variable on a behavior. Essentially two or more people examine a data set and offer an analysis. A meta-analysis found interrater agreement (IRA) across single case designs at .76 (Ninci et al., 2015). The researchers determined that varying designs produce higher agreement (e.g. multielement, AB). However, the researchers noted the analysis of effects may have caused an increase in agreement. Specifically, studies that examined multielement designs included raters who had extensive experience in single case research and the design uses visual aids which
has proven to assist to determine agreement. Visual supports may assist to increase IRA (Ninci et al., 2015).

As previously identified, one article found IRA appears higher for some single case designs (Ninci et al., 2015). However, another article examined the reliability of visual inspection with hypothetical data presented in a multielement design (Diller et al., 2016). Participants included 90 Board Certified Behavior Analysts (BCBAs) and 19 editors from JABA and the Journal of Experimental Analysis of Behavior (JEAB). The experimenters manipulated variability, trend, and mean shift of the data. Participants had to rate experimental control (Diller et al., 2016).

Across both sets of participants the mean correlation of agreement came to .43. When considering just editors the mean agreement across graphs rose to .69. Examining experimental control through statistical testing BCBAs had a mean correlation of .52 and the journal editors reported .44 (Diller et al., 2016). Ninci and colleagues (2015) hypothesized the impact of experimental control increased agreement in their article and the present study did not find significant results. Results from the study had lower values than other research. One possible explanation includes the lack of environmental context (Diller et al., 2016). The results suggested additional training in visual analysis may assist in determining effects (Diller et al., 2016). Also, providing environmental context may have helped increase scores.

**Supporting Viewpoints**

Many prominent behavior analysts use visual analysis and classify it as a conservative tool for examining behavior data (Baer, 1977; Michael, 1974; Parsonson & Baer, Kratochwill, & Levin 1992). However, the previously mentioned analysts did not conduct experiments to test their hypothesis. Two studies produced higher rates of agreement than previous studies indicating that visual analysis may prove reliable (Bobrovitz & Ottenbacher, 1998; Khang, Chung, Gutshall, Pitts, Kao, & Girolami, 2010). The importance of the findings suggest other issues may cause low agreement and high Type 2 error. Another possibility for higher rates could
lie in the experience of the participants. As suggested in previous research, experience in analyzing single case data may increase agreement (Ninci et al., 2015). For example, Khang and colleagues had participants that held editor roles in behavior analytic journals (2010). The authors found .89 agreement across displays.

The study replicated DeProspero and Cohen (1979) and demonstrated higher rates of agreement across raters (Khang et al., 2010). Khang and colleagues found a .89 absolute agreement which falls into the acceptable range (Müeller & Büttner, 1994). The participants included journal editors with a high level of educational attainment and years of experience. The findings suggest that visual analysis has reliability as a method of analysis. However, a key difference between the two manuscripts occurred with the concept of experimental control (Ninci et al., 2015). When considering experimental control, raters more clearly understood what they needed to make valid judgments. The authors demonstrated that better understanding of the procedures lead to enhanced decision making.

A study that investigated the difference between visual inspection and statistical tests further supported the reliability of visual analysis (Bobrovitz & Ottenbacher, 1998). The study used rehabilitation and health care workers analyzing data and then compared the results to statistical analysis. Agreement occurred in 86% of cases (Bobrovitz & Ottenbacher, 1998), supporting that visual analysis represents an adequate tool for interpretation of data. Namely, agreement occurred at higher levels with medium to large effects. The experimenters determined that effects can occur from both visual and statistical analysis. The manuscript tested ABAB designs and did not examine other types of single case displays. For example, would analysts determine similar conclusions with statistical testing with a multiple baseline design? Additionally, variability remained low across graphs in the experiment (Bobrovitz & Ottenbacher, 1998), supporting the view that when data remain stable, visual analysis proves reliable.
Supporting Visual Analysis

Supports exist to improve visual analysis. The literature proposes use of statistical methods, visual aids, and structured criteria for decision making (e.g., Brossart, Parker, Olson, Mahadevan, & 2006; Nelson et al., 2016; Roane, Fisher, Kelley, Mevers, & Bouxsein, 2013). Experimenters across studies recommend additional methods should support visual analysis and not replace it. The use of statistical methods and visual aids has largely appeared within the context of AB designs (Brossart et al., 2006; Parker, Cryer, & Byrns, 2006). The development of structured criteria specifically applies to FA and multielement designs (e.g. Hagopian, Fisher, Thompson, Owen-DeSchryver, Iwata, & Wacker, 1997).

Statistical Analysis

Statistical analysis of single case data has received attention in the literature. However, not all analyses achieve acceptable results. For example, a statistical test that violates the assumption of the independence of each observation would not provide effective analysis. Four manuscripts paired statistical analysis and visual analysis (Brossart et al., 2006; Davis, Gangè, Frederick, Alberto, Waugh, & Haardörger, 2013; Park, Marascuilo, & Gaylord-Ross, 1990; Parker, Cryer, & Byrns, 2006).

A group of experimenters supplemented visual analysis with Hierarchical Linear Modeling or HLM. HLM produced promising results as a support to visual analysis (Davis et al., 2013). The statistical test provides an effective size based on the series of data. For instance, once baseline and intervention data appear visually an effectiveness of an intervention can establish overall treatment impact. By applying an effect size to visual display, it may help provide added levels of analysis. Davis and colleagues demonstrated that HLM can support visual analysis.

Whereas HLM has produced a supplement to visual analysis, other statistical tests produced mixed results (Brossart et al., 2006). Researchers tested five different statistical analyses and found two significant results. Each comparison relied on ten data points per phase
and used AB designs. The insignificant tests may have resulted from using a design that does not conclusively demonstrate a functional relation. Statistical analysis that violates assumptions with regard to independence of data may not provide appropriate methods to supplement single case research designs.

Another study that tested rater decision making using AB designs found promising results but used subjective language to determine effects (Park et al., 1990). The terms used to describe effects included significant, nonsignificant, and unclear. The experimenters had judges determine effects visually and then ran statistical analyses, comparing results. The judges that participated in the study included one Ph.D. holder and four masters level behavior analysts. Each judge had submitted one behavior analytic manuscript to a peer reviewed journal (Park et al., 1990). Similar to Brossart and colleagues’ significant effects visually matched well with the results of the statistical testing. As level, trend, and stability became less clear, the statistical tests and raters did not match well (Park et al., 1990) further demonstrating that replacing visual analysis with statistical testing may improve effectiveness.

When examining data on graphical displays, it is important to consider baseline data before the implementation of an intervention. An increasing trend during a baseline phase may skew results, making interpretation of data difficult if not impossible (Parker et al., 2006). Parker et al. implemented two statistical tests to baseline trend and found significance. The Allison & Gorman (ALLIS) and Mean and Slope Adjustment (MASAJ) methods may help with decision making as a supplement to visual analysis. The experimenters used AB designs, which may have influenced results because of the design type (Ninci et al., 2015).

Statistical analyses demonstrate usefulness as a support to visual analysis. When comparing statistical tests such as HLM results to judge variability in the data, moderate differences between baseline and intervention, and variability in the intervention may cause a misrepresentation of results. Additionally, using HLM after an intervention may summarize the effects of an intervention but will not help during the intervention because the statistical test
measures overall effects. Therefore, the emerging data from statistical tests like the application of HLM may find their greatest utility to completed interventions rather than formative analyses.

**Visual Aids**

Visual aids may provide guidance to an analyst by summarizing a series of data. For instance, a goal line would highlight where data should appear on a display when a behavior has met a criterion. A trend line illustrates the direction of the data and the degree of change. The behavior analytic literature has examined the effectiveness of visual aids (e.g. Normand & Bailey, 2006; Ottenbacher & Cusik, 1991). Currently, more evidence suggests that visual aids improve accuracy in decision making (Fisher, Kelley, & Lomas, 2003; Hojem & Ottenbacher, 1988; Nelson et al, 2016; Ottenbacher, 1993; Ottenbacher & Cusik, 1991; Van Norman, Nelson, Shin, & Christ, 2013).

The studies that examined the effects of visual aids on decision making all indicated trend lines assist with accuracy in interpretation. Extreme variability in a data set negatively affects agreement and hinders effective decision making (e.g. Nelson et al., 2016). The utility of visual aids has mostly received attention with AB and ABAB designs (Fisher, et al., 2003; Hojem & Ottenbacher, 1988; Nelson et al., 2016; Ottenbacher, 1993; Van Norman et al., 2013). An analysis on the reliability of visual analysis recommended using visual aids to increase agreement (Ninci et al., 2015). Nelson and colleagues further support more training in how to omit extreme values, which includes the use of visual aids. While AB and ABAB designs remain prominent, consideration of other designs may produce different results (e.g. Alternating Treatments).

A study that focused on three experiments to increase agreement relating to treatment effects found behavior therapists benefited from visual aids (Fisher et al., 2003). First, the experimenters increased accuracy of visual inspection by training participants using the split-middle (SM), dual-criteria (DC), and conservative dual-criteria (CDC). The use of the techniques allowed for the placement of visual aids without having to use quantification of a data set. The ability to detect trends increased accuracy with visual supports as a tool for decision making with
the DC and CDC methods (Fisher et al., 2001). Second, five staff received training in the DC method. Third, the methods used to train the small group were replicated with a larger group. Mean effects increased 24-44%, supporting the use of visual aids.

Hojem and Ottenbacher demonstrated an increase in accurate inspection across two groups of participants with the use of trend lines for four of five displays (1988). Further, the experimenters support the use of quantification to support visual inspection. The participant group that used quantification had more extreme scores in agreement suggesting that although some variability existed in the data sets, the numerical values augmented confidence with responding. Although the graphical displays included used AB designs, the ability to detect effects with added supports can make visual analysis a more reliable method of inspection.

Most that examine effectiveness of visual aids use a group of participants and test their ability to determine effects. A literature review that examined agreement across fourteen studies found that trend lines produced a mean value of .64 and data without trend lines .59 (Ottenbacher, 1993) indicating that across years a higher rate of agreement occurred with trend lines. Controlling for hypothetical/real data, single case designs, and rater expertise, the study revealed a .58 agreement across 789 raters (Ottenbacher, 1993).

Last, a group of experimenters examined the use of visual aids on continuous measures (Van Norman et al., 2013). Participants examined graphs that had visual aids and no visual aids. The data revealed that visual aids and trend magnitude influenced correct identification of trend. Additionally, the degree to which progress appeared reduced correct responding. The ability to identify relations between interventions and behavior reduce when a dramatic effect does not appear on the display.

**Structured Criteria**

While visual aids assist with decision making, guidelines to make data based decision may further support visual analysis. Structured criteria offer specific criteria to assist in accurate interpretation of functional analysis (FA) data. Originally developed in 1997, the criteria
demonstrated that agreement increases across raters regardless of experience and educational attainment (Hagopian et al.). Two studies have examined the use of structured criteria for FA data (Hagopian et al., 1997; Roane et al., 2013). The criteria include several key features for decision making. First, each graph has standardized construction features such as scaling (15.3 cm x 21.6 cm). Second, the control condition determines the placement of criterion lines that assist in determination of function. Third, the number of data points in a test condition that fall above the criterion lines determines function. Fourth, specific guidelines exist for automatically reinforced functions (Hagopian et al., 1997).

One study tested the criteria on graduate students and found an increase in agreement after implementation (Hagopian et al., 1997). The experimenters used 10 data points per condition in the data, which closely mirrors traditional FA procedures (e.g. Iwata et al., 1994). Additionally, the data appeared on equal interval line graphs that required the same construction features to implement the criteria. The ability to provide standard construction of graphs, visual aids (i.e. criterion lines), and a quantitative value of data points above or below a line impacted agreement.

While ten data points per condition commonly occurs in practice, many behavior analysts use fewer points per condition depending on the type of FA conducted or individual client need. For instance, Trial Based Functional Analysis (TBFA) often uses less than ten data points per condition to determine function (Ruiz & Kubina, 2017). Further, some clients do not require the ten points per condition rule as the function of the behavior becomes apparent in five to six data points. Thus, the structured criteria tested with FA data included less than ten data points per condition (Roane et al., 2013).

Roane and colleagues used a similar procedures only modifying how many data points appeared per condition. The criteria produced similar results as Hagopian et al. (.92 agreement) suggested that guidelines in regard to decision making will increase agreement even with less data to make a decision. Similarly, Roane and colleagues had to develop non-standard equal
interval graphs with the same construction features to implement the criteria. The experimental design elements of standard graphical features further demonstrated the importance of appropriate line graph construction.

An important element to the replication of the structured criteria to consider includes the ability of visual aids and quantification to assist with agreement. With less information per condition, agreement increased with added tools, supporting the use of standardized construction, visual supports, and quantification. As a result, raters will more likely agree as to the relation between the target behavior and what maintains it.

Both studies found increased agreement regardless of educational attainment and experience (Hagopian et al., 1997; Roane et al., 2013). However, the structured criteria included other benefits assisting decision making. In other words, each graph in both studies had the same physical dimensions. For example, scaling of the axes remained constant. Another example, included the same number of graphical displays tested in each session. Controlling for the visual display elements may have contributed to the increased agreement as each graph provided standardized features and rating the same number of graphs prevented potential fatigue.

Errors in Graphical Construction in Multielement Designs

The ability to analyze behavioral data relies on appropriate graphical construction. For instance, if a graph does not scale appropriately, a trend line or structured criteria could lead to an incorrect conclusion (e.g., moderately decreasing instead of rapidly decreasing). Three of the seven dimensions of behavior analysis (Analytic, Effective, Generality) have a direct link to visual displays (Kubina, Kostewicz, Brennan, & King, 2017). Further, errors in graphical construction lessen the ability of supports for visual analysis (e.g. trend lines) and can affect supports of some statistical tests (e.g. PND).

A review of line graph construction in behavior analytic journals demonstrated a large amount of error (Kubina et al., 2017). The experimenters scrutinized the ratio of axes (i.e., proportional construction rule), data paths, tick marks, condition change lines, figure captions,
legibility of data points, label of each axis, and condition labels. The review examined varying designs and did not exclusively focus on multielement designs. While previous research has identified multielement designs as having higher rates of IRA, the construction of the graphs studied all used a standardized construction guideline (e.g. Hagopian et al., 1997).

Previously published literature reviews on FA identified the *Journal of Applied Behavior Analysis (JABA)* as the most frequent publisher of FA data (Beavers et al., 2013; Hanely et al., 2003). A hand search of the journal from 2003-2016 produced 135 articles that included an FA data set (See Appendix D). The year 2003 marked the initial publication of a literature review that reported JABA as the primary publisher of FA data (Hanley et al.) while 2016 represented the last complete calendar year where all issues of JABA were eligible for review.

Applying the construction features used by Kubina and colleagues’ each graph underwent inspection for construction errors. If a graph included used FA methodology it met inclusion criteria. Studies that targeted adaptive skills (e.g. exercise) did not meet inclusion criteria. Multielement and ABAB designs met inclusion. However, not all FAs use condition change lines thereby eliminating the feature from scoring. FA graphs then underwent inspection and each feature had a score of correct or incorrect. Each graphical display could achieve a maximum of 10 correct construction features and a minimum of zero correct. For example, if a data set did not include a data path that feature scored as incorrect for that display. Additionally, each feature has a percentage of error across graphs (See Table 1).

FA graphs had a lower error than the overall literature examined by Kubina and colleagues (See Table 1). Most common construction errors included the 2:3 proportional construction rule of the axis and spacing of tick marks. The 2:3 rule represents the scaling of the y-axis in comparison to the x-axis. Tick marks should have even spacing and appear on the outside of the graph. Of the 135 manuscripts reviewed 118 included 1-2+ errors (See Table 2). Seventeen of the manuscripts contained no error. The overall results indicate that an FA graph will more likely contain an error in construction than not.
FA data sets differ from those reviewed by Kubina et al. by using multielement designs and provide an assessment rather than experimental effects. Multielement designs use comparisons to demonstrate variables maintaining challenging behavior. Essentially, FA data compare test conditions against a control condition. When a test condition appears higher than a control, analysts can determine the consequence delivered maintained a challenging behavior (Betz & Fischer, 2011).

While error rates remain high with multielement designs the lower error rate in comparison to the results of Kubina et al. should receive further examination. Other than the purpose of the design in comparison to other types of single case graphs, the construction features of each display should follow the same rules. Further examination may assist in creating a lower error rate across designs.

**Conclusion**

The present literature review has three important conclusions. First, visual analysis can provide accurate conclusions with respect to effects of an independent variable on a dependent variable. However, when extreme values and/or variability occur in a data set then resolution of effects becomes compromised. Further, the type of design used impacts determination of effects. As suggested by one team of experimenters, some single case designs have increased agreement (Ninci et al., 2015).

Second, while some believe visual analysis functions effectively, the literature largely supports the need for improvement. Namely, using visual aids and quantification to increase agreement. Additionally, having criteria with quantification and standardized construction will increase agreement. The criteria have visual aids in the form of lines that help analysts determine the location of one datum in comparison to other data points.

Third, when reviewing FA graphs in *JABA* error rates in construction occur frequently. Based on the previously sourced literature it would appear as if using visual aids and criterion lines becomes difficult if manipulations of graphical displays occur. Further, the manuscripts
included in peer reviewed journals serve as models for those entering the field of behavior analysis. If those models violate construction rules future behavior analysts will most likely continue creating incorrect displays.

A graphical display that incorporates standardized features, visual aids, and quantification may provide more clarity without the need to train on additional procedures (e.g. statistical formulas). The field of behavior analysis continues to grow with new analysts. The multielement design lends itself well for assessment of challenging behavior and improving the likelihood of correct identification of results is critical.
Reference

*Indicates study was included for review


References for Question 3


delivery of food items and praise on problem behavior reinforced by escape.
http://doi.org/10.1901/jaba.2010.43-425

of variable-time versus contingent reinforcement delivery on problem behavior
maintained by escape. *Journal of Applied Behavior Analysis, 47*(2), 277-292.
DOI: 10.1002/jaba.110

stimulation and response interruption and redirection on vocal stereotypy. *Journal
of Applied Behavior Analysis, 45*(3), 549–564.
http://doi.org/10.1901/jaba.2012.45-549

analysis and supplemental feeding for postmeal rumination in children with
http://doi.org/10.1901/jaba.2007.743-747

schedules on resurgence of problem behavior. *Journal of Applied Behavior
Analysis, 47*(3), 455-469. DOI: 10.1002/jaba.134

Functional analysis and treatment of human-directed undesirable behavior
exhibited by a captive chimpanzee. *Journal of Applied Behavior Analysis, 44*(1),


http://doi.org/10.1901/jaba.2006.160-04


http://doi.org/10.1901/jaba.2010.43-47


## Tables

**Table 1: Percentage of error per feature.** *Not all FA’s use Condition Change lines.*

<table>
<thead>
<tr>
<th>Graphical Construction Feature</th>
<th>Percentage of Manuscripts with Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertical axis labeled with a quantitative measure; horizontal axis with time unit</td>
<td>4%</td>
</tr>
<tr>
<td>Vertical axis length has 2:3 ratio of length to the horizontal axis</td>
<td>61%</td>
</tr>
<tr>
<td>Tick marks point outward</td>
<td>2%</td>
</tr>
<tr>
<td>A minimal number of evenly spaced tick marks</td>
<td>43%</td>
</tr>
<tr>
<td>Tick marks have labels</td>
<td>13%</td>
</tr>
<tr>
<td>Data Points clearly visible</td>
<td>1%</td>
</tr>
<tr>
<td>Data path clearly visible</td>
<td>0.7%</td>
</tr>
<tr>
<td>Condition Change lines*</td>
<td>N/A</td>
</tr>
<tr>
<td>Condition labels</td>
<td>0.7%</td>
</tr>
<tr>
<td>Figure caption</td>
<td>5%</td>
</tr>
</tbody>
</table>

**Table 2: Number of manuscripts with 0, 1, and 2+ errors.**

<table>
<thead>
<tr>
<th>Number of Errors</th>
<th>Manuscripts</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Error</td>
<td>17</td>
</tr>
<tr>
<td>1 Error</td>
<td>59</td>
</tr>
<tr>
<td>2+ Errors</td>
<td>59</td>
</tr>
</tbody>
</table>
Appendix E
Recruitment Materials

Recruitment Email:

Hello BACB Certificate Holder,

Thank you for taking the time to view this email. We are seeking BCBAs to participate in a survey that examines the effects of a standardized graphical display on decision making for Board Certified Behavior Analysts. The data displayed are from Functional Analyses published in peer-reviewed manuscripts. The findings will be part of a doctoral dissertation that examines issues in decision making and graphs. Should you choose to participate please find the following information: link to tutorial, information regarding the survey, and a link to the survey, which is completed online. Please click or copy and paste the links into your browser should you decide to participate. Please contact the principal investigator should you have any questions.

Sincerely,

Salvador “Sal” Ruiz

Link to Tutorial: https://www.youtube.com/watch?v=XPqRYFE5n8A&feature=youtu.be

Link to Survey: https://pennstate.qualtrics.com/jfe/form/SV_2nQewLlGHo9m6Z7
Informed Consent:

Penn State University

Special Education Department

Principal Investigator: Sal Ruiz M.A., BCBA

Faculty Advisor: Richard M. Kubina Jr., Ph.D, BCBA-D

Title of Study: Effects of the Functional Analysis Celeration Chart on Decision Making for Board Certified Behavior Analysts

You’ve been invited to participate in a research project entitled “Effects of the Functional Analysis Celeration Chart on Decision Making for Board Certified Behavior Analysts.” The project is supervised by Dr. Richard M. Kubina. This consent document will explain the purpose of this research project and will go over all of the time commitments, the procedures, risks and benefits, and contact information. Please review the information provided and feel free to contact the PI with any additional questions.

Purpose of the Study

The study is designed to determine if a standardized graphical display called the Functional Analysis Celeration Chart (FACC) can assist Board Certified Behavior Analysts interpret Functional Analysis results. Additionally, demographic information will be collected to determine preferences and experience across participants.

Participation Eligibility

All Board Certified Behavior Analysts over 18 years old can participate in the survey. Further, all participation will be anonymous.

Time Commitment

If you decide to participate, it will take approximately 30 minutes to view the tutorial and complete the survey. There will be no further action on your behalf after completion of the survey.
Tutorial & Questions

You will be asked to view a brief (~5 minutes) tutorial and answer survey questions. The tutorial will be available via youtube.com and a link will be embedded in an email. The survey will include 18 graphical displays with varying types of results. Further, ten additional questions inquiring about experience and preferences will be included in the beginning and end of the survey. One question will request an email address if you would like to be entered for a chance to win a Visa Gift Card™.

Risks & Benefits

There are no known risks to participate in this study. Benefits include entering to win a $50 Visa Gift Card™. Participants that enter their email address will be randomly selected and contacted.

Access to Data

The survey and tutorial are anonymous. If you choose to enter to win the Visa Gift Card™ your email address will be stored in a separate password protected folder. Your responses to questions will not be associated with your email address. The PI and faculty advisor will be the only people to have access to the data. All data will be kept in a password-protected file on the principal investigator’s computer. The data may be used in conference presentations and/or manuscripts for publication in peer-reviewed journals. Your identity will not be disclosed.

Participation

You may choose to stop participation at any time for any reason. You will not be penalized in any way by stopping participation. You will not receive any punitive consequences if you choose to withdraw from this study.

If you have any further questions at any point you may contact the principal investigator, Salvador Ruiz M.A., BCBA at 201-575-7470 or sur190@psu.edu.

Participation in this survey indicates your consent for use of the answers you provide.
VITA

Salvador Ruiz

EDUCATION

Ph.D. in Special Education
The Pennsylvania State University 2018

M.A. in Psychology
The Chicago School of Professional Psychology 2014

B.A. in Sociology
The William Paterson University of New Jersey 2008

PROFESSIONAL CERTIFICATIONS

Approved BACB Supervisor, Behavior Analyst Certification Board
Board Certified Behavior Analyst (BCBA) Certificate No. 1-14-10394
Certificate for Online Teaching, The Pennsylvania State University (World Campus)

PROFESSIONAL EXPERIENCE

Feb 2017- Present  Behavior Analyst. Chartlytics LLC. State College, PA.
Aug. 2014- Present  Graduate Assistant (Special Education), The Pennsylvania State
University. University Park, PA.
Jan. 2012- Jul. 2014  Behavior Specialist, Bergen County Special Services, Paramus, New
Jersey.

PUBLICATIONS

behavior and training: A review of the literature. Behavior Analysis Research and Practice, 17,
347-356. doi: 10.1037/bar0000079

single-case treatment effects on self-injurious behavior for individuals with autism
and intellectual disabilities. Autism and Developmental Language Impairments, 2, 1-26. doi:
10.1177/239694151668839