Exploring the utility of differing methodological approaches to measure meaningful change in treatment and intervention scenarios

Bailey D. Ross, B.A.Sc.

Applied Disability Studies

Submitted in partial fulfillment of the

requirements for the degree of

Master of Arts

Faculty of Social Sciences, Brock University

St. Catharines, Ontario

Abstract

A large focus in social science research is geared towards establishing the effectiveness of treatments to help and support the population and provide best empirical data to researchers, practitioners, and policy makers. To support evidence-based practice and policy, findings from studies are synthesized in reviews and meta-analyses to verify the most effective treatments. However, due to differing metrics to describe effects found, single-case designs (SCD) have been excluded in such reviews and meta-analyses. This hinders the dissemination of valuable findings from studies that use SCD methodologies. The present study employs a unique dataset to exemplify differing methodological approaches to measure meaningful change from a treatment. The dataset, obtained from Vause and colleagues (2018), contained both a group-based design in the form of a randomized controlled trial and SCD methodologies on the same participants undergoing treatment. Thus, a d-statistic was calculated from the SCD methodology, also referred to as a between-case standardized mean difference effect size (ES_{BC}) and was compared with the group-based effect sizes originally found by Vause and colleagues. The effect sizes corroborated with each other, such that a large effect was deduced from both the SCD analysis (g = 1.22) and the average effect from the group-based analyses (g = 0.99). In addition, the ES_{BC} was found per participant allowing comparisons between individual effects and the overall outcome. Furthermore, this study explored how the acquired ES_{BC} estimates complements traditional SCD methodologies including visual analysis and overlap statistics. By utilizing statistical techniques such as this software, many behaviors, participants, and data points can be analyzed simultaneously. Moreover, a forest plot can be generated with the results, providing a different perspective than what is normally available to SCD researchers. Finally, the most valuable consequence to note is the acquisition of a ES_{BC} that results in the form of

Hedges' *g*, which can be compared across SCDs and between-group experimental designs. This is the first known study to explore and compare the effect size estimates of a treatment on participants' behaviors that was evaluated as a SCD as well as a group-based design.

Keywords: single case design, effect size, between-case standardized mean difference effect size, hedges' g, d-statistic

Acknowledgements

What is not seen within the words of this thesis are the new friends, the numerous conversations had, the countless coffees consumed, the cross Canada trips, the tears and laughs, the inevitable technology battles, and the certain co-dependency with my dog, Pi. Through all of this, there are particular people that must be acknowledged that supported this thesis!

Thank you to Dr. Jan Frijters and our lab. Over a cup of coffee, of course, you listened to my initial curiosity and speculation in this topic and you were able to set me on a path of exploration and discovery. The lab provided me with motivation to keep writing. Whether it was walks to Tims or our Goal Meetings throughout the pandemic, you inspired me and lifted me up!

Thank you to my committee members, Dr. Alison Cox, Dr. Tricia Vause, and Dr. Adrienne Perry for your interest, feedback, and support on this project. A big thanks to Dr. Vause for providing me the opportunity to utilize data from her previous research.

Thank you to my people. Mike for your support, cross Canada visits, and unwavering encouragement. The Brock Behaviour Babes for always answering my questions and validating me. Mom and Dad, thank you for helping me tackle new situations, engaging in furniture hunts, and always being up for an adventure. Tanner and Maymie, thank you for those long talks and inspiring me to keep balance in my life. The Entire Evans Family, thank you for always believing in me and being my second home and Uncle Clay for checking in and keeping me on track! And a final thanks to my friends and family across the world that have provided me with virtual laughs and love along the way.

Table of Contents

List of Tables

List of Figures

Introduction

The purpose of science is to obtain knowledge and understanding of the topic that is being researched and studied (Cooper et al., 2020). As well as identifying new information, researchers attempt to fill gaps in the current literature, replicate or contradict previous results, and disseminate the findings. A large focus in social science research is geared towards determining if treatments and interventions are effective at helping and supporting the population. Some examples can include pharmaceutical interventions, types of therapies, or behavioral interventions (Borenstein et al., 2009). The common objective among these studies is to provide practitioners with the best empirical data to support evidence-based practice (EBP).

To discover best practice, evidence and results of studies must be compared to substantiate what would be most effective (Borenstein et al., 2009). The results of a single study are not conclusive, but can add to an existing body of research, or can start a new body. It is only when studies are taken together that robust conclusions can be drawn about a phenomenon (Cumming, 2014). Therefore, methods have developed over time to effectively compare many findings to provide other researchers and practitioners with current and crucial information. Prior to the 1990s, an expert within a field would familiarize themselves with studies that addressed a particular question and then summarize their findings to verify if the treatment was effective or not (Borenstein et al., 2009). However, this was very subjective and cannot be maintained as literature grows. As more information becomes available, it becomes increasingly difficult for a reviewer to compare treatment effects and consider other contextual variables. As such, in the 1990s, researchers sought other methods to objectively and efficiently compare and contrast findings across numerous studies. These methods included systematic reviews and meta-analyses.

The use of a systematic review locates relevant studies to address a similar topic or a research question (Siddaway et al., 2019). To complete a systematic review, a methodological, replicable, and transparent technique is used, and inclusion and exclusion criteria are clearly stated. Once data is collected from the selected studies, a statistical synthesis, referred to as a meta-analysis, is performed (Borenstein et al., 2009). To be included in a meta-analysis, the evidence and results included by studies must be in a standardized metric. The standardization of the results in a comparable metric is what allows differing studies to be compared. However, some studies do not produce a standardized metric and thus are excluded from these reviews. For example, single case designs (SCD) generally do not result in a standardized metric comparable with other results, therefore, these results may be excluded from meta-analyses (Shadish & Rindskopf, 2007). This unfortunate exclusion hinders the potential for SCD results to inform EBP alongside other results (Shadish et al., 2015).

Single case designs have the potential to provide important information regarding the effectiveness of interventions and treatments that could contribute to EBP; however, the requirements to be included in meta-analyses to inform EBP have hindered this dissemination. Finding a way to merge SCD results with results more commonly included in meta-analyses, such as those from randomized controlled trials (RCTs) or other group-based designs, will in turn promote dissemination of findings and inform EBP and policy. Thus, the following pages will review the importance of EBP, strengths and limitations of selected research designs, and how results from different methodological designs may be merged.

Evidence Based Practice

The application of EBP is based on a model of professional decision making that aids in structuring professional practice and may influence policy (Slocum et al., 2014). The approach of

EBP includes a comprehensive collection of information and evidence based on three major components: practitioner expertise, acknowledgement of the values, differences, and circumstances of the target sample, and the best available evidence from research (APA Presidential Task Force on Evidence-Based Practice, 2006; McKibbon, 1998; Slocum et al., 2014). It is recognized, and at times mandated by law, that fields such as education, psychology, and healthcare should utilize practices that have previously been verified to be effective (American Psychological Association [APA] Presidential Task Force on Evidence-Based Practice, 2006; Cooper et al., 2020; Shadish & Rindskopf, 2007). Such practices include consultations, assessments, diagnosis, prevention, interventions, and treatments.

Practitioner Expertise

Although there is not a recognized operational definition of practitioner expertise, it has been found that there are significant differences between experts and novices within a field (APA Presidential Task Force on Evidence-Based Practice, 2006; McKibbon, 1998). Experts possess several characteristics that they have acquired through practice that aid them in their decision-making. They often can recognize and disregard irrelevant information, acquire, retrieve, and organize their knowledge to effectively represent their deep understanding of practice and utilize it functionally. They are flexible and adapt in situations, self-monitor their own practice, and achieve outcomes in their practice that appropriately reflect their expertise (APA Presidential Task Force on Evidence-Based Practice, 2006).

Expertise in one's practice is developed overtime from education and theoretical understandings, training, professional experiences, self-monitoring and reflection, background knowledge of research, and continuous education of current information (APA Presidential Task Force on Evidence-Based Practice, 2006). In some fields, such as applied behavior analysis

(ABA), there is an ethics code to promote practitioners' boundaries of competence and maintaining competence through professional development (Behavior Analyst Certification Board [BACB], 2014). This is a way in which one's expertise can develop in their field.

Some practitioners may find that relying on their own knowledge and experiences is quicker when they need to plan within their practice (McKibbon, 1998). However, to effectively utilize the approach of EBP, the practitioner will use their expertise which can include their ability to evaluate an individual's circumstances and integrate research findings into their practice (Babione, 2010).

Individual Differences and Circumstances

In accordance with EBP, an individual's characteristics, culture, and preferences should influence practitioner's decision making about applying practices such as assessments, interventions, or treatments (APA Presidential Task Force on Evidence-Based Practice, 2006; Babione, 2010; BACB, 2014; McKibbon, 1998). Such characteristics include, but are not limited to, biological factors, familial factors, socioeconomic barriers, personal supports, cultural components, environmental variables, diagnoses, personal values and beliefs, and personal history.

Recognition of individual differences and how this may influence practice is a key component to effectively utilize EBP (APA Presidential Task Force on Evidence-Based Practice, 2006; Babione, 2010; McKibbon, 1998). The BACB (2014) state in their ethical code that "behavior analysts must tailor behavior-change programs to the unique behaviors, environmental variables, assessment results, and goals of each client" (p.12). As well as adapting their practice based on the individual, they recognize that their clients have a right to effective treatment that is based on research literature.

Available Evidence from Research

The final component of EBP is basing practice on relevant and up to date peer-reviewed scientific literature (Babione, 2010; Cooper et al., 2020; Dozois et al., 2014). Scientific results that are published in research literature support the best evidence of intervention and treatment strategies, assessments, different clinical problems, and samples and populations in controlled and natural settings (APA Presidential Task Force on Evidence-Based Practice, 2006). The use of scientific literature to inform one's professional practice is often expected and is even written into the ethical code for behavior analysts (BACB, 2014).

Often, scientific evidence will support the efficacy of one type of intervention or treatment for a specific problem under precise conditions (APA Presidential Task Force on Evidence-Based Practice, 2006). Thus, research does not always address all issues or circumstances a practitioner comes across. In this situation, practitioners are expected to investigate literature that is relevant to the individual's characteristics, potential interventions and treatments, and the main concern.

When investigating literature, Dozois and colleagues (2014) described a hierarchy of research evidence. Weaker evidence is placed lower on the hierarchy and is described as unpublished data, opinions, or past experiences. However, they advise that such evidence should only be considered if better evidence is not available or does not exist. At the top of the hierarchy is the strongest evidence which is a systematic synthesis. This includes systematic reviews and meta-analyses. This is considered the best and strongest evidence because it uses standardized statistics to compare across multiple studies. Hence, to utilize this approach it requires that the included studies report their findings as a standardized result (Borenstein et al., 2009). In addition, the ideal results that are included have been replicated across studies and have utilized

methodologies that consider threats to validity such as internal validity, external validity, generalizability, and transferability. Nonetheless, when considering research evidence, the overall methodological quality of a scientific study is dependent on the research design type and how the research is conducted (Meline, 2006).

Research Designs

In behavioral science, correlational and experimental research designs are utilized (Morgan & Morgan, 2009). Researchers compare control and experimental conditions to determine if there has been an effect of the independent variable on the dependent variable (Johnston & Pennypacker, 2009). The most commonly used research designs in behavioral science are between-group designs and within-subject designs, also known as SCD.

Between-Group Design

A between-group design divides participants into a control condition group and an experimental condition group to evaluate if a treatment works (Johnston & Pennypacker, 2009; Price et al., 2015). The participants within this type of design are randomly assigned to either a treatment or control condition which controls extraneous variables across conditions. Pretest scores are collected from each group of the dependent variable and then once treatment is completed, post-test scores are collected (Cooper et al., 2020). To evaluate the effect of the treatment, the researchers consider the change in post-test to pretest scores of each group. This type of design, an RCT, is considered the *gold standard*, and is the strongest design to answer a research question concerning causality (Meline, 2006).

Sometimes it is not feasible or ethical to employ an RCT and so researchers will utilize a quasi-experimental design (Harris et al., 2006). These types of designs evaluate interventions or treatments, but they do not employ randomization. This design may be more appropriate for

ethical reasons, difficulty randomizing participants or locations, or there may not be a sufficient sample size. There are a few different types utilized such as using control groups (comparison groups) with or without pretest, not using a control group, or interrupted time-series designs.

Strengths. Between-group designs such as RCTs are known as having the ideal methodology for strong internal validity which establishes causality (Deaton & Cartwright, 2018; Meline, 2006). Although quasi-experimental designs may have less internal validity due to lack of randomization, they are considered to be between correlational studies and true experiments (Harris et al., 2006; Morgan & Morgan, 2009). In addition, some researchers assume that the averaging of group scores of many participants helps to control for inter-subject variability (Cooper et al., 2020). Another assumption is that the larger number of participants within a study will increase the external validity of the results.

Another strength of between-group studies is that often the statistical significance and effect size will be reported as a standardized statistic. The value in reporting an effect size is that regardless of the measurement scale used in the study, the effect size allows different studies to be compared (Hedges, 2008). Generally, the between-group effect size reported is the standardized mean difference effect size, Cohen's *d*. This effect size compares the mean from a treatment group to the mean of a comparison group and that mean difference is standardized (Shadish et al., 2015; Sullivan & Feinn, 2012). This effect quantifies the strength of the relation between the dependent and independent variables in the experiment (Borenstein et al., 2009; Hedges, 2008). Thus, this universal measure is the common unit among studies that allows them to be compared and analyzed in a meta-analysis to establish an overall summary of effect of the treatment of interest. The reported effect size of a treatment eventually informs EBP through dissemination in meta-analyses.

Limitations. Although some believe many participants increases the external validity of results, Meline (2006) suggests that between-group designs have weaker external validity due to restrictive selection criteria. This suggests the results from these studies are unable to generalize across other clinical populations, settings, or behaviors (Johnston & Pennypacker, 2009).

Groups-comparison approach has been commonly utilized in behavioral research and does work well for some form of treatment effects; however, it may not be appropriate or support all types of treatments that practitioners utilize (Cooper et al., 2020). For example, when the focus is on the behavior of individual subjects, such as in the field of ABA, a group-based design is difficult to employ due to the nature of targeting individual subjects and behaviors. When responses are pooled in a group-based design, it does not allow information on an individual's behavior. Without individual measurements, it is challenging for researchers to determine any extraneous factors that could be influencing an individual or a group of individuals (Johnston & Pennypacker, 2009). Thus, a different type of design, such as SCD, should be employed.

Within-Subject Design

A within-subject design, also referred to as a SCD, is characterized by three main features: a case of a single participant or a small cluster of participants, this case provides its own experimental control within the design, and this case is compared within and across different conditions such as with or without intervention or treatment (Kratochwill et al., 2010). Based on these features, the goal is to gain knowledge of the cause-effect relationship between an environmental stimulus such as an intervention or treatment and the behavior of interest (Bailey & Burch, 2002). The experimental research design of SCDs have different variations; however, they all include repeated and systematic measurement of the dependent variable, such as a

behavior, before, during, and after, the application of a treatment or intervention (Kratochwill et al., 2010; Poling & Grossett, 1986).

To establish experimental control in a SCD a replication of the effect of the intervention or treatment must be demonstrated (Kazdin, 2011). Replications in SCD occur when an experiment can reproduce an observed behavior change by manipulating the same independent variable (Cooper et al., 2020). Replications of behavior changes can be demonstrated within a participant and between participants (Lammers & Badia, 2005). This is referred to as intra-participant replication and inter-participants replication, respectively. The type of replication depends on the SCD variation employed. Such design variations include reversal, alternating treatment, multiple-baseline, and changing criterion (Cooper et al., 2020). The reversal design, alternating treatment design, and changing criterion design attempt to replicate effects within a participant. The multiple baseline design can demonstrate replication across participants, behaviors, and settings which allows evaluation of the independent variable when it is unable to be withdrawn or reversed. In addition, SCDs can be employed as a combination. For example, it is possible to have a reversal design embedded within a multiple baseline design. This combination of analytic tactics may provide an even stronger case for experimental control because the effect is demonstrated within and across a participant, behavior, or setting.

Within the SCD variations, rather than taking measures of the behavior once or a few times in a large group, like in between-subjects design, repeated measurements of the behavior of interest are taken during and without intervention or treatment (Poling & Grossett, 1986). In the condition without intervention or treatment, it is assumed that the behavior will continue without change and there would be an absence of any trend or slope in the data (Kazdin, 2011). With an intervention or treatment, a researcher is looking for a behavioral change. When the data of the

behavior in intervention or treatment differs from the predicted trend in the prior condition without intervention or treatment, this establishes a behavior change (Odom et al., 2018). Once this behavior change has been established, a replication of this effect will reduce the probability of another variable being responsible for this change and it determines that the behavior change is reliable (Cooper et al., 2020; Lammers & Badia, 2005).

Strengths. The SCD involves steady state strategy and baseline logic which is made up of three elements: prediction, verification, and replication (Cooper et al., 2020). With this strategy and elements, threats to internal validity can be controlled. The baseline phase of SCD acts as a form of prediction such that when the independent variable is not applied, the consistency of the data collected will determine or predict subsequent data points. Then once the independent variable is applied, the researcher may determine that the change in behavior is functionally related to the independent variable by removing the independent variable and verifying the first prediction of baseline levels. This return to baseline levels verifies the accuracy of the prediction. Next, to confirm the functional relation between the independent variable and the behavior, the independent variable is once again introduced. If the behavior change is observed again, this replication establishes that it is reliable. Thus, the experiment is considered internally valid because the overall effect can be attributed to the independent variable based on these three elements.

Other considerations that can be made is the type of variation of SCD and how that can influence internal and external validity. Depending on the type of SCD, it may have stronger internal and external validity contingent on its structure and replicability of the intervention effects (Kazdin, 1981; Kratochwill et al., 2010). If the research design factors in phase repetition and replication of the effect, then it is considered to have strong internal validity, as mentioned

above. The replication of the effect can occur through within-case or multiple-case replication in one experiment. Replication can also occur by conducting more than one experiment that is like the one being replicated. This may be across participants, settings, target problems, or other variables of interest (Kazdin, 1981). When replication of an effect occurs across different conditions, this increases the external validity.

Lastly, the traditional analyses applied to SCD provides a clear answer for ABA practitioners. Within the field of ABA, there is a commitment towards improving people's lives by improving their behavior (Cooper et al., 2020). Thus, SCDs are employed to confirm the functional relationship between the intervention or treatment and the potentially changed behavior. The change in behavior is confirmed by employing visual analysis which is dependent on graphing (Kazdin, 2011; Parsonsons & Baer, 1986; Poling & Grossett, 1986). Often the effect being sought after, a socially significant change of behavior, is potent and apparent enough via visual analysis (Cooper et al., 2020; Kazdin, 2011). A socially significant change means the intervention or treatment affects the behavior in a way that improves the individual's life. Visual analysis is an appropriate tool for this design when the effects of the intervention or treatment are of interest for one or a few cases.

Limitations. Although threats to internal and external validity can be dealt with, there are times when a SCD is unable to address these threats. Such threats that can occur include a lack of replication, maturation as a confounding variable, undesirable conditions in which selection occurs, and attrition (Kratochwill et al., 2010). These threats can be difficult to control for but will influence the outcome. A lack of replication decreases the probability that the effect is reliable. Hence, the functional relationship between intervention or treatment and the target behaviors cannot be confirmed.

When confirming a functional relationship between the intervention or treatment and target behaviors, traditionally SCDs employ visual analysis. Visual analysis interprets graphically displayed data to conclude if the behavior changed in a meaningful way for the participant and how much that change is from the intervention or treatment (Cooper et al., 2020). Although this type of analysis is suitable to help determine if the change is socially significant for that participant (as described above), concerns can be raised regarding the consistency across and even within raters, sensitivity, and specificity of this form of analysis (Kazdin, 2011; Lammers & Badia, 2005).

Admittedly, visual analysis provides information that is appropriate for practitioners or researchers of that study to establish if the behavior change was the desired effect for the participant. Thus, researchers and practitioners should not refrain from the use of visual analysis since it does provide valuable information. However, one aspect of visual analysis that has been brought up as a concern is that there is a lack of concrete decision rules when visually analyzing the graphic display of data (Kazdin, 2011). This can be an issue because researchers may differ in how they weigh the impact of the graphic features to verify an effect. This means that the visual analysis of data can be subjective and produce unequivocal outcomes between researchers and practitioners (DeProspero & Cohen, 1979).

Some past studies have explored how variable responses can be when employing visual analysis. To examine variability of responses, DeProspero and Cohen (1979) and Ottenbacher (1990) recruited participants to visually analyze a set of simulated graphs. Both studies found that there was variability in the interpretations of the graphs. The discrepancy in interpretations may call into question the reliability of the results. Another study completed by Danov and Symons (2008) examined rater reliability when utilizing visual analysis on functional analysis

graphs. They recognized that raters of the graphs may not agree on the presence or absence of differentiated response patterns and their corresponding behavioral function. Their overall results indicated that there was only a moderate interrater agreement when using visual inspection suggesting a low level of reliability of the interpretations of results.

Although the SCD community values visual analysis to verify treatment effects on target behaviors, other scientific fields may not be familiar with these standards. The effect that is reported based on visual analyses is not in a metric that will specifically quantify the treatment effects (Odom et al., 2018). This is a significant limitation. In addition to not reporting an effect that can be recognized by other researchers, this effect is also not in a metric that can be compared with other findings, such as those from group-based designs. A lack of reporting a universally understood standardized effect size that communicates the strength of the treatment leads to these findings not being merged or compared with other studies (Hedges, 2008). Hence, this conflicts with the goal of dissemination and inadvertently excludes SCD results from EBP reviews and meta-analyses (Shadish et al., 2015).

Exclusion of Single-Case Designs from Research Syntheses and Meta-Analyses

A research synthesis is recognized as being the most common type of literature review within social science research (Cooper, 2017). Research syntheses consolidate and compare empirical research to make generalizations about the findings, analyze the research, and determine other issues to investigate in future research projects (Cooper et al., 2009). From this synthesis of literature, a meta-analysis can be conducted to statistically analyze the results, in the form of effect sizes, from the compiled studies (Card, 2011). This type of analysis synthesizes all findings to draw main conclusions about the question being asked and will in turn inform future researchers and practitioners about the common effect found.

There are several steps that occur when a research synthesis is conducted. Cooper (2017) specifically lists seven steps as: 1) Formulate the problem, 2) Search through the literature, 3) Gather information from the selected literature, 4) Evaluate the quality of the studies, 5) Combine and analyze overall results form the studies, 6) Interpret the findings, and 7) Display the results. These steps, although necessary to complete the synthesis, also contribute to why some studies are excluded.

Although these steps are systematic, based on the scope of the synthesis, the process can still be considered biased (Cooper et al., 2009). The overall findings will be influenced based on the types of literature considered; published or unpublished (McKenzie et al., 2019). To begin searching for literature, often reference databases are used, such as Google Scholar, Science Direct, PsycINFO, ERIC, and Medline. These sources are known to be inclusive; however, a lot of unpublished literature is excluded in these systems (Cooper et al., 2009). Thus, some researchers will attempt to access their colleagues or other researchers unpublished work in that area of interest.

Other factors that may influence which literature is collected include particular populations of interest, specific interventions or treatments used, an overall result that is of interest, the length of time the topic has been researched, and a focus on certain research designs (Cooper et al., 2009; McKenzie et al., 2019; Shadish & Rindskopf, 2007). Of course, a specific research question will impact which literature is sought. For example, perhaps a research question is considering the most effective treatment to reduce obsessive compulsive behaviors (OCBs) in children and adolescents. Hence, the search parameters would most likely exclude studies involving an adult population and would include OCB treatments and interventions. Furthermore, to find a more precise answer to the question, the researchers may consider studies

about treatments that utilize specific strategies. Or perhaps researchers may consider studies in which participants have a specific type of diagnosis or other criteria. And finally, researchers may exclude certain study designs.

The exclusion of particular study designs may be based on the preference some scientists have towards randomized experimental designs and as Shadish and Rindskopf (2007) stated, "the statistical theory that guarantees that the average effect size over a large number of randomized experiments is a better approximation to the population parameter than is possible in any single experiment" (p. 96). Although this preference may exist, it is acknowledged that other research designs such as SCDs can provide useful information regarding treatment effects (Sandbank et al., 2020). Even though both types of designs, between-group and within-case, can provide valuable information regarding treatment or intervention effects, currently not all results are compatible to be synthesized in a meta-analysis.

Unfortunately, there is a lack of research comparing results of group-based randomized experiments to SCD (Shadish, Hedges, & Pustejovsky, 2014). This gap in literature exacerbates the problem that important findings from SCD research and group-based research are not being merged. This directly affects future research and EBP. Single-case designs are often used to determine which interventions and treatments are effective; however, Shadish and Rindskopf (2007) found that, out of thousands of meta-analyses completed over the past three decades that considered experiments of interventions and treatments, there were only 24 meta-analyses of SCDs. In addition, Burns (2012) completed a PsychINFO search that combined the term "meta-analysis" with articles published in the *Journal of Behavioral Education* and the *Journal of Applied Behavior Analysis* and found only two articles out of 3,077. Thus, the exclusion of SCDs is a limitation in meta-analytic research (Allison & Gorman, 1993).

To further exemplify this issue, a recent meta-analysis competed by Sandbank and colleagues (2020) focused on nonpharmacological early interventions for children with ASD. The selected studies met an inclusion criterion that specified they must be a "group design that included both an intervention and control group" (p. 7). Their reasoning to focus on only group-based designs is written in a public significance statement: "Behavioral intervention approaches also show evidence of effectiveness, but methodological rigor remains a pressing concern in this area of research" (p. 2). They acknowledge later in the paper that by excluding SCDs they have had to omit a large body of research that does contribute to EBP using effective behavior analytic interventions and strategies.

The above example clearly represents how the exclusion of important and effective findings from SCD research is occurring within the scientific community. The most prominent issue is the compatibility of synthesizing evidence from group-based designs and SCDs. This is highlighted by Sandbank and colleagues (2020) by stating that their main reasoning to exclude SCDs is that they do not have a comparable effect size metric to synthesize all effects. Thus, to effectively combine and compare results from all methodologies a common standardized effect size must be utilized by SCDs.

Effect Size

The American Psychological Association (2020) suggests that for a reader to understand a study's findings, there should be an effect size included in the results. As such, the reporting of a standardized effect size is imperative to the inclusion in meta-analyses. However, some research designs, like SCDs, have not been able to calculate a standardized effect size because they are not based on the comparison of two groups, such as treatment and control. For this

reason, results of SCDs do not include a metric that is comparable to group-based designs (Vannest et al., 2018).

Within-Subject Design Effect Size

Despite the difficulty of calculating a standardized effect size, there have been other methods to reporting SCD findings from fields of research such as psychology, social work, speech and language fields, special education, and ABA (Kratochwill & Levin, 2010). In SCDs, some effects are reported using regression-based statistics or overlap statistics (Shadish et al., 2015). These methodologies allow conclusions to be drawn about each separate case within the study. An example of such methodologies includes the calculations of percentage of non-overlapping data (PND) and percentage of data points exceeding the median of baseline phase (PEM). Although these methods can determine an estimate of the effect found, the outcomes are not a comparable metric to between-group design effect size estimators, such as Cohen's *d* (Shadish et al., 2008).

The standardized mean difference effect size, Cohen's *d*, can be referred to as a between-case effect size. This effect size compares the mean from a treatment group to the mean of a comparison group and that mean difference is standardized (Shadish et al., 2015). This form of effect size does not depend on sample size and only uses the underlying population parameters (Hedges, 2008). Additionally, the effect size is not influenced by the scale that is used to measure the outcome variable. This allows effect sizes from different studies to be compared even if the studies used different measurement scales. However, in SCDs the focus is on examining individual behavior change rather than average effects across groups of participants (Odom et al., 2018). This affects how statistical analyses can be performed.

There are methodological considerations that must be made when analyzing the effect in SCD. Due to the repeated measures of examining individual behavior change, it can be challenging to utilize statistical analyses to assess treatment effects (Hedges et al., 2012). Repeated measures are not independent and involve an autocorrelation or serial dependence structure which is complicated to include in statistical analyses (Kazdin, 2011). Furthermore, due to the differences in data stability in each phase, different statistical analyses would be needed to compare among each phase (Hedges et al., 2012). These considerations lead researchers to explore new forms of analyses.

Some researchers view new strategies to analyze SCDs with skepticism, whereas others are intrigued (Shadish, Hedges, Pustejovsky, Rindskopf et al., 2014). Thus, the evident dilemma of being unable to merge methodologies and appropriately analyze results has motivated SCD methodologists and statisticians to develop new statistical methods. In the last few years there have been developments towards analyses involving effect size indices or methods that are based on general linear modelling. The methods that involve general linear model approaches do not provide a common metric with group-based effect sizes and the effect size indices have had issues with statistical properties. Hence, Hedges and colleagues (2012) proposed a new effect size measure that would focus on a single type of stable pattern. These estimates of effect sizes are the same parameter as between-subject design effect sizes (Cohen's *d*). Therefore, SCD research can be included in meta-analyses and EBP reviews with results that were produced by other experimental methods (Shadish, Hedges, Pustejovsky, Boyajian et al., 2014).

This effect size metric, a *d*-statistic, can be referred to as the between-case standardized mean difference statistic due to its ability to compare across two types of study designs, SCDs and between-group experimental designs (Valentine et al., 2016). This statistical technique can

be applied to reversal designs (AB^K, where K = the number of AB phases completed) and multiple baseline designs across cases. To calculate this metric, there is a requirement of the design including at least three individual cases. In addition, the following assumptions are made: no time trend, the effect of the treatment is constant across cases, and the data is normally distributed (Shadish, Hedges, Pustejovsky, Boyajian et al., 2014).

Shadish, Hedges, Pustejovsky, Rindskopf, and colleagues (2014) express that the statistics involved in obtaining this metric are more complex than other statistical analyses that have been applied to SCDs. Complex formulas and statistical expertise can be a hinderance when SCD researchers may not have knowledge or training in such approaches. Thus, to make this technique more accessible, an SPSS macro titled DHPS.sps and its corresponding manual (Marso & Shadish, 2015) were developed to compute the *d*-statistic. Shadish, Hedges, and Pustejovsky (2014) note that this *d*-statistic considers the "number of cases in the study, the number of measurements made per case, first-order autocorrelation, intraclass correlation that measures the ratio of between-case variance to the sum of between and within-case variance, the number of repetitions, and the number of phases in a reversal design" (p. 126). With the use of this macro, the *d*-statistic and its variance are corrected for small sample bias and produce Hedges' *g* (Hedges, 1981) and its variance. In summary, by utilizing such a statistical tool on applicable SCD data, it will produce Hedges' *g* which is a comparable metric to effect size estimates from group-based designs. Hence, results from SCDs and between-group designs can be synthesized and analyzed with the comparable metric.

Present Study

Based on the literature reviewed above and the concerns described regarding a lack of research on how SCD results can be compared with those from group-based designs, it seems there is a need to explore and exemplify how this may be achieved. In addition, it would be interesting and paramount to consider the effect size estimate of an intervention or treatment evaluated as a SCD as well as a group-based design on the same participants. Although a study like this is extremely rare, they do exist. Such a study was completed by Vause and colleagues (2018). The study was unique such that it contained both a group-based design in the form of a RCT and SCD methodologies on the same participants that received the same treatment. Thus, the present study will take advantage of this unique opportunity to explore what can be found when analyzing data from the same participants as a group-based design and SCD.

The purpose of this study was two-fold. First, was to explore the utility of a statistical tool to compute between-case standardized mean difference effect sizes (ES_{BC}), in the form of Hedges' *g*, from SCD data. Second, to compare those findings with results of the treatment already determined through group-based design analyses. Therefore, the Vause et al. (2018) study was ideal for these purposes.

The goal of this study is not to undermine the use of traditional SCD methods such as visual analysis or general linear model approaches, but to explore what kind of effect can also be found using the featured statistical method. Consequently, the effect size estimate found may produce a comparable metric to the effects found using group-based analyses. Thus, it will be intriguing to explore how the effects compare and whether they corroborate one another. In addition, the comparability of effects using such a metric offers a unique presentation of how

SCD methods may be analyzed to provide a metric that can be included in meta-analyses so that all empirical evidence can be considered regardless of the experimental design.

Research Questions

Based on the information and rationale provided above, it is necessary to promote the importance of SCD methodologies to return a result that is in a comparable metric that can be included in research syntheses. In addition, it is crucial to determine the comparability and accuracy of the effect found. Hence, the following research questions were asked:

1. Does the between-case standardized mean difference effect sizes found corroborate with the effects found using group-based analyses on the same participants that underwent the same treatment?
2. What comparisons can be made between the overall between-case standardized mean difference effect size found across all behaviors with the effect sizes found per participant?
3. Can additional information of treatment effects be described by utilizing statistical tools to acquire between-case standardized mean difference effect sizes from SCD research?
4. Can the additional information of treatment effects found using statistical tools to acquire between-case standardized mean difference effect sizes further complement traditional SCD methodologies?

Method

For the purposes of the present study, a dataset was obtained from Vause et al. (2018) to illustrate acquiring a *d*-statistic and compare with the previous findings. The initial section of the method of the current study describes the Vause et al. (2018) study to clarify and emphasize the reasoning behind the use of the dataset. Following this description, information is provided on the data analytic process utilized on the selected dataset.

Description of Vause et al. (2018)

Prior to the Vause et al. (2018) study, the researchers completed a preliminary randomized trial of the Function-Based Cognitive-Behavioral Therapy (Fb-CBT) to treat OCB in children with autism spectrum disorder (ASD; Vause et al., 2017). Based on their review of previous research, both cognitive behavior therapy (CBT) and ABA can reduce symptoms of OCB in children with ASD (Leon et al., 2013; Neil & Sturmey, 2014; Rodriguez et al., 2012; Sigafoos et al., 2009; Sofronoff et al., 2005; Sukhodolsky et al., 2013). Thus, the researchers designed a manualized therapy referred to as Fb-CBT. The clinician's manual that was employed for the treatment was "*I Believe in Me, Not OCB!*" by Vause et al. (2013a) and a children's workbook (2013b). The preliminary randomized trial involved 14 children assigned to a treatment group or a treatment as usual (TAU) group. The overall findings provided preliminary evidence that validated the manualized treatment package. Therefore, Vause and colleagues went forward with a larger project to further evaluate and establish efficacy of the treatment.

Methods of Vause et al. (2018)

Participants. Researchers recruited participants through referrals from private clinics and non-profit agencies through advertisements and presentations. In total, 37 participants who entered the study with a diagnosis of ASD, had a presence of at least three OCBs, a Full-Scale IQ

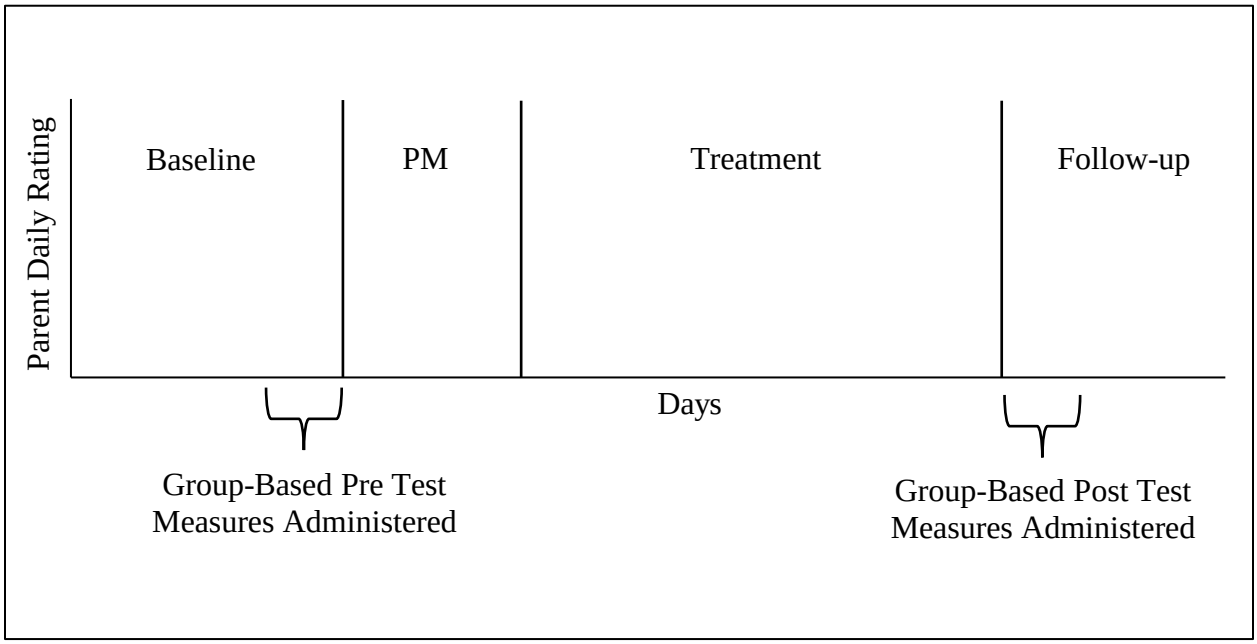
≥ 70, and had no planned medication changes during the study unless necessary. Out of the 37 participants, there were over 100 different OCBs treated.

Using a computer-generated algorithm, the participants were randomly assigned in a 1:1 ratio in either the Fb-CBT group or to the TAU group. Those that received TAU were the control group that was compared to the treatment group. Those that were in the TAU group were able to complete treatment post-TAU.

Procedure. The participants and families of the Fb-CBT group each selected up to 10 OCBs to target. The general structure of the procedure completed by this group can be viewed in figure 1. The Fb-CBT group completed the following phases: baseline, psychoeducation and mapping (PM), treatment, and follow-up at 1 and 6 months. Throughout all the phases, parents completed a daily rating scale of individual OCBs.

Figure 1

A visual representation of Vause et al. (2018) Fb-CBT procedure



Treatment was withheld during the baseline phase and PM phase. However, throughout PM the researchers focused on rapport building, session rules, and operationally defining OCB and triggers. Furthermore, some exercises were introduced to the participants that included concepts of interference, choosing an alternative activity rather than engaging in a compulsion, and learning to use a fear thermometer to rate level of distress. Additionally, the participants engaged in visually mapping their OCB based on a hierarchy of complete control of the compulsion to no control of the compulsion. The mapping of OCBs was used to aid researchers in determining the order of OCBs in the treatment phase. Thus, the first OCB targeted would be one in which the participant has some control of the compulsion and then move to OCBs in which they have less control.

The treatment phase consisted of function-based assessment and intervention, cognitive behavioral therapy strategies, and parent training. Prior to an OCB receiving treatment, the Questions About Behavioral Function (Matson & Vollmer, 1995) was completed by the parents and a function-based assessment (Matson & Williams, 2014) was utilized to help identify perceived singular or multiple functions. Additionally, descriptive functional assessment data was collected, and observations of the target behaviors were completed in contrived and natural settings. The researchers trained in ABA completed analyses of the collected data, antecedents, consequences, and potential functions to support the individual treatment plan. Based on the data collected and the perceived function of the OCB, function-based assessment and intervention plans (FBAI) were created, and parents were trained to implement them using behavioral skills training.

The overall treatment structure consisted of nine 2-hour weekly sessions that included groups of three to four participants and two therapists. The therapy sessions included group

activities, individual work, and group parent training. The sessions included individualized cognitive restructuring, learning replacement behaviors, creation of a plan for exposure to stimuli associated with the OCB, and positive reinforcement. Moreover, the treatment structure was family-based which meant that 30 minutes of the session was specifically for parent training to encourage independence in implementing the FBAI in therapy sessions and at home.

Within the TAU group, the participants were able to continue services that they were currently receiving, and they were also able to access new services. To assess treatment effects, the two primary measures and two secondary measures were administered for both the Fb-CBT group and the TAU group. The primary measures were Repetitive Behavior Scale-Revised (RBS-SCR) and Children's Yale-Brown Obsessive-Compulsive Scale (CY-BOCS). The secondary measures were The Child Obsessive-Compulsive Impact Scale-Revised Parent (COIS-RP) and Repetitive Behavior Scale-Revised 1-100 (RBS_100). These four measures were administered prior to treatment and after treatment. All the measures, except the CY-BOCS due to parent constraints, were administered at follow-up.

Selection of Dataset

The Vause et al. (2018) dataset was selected for the current study due its unique design. Specifically, it involved a nested structure. The daily parent rating data were collected on individual OCBs, the OCBs were nested within the participants, and the participants were nested in the randomly assigned Fb-CBT group and TAU group. This design provided a natural opportunity to analyze the data found as a SCD, using the parent rating data, and a between-group design, using the measures administered prior to treatment and after.

Measures

There were several measures utilized in the Vause et al. (2018) study to determine the effect of the treatment between groups. However, for the purpose of the current project, the only measure of relevance to the dataset is the Parent OCB Rating Scale.

Parent OCB Rating Scale

This rating scale was used to rate the individual OCBs daily (Vause et al., 2018). In this scale, the parents rated their children's OCBs using a Likert scale that ranged from 1 being *desired post-treatment levels of OCBs* to 3 being *partial improvement from pre-treatment levels* to 5 being *pre-treatment levels of OCBs*. Daily ratings were recorded by the parents during baseline, treatment, and follow-ups of 1 month and 6 months.

The Dataset

For the purpose of this project, only data from the participants assigned to the Fb-CBT were analyzed. This selection was based on necessary requirements when utilizing the DHPS.sps macro. To compute an effect size using this technique, the data had to be from an AB design, meaning data has been collected from a baseline phase and a treatment phase. Thus, the TAU participants were excluded. Within the treatment group, data obtained from 110 OCBs nested within 17 participants had been collected.

The Fb-CBT participants OCBs were rated by their parents throughout baseline, PM, and treatment. Psychoeducation and mapping is a phase in which the researchers and participants select the order of OCBs for treatment. Within this phase, the participants are not receiving treatment and the parents are providing ratings. Therefore, these additional observation points were included as part of the baseline phase to run analyses. This decision was made based on findings from Jaksic et al. (2018). They completed a growth curve analysis to test if baseline was

different from PM and found no statistical difference. Thus, this dataset merged baseline with PM and was coded all as the baseline phase to comply with the macro requirements.

Within the dataset there were 24.83% of expected responses that had been coded as missing. Another requirement of the macro is that the data must be complete and consecutively ordered. The researchers confirmed this was due to parents forgetting to fill out the daily rating scale. Therefore, since a response was not provided for 24.83%, these lines of data were deleted to adhere to the macro structure.

Procedure

Data Analysis

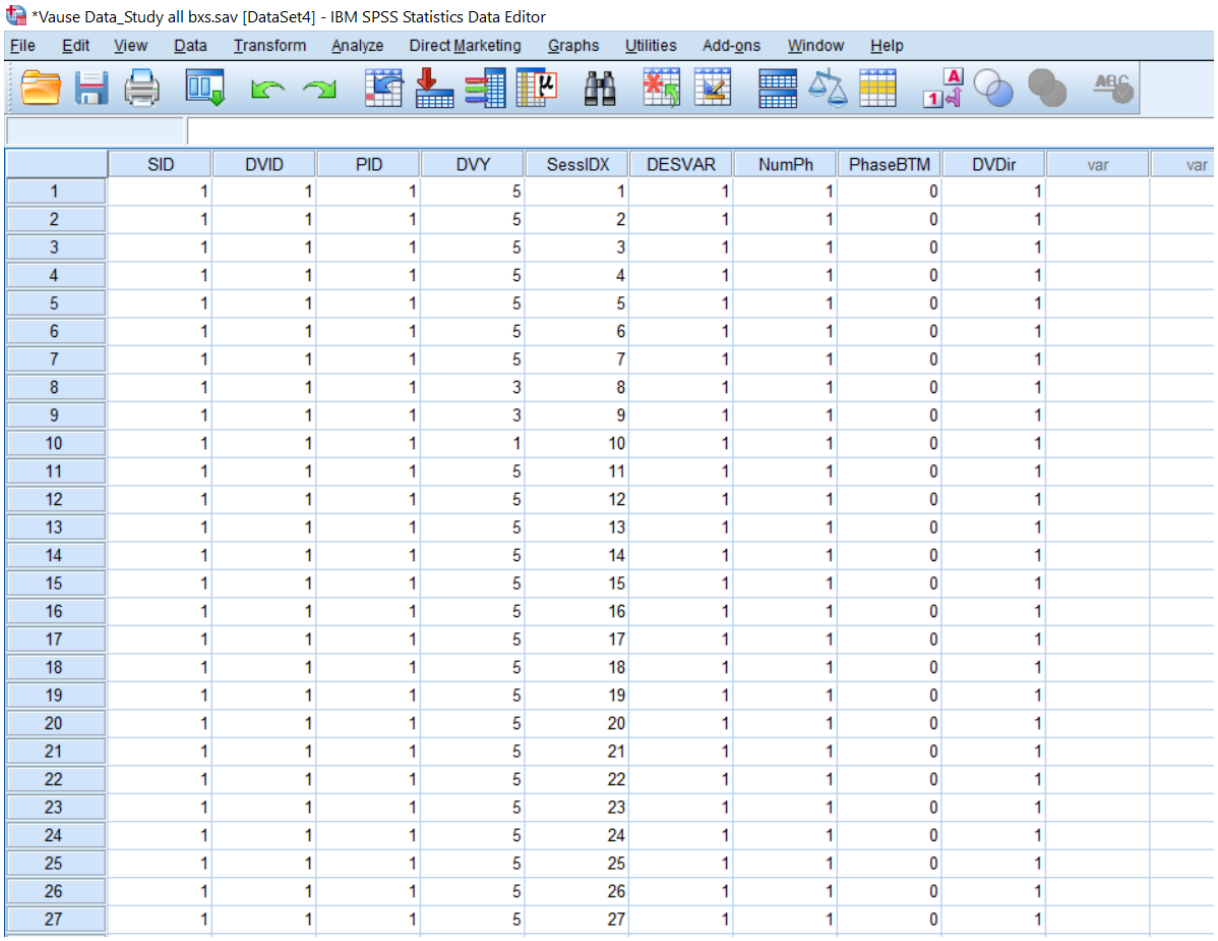
The analyses of the dataset were conducted using IBM SPSS Statistics 23. In order to compute a ES_{BC} from SCD data using the DHPS.sps macro, the dataset must adhere to an explicit structure that is described in the User Guide for DHPS (Marso & Shadish, 2015). This involves specific variable labels, clear-cut arrangement of the variables, and a particular coding structure.

The DHPS.sps macro includes the assumptions that the data has no trend and is normally distributed. The data that is analyzed must be either an AB^K design or a multiple baseline design. In addition, the macro requires a minimum of three cases in a study.

The dataset was arranged in two different ways to calculate an overall effect and to demonstrate the effect of treatment per participant. The first arrangement considered the dataset as one study to evaluate the effect of treatment on all behaviors. An example of this arrangement is shown in figure 2. The second arrangement considered the participants as separate studies to determine the effect of treatment per participant. An example of this arrangement is displayed in figure 3.

Figure 2

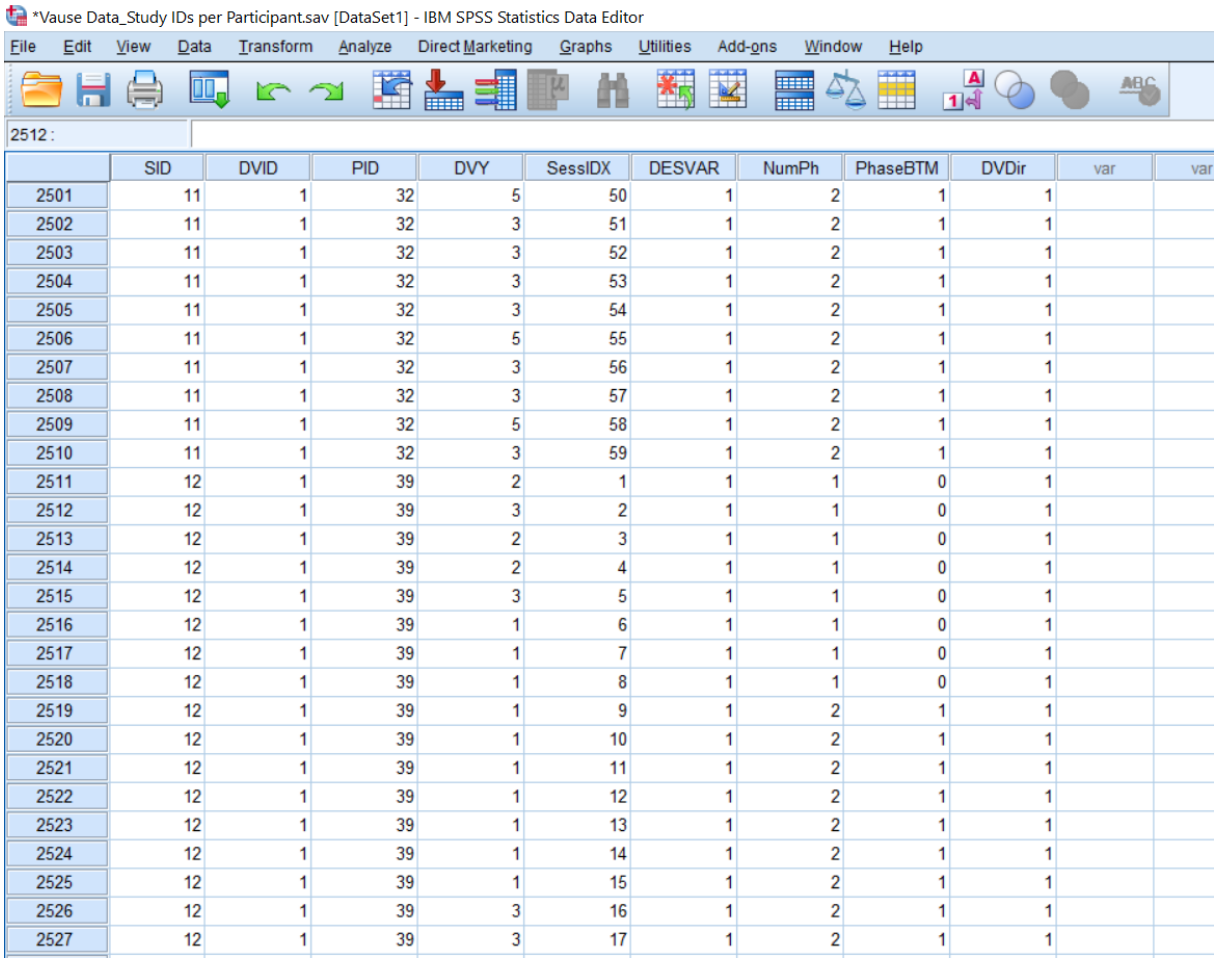
A screenshot of the SPSS Data Editor that displays the coded variables necessary for the macro. Each variable and the required coding are explained in the User Guide. This dataset is coded such that it is one study with each behaviour considered a separate case.



	SID	DVID	PID	DVY	SessIDX	DESVAR	NumPh	PhaseBTM	DVDir	var	var
1	1	1	1	5	1	1	1	0	1		
2	1	1	1	5	2	1	1	0	1		
3	1	1	1	5	3	1	1	0	1		
4	1	1	1	5	4	1	1	0	1		
5	1	1	1	5	5	1	1	0	1		
6	1	1	1	5	6	1	1	0	1		
7	1	1	1	5	7	1	1	0	1		
8	1	1	1	3	8	1	1	0	1		
9	1	1	1	3	9	1	1	0	1		
10	1	1	1	1	10	1	1	0	1		
11	1	1	1	5	11	1	1	0	1		
12	1	1	1	5	12	1	1	0	1		
13	1	1	1	5	13	1	1	0	1		
14	1	1	1	5	14	1	1	0	1		
15	1	1	1	5	15	1	1	0	1		
16	1	1	1	5	16	1	1	0	1		
17	1	1	1	5	17	1	1	0	1		
18	1	1	1	5	18	1	1	0	1		
19	1	1	1	5	19	1	1	0	1		
20	1	1	1	5	20	1	1	0	1		
21	1	1	1	5	21	1	1	0	1		
22	1	1	1	5	22	1	1	0	1		
23	1	1	1	5	23	1	1	0	1		
24	1	1	1	5	24	1	1	0	1		
25	1	1	1	5	25	1	1	0	1		
26	1	1	1	5	26	1	1	0	1		
27	1	1	1	5	27	1	1	0	1		

Figure 3

A screenshot of the SPSS Data Editor that displays the coded variables necessary for the macro. Each variable and the required coding are explained in the User Guide. This dataset is coded such that each participant is designated as a separate case.



	SID	DVID	PID	DVY	SessIDX	DESVAR	NumPh	PhaseBTM	DVDir	var	var
2501	11	1	32	5	50	1	2	1	1		
2502	11	1	32	3	51	1	2	1	1		
2503	11	1	32	3	52	1	2	1	1		
2504	11	1	32	3	53	1	2	1	1		
2505	11	1	32	3	54	1	2	1	1		
2506	11	1	32	5	55	1	2	1	1		
2507	11	1	32	3	56	1	2	1	1		
2508	11	1	32	3	57	1	2	1	1		
2509	11	1	32	5	58	1	2	1	1		
2510	11	1	32	3	59	1	2	1	1		
2511	12	1	39	2	1	1	1	0	1		
2512	12	1	39	3	2	1	1	0	1		
2513	12	1	39	2	3	1	1	0	1		
2514	12	1	39	2	4	1	1	0	1		
2515	12	1	39	3	5	1	1	0	1		
2516	12	1	39	1	6	1	1	0	1		
2517	12	1	39	1	7	1	1	0	1		
2518	12	1	39	1	8	1	1	0	1		
2519	12	1	39	1	9	1	2	1	1		
2520	12	1	39	1	10	1	2	1	1		
2521	12	1	39	1	11	1	2	1	1		
2522	12	1	39	1	12	1	2	1	1		
2523	12	1	39	1	13	1	2	1	1		
2524	12	1	39	1	14	1	2	1	1		
2525	12	1	39	1	15	1	2	1	1		
2526	12	1	39	3	16	1	2	1	1		
2527	12	1	39	3	17	1	2	1	1		

Results

Effect Sizes

The dataset was arranged as per the requirements of the statistical tool to determine the overall ES_{BC}, in the form of Hedges' g, of all behaviours nested in the participants that received treatment. In addition, using the macro, the standardized mean difference effect size per participant was calculated. The output results from the macro provided the effect size in the form of Hedges' g and the variance of Hedges' g. Using these values, 95% confidence intervals were manually calculated based on the suggested equation by Hedges and Olkin (1985), $CI = g \pm z_{.05}\sqrt{Varg}$, where the critical value $z_{.05} = 1.96$. If the confidence interval $CI_L \leq g \leq CI_U$ excludes zero, it indicates the effect size is statistically significant. The results are displayed in Table 1.

The *d*-statistic that was found from the analysis of all 110 behaviors from 17 participants suggested that the treatment reduced the scores of the behaviors by 1 and a fifth standard deviations (g = 1.22, VarG = .004), 95% CI [1.09, 1.36]. In addition, when considering the effect per participant, 16 of the 17 participants' confidence intervals excluded zero, which suggests the effects found were statistically significant at $p < .05$.

Table 1

Effect size results of the prepared Vause and colleagues (2018) dataset

	Number of behaviors	Hedges' *g*	Variance of Hedges' *g*	95% CI Lower	95% CI Upper
ES_{BC}_Participant 1	7	1.40	.036	1.03	1.77
ES_{BC}_Participant 3	10	1.75	.028	1.42	2.07
ES_{BC}_Participant 5	5	0.47	.028	0.14	0.80
ES_{BC}_Participant 7	7	0.68	.024	0.37	0.98
ES_{BC}_Participant 11	3	0.36	.044	-0.05	0.77
ES_{BC}_Participant 12	4	1.27	.160	0.49	2.05
ES_{BC}_Participant 13	6	0.66	.046	0.24	1.08
ES_{BC}_Participant 17	7	1.98	.086	1.41	2.56
ES_{BC}_Participant 18	5	1.53	.040	1.14	1.92
ES_{BC}_Participant 19	9	0.99	.079	0.44	1.54
ES_{BC}_Participant 20	7	0.70	.030	0.37	1.04
ES_{BC}_Participant 21	10	1.07	.056	0.60	1.53
ES_{BC}_Participant 22	6	0.72	.046	0.30	1.14
ES_{BC}_Participant 25	9	1.53	.021	1.25	1.81
ES_{BC}_Participant 28	6	0.90	.025	0.59	1.21
ES_{BC}_Participant 29	4	1.08	.050	0.64	1.52
ES_{BC}_Participant 31	5	1.31	.059	0.84	1.79
ES_{BC}_All Behaviours	110	1.22	.004	1.09	1.36

For comparison purposes of the standardized mean difference effect sizes, the between-group results found in the original Vause and colleagues (2018) study are presented in Figure 4. The results displayed in Table 1 and Figure 4 are compiled and depicted in a forest plot displayed in Figure 5.

Figure 4

The published table of results from the Vause and colleagues (2018) paper that presented effect sizes in the form of Hedges' g and the corresponding confidence intervals to represent the effects of the treatment.

Journal of Autism and Developmental Disorders

Table 3 Results of mixed-effects regression models representing adjusted post-test treatment effects

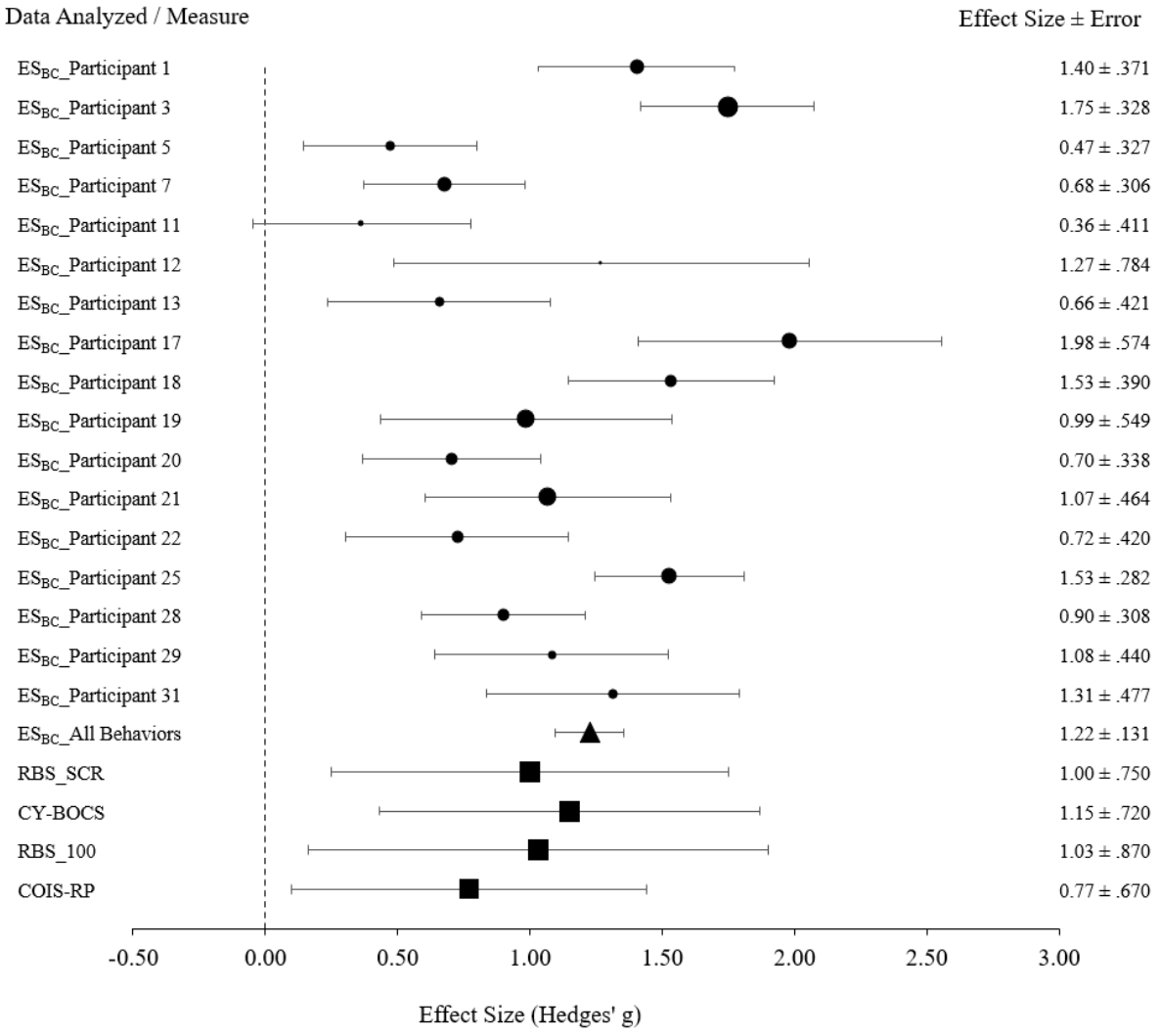
Outcome	Model ICC	Post-pre difference	SE	t	p	95%CI lower	95%CI upper	Followup-post difference	Effect Sizes			
									Hedge's g	SE	95%CI Lower	95%CI Upper
Primary												
RBS_SCR	0.11	10.47	3.80	2.76	0.026	1.62	19.32	-1.67 ns	1.00	0.39	0.24	1.75
CY-BOCS	0.43	2.97	1.09	2.74	0.023	0.51	5.43	n/a	1.15	0.37	0.43	1.87
Secondary												
RBS_100	0.13	24.68	9.24	2.67	0.026	3.79	45.57	-9.04 ns	1.03	0.44	0.17	1.90
COIS-RP	0.12	8.26	7.37	1.12	0.289	-8.14	24.66	3.83 ns	0.77	0.34	0.10	1.44

CY-BOCS Children's Yale-Brown Obsessive Compulsive Scale, *COIS-RP* The Child Obsessive–Compulsive Impact Scale-Revised Parent, *RBS_SCR* Repetitive Behavior Scale-Revised (sameness, compulsive, and ritualistic subscales), *RBS_100* Repetitive Behavior Scale-Revised 1–100 overall severity score

The forest plot was created to display all effect sizes and their corresponding standard errors. The forest plot indicates some variability in effect size estimates; however, the confidence intervals of the SCD effect size estimates all overlap with the confidence intervals of the original group-based effect size estimates. Of particular note is the effect size estimate of ES_{BC}_All Behaviours found using the DHPS.sps macro. This ES_{BC} (g = 1.22) is comparable to the between-case effect sizes found by Vause and colleagues (2018). The average effect size across the four outcomes found by Vause and colleagues was g = 0.99.

Figure 5

A forest plot which includes the ES_{BC}s of Vause and colleagues (2018) data analyzed using the DHPS.sps macro (Table 1) and the effect sizes originally found from the randomized controlled trial analyses (Figure 4).



Note. The size of the individual participants' ES_{BC} data points (closed circles) represents the number of behaviors analyzed per participant. The ES_{BC} of all behaviors across participants is represented by the closed triangle. The closed squares represent the effect sizes of the following measures: RBS_SCR Repetitive Behavior Scale-Revised, CY-BOCS Children's Yale-Brown

Obsessive Compulsive Scale, RBS-100 Repetitive Behavior Scale-Revised, COIS-RP The Child Obsessive-Compulsive Impact Scale Revised Parent.

Exemplification of Visual Analysis

The following figures show daily parent ratings (based on the Likert-type scale from 1 to 5) of target behaviors across baseline (including PM) and treatment. Three behaviors each from participant 3 and participant 7 were selected to display visually. These participants were selected to graphically display because they had a similar number of target behaviors but differed in the overall ES_{BC} found (Participant 3 g = 1.75, 95% CI [1.42, 2.07]; Participant 7 g = 0.68, 95% CI [0.37, 0.98]). Thus, although they had a similar number of behaviors undergo treatment, the treatment effects were larger for Participant 3. Furthermore, both participants' confidence intervals do not overlap with the confidence intervals of ES_{BC}_All Behaviours, Participant 3 is above the upper interval and Participant 7 is below the lower interval. Therefore, it was of interest to investigate further using visual analysis.

There are two sets of graphs displayed per participant. The first set displays the data based on how it was arranged according to the specifications of the DHPS.sps macro. Thus, the data is presented complete and consecutively in the first set. The second set displays the raw data which includes spaces where parents had forgotten to provide a response.

Using traditional visual analysis techniques to consider the functional relationship between the treatment and the parent responses of the target behaviors, the following considerations were made: the number of data points, the variability of data, the levels of the data, and the direction and degree of trends in the data (Cooper et al., 2020). For Participant 3, the parent responses recorded in baseline are very stable for behaviors 1 and 6 and somewhat variable for behavior 3. The level recorded in baseline suggests the target behaviors are

occurring at pre-treatment levels. When the treatment is introduced, the graphs for all three behaviors suggests that based on the parent ratings, the behaviors are reduced after approximately three days to partial improvement from pre-treatment levels and the desired post-treatment level of OCB. When inspecting participant 7's three graphs a lot of variability is observed. Two of the three behaviors have variable responses in baseline and in treatment without a clear representation of level or trend. However, the graph of behavior 22 displays an eventual stable trend in baseline such that parent ratings were at pre-treatment levels and once treatment is introduced the ratings quickly go down to the desired post-treatment level of OCBs. Thus, on average across behaviors visual analysis suggests parent ratings of the OCBs are reduced.

Figure 6

A sample of participant 3 behaviors of which parent rating responses were collected. The graphs have been developed based on the data inputted into the DHPS.sps macro, thus data is consecutive. Participant 3 parent rating responses of behavior 1 (top panel), behavior 3 (middle panel), and behavior 6 (bottom panel).

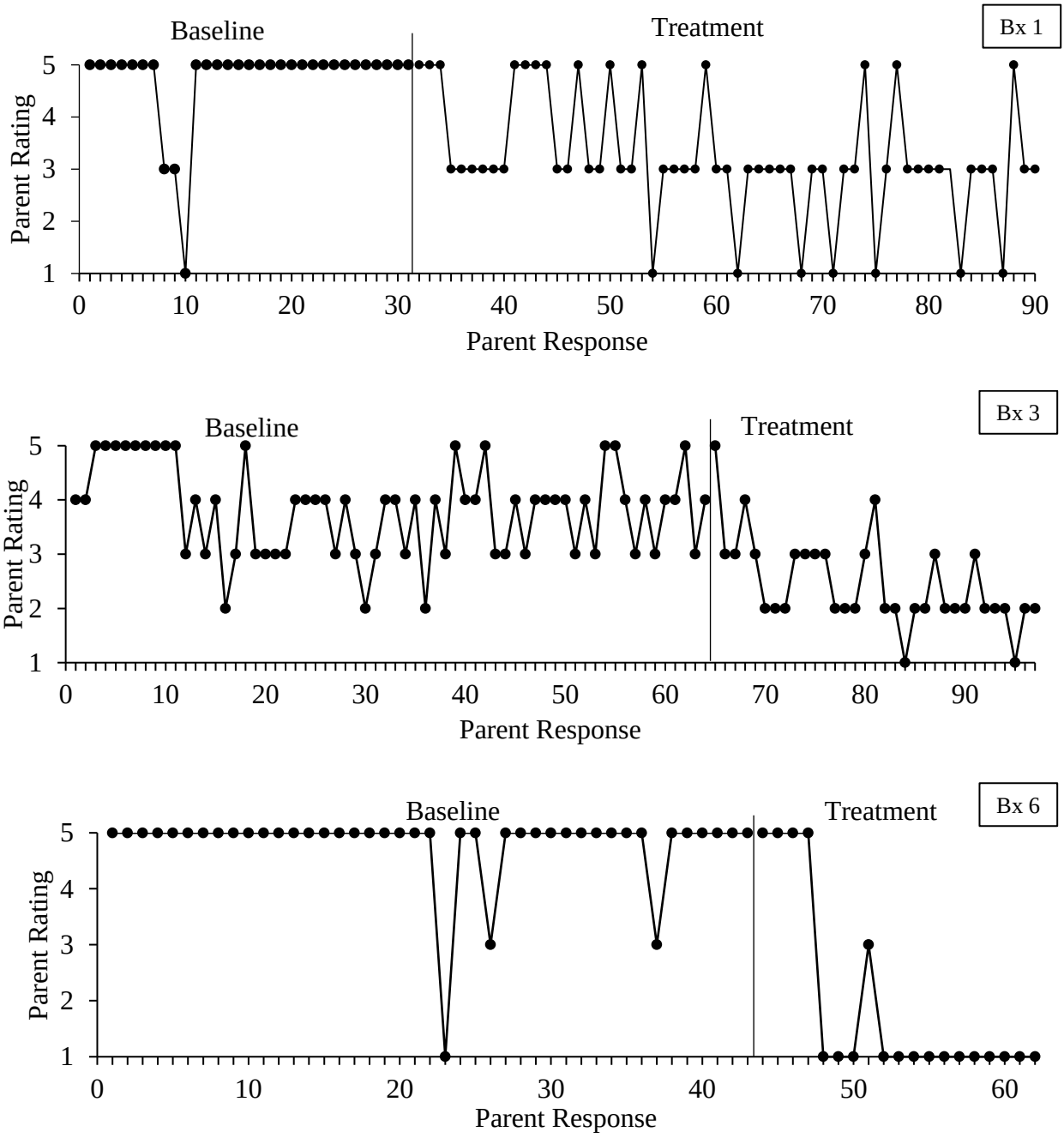


Figure 7

A sample of participant 3 behaviors of which parent rating responses were collected. The graphs have been developed based on the original data, thus including spaces where parents had forgotten to provide a response. Participant 3 parent rating responses of behavior 1 (top panel), behavior 3 (middle panel), and behavior 6 (bottom panel).

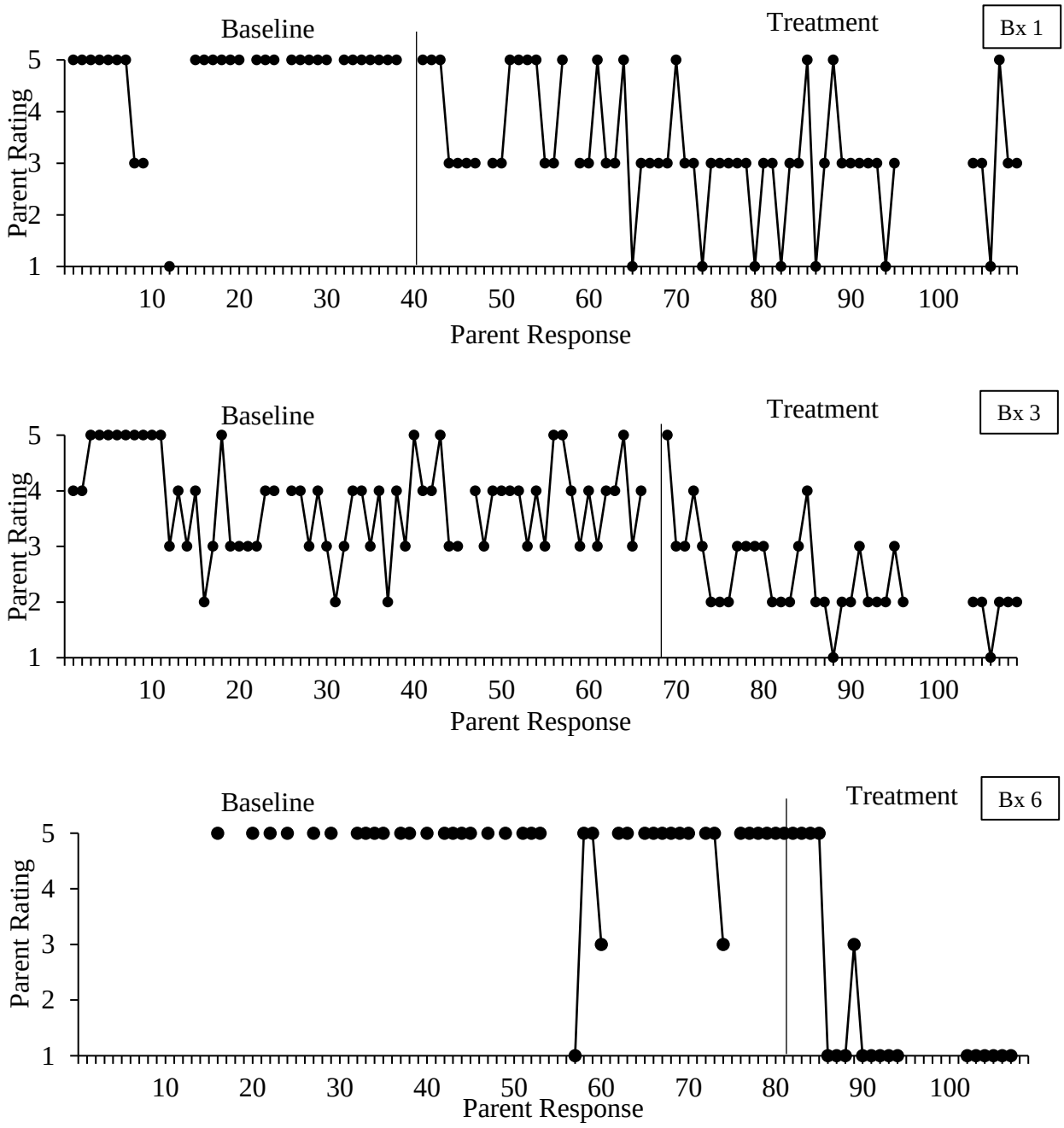


Figure 8

A sample of participant 7 behaviors of which parent rating responses were collected. The graphs have been developed based on the data inputted into the DHPS.sps macro, thus data consecutive. Participant 7 parent rating responses of behavior 21 (top panel), behavior 22 (middle panel), and behavior 23 (bottom panel).

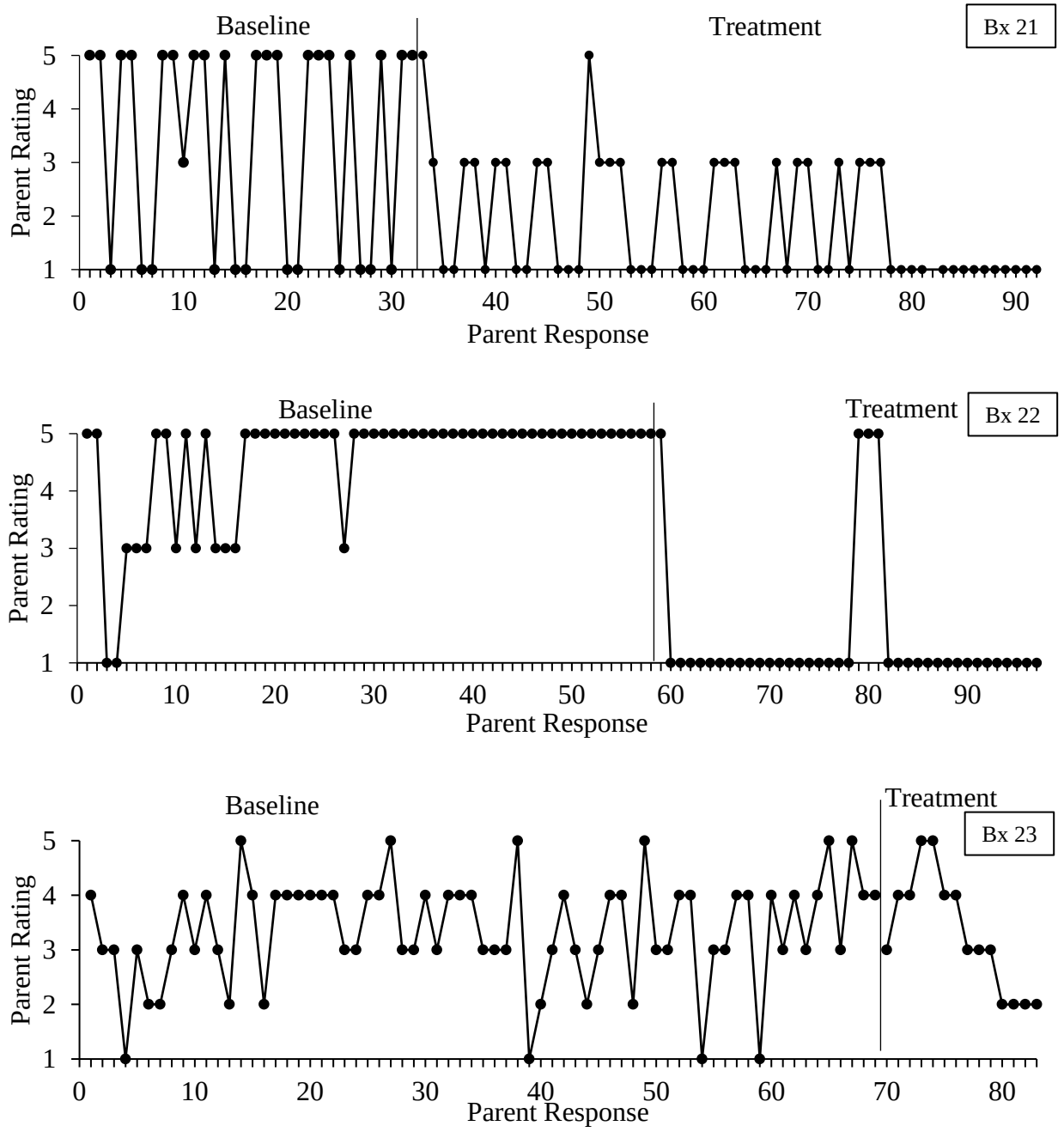
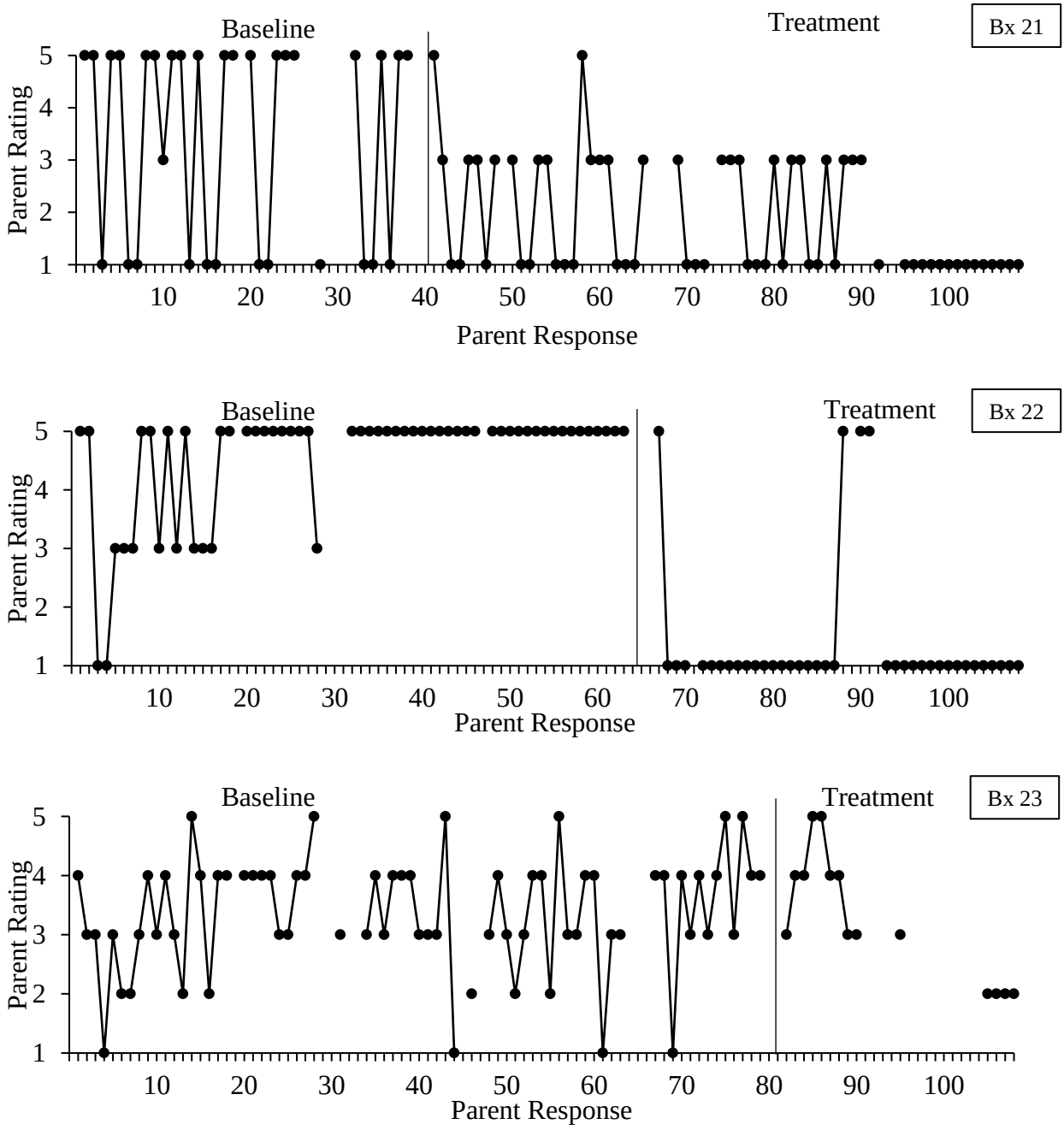


Figure 9

A sample of participant 7 behaviors of which parent rating responses were collected. The graphs have been developed based on the original data, thus including spaces where parents had forgotten to provide a response. Participant 7 parent rating responses of behavior 21 (top panel), behavior 22 (middle panel), and behavior 23 (bottom panel).



For discussion purposes and to provide further information on the effects found in the selected graphs, traditional effect size calculations for SCDs were utilized. The two approaches used were PND and PEM. The two equations used are

$$PND = \frac{\#\ of\ treatment\ data\ points\ below\ the\ lowest\ point\ of\ distribution\ in\ baseline\ phase}{Total\ \#\ of\ treatment\ data\ points}$$

and

$$PEM = \frac{\#\ of\ treatment\ data\ points\ below\ the\ median}{Total\ \#\ of\ treatment\ data\ points}$$

from the paper by Ma (2006). The results from these calculations are displayed in table 2.

Table 2

Results of PND and PEM calculations for the six selected behaviors of Participant 3 and Participant 7.

Participant	Behavior	PND	PEM
3	1	.000	.763
	3	.061	.909
	6	.000	.789
7	21	.000	.967
	22	.000	.897
	23	.000	.286

Discussion

The present study is the first known to explore and compare the effect size estimates of a treatment on participants' behaviors that was evaluated as a SCD as well as a group-based design. The aim of the present study was to analyze treatment effects from SCD data and compare them to the effects found from group-based analyses of the same participants undergoing the same treatment. More specifically, the following research questions were asked: 1) Does the between-case standardized mean difference effect sizes found corroborate with the effects found using group-based analyses on the same participants that underwent the same treatment? 2) What comparisons can be made between the overall between-case standardized mean difference effect size found across all behaviors with the effect sizes found per participant? 3) Can additional information of treatment effects be described by utilizing statistical tools to acquire between-case standardized mean difference effect sizes from SCD research? 4) Can the additional information of treatment effects found using statistical tools to acquire between-case standardized mean difference effect sizes further complement traditional SCD methodologies?

One of the most intriguing questions of the present study was if the ES_{BC} across all behaviors and per participant would corroborate with the group-based effect sizes found in the original Vause et al. (2018) study. It was found that the effects were quite comparable. The parent rating responses were collected and analyzed as an AB design. The overall ES_{BC} of the treatment was found to be large at $g = 1.22$, suggesting that the treatment across behaviors reduced parent rating scores by one and a fifth standard deviations. This analysis confirms what was already found using group-based analyses in the Vause et al. (2018) study. The original group-based analyses concluded that across all the outcomes, the mean effect size was large at *g*

= 0.99. Although these effects differ by approximately a fifth of a standard deviation, they both suggest a large effect.

In addition to the overall ES_{BC} found, the effects of treatment across behaviors per participant were also calculated. Of the 17 participants, the ES_{BC} estimates in the form of Hedges' *g* ranged from 0.36 to 1.98. This suggests, depending on the participant, there were small, medium, and large effects found (according to Cohen's 1988 interpretations of effect sizes). Although these effects appear to be variable, there are some considerations to be made.

When completing analyses, the number of cases analyzed, and the amount of data will influence the effects found. From the Vause et al. (2018) data the largest number of behaviors that received treatment per participant was 10 and the least amount was three. The other consideration is that out of the 8039 data points used to analyze the overall ES_{BC} of treatment across all behaviors, the largest amount of data points involved in the analysis per participant were 824 parent responses and the least amount were 117 parent responses. To more specifically exemplify how this will affect outcomes, we will consider the outcomes of Participant 3 and Participant 11. Participant 3 received treatment for 10 behaviors and the total amount of data included in the analyses were 824 parent responses. The ES_{BC} estimate of Participant 3 was found to be $g = 1.75$. Whereas Participant 11 received treatment for only three behaviors and the total amount of data included in the analyses were 162 parent responses. The ES_{BC} estimate of Participant 11 was found to be $g = 0.36$. Although treatment effects may differ based on contextual variables such as setting or treatment integrity, it is important to also recognize how sample size and total number of data points can influence an overall outcome.

The statistical technique to analyze the SCD data can be applied to explore the overall effect of a treatment within one study or many. Based on the structure of this design, the present

study exemplifies how an overall stable ES_{BC} can be generated across multiple participants, like a meta-analysis. Also, a ES_{BC} per participant can be generated once the data is rearranged in accordance with the statistical tool's requirements.

This approach may allow researchers to verify an overall treatment effect and treatment effects per participant. This will provide SCD researchers with the natural opportunity to evaluate how variables such as sample size, number of data points, or other contextual variables might affect the outcomes of the treatment across studies. This will allow researchers to dive deeper into interpretations of how a treatment can affect certain behaviors, which has the potential to drive future research projects and influence EBP. Furthermore, the use of a visual display using these results, such as a forest plot, can provide even more valuable information.

To exemplify what kind of information can be discovered with a visual aid, a forest plot of the findings was created (figure 5). The number of behaviors analyzed per participant is represented by the size of the data point indicating the effect found. The display of all effects presents the opportunity to compare results and reveal any variation of the outcomes. When viewing figure 5, there are a few immediate observations to note. The first is that the effect size of 'ES_{BC}_All Behaviours' is only slightly higher on the x-axis than the four effect sizes from Vause et al. (2018) study. The proximity of these effects confirms what has already been deduced: both types of analyses, SCD and group-based, obtained similar effects of the treatment on the behaviors. The second noteworthy finding from figure 5 is that all but one set of confidence intervals (Participant 11) excludes zero. When a confidence interval excludes zero it suggests that the effect found was statistically significant at $p < 0.05$.

The length of the confidence intervals per outcome provides this study with important information. Firstly, the width of the confidence interval often is influenced by the sample size or

the observation points involved in the analysis. The intervals can be further interpreted such that the narrower the confidence interval, the larger the sample size or observation points. The confidence interval will be narrower because it is a more precise estimate of the effect found. Furthermore, if the confidence intervals of the outcomes overlap, this suggests they are not significantly different. Thus, when considering the overall ES_{BC} found using the SCD data and the original group-based effect size estimates, they fully overlap suggesting they are not significantly different. In addition, this strategy allows further exploration of the individual ES_{BC}. There are some outcomes found that do not overlap with the 'ES_{BC}_All Behaviours' which suggests there are significant differences; however, all individual outcomes display overlap with the group-based outcomes suggesting there are no significant differences.

To further explore how the SCD ES_{BC} outcomes found can complement traditional SCD methodologies, single-case graphs were generated (figures 6-9), visual analysis was applied to the graphs, and corresponding calculations were completed (table 2). Although it would be interesting to view and visually analyze all 110 behaviors plotted in graphs, this would be excessive. Thus, two participants and three of their behaviors were selected to provide a visual representation of traditional SCD methodologies and calculations. The graphs were generated with and without the spaces that represent parents forgetting to respond. This allowed a visual representation of how data was arranged consecutively for the DHPS.sps macro and how it was originally collected. Based on the visual analysis conducted, on average across the six behaviors analyzed, the parent ratings of the OCBs reduced to post-treatment levels of OCBs. This analysis corresponds with the ES_{BC} outcomes of Participant 3 and 7.

In addition to visual analysis, PND and PEM calculations were completed. Although the effect from these calculations are not in a comparable metric that can be included in meta-

analyses, the calculations were completed to exemplify how a statistical approach can further complement these more traditional analyses. On average, the PND calculations across all behaviors suggested there was no effect. However, considering the visual analysis completed, it is clear this is not the case. The PND calculation is vulnerable to outliers which may promote a Type 2 error (Lenz, 2013). Therefore, another calculation was completed using PEM. Percentage of data exceeding the median accommodates the presence of outliers in baseline, such as the outliers that can be viewed in several of the graphs. Also, PEM can be utilized when there is variability over time, such as the variability seen in the graphs of behaviors 3, 21, and 23. Overall, the PEM calculations suggest there was a moderate to high effect. However, the PEM calculation of behavior 23 suggests the treatment was ineffective.

From the exploration of these differing methodologies, it is discernable that the use of a statistical technique can complement traditional SCD interpretations of treatment effects. The application of the DHPS.sps granted the opportunity to analyze a very large dataset. Additionally, depending on the design and arrangement of data, the data can be analyzed in a few different ways. This may provide researchers with different perspectives on their overall treatment outcomes across participants and behaviors. Furthermore, by utilizing this statistical technique, the outcomes found can be used to generate a forest plot. As well as comparing the numerical effect size estimates with each other, a visual representation can aid in comparisons. In addition, the confidence intervals that were calculated using the computed Hedges' g and corresponding variance assists in establishing significant differences between the effects. Thus, not only can researchers quickly recognize small, moderate, or large effects per case, depending on the design, they can also be compared with the overall effect.

Finally, the most significant benefit of utilizing the macro is that the analyzed effects are in a comparable metric to group-based intervention or treatment effects. This outcome is more recognizable to researchers and policy makers that are most familiar with group-based measures. Thus, SCD outcomes and group-based outcomes can be merged and compared. Not only can SCD results be included in research syntheses but they can be synthesized in meta-analyses. Consequently, SCD findings will influence EBP and policy alongside group-based research findings.

Limitations and Future Directions

While this study successfully exemplified how a *d*-statistic can be acquired using a statistical tool to analyze the SCD dataset, there are some limitations to acknowledge.

Statistical Tool Software

The advancement of SCD methodology and statistical techniques have been in the works for a number of years for the purpose of ensuring that validated findings from SCD research methods are disseminated and included in the wealth of knowledge within the scientific community (Maggin et al., 2017). One researcher that has contributed to this pursuit was William Shadish. Based on the proposed standardized mean difference effect size for SCDs (Hedges et al., 2012), Shadish (2015) developed the DHPS.sps macro software with the intention of it being user friendly. There were three separate software releases to the public between 2013 and 2015.

The present study employed the most recently released version from 2015. Although the use of this software successfully produced results, there were some technical difficulties that had to be dealt with. Some compatibility issues were difficult to identify such that the software only computed when employed on a Mac computer versus a PC computer. However, further attempts

included the use of both types of computers and different SPSS versions. Eventually, one main computer that evidently provided reliable output from the macro was used for all analyses.

Along with the links to download the software packages, there is a description of the software on Shadish's webpage. Within this description Shadish states "we will update the software as needed". Unfortunately, Shadish passed away in 2016, thus leaving this task to his colleagues. However, there have yet to be anymore updates to the software since the last release in 2015. Furthermore, within the corresponding DHPS.sps manual (Marso & Shadish, 2015), it states that "future releases of DHPS will allow trend and nonnormally distributed data" (p. 5).

The frequent updates of technology and statistical software may require the macro to be updated. If the macro software is not being maintained and updated this will limit future uses and hinder the pursuit to disseminate SCD findings. Thus, in the future, it would be advantageous for a professional to update and distribute an accessible SPSS package in order for this statistical approach to receive broad appreciation.

Although this statistical package is available to use with single-case reversal designs and multiple baseline designs, it unfortunately is unable to accommodate other variations such as alternating treatment designs, changing criterion designs, and combined designs that involve reversal and a multiple baseline.

Future Directions

Future research and exploration of SCD effect sizes could build on the present findings. This study exemplified how a large dataset can be analyzed efficiently while also focusing on individual effects. Thus, future studies should consider using a statistical tool to analyze their data to complement traditional methods. There are two other statistical packages that are available to calculate a SCD effect size that should be highlighted to aid in future research. The

first is the Single-case effect size calculator created by Pustejovsky and Swan (2018). This package involves functions to calculate basic effect size indices for SCDs. Although it does not provide an output directly comparable to group-based effect sizes, the creation and accessibility of such a package encourages SCD researchers to explore and recognize the potential in statistical packages. The second R package is the scdhlm created by Pustejovsky (2020). This package can be used to estimate hierarchical linear models and effect sizes from SCD data. The package involves functions that will estimate standardized mean difference effect sizes from reversal designs, multiple baseline designs, and multiple probe designs.

Not only is the highlighted software user friendly but it is open-source and freely available. By utilizing statistical tools to acquire an effect size estimate it could help to further substantiate and describe the effects found from SCD research. Furthermore, researchers conducting group-based experiments could also consider collecting individual data like the nested structure used by Vause and colleagues (2018). Although group-based research provides an overall effect, some researchers may be interested in individual effects as well. Therefore, with a comparable metric that can be calculated from SCD data, researchers may be more inclined to explore effects on different levels.

And finally, if there is another study structured like the Vause and colleagues (2018) study, it would be valuable to conduct explorative analyses to see if the same corroborative outcome is found.

Conclusion

This study explored the utility of differing methodological approaches to measure meaningful change in a treatment scenario. This is an important task such that by demonstrating how different methodologies can be analyzed to produce a comparable metric, findings from

differing experimental designs could be compared. However, there seems to be a lack of research and evidence specifically about how SCD effects compare with group-based effects. Therefore, to address this literature gap, the results of the present study have indicated that the analysis of differing methodologies using the same participants that underwent the same treatment, produced similar effects and were in a comparable metric. Not only do these findings support the ability to compare SCD effects with group-based design effects, but also the effect found using different methodologies were analogous. By presenting these results, it demonstrates the overwhelming value in utilizing such software and reinforces how it can complement traditional analyses. Additionally, perhaps more fulsome meta-analyses will be completed and lead to further dissemination of important findings that may eventually influence EBP and policy. Thus, I invite you to explore statistical tools to further demonstrate overall effects and disseminate your findings.

References

Allison, D. B. & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: the case of the single case. *Behavioral Research and Therapy, 31*(6), 621-631. https://doi.org/10.1016/0005-7967(93)90115-B

American Psychological Association. (2020). *Publication manual of the American Psychological Association: the official guide to APA style* (7th ed.). American Psychological Association.

American Psychological Association Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*(4), 271-285. https://doi.org/10.1037/0003-066X.61.4.271

Babione, J. M. (2010). Evidence-based practice in psychology: An ethical framework for graduate education, clinical training, and maintaining professional competence. *Ethics & Behavior, 20*(6), 443-453. https://doi.org/10.1080/10508422.2010.521446

Behavior Analyst Certification Board. (2014). *Professional and ethical compliance code for behavior analysts*. https://www.bacb.com/wp-content/uploads/BACB-Compliance-Code-english_190318.pdf

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Wiley.

Burns, M. K. (2012). Meta-analysis of single-case design research: Introduction to the special issue. *Journal of Behavioral Education, 21*(3), 175-184. https://doi.org/10.1007/s10864-012-9158-9

Card, N. A. (2011). *Applied meta-analysis for social science research*. Guilford Publications.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cooper, H. (2017). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). Sage.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis* (3rd ed.). Russell Sage Foundation.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Pearson Education.

Cumming, G. (2014). The new statistics: why and how. *Psychological Science, 25*(1), 7-29. https://doi.org/10.1177/0956797613504966

Danov, S. E. & Symons, F. J. (2008). A survey evaluation of the reliability of visual inspection and functional analysis graphs. *Behavioral Modification*, *32*(6), 828-839. https://doi.org/10.1177/0145445508318606

Deaton, A. & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2-21. https://doi.org/10.1016/j.socscimed.2017.12.005

DeProspero, A. & Cohen, S. (1979). Inconsistent visual analyses of intrasubject data. *Journal of Applied Behavior Analysis*, *12*, 573-579. https://doi.org/10.1901/jaba.1979.12-573

Dozois, D. J. A., Mikail, S. F., Alden, L. E., Bieling, P. J., Bourgon, G., Clark, D. A., Drapeau, M., Gallson, D., Greenberg, L., Hunsley, J., & Johnston, C. (2014). The CPA presidential task force on evidence-based practice of psychological treatments. *Canadian Psychology*, *55*(3), 153-160. https://doi.org/10.1037/a0035767

Harris, A. D., McGregor, J. C., Perencevich, E. N., Furuno, J. P., Zhu, J., Peterson, D. E., & Finkelstein, J. (2006). The use and interpretation of quasi-experimental studies in medical

informatics. *Journal of the American medical Informatics Association, 13*(1), 16-23. https://doi.org/10.1197/jamia.m1749

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics,* 6(2), 107-128. https://doi.org/10.3102/10769986006002107

Hedges L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* Academic Press.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods, 3,* 224-239. https://doi.org/10.1002/jrsm.1052

Jaksic, H., Vause, T., Frijters, J. C., & Feldman, M. (2018). A comparison of a novel application of hierarchical linear modeling and nonparametric analysis for single-subject designs. *Behavior Analysis: Research and Practice, 18*(2), 203-218. https://doi.org/10.1037/bar0000091

Johnston, J. M. & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research* (3rd ed.). Routledge.

Kazdin, A. E. (1981). External validity and single-case experimentation: issues and limitations (A response to J. S. Birnbrauer). *Analysis and Intervention in Developmental Disabilities, 1*(2), 133-143. https://doi.org/10.1016/0270-4684(81)90027-6

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). Oxford University Press, Inc.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *What works clearinghouse: Single-case designs technical documentation*. https://ies.ed.gov/ncee/wwc/Document/229

Lammers, W. J. & Badia, P. (2005). *Fundamentals of behavioral research*. Thomson/Wadsworth.

Lenz, A. S. (2013). Calculating effect size in single-case research: A comparison of nonoverlap methods. *Measurement and Evaluation in Counseling and Development, 46*(1), 64-73. https://doi.org/10.1177/0748175612456401

Leon, Y., Lazarchick, W. N., Rooker, G. W., & DeLeon, I. G. (2013). Assessment of problem behavior evoked by disruption of ritualistic toy arrangements in a child with autism. *Journal of Applied Behavior Analysis*, *46*(2), 507-511. https://doi.org/10.1002/jaba.41

Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification, 30*(5), 598-617. https://doi.org/10.1177/0145445504272974

Maggin, D. M., Lane, K. L., & Pustejovsky, J. E. (2017). Introduction to the special issue on single-case systematic reviews and meta-analyses. *Remedial and Special Education, 38*(6), 323-330. https://doi.org/10.1177/0741932517717043

Marso, D. & Shadish, W. R. (2015). *User guide for DHPS, D_Power, and GPHDPwrSPSS Macros* (Version 1.0). https://faculty.ucmerced.edu/wshadish/software/software-meta-analysis-single-case-design/dhps-version-march-7-2015

Matson, J. L. & Vollmer, T. R. (1995). *User's Guide: Questions About Behavioral Function (QABF). Baton* Rouge. Scientific Publishers, Inc.

Matson, J. L. & Williams, L. W. (2014). Functional assessment of challenging behavior. *Current Developmental Disorders Reports, 1*(2), 58-66. https://doi.org/10.1007/s40474-013-0006-y

McKenzie, J. E., Brennan, S. E., Ryan, R. E., Thomson, H. J., Johnston, R. V., & Thomas, J. (2019). Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions version 6.0*. Cochrane. https://training.cochrane.org/handbook/current/chapter-03

McKibbon, K. A. (1998). Evidence-based practice. *Bulletin of the Medical Library Association, 86*(3), 396-401.

Meline, T. (2006). Selecting studies for systematic review: inclusion and exclusion criteria. *Contemporary Issues in Communication Science and Disorders, 33*, 21-27.

Morgan, D. L. & Morgan, R. K. (2009). *Single-case research methods for the behavioral and health sciences*. SAGE Publications, Inc. https://doi.org/10.4135/9781483329697

Neil, N. & Sturmey, P. (2014). Assessment and treatment of obsessions and compulsions in individuals with autism spectrum disorders: A systematic review. *Review Journal of Autism and Developmental Disorders, 1*(1), 62-79. https://doi.org/10.1007/s40489-013-0006-1

Odom, S. L., Barton, E. E., Reichow, B., Swaminathan, H., & Pustejovsky, J. E. (2018). Between-case standardized effect size analysis of single case designs: Examination of the two methods. *Research in Developmental Disabilities, 79*, 88-96. https://doi.org/10.1016/j.ridd.2018.05.009

Ottenbacher, K. J. (1990). Visual inspection of single-subject data: An empirical analysis. *Mental Retardation, 28*(5), 283-290.

Parsonson, B. S. & Baer, D. M. (1986). The graphic analysis of data. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis: Issues and advances* (pp. 157-186). Plenum Press. https://doi.org/10.1007/978-1-4684-8786-2_8

Poling, A. & Grossett, D. (1986). Basic research designs. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis: Issues and advances* (pp. 7-27). Plenum Press. https://doi.org/10.1007/978-1-4684-8786-2_2

Price, P., Chiang, I-C. A., & Jhangiani, R. (2015). *Research methods in psychology* (2nd Canadian Ed.). BC Open Textbook Project. https://opentextbc.ca/researchmethods/

Pustejovsky, J. E. (2020). *scdhlm: A web-based calculator for between-case standardized mean differences* (Version 0.4.2.9000) [Web application]. https://jepusto.shinyapps.io/scdhlm

Pustejovsky, J. E. & Swan, D. M. (2018). *Single-case effect size calculator*. (Version 0.5) [Web application]. https://jepusto.shinyapps.io/SCD-effect-sizes/

Rodriquez, N. M., Thompson, R. H., Schlichenmeyer, K., & Stocco, C. S. (2012). Functional analysis and treatment of arranging and ordering by individuals with an autism spectrum disorder. *Journal of Applied Behavior Analysis, 45*(1), 1-22. (1), 1–22. https://doi.org/10.1901/jaba.2012.45-1

Sandbank, M., Bottema-Beutel, K., Corwley, S., Cassidy, M., Dunham, K., Feldman, J., Crank, J., Albarran, S., Raj, S., Mahbub, P., & Woynaroski, T. (2020). Project AIM: Autism intervention meta-analysis for studies of young children. *Psychological Bulletin, 146*(1), 1-29. https://doi.org/10.1037/bul0000215

Shadish, W. R. (2015). *Software for meta-analysis of single-case design*. (Version March 7, 2015) [Software]. https://faculty.ucmerced.edu/wshadish/software/software-meta-analysis-single-case-design

Shadish, W. R., Hedges, L. V., Horner, R. H., & Odom, S. L. (2015). *The role of between-case effect size in conducting, interpreting, and summarizing single-case research.* National Center for Education Research, Institute of Education Science, U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED562991.pdf

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology, 52*(2), 123-147. https://doi.org/10.1016/j.jsp.2013.11.005

Shadish, W. R., Hedges, L. V., Pustejovsky, J. E., Boyajian, J. G., Sullivan, K. J., Andrade, A., & Barrientos, J. L. (2014). A *d*-statistic for single-case designs that is equivalent to the usual between-groups *d*-statistic. *Neuropsychological Rehabilitation, 24*(3-4), 528-553. https://doi.org/10.1080/09602011.2013.819021

Shadish, W. R., Hedges, L. V., Pustejovsky, J. E., Rindskopf, D. M., Boyajian, J. G., & Sullivan, K. J. (2014). Analyzing single-case designs: *d*, *G*, multilevel models, Bayesian estimators, generalized additive models, and the hopes and fears of researchers about analyses. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and data-analysis advances* (pp. 247-281). American Psychological Association. https://doi.org/10.1037/14376-009

Shadish, W. R. & Rindskopf, D. M. (2007). Methods for evidence-based practice: Quantitative synthesis of single-subject designs. *New Directions for Evaluation,* 113, 95-109. https://doi.org/10.1002/ev.217

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment & Intervention, 2*(3), 188-196. https://doi.org/10.1080/17489530802581603

Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology, 70*(1), 747-770. https://doi.org/10.1146/annurev-psych-010418-102803

Sigafoos, J., Green, V. A., Payne, D., O'Reilly, M. F., & Lancioni, G. E. (2009). A classroom-based antecedent intervention reduces obsessive–repetitive behavior in an adolescent with autism. *Clinical Case Studies, 8*(1), 3-13. https://doi.org/10.1177/1534650108327475

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

Sofronoff, K., Attwood, T., & Hinton, S. (2005). A randomised controlled trial of a CBT intervention for anxiety in children with Asperger syndrome. *Journal of Child Psychology and Psychiatry, 46*(11), 1152–1160. https://doi.org/10.1111/j.1469-7610.2005.00411.x

Sukhodolsky, D. G., Bloch, M. H., Panza, K. E., & Reichow, B. (2013). Cognitive–behavioral therapy for anxiety in children with high-functioning autism: A meta-analysis. *Pediatrics, 132*(5), 1341–1350. https://doi.org/10.1542/peds.2013-1193

Sullivan, G. M. & Feinn, R. (2012). Using effect size – or why the p value is not enough. *Journal of Graduate Medical Education, 4*(3), 279-282. https://doi.org/10.4300/JGME-D-12-00156.1

Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (2012). *Test of word reading efficiency – Second edition (TOWRE-2)*. Pro-ed.

Valentine, J. C., Tanner-Smith, E. E., Pustejovsky, J. E., & Lau, T. S. (2016). Between-case standardized mean difference effect sizes for single-case designs: a primer and tutorial using the scdhlm web application. *Campbell Systematic Review*, *12*(1), 1-31. https://doi.org/10.4073/cmdp.2016.1

Vannest, K. J., Peltier, C., & Haas, A. (2018). Results reporting in single case experiments and single case meta-analysis. *Research in Developmental Disabilities, 79*, 10-18. https://doi.org/10.1016/j.ridd.2018.04.029

Vause, T., Jaksic, H., Neil, N., Frijters, J. C., Jackiewicz, G., & Feldman, M. (2018). Functional behavior-based cognitive-behavioral therapy for obsessive compulsive behavior in children with autism spectrum disorder: A randomized controlled trial. *Journal of Autism and Developmental Disorders*, 50(7), 2375-2388. https://doi.org/10.1007/s10803-018-3772-x

Vause, T., Neil, N., Jaksic, H., & Feldman, M. (2013a). *I Believe in Me Not OCB!: Clinician's Manual*. St. Catharines, ON: Brock University (Unpublished manuscript).

Vause, T., Neil, N., Jaksic, H., & Feldman, M. (2013b). *I Believe in Me Not OCB!: Workbook*. St. Catharines, ON: Brock University (Unpublished manuscript).

Vause, T., Neil, N., Jaksic, H., Jackiewicz, G., & Feldman, M. (2017). Preliminary randomized trial of function-based cognitive-behavioral therapy to treat obsessive compulsive

behavior in children with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities*, *32*(3), 218-228. https://doi.org/10.1177/1088357615588517

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Riverside Publishing.