Summer 6-9-2022

Meta-Analysis of Daily Behavior Report Cards

Meleah Ackley

META-ANALYSIS OF DAILY BEHAVIOR REPORT CARDS

by

Meleah M. Ackley

A Dissertation
Submitted to the Graduate School,
the College of Education and Human Sciences
and the School of Psychology
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved by:

Dr. Brad Dufrene, Committee Chair
Dr. Joe Olmi
Dr. Crystal Taylor
Dr. Richard Mohn

August 2022

Published by the Graduate School

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

ABSTRACT

Daily Behavior Report Cards (DBRCs) have proven effective across a vast array of behaviors and students by provide a rating system for school staff to report students' behavior progress. The current meta-analysis included eleven studies between 2007-2022, wherein participants ranged from preschool to sixth grade. The current study further investigated the evidence base by including the current What Works Clearinghouse standards (2020) to determine methodological rigor of single-case designs employing DBRCs. Standardized mean difference calculations for omnibus effect showed that DBRCs significantly improve student outcomes. Moderators of DBRC IOA, multiple baseline design type, and publication status explained some variability within the studies. Results are interpreted within their limitations.

ACKNOWLEDGMENTS

I would like to thank Dr. Brad Dufrene, my committee chair, for your mentorship. Your continued support has enabled me to develop as a professional. I also thank my committee members, Dr. Joe Olmi, Dr. Crystal Taylor, and Dr. Richard Mohn for your guidance in this project. To Caitlyn Weaver, I appreciate the time and guidance you provided with your statistical knowledge. To Lauren Layman, thank you for your support in this process.

DEDICATION

This project is dedicated to my friends and family, whose support enabled me to successfully complete this project. First and foremost, thank you to my husband, Dustin Ackley, for your encouragement and belief in me and my endeavors. It will finally be your turn to pick our next destination. To my mother, Deborah Aldridge, I know that you are so proud of my accomplishments and fulfilling my legacy. Thank you to my father, Sky Aldridge, for ensuring I always pick myself up again to continue the journey. Thank you to my friends for making me feel so supported.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

ASD	Autism Spectrum Disorder
ADHD	Attention Deficit Hyperactivity Disorder
BCT	Baselined Corrected Tau
CICO	Check-in/Check-out
DBRC	Daily Behavior Report Card
DBR	Direct Behavior Rating
DBR-MIS	*Direct Behavior Rating-Multi-Item Scales*
DBR-SIS	*Direct Behavior Rating-Single Item Scales*
IEP	Individual Education Plan
IOA	Inter-Observer Agreement
IRD	Improved Rate Difference
MTSS	Multi-Tiered Systems of Support
SCD	Single Case Design
SLD	Specific Learning Disability
WWC	What Works Clearinghouse

CHAPTER I – INTRODUCTION

Multi-tiered Systems of Support (MTSS) offers a framework for teachers and school personnel to support students' academic and behavioral well-being. Broadly, MTSS include a continuum of gradually intensifying supports to meet students' academic and behavioral needs. Additionally, MTSS includes universal screening and progress monitoring so that deficits are detected quickly and students' responses to supports are gauged. Thus, there is a need for instruments that measure students' responses to supports at each tier (Messick, 1995). School psychologists may use systematic direct observations for such. While such tools are considered the gold standard because they are direct measures of student performance, they have not been widely used in applied settings (Chafouleas et al., 2002).

Direct observations are frequently recommended for collecting progress monitoring data because they are direct measures of behavior (Hintze & Matthews, 2004). Although teachers and staff may be trained to collect such data, observation data are most often collected by graduate students and other professionals (e.g., school psychologists, behavior analysts) in the context of research. Therefore, it may be considered less practical and more resource intensive than indirect measures (Chafouleas et al., 2002; Nolan & Gadow, 1994). This is especially true for school psychologists and behavior analysts who have a high case load and may not have the resources to directly observe every referred student and their teachers repeatedly.

Fortunately, researchers have created measures such as Direct Behavior Ratings (DBRs), which combine systematic direct observations with behavior rating scales (Chafouleas, et al., 2007) such that immediately after a predetermined duration of direct observation, a behavior rating scale is completed. Numerous behavior rating scales fall under the umbrella of DBRs, such as home-school notes and Daily Behavior Report Cards (DBRC; Riley-Tillman, et al., 2009). The duration of time between an observation of behavior and rating of behavior varies greatly across DBRCs.

DBRCs may be useful to school personnel because they require less resources than systematic direct observations. DBRCs can be completed quickly because they contain one to a few items. In contrast, direct observations may be 20 minutes or more for each observation, and teachers complete the rating, and the school psychologist or behavior analyst receives the information later. DBRCs may be completed via pencil and paper or electronically (Burke & Vannest, 2008). Given these features, DBRCs are time and resource efficient relative to systematic direct observations and allow professionals to complete other tasks while teachers complete the rating. In addition to being an observation tool, DBRCs can serve as an intervention ((Fabiano et al., 2009; Fabiano et al., 2017; Riden et al., 2018; Vannest et al., 2010). Teachers can use them to observe an individual and reward them for appropriate behaviors.

Psychometric Properties

In addition to the feasibility of a measurement instrument or its ability to operate as an intervention, it is important that the measurement instrument demonstrate technical adequacy via its psychometric properties. Fabiano and colleagues (2009) evaluated temporal stability, criterion validity, and content validity of DBRCs for special education students with Attention-Deficit/Hyperactivity Disorder (ADHD). Results indicated that there was substantial temporal stability between administrations on odd days and even days. Moderate correlations were found between direct observations and DBRCs completed. For content validity, academic goals on Individualized Education Plans (IEP) and academic DBRC targets, indicated fair to moderate agreement, while social behaviors' agreement between IEP and DBRC goals suggested no agreement between raters. Although this study found content validity to be poor, it is important to note that DBRCs are flexible in terms of the content of behavioral items. Therefore, poorly chosen behavioral targets that do not match IEP goals are the product of individuals that create the DBRC and not the instrument.

Miller, Dufrene, Olmi, et al. (2015) and Miller, Dufrene, Sterling, et al. (2015) used DBRCs in the context of tier two interventions and evaluated maintenance by gradually eliminating the teacher completed DBRC and replacing it with a student completed DBRC. Both studies reported correlations between teacher and student raters. Results from both studies indicate moderate to strong, significant correlations between teacher and student raters, which further supports the interrater reliability of DBRCs. In addition to traditional psychometric properties, other researchers have evaluated treatment sensitivity of DBRCs.

Volpe and Gadow (2010) demonstrated that DBRCs are sensitive to changes in treatment for a sample of 65 children with ADHD. Children in this study also had comorbid diagnoses of Oppositional Defiant Disorder and Chronic Multiple Tic Disorder. They evaluated three constructs: inattention-overactivity, aggression, and peer conflict. These constructs were evaluated with the IOWA Conners Teacher's Rating Scale (Loney & Milich, 1982), the Peer Conflict Scale (Gadow, 1986), the Child Symptom-Inventory-4 (CSI-4; Sprafkin et al., 2002), and individualized scales for each participant. The individualized scales had three items to score per participant. Results found that the individualized DBRCs were more sensitive to medication effects of methylphenidate compared to the other measures for constructs of inattention-overactivity, aggression, and peer conflict. Given the role of DBRCs in MTSS to gauge students' response to intervention, results from Volpe and Gadow (2010) are important as they demonstrate that DBRCs may be appropriate for gauging a student's response to intervention.

To contrast, Iznardo et al. (2017) evaluated treatment sensitivity of DBRCs within the context of an intervention for a sample of 272 participants with ADHD. Results indicated that direct observations may be more sensitive to ADHD symptoms compared to teacher ratings. However, it may be that the broad operational definitions of target behaviors on the DBRC in Iznardo et al. negatively impacted sensitivity of the DBRC items relative to the operational definitions used for direct observations. Regardless, it would behoove researchers to conduct additional research evaluating the treatment sensitivity of DBRCs. Relative to DBRCs, there has been more research conducted evaluating the psychometric properties of DBRs.

Effectiveness as an Intervention

In addition to being used as an assessment measure, DBRCs have also been used to alter students' behavior. Results of research testing DBRCs suggest that they may result in increases in academic and social behaviors for students with disabilities (Atkeson & Forehand, 1979; Barth, 1979; Burke & Vannest, 2008; Chafouleas et al., 2002; Riden et al., 2018; Smith et al., 1983; Vannest et al. 2010) and general education students (Riley-Tillman et al., 2009; Volpe & Briesch, 2012).

Further, DBRCs are implemented across grades, though primarily implemented with elementary students, and with a variety of problem behaviors and diagnoses (Riden et al., 2020). Thus, the current study included grades from Pre-K through sixth grade, with any type of behavior targeted. Such problem behavior targets typically include but are not limited to work completion (Jurbergs, et al., 2007), and increasing on-task behaviors (Kelley & McCain, 1995). However, DBRC intervention effectiveness has varied (Riden et al., 2018; Vannest et al., 2010).

Schumaker et al. (1977) investigated DBRCs in a secondary setting within three experiments that targeted several behaviors, such as rule following, classwork completion, grades, and teacher satisfaction. In the first experiment, three students' behaviors were monitored. When a student met classroom expectations in a class, the student earned two points, while if a student broke more than one rule per class, zero points were awarded for that class. Parents and interventionists communicated once per week which included a discussion regarding the previous week's DBRC. For all students, behaviors improved. In the second experiment, a DBRC was used to determine the necessity of parent praise for the improvement of school performance. For one student, the combination of praise and DBRC resulted in increasing appropriate behaviors, while another student required home consequences related to the DBRC scores to then improve school behaviors. In the third experiment, school guidance counselors were provided with a DBRC instructional manual, and in using the DBRCs described within, they reported an increase in school performance for two students.

Yeo et al. (2018) tested an online DBRC to decrease off-task behavior for students with ADHD in Singapore. Immediately following each class, teachers rated students' off-task behaviors that were then converted to percentages of on-task behaviors. For one student, off-task behavior averaged 38% in baseline, and the DBRC intervention reduced the behaviors to an average of 21% across sessions. For another participant, off-task behavior originally averaged 69%, and dropped to 50% on average. For the third participant, off-task behavior occurred an average of 22% across sessions, and with a DBRC, behaviors ranged from 0-28%. With the third participant, it should be noted that only five intervention data points were collected, and there was only one session in which off-task behavior was at 28%; all other sessions were at or near 0%. The findings from Yeo and colleagues suggest that, even for students with low percentages of off-task behavior, DBRCs may be effective in further reducing off-task behavior, although these differences may not be represented as a statistically significant difference. Thus, it may be useful to determine the extent of DBRC effectiveness studies; meta-analyses are one way for such a determination.

Recent Meta-Analyses

Two relatively recent meta-analyses pertaining to DBRCs have been conducted by Vannest et al. (2010) and Riden et al. (2018). Vannest and colleagues (2010) conducted a meta-analysis of single-case research studies that tested the effectiveness of DBRCs as an intervention tool. The meta-analysis included 17 studies published between 1970 and 2007. The 17 studies included 107 participants. Their inclusion criteria limited their search results to include DBRCs as interventions. They omitted studies that related to academic performance, such as reading fluency, but the caveat was that skills such as study skills, time on task, work attempts, and work completion were included. For inclusion, studies needed to share information across stakeholders with parents involved in either the intervention planning, reinforcement, or feedback. Studies that included only self-monitoring interventions, or those without home involvement were not included. The current study varied in that we included interventions with and without caregiver involvement to determine if that was an effective component for DBRCs.

Their findings were analyzed with an improved rate-difference (IRD) as the overlap-based effect size due to its ability to account for unstable baseline data. The IRD effect size represents the proportion of higher scores between baseline and intervention. Scores after the B phase were not included in the analysis. For their study, an 83% confidence interval was used to test if there was a statistically significant difference (p = .05) between IRD scores in baseline and intervention. If there was no overlap between the upper and lower limits of the 83% confidence interval for the two IRDs, then a statistically significant difference was determined. Overall, the effect sizes widely ranged from -0.14 to 0.97.

Vannest et al. (2010) examined six potential moderators, including student characteristics (i.e., age, primary vs. secondary settings, targeted behaviors), home-school collaboration (i.e., home training, reinforcement collaboration, quality of student feedback), breadth of use across the school day (i.e., multiple hour durations vs. an hour or less), scale construction (i.e., frequency/duration scales, qualitative scales, combination scales), and lastly, reliability measurement (i.e., comparison of implementation via school personnel, researchers, or a combination). Student characteristics analyses did not find any significant differences between primary vs. secondary settings, and no significant differences were found between targeted behaviors. However, there were moderate effects when there was a high degree of home involvement across procedures. Interestingly, in moderators for breadth of use, Vannest and colleagues found that DBRCs used for observing behaviors across the day were more effective than those used for an hour or less per day. Qualitative scaling or qualitive and quantitative scaling combinations were more effective than solely quantitative scaling. Also, when a teacher and researcher collaborated on reliability measurement, DBRCs were found to be more effective.

Vannest et al. (2010) also investigated methodological rigor, as defined by Horner et al. (2005). Vannest et al. (2010) created rankings to determine the extent to which studies met the methodological standards, ranking from weak to very strong. For a multiple baseline design to be considered very strong, it needed three phases across three participants. Reversal and changing criterion designs were very strong if they included seven or more phases. Other methodological rigor indicators included examining changes between A-B phases for effect, which Vannest et al. (2010) calculated using IRD. After data extraction, the authors used IRD to compare over-lapping data for each A-B contrast. The authors reported that an average DBRC intervention study resulted in an IRD of 61% compared to baseline, with an estimated range of improvement between 56% to 66%; thus, in comparison to baseline, DBRCs appear to be effective in improving behaviors. However, overall, IRDs ranged from -0.14 to 0.97, and such a range suggests further examination of effectiveness is warranted, and there may be unknown moderators that influence the effectiveness range.

Riden et al. (2018) conducted a synthesis and meta-analysis of both single-case and group design DBRC studies published between 2007 and 2013. Unlike Vannest (2010), Riden and colleagues did not include any moderator analyses. However, they did anecdotally report research design quality indicators, such as clear descriptions of participants, settings, dependent variables, independent variables, and baseline data.

Quality design indicators also included internal, external, and social validity (Horner et al., 2005). Two studies (LeBel et al., 2013 and Sanetti et al., 2016) met all such indicators. Although IOA was indicated as an important factor of design quality, it was not included in a moderator analysis; thus, the current study will include DBRC IOA as a potential moderator.

Their meta-analysis included a total of 11 studies; three of which were single-case designs, and eight were group designs. With a total of 390 participants, 11 participants were in single-case design studies, whereas the remaining 379 were in group design studies. All participants were identified as at-risk, eligible for special education services, and/or having a disability, such as ADHD or emotional and behavioral disorders (EBD). Inclusion criteria included using a DBRC intervention for students with challenging behaviors and at risk or with a disability. Interventions had to occur in a Pre-K through 12th grade setting with special education teachers. Their study did not include Check-In/Check-Out (CICO) usage of DBRCs.

For analysis, they utilized a Tau-U effect size for single-case effect analysis to indicate the percentage of non-overlap between phases for each study. Maintenance sessions were not included in the analysis. Tau-U values between 0 and .65 were considered weak or small effects, values between .66 and .92 were considered medium to large effects, and values between .93 and 1.00 were considered medium to large effects. The Tau-U values for this study resulted in a medium effect (.66). In contrast, the eight group research designs were reported with Hedges' *g*. Scores were interpreted with a range from 0-1. Hedges' *g* values between 0 and 0.50 were considered small effects, values between .50 and .80 were considered medium effects, and values between .80 and

1.0 were considered large effect sizes (Cohen, 1988). Hedges' *g* reported a low to medium effect (0.03-0.72). Therefore, across two meta-analyses, researchers have found a large range of effect sizes across studies.

Gaps in the DBRC Meta-Analysis Literature

The previous meta-analyses both determined methodological quality based upon Horner et al. (2005). Thus, there have yet to be meta-analyses on DBRCs that have included the most recent WWC standards (2020). Previous WWC standards (i.e., 2.1 and 3.0) had the same single-case design standards as one another, wherein to fully meet such standards, there must be systematic manipulation of the independent variable, IOA must be collected for at least 20% across phases, and at least three effect demonstrations (WWC, 2013). The newest WWC standards require at least 20% IOA per phase, graphical representation of data, and no residual treatment effects. With such increasingly rigorous standards, it is vital that re-interpretation of DBRC studies are done to determine if they meet the newest WWC standards. Of Moreover, Vannest et al. (2010) used IRD as the effect size, and IRD is problematic in that outcome values may be less representative of behavior change and more reflective of procedural variations such as number of data points per condition (Zimmerman et al., 2018). This study will use BCT to account for any baseline trends. Finally, the available meta-analyses have limited their inclusion criteria to only include students with disabilities and social behavioral referral concerns. The current study is more inclusive and includes students with and without disabilities, and any type of behavioral concern. Due to the addition of students without disabilities, disability status was added as a potential moderator.

Purpose of the Present Study

DBRCs have been used as assessment tools and interventions, and have empirical support for psychometric properties, as well as intervention effectiveness. There have been two meta-analyses examining the effectiveness of DBRCs as an intervention. We aim to expand these by also examining DBRC intervention WWC standards (2020), and a moderation analysis regarding experimental design, DBRC IOA, publication status, disability status, and home/caregiver communication. These moderators were chosen due to multiple-baseline designs being the primary single-case design in the DBRC literature, the importance of IOA in relation to a study's reliability, publication bias issues, previous studies did not include those without disabilities, and caregiver communications are considered to be important for the success of a DBRC. The previous meta-analyses both only examine methodological rigor based on Horner, Riden limits the interventionists to special education teachers only (2018), and Vannest referral behaviors did not include academic behaviors (2010). Also, the DBR-MIS and DBR-SIS literature is relatively new, and as a result, meta-analyses have not included DBR studies that included the DBR as an intervention tool (Riden et al., 2018; Vannest et al., 2010). Previous meta-analyses only included studies that have been published in peer-reviewed journals and did not include unpublished theses and dissertations, which may reflect publication biases. Lastly, the most recent meta-analysis on DBRCs included studies up to 2017, so there is a gap of approximately 5 years in the meta-analysis literature. Notably, only two of their studies were published after 2013. Therefore, the purpose of this study is to address these gaps in the literature as they pertain to meta-analyses of studies testing the effectiveness of DBRCs as an intervention tool.

Research Questions

The following research questions were addressed:

1. What is the effect of using DBRCs as interventions to improve student behavior?
2. What is the extent to which studies including DBRCs as interventions met design standards set forth by What Works Clearinghouse (WWC; 2020)?
3. Do moderators such as research design characteristics (i.e., design type, DBRC IOA), publication status, participant disability status, or home communication moderate treatment effect?

CHAPTER II - METHOD

Search Procedures

A systematic literature search was the first step for inclusion of studies. Databases were searched for single-case research studies published between 2007 and 2022. This date range was selected to include studies from the most recent meta-analysis to present day. The primary investigator and another graduate student conducted independent literature reviews. This study's initial searches included the following electronic databases: ERIC, ProQuest Educational Journals, ProQuest Dissertation Abstracts, Psycinfo, PsycArticles, EBSCO, and Google Scholar. Search terms included those from Vannest et al. (2010) and Riden et al. (2018): "Daily Behavior Report Card(s)," "Daily Behavior Report(s)," "DBRC(s)," "DBR(s)," "Home-School Note(s)," "Home School Communication(s)," "Home-based reinforcement(s)." The current study added the following search terms: "daily report card(s)," "daily behavior form(s)," "Direct Behavior Rating(s)," "Direct Behavior Rating Scale(s)," "Daily Behavior Rating(s)," and "Daily Behavior Rating Scale(s)." The Boolean operator "AND" were used to include only studies for "elementary" aged participants. The Boolean operator "OR" was used between all other initial search terms.

Inclusion and Exclusion Criteria

For inclusion, studies must have met the following criteria: a) single-case research design, b) published between 2007 and 2022, c) have student(s) identified in preschool through 6th grade, d) describe DBRCs to increase or decrease student behaviors (could target social, emotional, academic behavioral problems), e) published in an English language, and f) published in a peer reviewed journal or a thesis/dissertation project.

If studies did not meet the above criteria, they were excluded from the study. In addition to inclusion criteria, there were additional exclusion criteria. If the study did any of the following, they were not included in the study: a) did not clearly state investigation of DBRC effectiveness in decreasing/increasing student behaviors (i.e., studies in which psychometric properties were the focus of investigation), b) included DBRCs as part of Check-In/Check-Out, or c) included a group design. Group design studies were excluded due to the current study's emphasis on the methodological rigor of single-case design studies, and the inclusion of only single-case design studies allowed for a single set of effect sizes and a focus on WWC design standards.

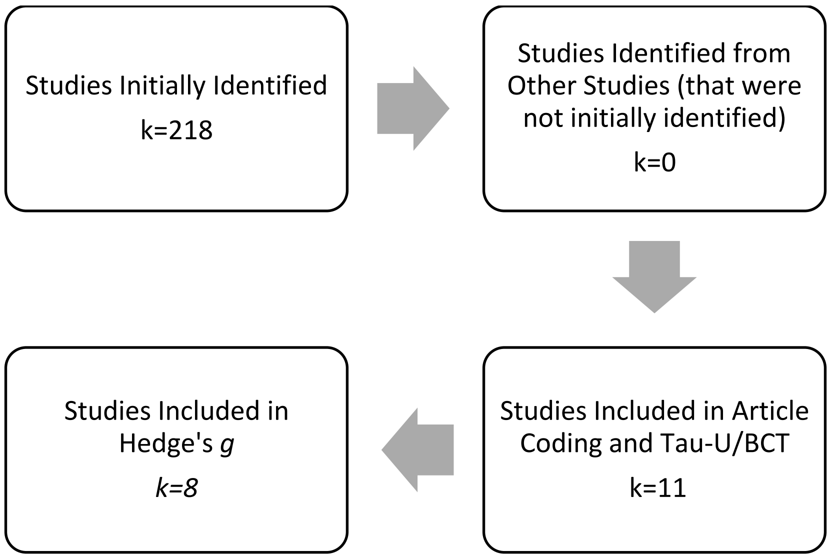
The search resulted in initially identifying 218 single case design studies testing home school notes, DBRs, or DBRCs. After the initial search and omitting duplicate articles, reference sections of articles were searched to identify other articles that meet inclusion criteria. No additional articles were identified.

After the initial literature search, the primary investigator applied the inclusion criteria to each study. Another graduate student independently coded 20% of the 218 studies. IOA for the initial inclusion phase of the study was equal to 90% using a total agreement method. When discrepancies were found regarding inclusion of studies, the investigators reached a verbal agreement regarding inclusion versus exclusion into the current study. After such verbal agreements, IOA for the inclusion phase was 100%.

During the inclusion phase, ninety-eight studies were either group designs, commentaries regarding DBRs or DBRCs, and/or Check-In/Check-Out studies. Of the remaining excluded studies, the majority of which focused on the psychometric properties of the development of DBRs and/or did not provide graphical representation of

DBRC associated data. This resulted in 11 studies for the TAU-U and BCT analyses. Three studies were excluded from Hedge's g analyses due to their not meeting assumptions, such as less than three participants.

Figure 1.
Studies Included by Phase



Article Coding

After a study met inclusion criteria, it was coded for descriptive variables, such as: participant characteristics (i.e., sex, grade, disability), setting characteristics (i.e., type of classroom, such as general education, self-contained classroom), research design characteristics (i.e., experimental design, interobserver agreement, treatment integrity), and if any home communication was included.

Evaluation of Methodological Rigor

To determine the extent to which a study's design and procedures met WWC SCD standards the following standards were also coded (2020; Appendix A): data availability, independent variable(s), inter-assessor agreement, residual treatment effects (if applicable), and other concerns such as confounding factors and if training phases overlapped. All of these standards must be met to determine the last standard, which varies according to design type and includes demonstrations of effects over time and data points per phase. For dichotomous standards (i.e., data availability, independent variables, inter-assessor agreement, and residual treatment effects), data were recorded as either yes for meeting standards or no for not meeting standards. For data availability, studies must provide data in a graphical or tabular display for visual analysis. Also, an independent variable must be systematically manipulated. For inter-assessor agreement, also known as interobserver agreement, there must be agreement between assessors in each phase and at least 20 percent of the data points in each condition. If measured by percentage of agreement, the minimal standard for agreement was at least 80%. If measured by Cohen's kappa, values must be at least 60%. The value of 80% or 60% must be met across all phases and participants but is not a requirement for each phase or participant.

Regarding residual treatment effects, those would only be judged if a study included three or more interventions, such as in an alternating treatment design. If a content expert and reviewer deemed that the intervention's effectiveness could not be exclusively due to the intervention, then the study would not meet standards. For the purposes of the current study, the content expert was a professional who has taught a single-case design course, and the reviewer was the principal investigator. If training phases were present and they overlapped among participants, then a study did not meet standards. For SCDs, if there was a different interventionist in the baseline compared to the intervention phase, that was a confounding variable, and the study did not meet standards.

The remaining standards were dependent on design type (i.e., reversal or withdrawal, multiple baseline and multiple probe, alternating treatment, changing criterion). The standards for reversal or withdrawal designs are identical to those for the changing criterion designs. Each design type varies in the criteria for the number of data points and phases required, and each design was coded as meeting standards without reservations, meeting standards with reservations, or not meeting standards. For instance, in a reversal or withdrawal design, to meet standards without reservations, there must be at least four phases per participant or subject, with at least five data points per phase. To meet standards with reservations, there must be at least four phases per participant or subject, with at least three data points in each phase. If a phase contains fewer than three data points, the reversal or withdrawal design would not meet standards. Further details for the other study design types may be found in Appendix A. Additionally, to determine if the interventions or studies are effective, the WWC recommends using effect sizes, such as Hedges' *g*, which the researcher employed as previously described for the current study.

A secondary observer coded three (27.27%) of the 11 included studies. A point by point agreement method was used, such that each study's variable was coded for agreement. Initial agreement was 96.8% (ranging from 92.3-100%) between two observers. After discussing discrepancies in coding, final agreement resulted in 100% between both observers.

Moderator Analysis

A moderator analysis was conducted for potential moderating variables to determine heterogeneity of effects. After the initial search and coding procedures, moderators were identified. Potential moderators included research design characteristics (i.e., experimental design, interobserver agreement), publication status, participant characteristics (i.e., disability), and if any home communication was included.

Following the procedures of Soares et al. (2016), Bowman-Perrot et al. (2014), and Briesch and Briesch (2016), the researcher analyzed moderator effects by dichotomously coding the moderator variables within studies and examining statistically significant differences using Tau-U within each category. Calculations for reliable difference (i.e., differences unaccounted by solely by chance) for each moderator pair overall and within DBRC type to determine if differences were statistically significant. The following formula was used: (L1 – L2)/√[(SETau12) + (SETau22)], where L1 is the first level of the moderator (e.g., setting) and L2 is the second level of the moderator (e.g., behavior type).

Data Extraction

DigitizeIt version 2.5.9 was the software program that extracted the raw data from studies' graphs (Bormann, 2020). DigitizeIt was chosen due to its high degree of reliability with the original studies' data (*r* = .990, range = .933–1.000; Rakap et al., 2016), and its availability for Windows, Mac, and Linux users (Bormann, 2020). Upon meeting inclusion criteria, raw data were extracted from each graph within the 11 studies to provide data for analysis regarding the magnitude of effects (Pustejovsky & Ferron, 2017). Nine graphs (out of 15 total) from Pyle (2018) were excluded due to their lack of graphical representation of their x-axes. A second observer recorded data for three (27.27%) of the eleven included studies. IOA for data extraction included a calculation for mean count per interval IOA. Average agreement across data points for all studies was 96.58% (ranging from 82.7-100%) across observers.

Effect Size Calculations

From the raw data provided by the DigitizeIt software (Bormann, 2020), effect sizes were calculated for each study (k=11). Baseline-corrected Tau (BCT) and Tau-U were employed for calculating effect sizes. BCT is a non-parametric effect size that evaluates overlap (or lack thereof) of data points in adjacent phases. The online calculator available at ktarlow.com was used to calculate BCT and Tau-U (Karlow, 2016). The online calculator allows for a test of a significant monotonic baseline trend and if there was a significant baseline trend then the Tau-U value was adjusted to account for trend. BCT values between 0 and 0.2 were described as small effects, values between 0.2 and 0.6 were considered moderate effects, and values larger than 0.8 were described as large effects (Vannest & Ninci, 2015).

To compare effect sizes between studies, Hedges' *g* was used for studies that included three or more participants to determine the magnitude of effects (k=8; Pustejovsky & Ferron, 2017). It also accounted for any overestimation that may occur due to small sample sizes (Hedges, 1981). Hedges' *g* is a member of the standardized mean difference family of effect size calculations and has been used previously in meta-analyses of single case research design studies (e.g., Riden et al., 2018). The calculation for Hedges' *g* is g=M1-m2/Pooled Standard Deviation (Ellis, 2009). To conduct the meta-analysis using Hedges' *g*, each effect size estimate and its standard error were entered into the R software to calculate an omnibus effect size (RStudio, 2020). R software is a free program available on multiple platforms, such as Windows, Mac OS, and Linux. Interpretations were as follows, with scores ranging from 0-1. Hedges' *g* values between 0 and 0.50 considered small effects, values between .50 and .80 considered medium effects, and values between .80 and 1.0 considered large effect sizes (Cohen, 1988).

CHAPTER III – RESULTS

Eleven studies with a total of thirty-nine participants were included. Seven studies were published in peer-reviewed journals and four were theses or dissertations. Articles were published between 2007 and 2022. Studies in the present meta-analysis included many single-case research designs, including multiple-baseline (k=7; 63.6%), AB (k=1; 9.1%), reversal (k=1; 9.1%), changing criterion (k=1; 9.1%), and multiple probe multiple baseline (k=1; 9.1%).

Race was specified for 37 of the participants across studies (two participants were unspecified). Within the meta-analysis, 23 students identified as Caucasian (62.1%), six students identified as African American (16.2%), six identified as Hispanic or Latino (16.2%), one as Pacific Islander (4.3%), and one as Asian American (4.3%). Participants were primarily male (n=31; 79.5%).

Table 1 *Demographic Information*

	Demographic Variable	*N*
Race/Ethnicity		
	Asian American	1 (2.6%)
	African American	6 (15.4%)
	Caucasian	23 (58.97%)
	Hispanic or Latino	6 (15.4%)
	Pacific Islander	1 (2.6%)
	Unspecified	2 (5.1%)
Sex		
	Female	6 (20.5%)
	Male	31 (79.5%)

Students within the present meta-analysis received services across preschool through sixth grade. Across all cases, 11 were in preschool, five were in kindergarten, three in first grade, six in second grade, two in third grade, five in fourth grade, six in fifth grade, and one in sixth grade. Of the participants, 19 did not have reported disability statuses, while 20 did have reported disability statuses. Of the reported disabilities, four had an Intellectual Disability, two had Specific Learning Disability (SLD), six with Autism Spectrum Disorder (ASD), four with Attention Deficit Hyperactivity Disorder, and three had more than one reported disability (one with ASD, ADHD, and SLD; one with ASD, ADHD; one with speech/language impairment and ADHD). Students received services in a variety of placements. The majority of studies did not specify school placement (15 students), while another 15 students were in general education, five were in inclusion classrooms, four in self-contained.

WWC Standards

Eleven studies were evaluated to determine the extent to which they met WWC standards for single-case design methodology (WWC, 2020). None of the studies met WWC standards. All of the studies (100%) provided graphical representation of data, systematically manipulated the independent variables, and included no residual treatment effects. However, none of them met criteria in their reports for collecting interobserver agreement, due to the lack of reporting IOA for at least 20% of observation per phase and case. Therefore, overall, none of the studies met WWC standards. More details in regard to the extent to which studies met design standards appear in Table 2. Since WWC standards became stricter in 2020 regarding IOA per phase and case, it is important to note that some studies may have previously met WWC standards.

Table 2 *WWC Standard by Study*

Design Standards	Study	Standards Not Met
Does Not Meet *k* = 11 (100%)	**Canfield & Cividini-Motta (2021)**	Interobserver agreement; Data points per phase
	Chafouleas, et al. (2007)	Interobserver agreement; Number of phases per case
	Daniels et al. (2021)	Interobserver agreement; Data points per phase

Table 2 Continued

Gilic (2016)	Interobserver agreement
Goldman (2016)	Interobserver agreement
Grant (2012)	Interobserver agreement; Number of phases per case
LeBel et al. (2013)	Interobserver agreement; Data points per phase
Lopach et al. (2018)	Interobserver agreement
Pyle (2018)	Interobserver agreement
Taylor & Hill (2017)	Interobserver agreement; Data points per phase

DBRC Interobserver Agreement and Integrity

Of the eleven studies in the present meta-analysis, only two collected IOA for DBRC data. Thus, seven out of the 37 students had IOA for their DBRC data. Regarding integrity data, 10 out of the 11 studies collected some form of integrity data, for a total of 37 out of thirty-nine students across studies. Of the studies including integrity data, data were collected for at least 20% of intervention phases and integrity ratings were at least 90%, with no rating below 84%.

Overall Effect

Tau-U was calculated for each A-B contrast in eleven studies within the present analyses. Effect sizes were only calculated for baseline to intervention contrasts, so an overall Tau-U value was not calculated for each study. Tau-U was calculated for 114 A-B contrasts. Each contrast was assessed for baseline trend and a significant baseline trend was present for four (3.5%) A-B contrasts. When baseline trends were present, baseline-corrected Tau-U was calculated. Values for Tau-U and baseline-corrected Tau-U for each A-B contrast are contained in Table 3. Tau-U values ranged from values of 0 to 0.937. The average Tau-U value was 0.508, indicating a moderate effect. When BCT was used (i.e., 4 contrasts), the data were bolded (see Table 3 below). For BCT values, they ranged from 0.392 to 0.640; the average BCT was 0.5165, indicating a moderate effect.

Table 3 *Baseline-Corrected Tau and Tau-U*

Study Dependent Variable	Participant and Phase Contrast	Tau	*p*	SE
Canfield, et al. (2021)				
Percent of Disruptive Behavior	Rick	0.361	0.100	0.320
	Jake	0.688	0.001	0.242
		0.514	0.036	0.324
	Nate	0.734	0.001	0.248
		0.603	0.049	0.357
DBRC Scores	Rick	0.541	0.018	0.289
	Jake	0.730	0.001	0.228
		0.532	0.039	0.320
	Nate	0.752	0.001	0.241
		0.632	0.046	0.346
Chafouleas et al. (2007)				
Teacher Rating	Mark	0.078	0.853	0.425
	Brian	0.530	0.114	0.379
	Jason	0.361	0.171	0.366
Observer Rating	Mark	0.281	0.357	0.409
	Brian	0.779	0.013	0.281
	Jason	0.392	0.132	0.361
Percent of On-Task Behaviors	Mark	0.295	0.315	0.407
	Brian	0.566	0.060	0.369
	Jason	0.299	0.252	0.374
Daniels et al. (2021)				
Academic Engagement	Veronica	0.109	0.697	0.376
		0.167	0.518	0.373
	Daniel	0.547	0.015	0.296
		0.070	0.806	0.364
	Ethan	0.138	0.616	0.374
		0.318	0.233	0.372
Disruptive Behavior	Veronica	0.017	1.000	0.378
		0.220	0.431	0.369
	Daniel	0.508	0.025	0.305
		0.569	0.016	0.300
	Ethan	0.254	0.339	0.366
		0.435	0.118	0.353
Gilic (2016)				
Frequency of Completed Homework	Student S	0.607	0.024	0.312
		0.824	0.003	0.231
	Student J	0.686	0.011	0.285
		0.885	0.003	0.190
Accuracy of Completed Homework	Student S	0.680	0.011	0.287
		0.775	0.004	0.258
	Student J	0.691	0.010	0.283
		0.937	0.002	0.143

Table 3 Continued

Goldman (2016)				
Off-Task Behavior from Teacher	Ryan	0.391	0.014	0.234
	Daniel	0.286	0.237	0.350
	Leo	0.264	0.194	0.305
	Emily	0.339	0.038	0.247
Off-Task Behavior from Researcher	Ryan	0.368	0.033	0.263
	Daniel	0.496	0.009	0.268
	Leo	0.009	0.978	0.283
	Emily	0.494	0.002	0.228
Grant (2012)				
Homework Accuracy	Michael	0.252	0.341	0.380
		0.307	0.269	0.389
	Ruth	0.606	0.032	0.339
		0.732	0.037	0.341
	Esther	0.577	0.039	0.309
		0.426	0.216	0.369
	Sarah	0.548	0.172	0.418
		0.739	0.026	0.318
Homework Completion	Michael	0.373	0.160	0.364
		0.667	0.015	0.304
	Ruth	0	1.085	0.426
		0.061	1.000	0.499
	Esther	0.160	0.596	0.373
		0.066	0.911	0.407
	Sarah	0.564	0.168	0.413
		0.762	0.024	0.305
Classwork Accuracy	Michael	0.551	0.033	0.327
		0.539	0.049	0.344
	Ruth	0.306	0.306	0.406
		0.732	0.037	0.341
	Esther	0.649	0.010	0.299
		0.252	0.341	0.380
	Sarah	0.732	0.037	0.341
		0.707	0.028	0.333
Classwork Completion	Michael	**0.394**	**0.124**	**0.361**
		0.392	**0.149**	**0.376**
	Ruth	0.621	0.045	0.334
		0.826	0.026	0.282
	Esther	0.502	0.059	0.339
		0.386	0.152	0.362
	Sarah	0.745	0.036	0.333
		0.775	0.022	0.298
On-Task Behavior	Michael	0.252	0.341	0.380
		0.579	0.033	0.333
	Ruth	0.661	0.019	0.320
		0.732	0.037	0.341
	Esther	0.468	0.067	0.347
		0.725	0.004	0.270

Table 3 Continued

		Sarah	0.732	0.037	0.341
			0.707	0.028	0.333
LeBel et al. (2013)					
	Disruptive Behavior	Robby	0.697	0.003	0.262
			0.707	0.079	0.378
		Wendy	0.703	0.000	0.231
			0.624	0.057	0.369
		Zander	0.676	0.001	0.233
			0.572	0.045	0.350
		Jake	0.691	0.000	0.209
			0.535	0.038	0.331
Lopach et al. (2018)					
	Percentage of On-Task Behavior	Participant 1	0.716	0.004	0.274
			0.732	0.037	0.341
		Participant 2	0.716	0.004	0.274
			0.732	0.037	0.341
		Participant 3	0.603	0.023	0.326
			0.049	1.000	0.499
		Participant 4	0.524	0.051	0.348
			0.149	0.764	0.494
Pyle (2018)					
	DBR Points for Follow Directions	Patrick	**0.640**	**0.001**	**0.249**
	DBR Points for Actively Involved	Patrick	0.619	0.005	0.269
	DBR Points for Safe Hands	Patrick	**0.640**	**0.001**	**0.249**
	Percentage of Academic Engagement	Henry	0.361	0.125	0.341
	Percentage of Disruptive Behavior	Henry	0.111	0.668	0.363
	Percentage of Social Engagement	Henry	0.340	0.220	0.343
Taylor et al. (2017)					
	Total DBR Points	Tim	0.650	0.004	0.261
		Willy	0.694	0.001	0.240
		Addy	0.758	0.002	0.238
		Art	0.642	0.003	0.256

An omnibus effect was calculated based upon data from eight studies. The other three studies were not included due to not meeting the assumptions of Hedge's g, such as 3 or more participants, or baseline data were zero. The Hedge's *g* value for these eight studies was 1.1632, p <0.0001. Within this model, tests of heterogeneity were significant. Therefore, the data were tested to see if significant outliers were present. Two effect sizes across eight studies were determined to be outliers and excluded from the meta-analysis. Effect sizes by study and variable, as well as outliers, are noted in Appendix A. When outliers are removed, the omnibus Hedge's *g* value was 1.0456, p =0.0001. This model accounted for 17.9% of variance in the data. Effect sizes were also calculated by dependent variable per each study that included three or more participants for a total of seven studies and 88 phase contrasts. Details of the effect sizes for each study are found in Table 4. Of note, in the Chafouleas et al. (2007) study, one participant's standard deviation was zero, so the second participant's data are not accounted for.

Table 4 *Effect Size by Dependent Variable Per Study*

Author	Dependent Variable	*n* (Contrasts)	Hedge's *g*	SE	95% Confidence Interval Lower Limit.	Upper Limit
	Overall	**88**	**1.0456**	**0.18**	**0.695**	**1.396**
Canfield & Cividini-Motta (2021)	Disruptive Behavior	5	2.8	1.077	0.6894	4.912
	DBRC Scores	5	2.4439	0.6295	1.2109	3.677
Chafouleas, et al. (2007)	Teacher Rating	2	0.5384	1.19	1.4581	2.535
	Observer Rating	2	0.7585	0.0738	0.6139	0.903
	% On-Task	3	0.6846	0.349	0.1981	1.567
Daniels et al. (2021)	Academic Engagement	6	0.3745	0.2624	0.1435	0.892
	Disruptive Behavior	6	0.8694	0.0066	0.3686	1.370
Goldman (2016)	Off-Task Teacher Rating	4	0.8753	0.2607	0.3643	1.386
	Off-Task Researcher Rating	4	0.6431	0.8121	0.9487	2.235
Grant (2012)	Homework Accuracy	8	0.9933	0.3774	0.2536	1.732
	Homework Completio	8	0.5	0.0443	0.1835	0.816
	Classwork Accuracy	8	1.2655	0.3953	0.4907	2.040
	Classwork Completion	8	1.1462	0.3333	0.4929	1.799
	On-Task Behavior	8	2.4723	0.8675	0.7719	4.173
LeBel et al. (2013)	Disruptive Behavior	8	3.126	0.5320	2.0819	4.170
Lopach et al. (2018)	% On-Task	8	1.7725	0.7723	0.2586	3.286

Moderator Analysis

Several variables were assessed to determine if they moderate the effects of DBRCs. Hedge's g was used to evaluate whether or not design type, publication status, DBRC IOA, disability status, or caregiver involvement influenced the effect of DBRCs. Table 5 contains moderator variables' effect size data.

Table 5 *Effect Size Results for Moderator Variables*

Moderator		*k* (studies)	Hedge's *g*	95% Confidence Interval Lower Limit	Upper Limit
Design					
	Multiple Baseline Design	10	0.3573	0.0241	1.8048
	Other Single Case Design	4	0.8903	0.0241	1.8048
Published					
	Yes	9	0.2731	0.6141	1.1603
	No	5	1.0226	0.3758	1.6694
DBRC IOA					
	Present	2	1.5519	0.0367	3.0671
	Absent	12	1.0035	0.6250	1.3820
Disability Status					
	Disability Reported	11	0.2529	1.2654	0.7597
	No Disability Reported	3	1.3573	0.4885	2.2262
Caregiver Involvement					
	Present	9	0.1019	1.1208	0.9169
	Absent	5	1.2507	0.3794	2.1219

Design

Studies were compared as either having a multiple-baseline design or another type of single case design. The majority of studies (k=10) were multiple baseline studies. Multiple baseline design had a small effect on studies (g=0.3573). Further, the effect of design on DBRC outcomes was not significant ($F_{1,14} = 0.5395$, $p = 0.4747$).

Published

Studies that were published in peer-reviewed journals were compared to studies that were unpublished theses or dissertations to determine if publication status impacted DBRC effectiveness. The majority of studies (k=9) were published in peer-reviewed journals. Publication in a peer-reviewed journal had a small effect on studies (g=0.2731);

in addition, the effect of publication on DBRC outcomes was not significant ($F_{1,14}$ = 0.4358, p = 0.5199).

DBRC IOA

Studies that included interobserver agreement for DBRCs were compared to studies that did not include DBRC IOA to determine if DBRC IOA impacted DBRC effectiveness. The majority of studies (k=12) did not include IOA for DBRCs. Studies that included DBRC IOA had a slightly larger effect size (g=1.5519) than those that did not. The effect of DBRC IOA had a statistically significant difference on DBRC outcomes ($F_{1,14}$ = 4.8254, p = 0.0454).

Disability Status

Studies that included students with reported disabilities were compared to studies that did not report student disabilities to determine if effectiveness was impacted dependent on student reported disability status. The majority of studies (k=11) reported student disabilities. However, disability status did not result in a larger effect size compared to other studies (g=0.2529). The effect of students' reported disabilities did not have a significant effect on DBRC outcomes ($F_{1,14}$ = 0.2869, p = 0.6006).

Caregiver Involvement

Studies that included caregiver involvement with DBRCs were compared to studies that did not to determine if caregiver involvement impacted DBRC outcomes. The majority of studies (k=9) included caregiver involvement. Caregiver involvement did not result in a larger effect size (g=0.1019). The effect of caregiver involvement did not have a significant effect on DBRC outcomes ($F_{1,14}$ = 0.0461, p = 0.8332)

CHAPTER IV – DISCUSSION

This meta-analysis of the DBRC literature was conducted to extend findings of previous systematic reviews and meta-analyses by evaluating methodological rigor of the literature based on WWC design standards. Additionally, inclusion criteria for this study expanded teacher and student populations and dependent variables relative to previous reviews of the DBRC literature by including non-disabled students and their teachers and academic behaviors.

With regard to the first research question, what is the effect of using DBRCs as interventions to improve student behavior?, results from this study indicate a significant and omnibus effect, as indicated by the Hedge's g value of 1.1632, $p < 0.0001$. This indicates that student outcomes increase by 1.1 standard deviations, which is larger than the Hedge's g findings for group designs in Riden et al. (2018)'s meta-analysis. In Riden and colleagues' meta-analysis, group design effect sizes ranged from very small to medium effects.

This study's Tau-U values ranged from values of 0 to 0.937. The average Tau-U value was 0.508, indicating moderate effects, which parallels with Riden's results (2018). With the three single case designs from Riden, two had moderate effects (Tau-U = .51 and .65), and one had a large effect (Tau-U = .81). In comparison to Vannest et al. (2010), IRD differences for single-case designs had a large variance (i.e., -0.15-0.97, mean = 0.61), similar to the current study.

With regard to the second research question, what is the extent to which studies including DBRCs as interventions met design standards set forth by What Works Clearinghouse (WWC; 2020)?, results from the current study also extended the DBRC meta-analysis literature by systematically evaluating the methodological rigor of single-case research design studies. In Riden et al. (2018), of the three single-case design studies included, quality indicators were reported (according to Horner et al., 2005 standards), but they were not reported according to WWC standards, or the most current WWC standards. Vannest and colleagues (2010) also used Horner et al. (2005) to determine methodological rigor, and of the 17 studies included in their meta-analysis, 47% of studies (8) were coded as low quality (average IRD = 0.56; CI_{83} [0.47, 0.65]), while 53% of studies (eight) were coded as either medium, high, or very high quality. Unfortunately, the eleven studies included in the current study did not meet all of the WWC standards. In fact, none of them met the criteria for IOA standards of at least 20% IOA per phase. This was in part potentially due to the authors not explicitly stating the percentage of IOA per phase, since the WWC standards were recently altered. Also, studies did have IOA (typically for direct observations). However, it still remains that the lack of DBRC IOA is a huge dilemma for the DBRC literature, especially considering that without those IOA data, it is difficult to determine the reliability, validity, and accuracy of the intervention data.

With regard to the third research question, do design type, DBRC IOA, publication status, participant disability status, or home communication moderate the effects of DBRCs as an intervention tool?, results from this study indicate inclusion of DBRC IOA, publication in a peer-reviewed journal, and a multiple-baseline design type resulted in a small effect size in DBRC outcomes. Home-school collaboration was not found to be a significant indicator of DBRC success, which varies from the Vannest et al. (2010) meta-analysis. The current study's findings also differ from Vannest in that there was a significant finding in relation to the design type. Vannest et al. (2010) found that rather than research design quality, other specific moderators were found to better explain outcomes. Neither of the previous meta-analyses included DBRC IOA as a potential moderator, though it makes theoretical sense that a measure of reliability may increase the likelihood of significant outcomes.

Limitations

There are several limitations to consider for the current study's findings, particularly related to effect sizes. Although standardized mean difference has been demonstrated as valid for reversal and multiple-baseline designs (Hedges, et al., 2012; Hedges et al., 2013), it remains unclear regarding its applications for other designs. In the current study, one of the included studies was a changing criterion design, one an AB design, and another a multiple probe multiple baseline design. In addition, the majority of studies included multiple dependent variables, particularly Goldman (2012), with five dependent variables per participant; thus, some effects are not truly independent, which may thereby affect standard error. An additional limitation is the limited number of studies included in the meta-analysis. Although at least five studies may prove sufficient enough power (Jackson & Turner, 2017), more studies would assist in determining the effectiveness of potential moderators.

Summary and Future Directions

DBRCs have long been used as a progress monitoring tool to gauge students' response to intervention within MTSS and have also been used as an intervention tool. Results from this study indicate that when DBRCs are used as an intervention tool, they produce beneficial effects for students. However, these results must be considered with caution given the findings of the review of methodological rigor of DBRC studies. In particular, results from this study indicate that none of the 11 studies fully met WWC design standards. Additionally, in regard to IOA and treatment integrity design standards, results from this study indicate a lack of DBRC IOA, which is a threat to internal validity of the studies. Finally, regarding participant characteristics, participants in DBRC studies represent a restricted age and grade range (i.e., preschool to sixth grade) and homogeneity in regard to reported participant sex;as a result, there are concerns about the external validity of findings. In research, it is important to know not only what works, but who does it work for and under what conditions. Given the totality of these findings, future DBRC research should include great attention to methodological rigor, particularly IOA per phase and treatment integrity standards, and a wider variety of participants in terms of grade, age, and sex.

References

Atkeson, B. M., & Forehand, R. (1979). Home-based reinforcement programs designed to modify classroom behavior: A review and methodological evaluation. *Psychological Bulletin, 86,* 1298–1308.

Barth, R. (1979). Home-based reinforcement of school behavior: A review and analysis. *Review of Educational Research, 49,* 436–458.

Bormann, I. (2020). DigitizeIt software (version 2.5) [Computer software]. Braunschweig, Germany. Available from https://www.digitizeit.de/

Briesch, A. M., & Briesch, J. M. (2016). Meta-analysis of behavioral self-management interventions in single-case research. *School Psychology Review, 45*(1), 3–18. doi:10.17105/spr45-1.3-18

Burke, M. D., & Vannest, K. J. (2008). Behavioral-progress monitoring using the electronic daily behavioral report card (e-DBRC) system. *Preventing School Failure: Alternative Education for Children and Youth, 52*, 51–60.

*Canfield, D., & Cividini-Motta, C. (2021). The Effects of a Daily Behavior Report Card Intervention: Inclusion of a Peer Mediator. *Journal of Behavioral Education*, 1-21.

Chafouleas, S. M., Riley-Tillman, T. C., & McDougal, J. L. (2002). Good, bad, or in-between: How does the daily behavior report card rate? *Psychology in the Schools, 39*(2), 157-169.

*Chafouleas, S. M., Riley-Tillman, T. C., Sassu, K. A., LaFrance, M. J., & Patwa, S. S. (2007). Daily behavior report cards: An investigation of the consistency of on-task data across raters and methods. *Journal of Positive Behavior Interventions, 9,* 30–37.

Chafouleas, S. M., Sanetti, L. M. H., Kilgus, S. P., & Maggin, D. (2012). Evaluating sensitivity to behavioral change across consultation cases using Direct Behavior Rating–Single-Item Scales (DBR-SIS). *Exceptional Children, 34,* 575–591.

Chafouleas, S. M., Volpe, R. J., Gresham, F. M., & Cook, C. R. (2010). School-based behavioral assessment within problem solving models: Current status and future directions. *School Psychology Review, 39,* 343–349.

Christ, T. J.; Riley-Tillman, T. C. & Chafouleas, S. M. (2009). Foundation for the development and use of Direct Behavior Rating (DBR) to assess and evaluate student behavior. *Assessment for Effective Intervention, 34*: 201–213. doi:10.1177/1534508409340390.

*Daniels, B., Bender, S. L., Briesch, A. M., Susilo, A., & Fallon, L. M. (2021). Expanding Daily Report Card Intervention to the Preschool Setting: Evaluation of Effectiveness and Usability for Teachers and Parents. *Contemporary School Psychology*, 1-14.

Ellis, P.D. (2009). *Effect size equations*. Retrieved from http://www.polyu.edu.hk/mm/effectsizefaqs/effect_size_equations2.html

Fabiano, G. A., Pyle, K., Kelty, M. B., & Parham, B. R. (2017). Progress monitoring using Direct Behavior Rating Single Item scales in a multiple-baseline design study of the daily report card intervention. *Assessment for Effective Intervention, 43,* 21–33.

Fabiano, G. A., Vujnovic, R., Naylor, J., Pariseau, M., & Robins, M. (2009). An investigation of the technical adequacy of a daily behavior report card (DBRC) for monitoring progress of students with attention-deficit/hyperactivity disorder in special education placements. *Assessment for Effective Intervention*, *34*(4), 231-241.

Fabiano, G. A., Vujnovic, R., Pelham, W. E., Waschbusch, D. A., Massetti, G. M., Yu, J., . . .Volker, M. (2010). Enhancing the effectiveness of special education programming for children with ADHD using a daily report card. *School Psychology Review, 39,* 219–239.

Fallon, L. M., Sanetti, L. M., Chafouleas, S. M., Faggella-Luby, M. N., & Briesch, A. M. (2018). Direct training to increase agreement between teachers' and observers' treatment integrity ratings. *Assessment for Effective Intervention*, *43*(4), 196-211.

Gadow, K. D. (1986). Peer conflict scale. Stony Brook, NY: State University of New York, Department of Psychiatry.

*Gilic, L. (2016). Increasing homework completion and accuracy rates with parental participation for young children with Autism Spectrum Disorder.

*Goldman, S. E. (2016). Using School-Home Communication to Reduce Problem Behaviors in Children with Autism Spectrum Disorders. Vanderbilt University.

*Grant, S. M. (2012). Parental involvement in improving academic success for students with ADHD: a comparison of daily behavior report cards and homework self-monitoring.

Gresham, F. M. (2005). Response to intervention: An alternative means of identifying students as emotionally disturbed. *Education and Treatment of Children, 28,* 328–344.

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

Hintze, J. M., & Matthews, W. J. (2004). The generalizability of systematic direct observations across time and setting: A preliminary investigation of the psychometrics of behavioral observation. *School Psychology Review, 33,* 258–270.

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

Iznardo, M., Rogers, M. A., Volpe, R. J., Labelle, P. R., & Robaey, P. (2017). The effectiveness of daily behavior report cards for children with ADHD: A meta-analysis. *Journal of Attention Disorders, 24*(12), 1623-1636.

Jackson, D. & Turner, R. (2017). Power analysis for random-effects meta-analysis. Research Synthesis Methods, 8, 290-302. doi: 10.1002/jrsm.1240

*LeBel, T. J., Chafouleas, S. M., Britner, P. A., & Simonsen, B. (2013). Use of a daily report card in an intervention package involving home-school communication to reduce disruptive behavior in preschoolers. *Journal of positive behavior interventions, 15*(2), 103-112.

Loney, J., & Milich, R. (1982). Hyperactivity, inattention, and aggression in clinical practice. In M. Wolraich & D. K. Routh (Eds.), Advances in developmental and behavioral pediatrics (Vol. 3, pp. 113–147). Greenwich, CT: JAI.

*Lopach, L., Jenson, W., Davis, J., Knorr, J., & O'Neill, R. (2018). The electronic home note program: an effective intervention package to improve parent program participation, on-task behavior, and academic performance. *Contemporary School Psychology, 22*(4), 488-502.

Miller, F. G., Crovello, N. J., & Chafouleas, S. M. (2017). Progress monitoring the effects of daily report cards across elementary and secondary settings using Direct Behavior Rating: Single Item Scales. *Assessment for Effective Intervention, 43,* 34–47.

Miller, L. M., Dufrene, B. A., Olmi, D. J., Tingstrom, D., & Filce, H. (2015). Self-monitoring as a viable fading option in check-in/check-out. *Journal of School Psychology, 53*(2), 121-135.

Miller, L. M., Dufrene, B. A., Sterling, H. E., Olmi, D. J., & Bachmayer, E. (2015). The effects of check-in/check-out on problem behavior and academic engagement in elementary school students. *Journal of Positive Behavior Interventions, 17*(1), 28-38.

Nolan, E. E., & Gadow, K. D. (1994) Relation between ratings and observations of stimulant drug response in hyperactive children. *Journal of Clinical Child Psychology, 23,* 78–90.

Parker, R. I., & Vannest, K. J. (2012). Bottom-up analysis of single-case research designs. *Journal of Behavioral Education*, *21*(3), 254-265. doi:10.1007/s10864-012-9153-1

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single case research: Tau-U. *Behavior Therapy, 42*, 284–299.

Pustejovsky, J. E., & Ferron, J. M. (2017). Research synthesis and meta-analysis of single-case designs. *Handbook of special education*, 168-186.

*Pyle, K. (2018). *Efficacy of the Daily Report Card Intervention for High-functioning Children with Autism Spectrum Disorder: A Multiple Baseline Study* (Doctoral dissertation, State University of New York at Buffalo).

Rakap, S. (2015). Effect sizes as result interpretation aids in single-subject experimental research: description and application of four nonoverlap methods. *British Journal of Special Education*, *42*(1), 11-33.

Rakap, S., Rakap, S., Evran, D., & Cig, O. (2016). Comparative evaluation of the reliability and validity of three data extraction programs: UnGraph, GraphClick, and DigitizeIt. *Computers in Human Behavior, 55*, 159–166. doi: 10.1016/j.chb.2015.09.008.

Riden, B. S., Taylor, J. C., Lee, D. L., & Scheeler, M. C. (2018). A synthesis of the daily behavior report card literature from 2007 to 2017. *The Journal of Special Education Apprenticeship*, *7*(1), 3.

Riden, B. S., Taylor, J. C., Ruiz, S., Lee, D. L., & Scheeler, M. C. (2020). Using a Daily Report Card to Reduce Off-Task Behaviors for a Student with Autism Spectrum Disorder. *Journal of Behavioral Education*, 1-20.

Riley-Tillman, T. C., Chafouleas, S. M., & Briesch, A. M. (2007). A school practitioner's guide to using daily behavior report cards to monitor student behavior. *Psychology in the Schools, 44,* 77–89. https ://doi.org/10.1002/pits.20207.

Riley-Tillman, T. C., Methe, S. A., & Weegar, K. (2009). Examining the use of direct behavior rating on formative assessment of class-wide engagement: A case study. *Assessment for Effective Intervention*, *34*(4), 224-230.

RStudio Team (2020). RStudio: Integrated development for R. [Computer Software]. PBC, Boston, MA. http://www.rstudio.com/.

Sanetti, L. M. H., Chafouleas, S. M., Fallon, L. M., & Jaffrey, R. (2014). Increasing teachers' adherence to a classwide intervention through performance feedback provided by a school-based consultant: A case study. *Journal of Educational and Psychological Consultation*, *24*(3), 239-260.

Schumaker, J. B., Hovell, M. F., & Sherman, J. A. (1977). An Analysis of daily behavior report cards and parent-managed privileges in the improvement of adolescents' classroom performance. *Journal of Applied Behavior Analysis, 10*(3), 449–464.

Sims, W. A., Riley-Tillman, C., & Cohen, D. R. (2017). Formative assessment using Direct Behavior Ratings: Evaluating intervention effects of daily behavior report cards. *Assessment for Effective Intervention, 43,* 6–20.

Smith, M. A., Williams, R. L., & McLaughlin, T. F. (1983). The daily report card as an intervention technique for classroom academic and social behavior: A review. *BC Journal of Special Education, 7,* 369–380.

Sprafkin, J., Gadow, K. D., Salisbury, H., Schneider, J., & Loney, J. (2002). Further evidence of reliability and validity of the Child Symptom Inventory-4: Parent checklist in clinically referred boys. *Journal of Clinical Child and Adolescent Psychology, 31*(4), 513-524.

Soares, D. A., Harrison, J. R., Vannest, K. J., & McClelland, S. S. (2016). Effect size for token economy use in contemporary classroom settings: A meta-analysis of single-case research. *School Psychology Review, 45,* 379–399. doi:10.17105/spr45- 4.379-399.

Tarlow, K. R. (2016). Baseline Corrected Tau Calculator. Retrieved from http://www.ktarlow.com/stats/tau

*Taylor, J. C., & Hill, D. (2017). Using Daily Behavior Report Cards during Extended School Year Services for Young Students with Intellectual and Developmental Disabilities. *Education & Treatment of Children (West Virginia University Press)*, *40*(4).

Vannest, K. J., Davis, J. L., Davis, C. R., Mason, B. A., & Burke, M. D. (2010). Effective intervention for behavior with a daily behavior report card: A meta-analysis. *School Psychology Review*, *39*(4), 654-672.

Vannest, K. J., Parker, R.I., Gonen, O., & Adiguzel, T. (2016). Single Case Research: Web-based calculators for SCR analysis. (Version 2.0) [Web-based application]. College Station, TX: Texas A&M University. singlecaseresearch.org

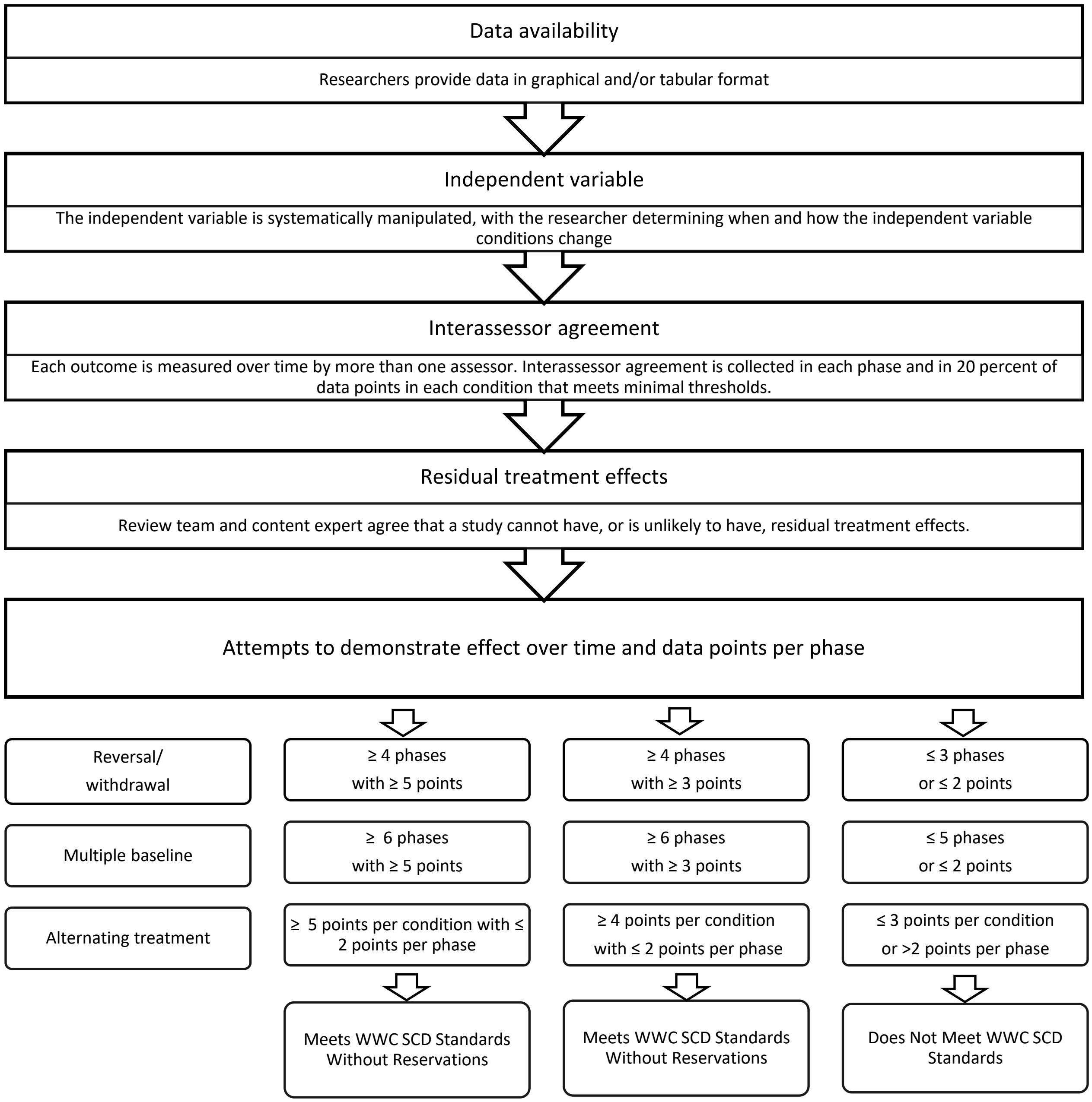
Volpe, R. J., & Briesch, A. M. (2012). Generalizability and dependability of single-item and multiple-item direct behavior rating scales for engagement and disruptive behavior. *School Psychology Review*, *41*(3), 246-261.

Volpe, R. J., & Gadow, K. D. (2010). Creating abbreviated rating scales to monitor classroom inattention-overactivity, aggression, and peer conflict: Reliability, validity, and treatment sensitivity. *School Psychology Review, 39,* 350–363.

What Works Clearinghouse. (2020). What Works Clearinghouse Standards Handbook, Version 4.1. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. This report is available on the What Works Clearinghouse website at https://ies.ed.gov/ncee/wwc/handbooks.

Yeo, W., Goh, A., & Tan, C. (2018). Efficacy of online daily behaviour report card to decrease off-task behaviour for adolescents with attention-deficit/hyperactivity disorder. *Australasian Journal of Special and Inclusive Education*, *42*(2), 143-157.

APPENDIX A - Methodological Rigor Determinants for Single-Case Designs



Note. From “Figure IV.1. Study rating determinants for single-case designs” by What Works Clearinghouse, 2020, *What Works Clearinghouse Standards Handbook* (Version 4.1), p. 85. https://ies.ed.gov/ncee/wwc/handbooks