Daily Report Card Intervention and Attention Deficit Hyperactivity Disorder: A Meta-Analysis of Single-Case Studies

Kellina K. Pyle, Gregory A. Fabiano

University at Buffalo, SUNY

Author Note

Correspondence concerning this article should be addressed to:
Kellina K. Pyle
331 Diefendorf Hall
University at Buffalo, SUNY
Buffalo, NY 14214
e: kellinal@buffalo.edu
p: (716) 829-5977
f: (716) 829-6052

Gregory A. Fabiano: fabiano@buffalo.edu

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A160133 to University of Buffalo. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. The authors would like to thank Stephanie Jerome for her assistance with reliability coding.

Manuscript submitted to *Exceptional Children* October 2016; accepted March 2017.

Abstract

The Daily Report Card (DRC) is a commonly employed behavioral intervention for treating Attention Deficit Hyperactivity Disorder (ADHD) in schools. Much of the support for the DRC comes from single-case studies, which have traditionally received less attention than group studies. This lack of attention to single-case studies results in an incomplete review of the literature for this intervention. The present study utilized meta-analytic techniques to examine the DRC as used in single-case design studies, with moderating variables explored through Hierarchical Linear Modeling (HLM). Fourteen papers, including data on 40 single subject cases, were included in the analyses. Effect sizes generally illustrated improvement using the DRC, with some differences across methods of effect size estimation. Study quality and class type moderated outcomes. Overall, the present study supports the use of the DRC with students who have ADHD, and provides guidance for using single-case design studies in meta-analyses of intervention effects.

Keywords: ADHD; Attention Deficit Hyperactivity Disorder; Daily Report Card; Meta-Analysis; Single-Case Design

A Meta-Analysis of Single-Subject Design Studies Utilizing the Daily Report Card Intervention for Students with ADHD

Attention-deficit/hyperactivity disorder (ADHD) is a prevalent and chronic mental health disorder, comprising the majority of students in the Emotional Disturbance (ED) and Other Health Impaired (OHI) categories in special education in the U.S. (Schnoes, Reid, Wagner, & Marder, 2006). The adverse outcomes of ADHD include severe disruptions in relationships (McQuade & Hoza, 2015), and academic problems throughout the school year (McConaughy, Volpe, Antshel, Gordon, & Eiraldi, 2011), which may lead to the poor academic, social, and school completion outcomes commonly seen for students with ADHD (Kent et al., 2011).

To address these significant difficulties within school settings, numerous behavioral interventions have been developed and evaluated for youth with ADHD. One of the most commonly employed behavioral interventions for children with ADHD is the Daily Report Card (DRC; Kelley, 1990; O'Leary, Pelham, Rosenbaum, & Price, 1976; Volpe & Fabiano, 2013). The DRC is an operationalized list of a child's target behaviors (e.g., interrupting, noncompliance, academic productivity), and includes specific criteria for meeting each behavioral goal (e.g., interrupts *three* or fewer times during math instruction). Teachers provide immediate feedback to the child regarding target behaviors on the DRC, and typically some reward is provided contingent on the child's performance. DRCs are commonly employed and acceptable interventions for school settings (Chafouleas, Riley-Tillman, & Sassu, 2006).

While there are numerous examples of the efficacy of the DRC when used as a component of a multi-modal treatment package (e.g., MTA Cooperative Group, 1999; Owens, Murphy, Richerson, Girio, & Himawan, 2008), there are fewer studies that have investigated the efficacy of the DRC as a stand-alone intervention for ADHD (e.g., McCain & Kelley, 1993).

Assessing efficacy of the DRC as a stand-alone intervention is important as it will begin to elucidate whether the DRC is a contributor to the positive effects within multi-modal studies, which fits within recent initiatives to identify the effective components of treatment (e.g. National Center on Intensive Intervention; What Works Clearinghouse).

Whereas group designs are most likely to include multi-modal interventions, single-subject designs more commonly employ stand-alone interventions. Further, single-subject designs make up a large proportion of the ADHD psychosocial treatment literature (Fabiano, et al., 2009). A recent review of meta-analyses of ADHD treatment by Fabiano et al. (2015) revealed that single-subject designs often generate large effect sizes for youth with ADHD, with some notable exceptions (DuPaul & Eckert, 1997; DuPaul, Eckert, & Vilardo, 2012; Fabiano et al., 2009). Thus, single-subject designs should be subjected to the same scrutiny as between group designs (i.e., see guidelines for the What Works Clearinghouse for both group and single-subject designs; What Works Clearinghouse, 2014) to better understand the efficacy of interventions such as the DRC.

In a recent meta-analysis, Vannest et al. (2010) examined the efficacy of the DRC as a stand-alone intervention across 17 single-subject design studies and showed variable but in general positive support for the intervention, with effect sizes ranging from -0.15 to 0.97, and an average effect of 0.61. To account for this range, the study examined several moderating variables, and found that greater home-school communication and greater use of the DRC (using it for more than one hour a day) produced significantly stronger effect sizes. One limitation of this meta-analysis, however, was that the focus was on the daily report card as an intervention, and not necessarily the presenting problems of the students. Thus, the students included in the studies demonstrated a wide variety of symptom profiles and impairment, making it difficult to generalize results to a particular group of students such as those with ADHD.

The present study aims to expand these results by examining those DRC effects specific to children diagnosed with ADHD. Although behavioral interventions such as the DRC have been identified as best practice for children with ADHD (DuPaul & Eckert, 1997; Evans, Owens, & Bunford, 2014), single-case design studies implementing the DRC with children who have ADHD have never been examined as a whole. Additionally, despite the commonalities of a DRC which include setting clear goals, providing contingent feedback, and establishing contingent rewards for goal attainment, there are many different parameters of the DRC that can be varied across students and settings. These differences include changes to the amount of home-school communication, the age and gender of the students it is used with, and the class type in which it is implemented (e.g., special versus general education). These factors may change the efficacy of the DRC, and further examination of their moderating influence is needed.

To date, there have been six between-group and one within-group design studies that have investigated the efficacy of the DRC as a stand-alone intervention, but the diversity amongst the aims of these studies precludes a meta-analysis (Blechman, Taylor, & Schrader, 1981; Fabiano et al., 2010; Leach & Byrne, 1986; Murray, Rabiner, Schulte, & Newitt, 2008; O'Leary, Pelham, Rosenbaum, & Price, 1976; Owens et al., 2012; Palcic, Jurbergs, & Kelley, 2009). Although single-subject design studies of DRC efficacy are relatively more numerous, they have not been systematically reviewed in a meta-analysis as a stand-alone intervention for individuals with ADHD (Chronis, Jones, & Raggi, 2006; DuPaul, Eckert, & Vilardo, 2012; Evans, Owens, & Bunford, 2014; Fabiano et al., 2009; Fabiano et al., 2015; Pelham & Fabiano, 2008). Thus, single-subject designs that use the DRC as a stand-alone intervention for students with ADHD will be the focus of the current investigation.

Approaches to Quantifying Single-Subject Study Results

It is important to acknowledge that the quantification of effects across single subject studies in a meta-analysis is an evolving area within the field of intervention research (What Works Clearinghouse, 2014). To measure effects within and across single-subject design studies, scholars have focused on examining graphed time-series data, both visually and quantitatively. These procedures reveal how effective the intervention has been at improving outcomes, and demonstrate how these outcomes may be moderated by student or study-level characteristics. While there is currently no “gold standard” for calculating effect sizes in single-case design research, there have been several recommendations to use nonparametric and parametric methods in tandem (Gage & Lewis, 2014; Kratochwill et al., 2010; Wolery et al., 2010). Non-parametric methods include non-overlap-based effect sizes such as the Percent of Nonoverlapping Data (PND; Scruggs, Mastopieri, & Casto, 1987), and the Improvement Rate Difference (IRD; Parker, Vannest, & Brown, 2009), while parametric methods include regression (Allison & Gorman, 1993), and Hierarchical Linear Modeling (HLM; Raudenbush & Bryk, 2002; Van Den Noortgate & Onghena, 2007).

In the interest of expanding the literature on effect sizes in single-case designs using a clinically relevant sample, the present study utilized several nonparametric effect size approaches in combination with HLM. Effect sizes included in the present study were selected based on their use in previous research on the DRC (Owens et al., 2012; Vannest et al., 2010), and their ability to address unique concerns, such as baseline trend (Tau-*U*; Parker, Vannest, Davis, & Sauber, 2010). HLM was chosen over regression due to the hierarchical structure of single-case data

(data points are nested within treatment phases, which in turn are nested within participants and separate studies), and because HLM analyses can account for complex data structures (missing data, varying intervention lengths) likely to be found in single-case-design studies (Gage & Lewis, 2014; Raudenbush & Bryk, 2002; Van den Noorgate & Onghena, 2007, 2008).

Efforts to increase the yield and precision of single-subject design study outcomes are critical, as these studies have been marginalized in systematic reviews and determinations of research evidence for particular interventions, including the DRC. Indeed, in contrast to What Works Clearinghouse (2014) evidentiary standards, the most recent criteria for determining evidence-based, child and adolescent treatments (Southam-Gerow & Prinstein, 2014) no longer include single-subject design studies as appropriate empirical evidence for determining the strength of evidence for a child treatment. These modified recommendations effectively remove the majority of studies on interventions like the DRC from further consideration (see Fabiano et al., 2015; Fabiano et al., 2009). Further development of appropriate methods for quantifying single-subject results may allow researchers and policy-makers to include evidence from these designs in decision-making, which will help bridge the gap between more traditional research designs (i.e. large randomized controlled trials), and applied practice.

Summary and Research Questions

Although scholars have identified classroom contingency management as an evidence-based intervention for ADHD (Evans et al., 2014; Pelham & Fabiano, 2008; Pelham, Wheeler, & Chronis, 1998), at the present time a systematic review of the DRC as a specific intervention is needed. Reasons for this include: (a) the majority of studies in the literature that utilized a DRC did so as a part of a multi-component intervention (e.g., MTA Cooperative Group, 1999; Owens et al., 2008), (b) prominent groups have stated interventions such as the DRC should be utilized as second line intervention for children with ADHD in elementary school (American Academy of Pediatrics, 2011), and (c) only a handful of controlled trials exist using the DRC alone (Fabiano et al., 2010; Murray et al., 2008). In addition, prior systematic reviews and meta-analyses of the DRC as an intervention in general yielded support for the DRC, with some differences in levels of support across moderators (e.g., Vannest et al., 2010). Given that the ADHD group may be one contributor to heterogeneity of effect within the studies examined to date, there is a need to investigate the DRC as a stand-alone intervention for students with ADHD, synthesizing single-case design research.

Based on the group literature supporting the DRC as a stand-alone intervention, the present study specifically hypothesizes that: (a) the DRC will show large treatment effects, as measured by several non-overlap-based effect sizes, (b) effect sizes will be strongly correlated with one another, and (c) student- and/or study-level variables, including age, gender, diagnostic criteria, level of home-school communication, study quality, and/or class type will moderate the effectiveness of the DRC.

Method

In conducting this meta-analytic search and synthesis, we followed recommendations made in standard texts on research synthesis (Cooper & Hedges, 1994; Schmidt & Hunter, 2014), meta-analytic reporting standards (MARS) criteria from the APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008), and papers written specifically for meta-analytic examination of single-subject design (Gage & Lewis, 2014; Wang, Parrila, & Cui, 2013; Wolery et al., 2010). First, literature searches using the databases PsycInfo, EBSCO, and ERIC were conducted. Search criteria entered into these databases included: daily report card, daily behavior report card, home-school note, home school note, school home note, and school-home note. There was no specific date range selected. Both peer-reviewed journal articles and dissertations were examined and selected for the present study. Following this literature search, each identified article's reference section was also systematically analyzed for additional articles. Studies within several meta-analyses of behavior modification interventions for ADHD were also reviewed (DuPaul & Eckert, 1997; DuPaul et al., 2012; Evans, Owens, & Bunford, 2014; Fabiano et al., 2009). The literature search was terminated in January of 2016.

Inclusion Criteria

A study was included in the initial collection based on specified search criteria: (a) the participants must be identified as having ADHD either through prior diagnosis or the collection of diagnostic information through standardized ADHD rating scales (e.g., Connors Teacher Rating Scales; Connors, Sitarenios, Parker, & Epstein, 1998); (b) the participants must be under 18 years of age; (c) the study must include information that would permit the calculation of effect sizes (i.e., graphed time series data across baseline and intervention phases); (d) studies must use a daily report card as a stand-alone intervention; (e) the daily report card must have been used in a school or primarily academic (e.g. after-school education program) setting.

In the first stage, 132 articles and dissertations were identified that met initial search criteria. In the second stage, the abstracts of these papers were reviewed to identify those papers that used a single-case design method. Using this criterion, 94 papers were excluded, and 38 papers were kept for more detailed analysis. Fourteen of these 38 papers met all of the inclusion criteria outlined above. Of these 14 papers, fewer than half (Cottone, 1998; Cowart, 1999; Jurbergs, Palcic, & Kelley, 2007; Kelley & McCain, 1995; McCain & Kelley, 1993; McCain & Kelley, 1994) were previously examined in a meta-analysis of the DRC (Vannest et al., 2010), which underscores the unique nature of the present collection of studies. When a paper examined more than one participant, each participant was counted as an independent case study. In total, 40 student participants were identified from fourteen separate studies, with publication dates ranging from 1975 to 2013. A reliability search was conducted by the second author using the same search terms, databases, and meta-analysis reference sections, and yielded 100% reliability with the original search. Of note was a single dissertation, identified in both the primary and reliability searches (Kraemer, 1994) that could not be obtained through inter-library loan or direct contact with the author, and is therefore not included in the present analyses.

Coding

All studies were coded at three levels, including individual data points, student-level characteristics, and study-level characteristics. All individual data points were also coded for phase (whether they were data points in baseline or intervention), and, if a reversal design was used, order (whether they came before reversal or after). Student- and study-level variables were coded to examine possible moderation of treatment effects. Student level variables included age and gender, while study-level variables included the level of home-school communication, classroom type, and quality of the research design.

Outcomes. Perhaps due to the nature of ADHD, or the common utilization of the DRC to manage disruptive behavior, almost all studies included in this meta-analysis examined observations of disruptive or on-task behaviors as their primary outcome. In total, five outcome variables were identified. These included: percent of time on-task, percent of time disruptive, number of activity changes, percent of time spent exhibiting hyperactive symptoms, and percent of homework completed. To allow for a common interpretation of effect, all outcomes were either kept as, or converted to percentages. The number of activity changes was converted to a percentage by dividing the total number of activity changes by the time of the observation period (50 minutes). Thus, if the student changed activities 10 times, the resulting percentage would be 10/50*100 = 20%. Additionally, all outcomes were categorized as "disruptive" or "on-task" targets. On-task outcomes included time on-task and percent of homework completed, while disruptive outcomes included time spent disruptive, number of activity changes, and percent hyperactivity. A summary of the outcomes for each study is provided in Table 1.

To minimize the confounding effects of medication on the DRC, data from phases that intentionally manipulated medication were excluded. Specifically, Atkins et al. (1990) and Ayllon, Layman, and Kandel (1975) both manipulated medication. In the Atkins et al. (1990) study, medication was implemented in the last phase of treatment in conjunction with the DRC. Data from this final phase were excluded. In the Ayllon et al. (1975) study, medication was given in a phase prior to implementing the DRC, with a three day "wash-out" period between the medication and DRC phases. Data from the medication phase were excluded, with data in the DRC phase kept and assumed to be free of medication effects due to the wash-out period.

Graphs depicting outcome data were scanned and imported into UnGraph 5 (Version 5.0.1; Biosoft, 2015) in order to accurately read the values of the data points from the figures. All data points included in the graphs were coded. In cases where a reversal ABAB design was used, both the first AB (baseline-intervention) and the second AB pair were coded. A special code was assigned to each pair to determine order, where 0 = first AB pair, and 1 = second AB pair. No studies used more than one reversal. In all, 1570 data points were coded.

Quality. An aggregate measure of quality, based on three broad What Works Clearinghouse (WWC) recommendations for single-case design studies, supplemented with two external indicators of validity, was created to examine how rigorously each study designed and implemented the DRC. The What Works Clearinghouse lists specific guidelines for single-case designs to meet evidence standards (Kratochwill et al., 2010; What Works Clearinghouse, 2014). For the present study, three of these criteria were chosen and coded as 1 = Meets Criterion, and 0 = Does not Meet Criterion. These included: (a) inter-observer agreement reported for at least 20% of the data points, with at least 80% agreement; (b) at least 5 data points within each phase; and (c) data within the baseline phase provide a sufficient demonstration of a clearly defined pattern of responding (e.g. small differences from day-to-day, compared to large peaks and valleys), determined through visual analysis. Therefore, for each WWC criterion, the study could receive a score of 1 or 0, with higher scores (up to 3) indicating greater quality.

In addition to these four WWC criteria, the present study also used two external indicators of study internal validity, including: (a) Treatment integrity reported (0 = No, 1 = Yes), and (b) Observers blind to treatment conditions (0 = No, 1 = Yes). Scores for the five indicators were added, and a total quality score was found, with higher scores indicating greater quality. As some participants evinced certain criteria (e.g. five data points in each phase) while others did not, quality scores were initially calculated at the individual level, and then averaged to provide a study-level quality score (see Table 1).

Level of home-school communication. Home-school communication was coded following a similar strategy to Vannest et al. (2010) in their meta-analysis of the DRC. Specifically, an aggregate score was calculated using three criteria: (a) Reinforcement, where 0 = no reinforcement planning, 1 = reinforcement determined by the researcher, and 2 = reinforcement determined collaboratively; (b) Home Training, where 0 = no home training, 1 = indirect training (e.g. with a handout), and 2 = direct parent training (e.g. in-person meeting); and (c) Feedback, where 0 = feedback on school behavior given at only one location (home or school), and 1 = feedback on school behavior given at both home and school. These scores were combined to yield a study-level communication score (see Table 1).

Classroom type. Differences between special education versus general education classes, such as the presumed greater availability of resources and supports in special education classes, may influence the efficacy of the DRC. Student's classroom placements (when available) were coded (0 = general education, 1 = special education).

Age. Age may be related to DRC effectiveness. For instance, older children attending middle school tend to have a highly varied schedule, with a number of different teachers. These changes may lead to less consistency. This speculation needs to be evaluated empirically as other studies have not documented a moderating effect for age on behavioral treatment (Pelham & Fabiano, 2000). Age was coded numerically for all participants (see Table 1 for summary).

Gender. The moderating effect of gender on DRC effectiveness is in need of exploration as girls may exhibit different profiles relative to boys (Gaub & Carlson, 1997; Pelham & Bender, 1982). All participants were coded for gender (0 = female, 1 = male; see Table 1 for summary).

Reliability

Data points from graphs and all moderator variables (predictors) were coded twice (once by the main author, and once by a trained graduate assistant blind to the previous coding) to ensure reliability. Training was held in an hour-long meeting with the main author, in which all articles and operational definitions for codes were reviewed. The reliability of the data point coding was examined using an intra-class correlation, while the reliability of all predictor-level coding was found using the formula: (agreements)/(agreements + disagreements).

Analysis

The analysis for the present study was conducted in two stages. First, well-supported effect sizes for single-case designs, including the Standard Mean Difference (SMD; Busk & Serlin, 1992); Percent of non-overlapping data (PND; Scruggs et al., 1987); percent of all non-overlapping data (PAND; Parker, Hagan-Burke, & Vannest, 2007); Percent Exceeding the Median (PEM; Ma, 2006); Improvement Rate Difference (IRD; Parker et al., 2009); and Tau-*U* (Parker et al., 2010) were calculated. For each goal type (disruptive or on-task), a separate effect size was calculated. Following the calculation of these effect sizes, the relationships between effect sizes were examined using Pearson correlations.

The second part of the analysis used HLM to examine the moderating influence of several student- and study-level variables on the efficacy of the DRC. In addition to these moderating effects, HLM was also used to estimate an overall effect size (Hedges *g*) across all studies included in the meta-analysis.

Standard mean difference (SMD). The SMD is sometimes referred to as the "No Assumptions Effect Size" (NAES; Busk & Serlin, 1992), and is calculated by subtracting the mean of the baseline from the mean of the intervention data, and dividing by the standard deviation of the baseline.

Percent non-overlapping data (PND). The PND is calculated by identifying the most extreme baseline point (highest, if an increase is desired, lowest if a decrease is desired), and determining how many intervention data points fall above or below that extreme, depending on the effect desired (Scruggs et al., 1987).

Percent of all non-overlapping data (PAND). The PAND is the percentage of data remaining after removing the fewest data points that would eliminate all overlap. PAND takes into account all data points within both treatment and baseline phases, rather than a single extreme data point, such as in PND. PAND is scaled from 0 to 100, with greater values being more desirable (Parker et al., 2007).

Percentage of data exceeding the median (PEM). The PEM is calculated by locating the median of the baseline phase and determining the percentage of intervention data points above or below that point (depending of the effect desired). PEM is advantageous in that it is not necessarily affected by extreme baseline values, and may therefore give a estimate of intervention efficacy less influenced by outlier values (Ma, 2006).

Improvement rate difference (IRD). The IRD examines the difference in improvement rates between the baseline and intervention phases. It was modeled after the "Risk Difference" concept used in medical research and reflects visual non-overlap well. To calculate the IRD, data points in the intervention phase that overlap with data points in the baseline phase are identified and counted. This number is considered the "minimum removed" needed to eliminate all overlap between the intervention and baseline phases. The minimum is then divided in half, and the intervention and baseline "rates" are found. The difference between the intervention and baseline rates is the IRD (Parker et al., 2009).

Tau and Tau-*U*. Tau and Tau-*U* examine the percentage of data that shows improvement across phases by comparing pairs of data points. By comparing the amount of non-overlap (desired) to the amount of overlap (not desired) a conservative effect size can be calculated. Tau-*U* has the added benefit of controlling for positive baseline trend, when present. Both tests show more statistical power than other nonoverlap-based effect sizes (Parker, Vannest, & Davis, 2011), and allow for the calculation of *p*-values and confidence intervals.

Hierarchical linear modeling (HLM). All statistical analyses were conducted with HLM 7 (Bryk, Raudenbush, & Congdon, 2011). In the present study, a 3-level, linear growth model was used to explore the treatment effects from baseline to intervention, and to examine the impact certain student- and study-level predictors had on this treatment effect. In these models, Level 1 represents the data points, or repeated measures within persons. There were 1570 data points. Level 2 represents the students, and those characteristics, such as age and gender, that may influence the mean of their data points, or the way in which their behavior changes from baseline to intervention. There were 40 student cases at Level 2. Level 3 represents those study characteristics, such as classroom type and study quality that may affect treatment outcomes. There were 14 study cases at level 3. To allow for a common interpretation of effects, all outcomes were coded so that higher percentages were always considered “more desirable” regardless of whether the goal was for disruptive or on-task behavior. HLM models were created sequentially to address four major goals, including: (a) order and measure-type effects, (b) treatment effect, (c) student-level variables, and (d) study-level variables.

Data considerations for HLM. Initial examination of the data revealed that the Cottone (1998) dissertation acted as a major outlier in our analyses, driving effects at both the student- and study-levels. These results were due in large part to the “disruptive” goal included in the dissertation, which suffered from significant floor effects at both baseline (where the most common amount of disruptive behavior was 0) and intervention. To create a more parsimonious model that better reflected the data as a whole, rather than an individual study, the Cottone (1998) dissertation data were removed from all analyses.

Results

Results for each outcome, including reliability of data and moderator coding, effect size calculations, estimates of publication bias, and Hierarchical Linear Modeling are each explored individually below.

Reliability

With regard to the data points coded from UnGraph, a high degree of reliability was found, as indicated by an intra-class correlation of .97, with a 95% confidence interval from .96 to .98, $F(609,610) = 70.16$, $p < .001$. Codes for the predictor variables were created separately and then compared. These codes ranged in reliability from 87% to 100%, with the greatest discrepancies in: (a) consistent baseline trend (Quality; 87% agreement); and (b) feedback at both home and school (Home-School Communication; 87% agreement).

Effect Sizes

Overall, effect sizes generally illustrated improvement from implementing the DRC, with some differences across methods. Of the methods used, the most varied effect sizes were produced using the SMD (-0.27 to 54.45; at the individual level). Effect sizes calculated using the, PND, PAND, PEM, IRD, and Tau-*U* methods were generally similar, with average effect sizes across all studies ranging from 0.59 to 0.94. Average effect sizes across participants for each study are listed in Table 2. Pearson correlations demonstrated that all effect sizes were significantly related, with the strongest relationships between PND, PAND, PEM, IRD, and Tau, and the weakest relationships with SMD. All correlations are listed in Table 3.

Publication Bias

The present study sought to limit errors based on publication bias by incorporating published and unpublished studies (dissertations). Additionally, a Fail Safe N (N_{fs}; Cooper, 1979) was calculated for each effect size. A criterion effect size of $d = 0.10$ was chosen to represent a "null" effect. For the smallest average effect size found in the present study (PND, Disruptive = 0.59), at least 68 studies would need to find a null effect to reduce the effect size to an insignificant level. For the largest average effect size (SMD, On-Task = 4.31), over 500 studies would need to find a null effect to reduce this effect size. These results suggest that publication bias is unlikely to have distorted the reported findings.

Hierarchical Linear Modeling (HLM)

Several initial models were created to examine the data. These models demonstrated: (1) the relative magnitude of variance between students versus between studies, (2) the differences in the treatment effect between first and second AB pairs in reversal studies, and (3) the differences in the treatment effect between outcomes (on-task versus disruptive). These models were designed in a similar manner to those outlined by Gage and Lewis (2014).

Initial models. We first examined a fully unconditional model, in which no predictors were entered. This model demonstrated that approximately 25% of the variance in behavioral outcomes lay between students, while 20% lay between studies. These results support our interest in examining the moderating effects of student- and study-level variables. Next, we examined the effect of the Level 1 Order predictor (0 = first AB pair, 1 = second AB pair). This model demonstrated that after reversal, students may show faster change from baseline to intervention, speeding up the change by approximately 6 percentage points, $\beta = 6.09$, $t(13) = 3.12$, $p < .01$. Finally, we examined whether there were differences in the treatment effect due to the goal type (on-task versus disruptive). No significant differences were found between the goal types, $\beta = 7.58$, $t(13) = 1.41$, $p = .18$.

Partially conditional model. The partially conditional model examined the effect of the Level 1 Phase predictor (0 = baseline, 1 = treatment) on the data points. This model indicated that the average treatment effect across studies was significant, with participants gaining approximately 30 percentage points from baseline to treatment, $\beta = 30.32$, $t(12) = 8.82$, $p < .001$. The partially conditional model also indicated that there is significant variability among students in their scores at baseline, $r = 67.91$, $\chi^2(23) = 130.05$, $p < .001$, and among students in their response to the intervention, $r = 23.04$, $\chi^2(23) = 41.19$, $p < .05$. At the study level, there is significant variation in both the baseline scores of participants, $u = 135.38$, $\chi^2(12) = 68.63$, $p < .001$, and treatment effects, $u = 128.53$, $\chi^2(12) = 89.21$, $p < .001$. These results suggest that there may be student- and study-level characteristics that moderate the treatment effect (see Table 4). Hedge's *g* was calculated for the partially conditional model, and was found to be 2.19.

Fully conditional model. Age (n = 37; range 4-14) and gender (n = 37; 25 male) showed no significant impact on baseline or the change from baseline to intervention (see Table 5).

Due to these findings, a more parsimonious model in which age and gender were excluded was used to examine the effects that study-level variables, including quality (n = 13), home-school communication (n = 13), and class type (n = 13), had on outcomes (see Table 1 for study-level details). In this fully conditional model, quality and class type moderated the treatment effect, but home-school communication did not. On average, higher quality studies demonstrated significantly greater change across the phases by approximately 13 percentage points, $\gamma = 13.17$, $t(9) = 3.04$, $p = .01$. Changes in class type yielded a similar increase, with studies completed in a special education classrooms gaining approximately 33 percentage points more across phases, $\gamma = 32.99$, $t(9) = 3.23$, $p = .01$. This result should be interpreted with caution as there was only one study included in these analyses that examined students in special education classrooms (Ayllon et al., 1975). Home-school communication was not significantly related to outcome, $\gamma = -2.21$, $t\ (9) = -1.02$, $p = .34$ (see Table 6).

Discussion

Overall, the results of the present study support the daily report card as an effective stand-alone intervention for students with ADHD based on the results of single-subject design studies. The implementation of the DRC significantly changes behavior, increasing desirable behavior by almost 30 percentage points from baseline to intervention. Using HLM, the moderating effects of class type and study quality were illustrated, with higher quality studies and special education classrooms associated with greater gains. The effects of the DRC are consistent and large, as indicated by non-overlap-based effect sizes that range from 0.59 – 0.94, and an overall Hedges *g* of 2.19. Additionally, the present study demonstrated that evidence for an intervention can be shown using a meta-analysis of single-subject design studies, particularly with the advent of statistical techniques like HLM. These findings support the utility and continued inclusion of single-subject designs in meta-analyses of treatment effects. Inclusion of these studies is especially important for the ADHD treatment literature, where the majority of studies are single-subject designs (DuPaul, 1997; 2012; Fabiano, 2009).

Although HLM is relatively new, it shows promise for addressing many of the criticisms levied against statistical analysis of single-subject designs (Kratochwill et al., 1974; Parsenson & Baer, 1992; Salzberg, Strain, & Baer, 1987; White, 1987), and meets proposed criteria for meta-analysis of single-subject designs (Wolery et al., 2010). In the present study, HLM analyses demonstrated that students with ADHD who were given a daily report card showed a mean improvement of approximately 30 percentage points from baseline to intervention. Given the initial baseline average of 51%, this shift resulted in students who were more than 80% on-task, and disruptive less than 20% of the time. This mean shift is consistent with the results of Gage and Lewis (2014), who used HLM to demonstrate that Functional Behavior Assessment (FBA)-based interventions increased desirable behavior by 34 percentage points from baseline to intervention for students with Emotional and Behavioral Disorders (EBD).

Although the benefits of the DRC were considerable, significant variability remained between students and studies in the treatment effect, suggesting that there were student- and study-level moderators. In the present study, age, gender, class type, home-school communication, and study quality were examined as potential moderators of the DRC. Neither age nor gender moderated the treatment effect. These results are positive, suggesting that students from different genders and age groups will benefit from the DRC intervention equally.

While age, gender, and home-school communication did not moderate outcomes, class type and study quality significantly moderated outcomes. As anticipated, higher quality studies were associated with greater gains from baseline to intervention. This result lends support to the use of certain guidelines (e.g., WWC; Kratochwill et al., 2010) in conducting single-case design research. Although class type was also anticipated to moderate outcomes, this result should be interpreted with caution, as only one study included in these analyses was conducted in special education classrooms (Ayllon et al., 1975). The non-significant moderation of home-school communication on outcomes was not anticipated, and deserves a more thorough investigation.

Greater home-school communication is theorized to be one of the lynchpins of the DRC, allowing teachers and parents to work collaboratively to improve a student's behavior (Fabiano et al., 2010; Kelley, 1990). Indeed, in a prior meta-analysis of the daily report card, Vannest et al. (2010) demonstrated that those with the highest home-school communication showed significantly stronger effect sizes when compared to those with the lowest home-school communication. Although the results of the present study appear to contradict these findings, they may suggest something unique about the population of students with ADHD. For instance, it is possible that an increased amount of communication between the home and the school may not always be beneficial to the student, and may in fact represent a more severe impairment that requires a greater concerted effort (home and school rewards, etc.) to address the problem. It is clear that there is a need for more research in this area to examine the influence of home-school communication on student behavioral outcomes. Particularly, future studies should endeavor to operationalize and clearly report the level and type of home-school communication used, as this will help future meta-analyses determine the moderating influence of changes in this variable.

The present study used values from the partially conditional HLM model to calculate an overall Hedges *g* of 2.19, which suggests that the DRC is very effective at increasing desirable behavior in students with ADHD. Although this effect size was very large, it is consistent with the significant changes demonstrated by the non-parametric effect sizes, which ranged, on average, from 0.59 to 4.31. This large range was due to the use of the SMD effect size, which is not based on percent of overlap from baseline to intervention (Busk & Serlin, 1992). Although the SMD yields effect sizes that are not interpretable by current standards (e.g., Cohen, 1992), research continuing to use this effect size and compare it to other effect sizes is greatly needed, especially to create new standards for judging the magnitude of these effect sizes, which are often very large (Gage & Lewis, 2014).

Although there was some variability in the non-parametric effect sizes, all methods were significantly correlated, suggesting that they largely agreed in illustrating improvement with the DRC. While there are no firmly established standards for these non-overlap-based effect sizes, suggested criteria list effect sizes of 0.70 - 0.90 as denoting moderately effective interventions, and effect sizes larger than 0.90 as highly effective (Ma, 2006; Parker et al.,2011; Scruggs & Mastropieri, 2001). By these criteria, the DRC intervention is supported as a moderate- to highly-effective stand-alone intervention for children identified as having ADHD. Additionally, although the Vannest et al. (2010) meta-analysis demonstrated conflicting support for the DRC using the IRD effect size, the present study did not find the same variability in the IRD, suggesting that the DRC is a particularly effective intervention for youth with ADHD.

Limitations

The present study has several limitations. First, although efforts were made to select statistical models that addressed the sample size issue inherent in single-subject design, the number of studies and participants included in this meta-analysis is still small. The sample size may limit the generalizability of these findings, especially with regard to the moderating effects of study-level variables.

The study was also limited by small sample sizes of subgroups, particularly with regard to girls (n = 11) and older children (above the age of 10; n = 9). This lack of diversity in gender and ages may limit the generalizability of the present findings. Additionally, the present study and was not able to account for the severity of ADHD symptoms, ethnicity of participants, the types of services offered to students within special education, or the presence of co-morbid conditions. These factors deserve further exploration in future studies of the DRC.

Conclusion

The present study supports the use of the DRC as an effective intervention for students with ADHD. While higher quality designs and special education classrooms led to more rapid behavioral change, greater home-school communication was not associated with outcomes. School psychologists, special educators, and clinicians are encouraged to use the daily report card to address both on-task and disruptive behaviors with students who have ADHD (e.g. Volpe & Fabiano, 2013). Future research is needed to address the elements of home-school communication as they relate to the DRC, particularly identifying the type and degree of home-school communication that influences outcomes.

References

*References preceded by * were included in the single-subject design study meta-analysis.*

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

American Academy of Pediatrics. (2011). ADHD: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. Retrieved from http://www.pediatrics.org/cgi/doi/10.1542/peds.2011-2654

APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, 63, 839–851. doi: 10.1037/0003-066X.63.9.839

*Atkins, M.S., Pelham, W.E., White, K.J. (1990). Hyperactivity and Attention Deficit Disorders. In M. Hersen & V.B. Van Hasselt (Eds.) *Psychological aspects of developmental and physical disabilities: A casebook* (137-156). Thousand Oaks, CA: Sage Publications.

* Ayllon, T., Layman, D., & Kandel, H. J. (1975). A behavioral-educational alternative to drug control of hyperactive children. *Journal of Applied Behavior Analysis, 8*(2), 137-146. doi: 10.1901/jaba.1975.8-137

Biosoft. (2015). Ungraph 5 [Computer software]. Retrieved from http://www.biosoft.com/w/ungraph.htm

Blechman, E. A., Taylor, C. J., & Schrader, S. M. (1981). Family problem solving versus home notes as early intervention with high-risk children.*Journal of Consulting and Clinical Psychology, 49*(6), 919. doi: 10.1037/0022-006X.49.6.919

Bryk, A.S, Raudenbush, S.W., & Congdon, R. (2011). HLM 7 for Windows [Computer software]. Chicago, IL: Scientific Software International, Inc.

Busk, P.L., & Serlin, R.C. (1992). Meta-analysis for single-case research. In T.R. Kratochwill & J.R. Levinn (Eds.) *Single-case research design and data analysis: New directions for psychology and education* (pp. 187-212). Hillsdale, NJ: Erlbaum.

Chafouleas, S.M., Riley-Tillman, T.C., & Sassu, K.A. (2006). Acceptability and reported use of daily behavior report cards among teachers. *Journal of Positive Behavioral Interventions, 8,* 174-182. doi: 10.1177/10983007060080030601

Chronis, A. M., Jones, H. A., & Raggi, V. L. (2006). Evidence-based psychosocial treatments for children and adolescents with attention-deficit/hyperactivity disorder. *Clinical psychology review*, *26*(4), 486-502. doi: 10.1016/j.cpr.2006.01.002

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155. doi: 10.1037/0033-2909.112.1.155

Conners, C. K., Sitarenios, G., Parker, J. D., & Epstein, J. N. (1998). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): factor structure, reliability, and criterion validity. *Journal of abnormal child psychology*, *26*(4), 279-291. doi: 10.1023/A:1022606501530

Cooper, H. M. (1979). Statistically combining independent studies: A meta-analysis of sex differences in conformity research. *Journal of personality and social psychology*, *37*(1), 131-146. doi: 10.1037/0022-3514.37.1.131

Cooper, H. & Hedges, L. V. (1994). *The Handbook of Research Synthesis.* New York, NY: Russell Sage Foundation.

*Cottone, E.A. (1998). Home-school collaboration: Evaluating the effectiveness of a school-home note program for children with ADHD. *Dissertation Abstracts International*, *59*, 2442A. (UMI No 9840485)

*Cowart, L.U. (1999). The efficacy of a school-home note in reducing disruptive classroom behavior: The differential effects of selecting social and academic targets. *Dissertation Abstracts International*, *60*, 4869B. (UMI No. 9947683)

DuPaul, G. J., & Eckert, T. L. (1997). The effects of school-based interventions for Attention Deficit Hyperactivity Disorder: A meta-analysis. *School Psychology Review*, *26*(1), 5-27. Retrieved from http://www.nasponline.org/publications

DuPaul, G. J., Eckert, T. L., & Vilardo, B. (2012). The effects of school-based interventions for attention deficit hyperactivity disorder: A meta-analysis 1996-2010. *School Psychology Review*, *41*(4), 387-412. Retrieved from http://www.nasponline.org/publications

Evans, S. W., Owens, J. S., & Bunford, N. (2014). Evidence-based psychosocial treatments for children and adolescents with attention-deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology*, *43*(4), 527-551. doi: 10.1080/15374416.2013.850700

Fabiano, G. A., Schatz, N. K., Aloe, A. M., Chacko, A., & Chronis-Tuscano, A. (2015). A systematic review of meta-analyses of psychosocial treatment for attention-deficit/hyperactivity disorder. *Clinical child and family psychology review*, *18*(1), 77-97. doi: 10.1007/s10567-015-0178-6

Fabiano, G. A., Pelham, W. E., Coles, E. K., Gnagy, E. M., Chronis-Tuscano, A., & O'Connor, B. C. (2009). A meta-analysis of behavioral treatments for attention-deficit/hyperactivity disorder. *Clinical Psychology Review*, *29*(2), 129-140. doi: 10.1016/j.cpr.2008.11.001

*Fabiano, G. A., & Pelham, W. E. (2003). Improving the Effectiveness of Behavioral Classroom Interventions for Attention-Deficit/Hyperactivity Disorder: A Case Study. *Journal of Emotional and Behavioral Disorders*, *11*(2), 122-128. doi: 10.1177/106342660301100206

Fabiano, G. A., Vujnovic, R. K., Pelham, W. E., Waschbusch, D. A., Massetti, G. M., Pariseau, M. E., ... & Volker, M. (2010). Enhancing the effectiveness of special education programming for children with attention deficit hyperactivity disorder using a daily report card. *School Psychology Review*, *39*(2), 219-239. Retrieved from http://www.nasponline.org/publications

Gage, N.A., & Lewis, T.J. (2014). Hierarchical Linear Modeling Meta-Analysis of Single-Subject Design Research. *The Journal of Special Education*, *48*(1), 3-16. doi: 10.1177/0022466912443894

Gaub, M., & Carlson, C. L. (1997). Gender differences in ADHD: A meta-analysis and critical review. *Journal Of The American Academy Of Child & Adolescent Psychiatry, 36*(8), 1036-1045. doi: 10.1097/00004583-199708000-00011

*Grady, E. (2013). *The Development of an Enhanced School Home Note Intervention: Applying Key Behavioral Parenting Training Components to Improve the Outcomes of School Based Behavioral Intervention.* (Doctoral dissertation, University of Washington). Retrieved from http://hdl.handle.net/1773/23607

*Jurbergs, N., Palcic, J., & Kelley, M. L. (2007). School-home notes with and without response cost: Increasing attention and academic performance in low-income children with attention-deficit/hyperactivity disorder. *School Psychology Quarterly*, *22*(3), 358. doi: 10.1037/1045-3830.22.3.358

Kelley, M. L. (1990). *School-home notes: Promoting children's classroom success*. New York: Guilford Press.

*Kelley, M. L., & McCain, A. P. (1995). Promoting Academic Performance in Inattentive Children The Relative Efficacy of School-Home Notes with and without Response Cost. *Behavior Modification, 19*(3), 357-375. doi: 10.1177/01454455950193006

Kent, K. M., Pelham Jr, W. E., Molina, B. S., Sibley, M. H., Waschbusch, D. A., Yu, J., ... & Karch, K. M. (2011). The academic experience of male high school students with ADHD. *Journal of abnormal child psychology*, *39*(3), 451-462. doi: 10.1007/s10802-010-9472-4

Kraemer, E. S. (1994). *Effectiveness of a home-school note procedure for increasing appropriate classroom behaviors exhibited by children diagnosed with attention-deficit hyperactivity disorder*. Retrieved from ProQuest Dissertations & Theses Global. (Order No. 9432810).

Kratochwill, T., Alden, K., Demuth, D., Dawson, D., Panicucci, C., Arntson, P., ... & Levin, J. (1974). A further consideration in the application of an analysis-of-variance model for the intrasubject replication design. *Journal of Applied Behavior Analysis*, *7*(4), 629-633. doi: 10.1901/jaba.1974.7-629

Kratochwill, T.R., Hitchcock, J., Horner, R.H., Levin, J.R., Odom, S.L., Rindskopf, D.M. & Shadish, W.R. (2010). Single-case designs technical documentation. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Leach, D. J., & Byrne, M. K. (1986). Some 'spill-over'effects of a home-based reinforcement programme in a secondary school. *Educational Psychology*, *6*(3), 265-276. doi: 10.1080/0144341860060304

*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. doi: 10.1177/1098300712440451

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. doi: 10.1177/0145445504272974

*McCain, A. P., & Kelley, M. L. (1993). Managing the classroom behavior of an ADHD preschooler: The efficacy of a school-home note intervention. *Child & Family Behavior Therapy*, *15*(3), 33-44. doi: 10.1300/J019v15n03_03

*McCain, A. P., & Kelley, M. L. (1994). Improving Classroom Performance in Underachieving Preadolescents: The Addictive Effects of Response Cost to a School-Home Note System. *Child & Family Behavior Therapy*, *16*(2), 27-41. doi: 10.1300/J019v16n02_02

McConaughy, S. H., Volpe, R. J., Antshel, K. M., Gordon, M., & Eiraldi, R. B. (2011). Academic and social impairments of elementary school children with attention deficit hyperactivity disorder. *School Psychology Review*, *40*(2), 200-225. Retrieved from http://www.nasponline.org/publications

*McCorvey, Z. J. (2013). *Addressing disruptive behaviors in an After School Program classroom: The effects of the Daily Behavior Report Card.* Retrieved from ProQuest LLC. (Order No. 3574909).

McQuade, J. D., & Hoza, B. (2015). Peer relationships of children with ADHD. In R. A. Barkley, R. A. Barkley (Eds.) *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment (4th ed.)* (pp. 210-222). New York, NY, US: Guilford Press.

*Miller, D. L., & Kelley, M. L. (1994). The use of goal setting and contingency contracting for improving children's homework performance. *Journal of Applied Behavior Analysis*, *27*(1), 73-84. doi: 10.1901/jaba.1994.27-73

MTA Cooperative Group. (1999). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, *56*(12), 1073. doi: 10.1001/archpsyc.56.12.1073.

Murray, D. W., Rabiner, D., Schulte, A., & Newitt, K. (2008). Feasibility and integrity of a parent–teacher consultation intervention for ADHD students. *Child & Youth Care Forum*, 37, 111-126. doi: 10.1007/s10566-008-9054-6

O'Leary, K. D., Pelham, W. E., Rosenbaum, A., & Price, G. H. (1976). Behavioral treatment of hyperkinetic children. An experimental evaluation of its usefulness. *Clinical Pediatrics*, *15*(6), 510-515. doi: 10.1177/000992287601500603

Owens, J. S., Holdaway, A. S., Zoromski, A. K., Evans, S. W., Himawan, L. K., Girio-Herrera, E., & Murphy, C. E. (2012). Incremental benefits of a daily report card intervention over time for youth with disruptive behavior. *Behavior therapy*, *43*(4), 848-861. doi: 10.1016/j.beth.2012.02.002

Owens, J. S., Murphy, C. E., Richerson, L., Girio, E. L., & Himawan, L. K. (2008). Science to practice in underserved communities: The effectiveness of school mental health programming. *Journal of Clinical Child & Adolescent Psychology*, *37*(2), 434-447. doi: 10.1080/15374410801955912

Palcic, J. L., Jurbergs, N., & Kelley, M. L. (2009). A comparison of teacher and parent delivered consequences: Improving classroom behavior in low-income children with ADHD. *Child & Family Behavior Therapy*, *31*(2), 117-133. doi: 10.1080/07317100902910513

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of All Non-Overlapping Data (PAND) An Alternative to PND. *The Journal of Special Education*, *40*(4), 194-204. doi: 10.1177/00224669070400040101

Parker, R.I., Vannest, K.J., & Brown, L. (2009). The Improvement Rate Difference for Single-Case Research. *Exceptional Children, 75*(2), 135-150. doi: 10.1177/001440290907500201

Parker, R.I., Vannest, K.J., & Davis, J.L. (2011). Effect Size in Single-Case Research: A Review of Nine Nonoverlap Techniques. *Behavior Modification, 20*(10), 1-20. doi: 10.1177/0145445511399147

Parker, R.I., Vannest, K.J., Davis, J.L., Sauber, S.B. (2010). Combining non-overlap and trend for single case research: Tau-U. *Behavior Therapy. 42*, 284–299. doi: 10.1016/j.beth.2010.08.006.

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

Pelham, W. E., & Bender, M. E. (1982). Peer relationships in hyperactive children: Description and treatment. *Advances in Learning & Behavioral Disabilities*, *1,* 365-436. doi: 1983-06085-001

Pelham Jr, W. E., & Fabiano, G. A. (2000). Behavior modification. *Child and adolescent psychiatric clinics of North America, 9*(3), 671-688. Retrieved from http://www.childpsych.theclinics.com/

Pelham Jr, W. E., & Fabiano, G. A. (2008). Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology*, *37*(1), 184-214. doi: 10.1080/15374410701818681

Pelham Jr, W. E., Wheeler, T., & Chronis, A. (1998). Empirically supported psychosocial treatments for attention deficit hyperactivity disorder. *Journal of clinical child psychology*, *27*(2), 190-205. doi: 10.1207/s15374424jccp2702_6

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage Publications.

Salzberg, C. L., Strain, P. S., & Baer, D. M. (1987). Meta-analysis for Single-Subject Research When Does It Clarify, When Does It Obscure? *Remedial and Special Education*, *8*(2), 43-48. doi: 10.1177/074193258700800209

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage publications.

Schnoes, C., Reid, R., Wagner, M., & Marder, C. (2006). ADHD among students receiving special education services: A national survey. *Exceptional Children*, *72*(4), 483-496. doi: 10.1177/001440290607200406

Scruggs, T.E., Mastropieri, M.A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and Validation. *Remedial and Special Education*, *8*(2), 24-33. doi: 10.1177/074193258700800206

Scruggs, T. E., & Mastropieri, M. A. (2001). How to summarize single-participant research: Ideas and applications. *Exceptionality*, *9*(4), 227-244. doi: 10.1207/S15327035EX0904_5

Southam-Gerow, M. A., & Prinstein, M. J. (2014). Evidence base updates: The evolution of the evaluation of psychological treatments for children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *43*(1), 1-6. doi: 10.1080/15374416.2013.855128

Van den Noorgate, W., & Onghena, P. (2007). The aggregation of single-case results using hierarchical linear models. *The Behavior Analyst Today*, *8*(2), 196-209. doi: 10.1037/h0100613

Van den Noorgate, W. & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence-Based Communication Assessment and Intervention*, *2*, 142-151. doi: 10.1177/0013164402251027

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. Retrieved from http://www.nasponline.org/publications

Volpe, R. J., & Fabiano, G. A. (2013). *Daily behavior report cards: An evidence-based system of assessment and intervention*. New York: Guilford Press.

Wang, S. Y., Parrila, R., & Cui, Y. (2013). Meta-analysis of social skills interventions of single-case research for individuals with autism spectrum disorders: Results from three-level HLM. *Journal of autism and developmental disorders*, *43*(7), 1701-1716. doi: 10.1007/s10803-012-1726-2

*Weakley, N. M. (2012). *Effects of Check-in/Check-out with and without access to feedback from a daily behavior report card on the levels of appropriate behavior in secondary school students.* Retrieved from ProQuest Dissertations & Theses Global. (Order No. 3534868).

What Works Clearinghouse. (2014). *What Works Clearinghouse: Procedures and Standards Handbook Version 3.0.* Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf

White, O. R. (1987). Some comments concerning "The quantitative synthesis of single-subject research". *Remedial and Special Education*, *8*(2), 34-39. doi: 10.1177/074193258700800207

Wolery, M., Busick, M., Reichow, B., Barton, E. (2010). Comparison of Overlap Methods for Quantitatively Synthesizing Single-Subject Data. *Journal of Special Education, 44*(1), 18-28. doi: 10.1177/0022466908328009

Table 1
Summary of Studies

Authors	Outcome	Method	Age(s)	# Male	# Female	Home-School	Quality	Class Type
1.Atkins et al. (1990)	1	Alternating	9	1	0	5	2	General
2. Ayllon et al. (1975)	3, 4	MB	8, 9, 10	2	1	2	2	Special
3. Cottone (1998)	1, 2, 3	MB	7, 11, 12	3	0	5	4	Special/General
4. Cowart (1999)	1, 2	MB	8, 11	2	0	3	3.5	General
5. Fabiano & Pelham (2003)	1, 2	MB	8	1	0	2	2	General
6. Grady (2013)	1, 2	MB	5, 6, 6	2	1	5	3	General
7. Jurbergs et al. (2007)	1	ABAB	6, 6, 8, 8, 8, 8	5	1	5	3.5	General
8. Kelley & McCain (1995)	1	Alternating	6, 6, 7, 7, 9	2	3	5	3.6	General
9. LeBel et al. (2012)	2	MB	4, 4, 4, 4	3	1	4	3.5	General
10. McCain & Kelley (1993)	1, 2, 5	ABAB	5	1	0	5	3	General
11. McCain & Kelley (1994)	1, 2	Alternating	11, 11, 11	3	0	5	3.33	General
12. McCorvey (2013)	1, 2	MB	9, 9, 9	1	2	3	2.33	General
13. Miller & Kelley (1994)	1, 3	MB, ABAB	9, 10, 11, 11	2	2	4	3.75	Unknown
14. Weakley (2012)	1	MB, Alternating	14	1	0	1	3	General

Note. 1 = Percent of time spent on-task; 2 = Percent of time spent disruptive; 3 = Percent of homework completed; 4 = Percent of time engaged in hyperactive behaviors; 5 = Number of activity changes made; Alternating = alternating treatment design; ABAB = a reversal design; MB = a multiple baseline design; N/A = Not Applicable. Quality and Home-School Communication scores represent an average across all participants.

Table 2
Effect Sizes Across Studies

Authors	Goal Type	SMD	PND	PAND	PEM	IRD	Tau/Tau-*U*	*p*-value
1. Atkins et al. (1990)	On-Task	1.32	0.30	0.68	0.78	0.59	**0.47**	< .01
2. Ayllon et al. (1975)	On-Task	8.89	1.00	1.00	1.00	1.00	1.08	< .001
	Disruptive	9.90	1.00	1.00	1.00	1.00	**1.00**	< .05
4. Cowart (1999)	On-Task	2.78	0.77	0.94	1.00	0.84	0.91	< .001
5. Fabiano & Pelham (2003)	On-Task	1.62	0.15	0.76	1.00	0.88	0.86	< .001
	Disruptive	1.34	0.31	0.77	1.00	0.75	0.90	< .001
6. Grady (2013)	On-Task	2.29	0.74	0.78	0.86	0.75	0.78	< .001
	Disruptive	3.26	0.71	0.80	0.85	0.67	0.76	< .001
7. Jurbergs et al. (2007)	On-Task	2.87	0.91	0.96	0.98	0.93	0.94	< .001
8. Kelley & McCain (1995)	On-Task	9.86	0.92	0.97	0.99	0.94	0.93	< .001
9. LeBel et al. (2012)	Disruptive	3.61	0.91	0.95	1.00	0.94	0.96	< .001
10. McCain & Kelley (1993)	On-Task	3.30	1.00	1.00	1.00	1.00	1.00	< .001
	Disruptive	4.03	0.79	1.00	1.00	0.89	1.00	< .001
11. McCain & Kelley (1994)	On-Task	10.17	0.95	0.97	1.00	0.95	0.98	< .001
	Disruptive	1.08	0.56	0.67	0.92	0.61	**0.73**	< .001
12. McCorvey (2013)	On-Task	0.79	0.35	0.71	0.72	0.54	0.44	< .05
	Disruptive	0.62	0.22	0.67	0.61	0.31	0.17	> .05
13. Miller & Kelley (1994)	On-Task	2.40	0.75	0.88	0.96	0.82	0.84	< .001
14. Weakley (2012)	On-Task	1.45	1.00	1.00	1.00	1.00	1.00	< .001
Average Across Studies	**On-Task**	**4.31**	**0.76**	**0.89**	**0.94**	**0.84**	**0.84**	
	Disruptive	**3.14**	**0.59**	**0.81**	**0.87**	**0.69**	**0.72**	

Note. The effect sizes shown in this table represent the average effect sizes across all participants or phases. The only exceptions to this rule are the Tau-*U* effect size, which represents a weighted average, and its related *p*-value, which represents the significance of improvement across phases. Additionally, all Tau effect sizes shown in bold are Tau-*U* effect sizes, and have been corrected for positive baseline trend. SMD = Standard Mean Difference, PND = Percent Nonoverlapping Data, PAND = Percent All Nonoverlapping Data, PEM = Percent Exceeding the Median, IRD = Improvement Rate Difference.

Table 3
Correlation Analysis of All Non-parametric Effect Sizes

	SMD	PND	PAND	PEM	IRD
PND	.36**				
PAND	.35**	.90**			
PEM	.26*	.69**	.71**		
IRD	.35*	.86**	.83**	.82**	
Tau-*U*	.31*	.81**	.78**	.90**	.95**

Note. ***p* < .01; **p* < .05. PND = Percent Non-overlapping Data; PAN D = Percent of All Non-Overlapping Data; PEM = Percent Exceeding the Median; IRD = Improvement Rate Difference; SMA = Standard Mean Difference

Table 4
HLM Partially Conditional Model Showing Average Change Across Phases

Fixed Effect	*Coefficient*	*se*	*t Ratio*	*p Value*
Mean at Baseline, γ_{000}	50.25	3.67	13.69	<.001
Mean Growth Rate (Treatment Effect), γ_{100}	30.32	3.43	8.82	<.001
Random Effect – Variability Among Participants (Level 2)	*Variance*	*df*	χ^2	*p Value*
Baseline, r_0	67.91	23	130.05	<.001
Treatment Effect, r_1	23.04	23	41.19	0.01
Level-1 error, *e*	233.75			
Random Effect – Variability Among Studies (Level 3)	*Variance*	*Df*	χ^2	*p Value*
Baseline, u_{00}	135.38	12	68.63	<.001
Treatment Effect, u_{10}	128.53	12	89.21	<.001

Note. All coefficient values are in percentage points. The average treatment effect refers to the average change that students showed from baseline to intervention (their average improvement).

Table 5

Fully Conditional Model Showing the Effects of Age, Gender, and ADHD Diagnosis

Fixed Effect	*Coefficient*	*se*	*t Ratio*	*p Value*
Effects on Baseline Behavior Averages (*intercepts*)				
Age	1.07	1.02	1.05	0.32
Gender	-1.28	4.54	-0.28	0.78
Effects on Treatment Effect (*slopes*)				
Age	-0.30	1.48	-0.20	0.85
Gender	0.29	3.24	0.09	0.93

Note. All coefficient values are in percentage points. Negative values indicate a decrease in percentage points associated with a 1-point increase in the moderating variable. The average treatment effect refers to the average change that students showed from baseline to intervention (their average improvement).

Table 6

Fully Conditional Model Showing the Effects of Home-School Communication, Study Quality, and Class Type

Fixed Effect	*Coefficient*	*se*	*t Ratio*	*p Value*
Effects on Baseline Behavior Averages (*intercepts*)				
Communication	-3.09	2.89	-1.07	0.31
Quality	-0.23	5.95	-0.04	0.97
Class Type	-28.06	13.54	-2.07	0.07
Effects on Treatment Effect (*slopes*)				
Communication	-2.21	2.18	-1.02	0.34
Quality	13.17	4.33	3.04	0.01
Class Type	32.99	10.22	3.23	0.01

Note. All coefficient values are in percentage points. Negative values indicate a decrease in percentage points associated with a 1-point increase in the moderating variable. The average treatment effect refers to the average change that students showed from baseline to intervention (their average improvement).