Article

Hammill Institute on Disabilities

Remedial and Special Education
2017, Vol. 38(6) 353–370

DOI: 10.1177/0741932517716900
rase.sagepub.com
SAGE

A Meta-Analysis of School-Based Group Contingency Interventions for Students With Challenging Behavior: An Update

Daniel M. Maggin, PhD[1], James E. Pustejovsky, PhD[2], and Austin H. Johnson, PhD[3]

Abstract

Group contingencies are recognized as a potent intervention for addressing challenging student behavior in the classroom, with research reviews supporting the use of this intervention platform going back more than four decades. Over this time period, the field of education has increasingly emphasized the role of research evidence for informing practice, as reflected in the increased use of systematic reviews and meta-analyses. In the current article, we continue this trend by applying recently developed between-case effect size measures and transparent visual analysis procedures to synthesize an up-to-date set of group contingency studies that used single-case designs. Results corroborated recent systematic reviews by indicating that group contingencies are generally effective—particularly for addressing challenging behavior in general education classrooms. However, our review highlights the need for more research on students with disabilities and the need to collect and report information about participants' functional level.

Keywords

evidence-based practice, single-subject, research methodology, meta-analysis, management, behavior

Classroom management remains among the greatest challenges confronted by school personnel (Hoglund, Klingle, & Hosan, 2015). Disruptive and off-task behaviors can have a broad impact on the academic and social performance of individual students as well as their classrooms (Pas, Cash, O'Brennan, Debnam, & Bradshaw, 2015). These concerns have contributed to systematic lines of classroom management research aimed at identifying effective methods for creating contexts to support student learning (Emmer & Sabornie, 2015). With the advent of the evidence-based practice movement within education, these classroom management methods have increasingly become the focus of experimental research, with the goal of establishing their effectiveness and determining for whom and under what conditions they are most likely to work (Kratochwill, Altschaefl, Bice-Urbach, & Kawa, 2015). Perhaps the most widely researched and experimentally tested classroom management intervention has been the group contingency (Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008; Stage & Quiroz, 1997). Indeed, research reviews dating back four decades have demonstrated that group contingencies are an effective intervention for addressing the classroom behavior of students (e.g., Litow & Pumroy, 1975; Theodore, Bray, Kehle, & Dioguardi, 2004). Despite these positive results, there remains a need to examine the group contingency research base to ensure that the most accurate and current information is being disseminated and provide more nuanced and detailed analysis of variables that influence their efficacy.

Group Contingency Interventions

The challenges associated with managing student behavior have led to the development of a number of strategies and methods to systematize the implementation of effective practices (Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2016). Among the most widely researched classroom management strategies available is the group contingency, in which students receive a predetermined preference item or activity contingent on the behavior of one or more students in a group (Little, Akin-Little, & O'Neill, 2015). More specifically, students receive these rewards based on whether a subset of students is able to meet predetermined behavioral expectations (Chow & Gilmour, 2016). For

[1]The University of Illinois at Chicago, USA
[2]The University of Texas at Austin, USA
[3]University of California, Riverside, USA

Corresponding Author:
Daniel M. Maggin, The University of Illinois at Chicago, 1040 W Harrison Street, M/C 147, Chicago, IL 60607, USA.
Email: dmaggin@uic.edu

instance, a group contingency intervention might provide students with an opportunity to engage in a preferred game if there are fewer than five disruptions during a particular period such as math or language arts.

From the principles of operant conditioning, group contingencies are effective because the behavior of individual students is conditioned through the delivery of a positive consequence for meeting a set of specified expectations (Hanley & Tiger, 2011). Moreover, group contingencies have a well-defined antecedent event through formally communicated start and end points, which signal that the behavioral expectations and potential for earning the positive consequences are in effect (Cooper, Heron, & Heward, 2007). Based on behavioral learning theory, therefore, students learn to respond in accordance with the expectations because doing so results in a positive consequence. The association between the initiation of the group contingency and the outlined expectations grows stronger over time as students learn the responses that will and will not lead to the reward. The implementation of an intervention with clear procedures can have the added benefit of enhancing the positive perception of the intervention by implementers as well (Briesch, Briesch, & Chafouleas, 2015).

As previously noted, the group contingency is among the most widely researched classroom management interventions (Simonsen et al., 2008). Several research teams have examined the corpus of research to evaluate the evidence base and to provide recommendations regarding the contexts in which group contingencies are most likely to work. Over time, these reviews have become increasingly systematic as methodologies have developed and perspectives have evolved to embrace an evidence-based perspective. One of the first, and most widely cited, reviews was by Litow and Pumroy (1975), who provided a narrative overview that sought to determine which specific procedures were most effective. A primary conclusion of this review was the need for additional research to inform the selection of particular intervention methods and to determine the contexts to which findings would generalize. Other narrative reviews on group contingencies have drawn similar conclusions—that the intervention has a number of inherent strengths and that the evidence was strong, but that more research was needed to make statements regarding for whom they work best and why (Theodore et al., 2004).

With the increasing emphasis on evidence-based practice, subsequent efforts to review the research on group contingencies have used increasingly systematic methods. Building on previous narrative approaches, recent reviews have provided careful consideration of both methodological issues and those related to determining the students and contexts to which group contingencies apply. For instance, Little et al. (2015) conducted a systematic review to evaluate the overall strength of the evidence. Using the within-case standardized mean difference (SMD) index (Busk & Serlin, 1992), the authors reported effect size information for each included study, as well as descriptive summaries of those effect size indices. Their findings indicated that group contingencies were effective for a range of student populations, outcomes, and contexts.

Despite the general support for group contingencies, however, there continues to remain a need for greater clarification regarding the particular functional and environmental characteristics that predict successful generalization (Maggin, 2015). In an earlier effort to address these issues, Maggin, Johnson, Chafouleas, Ruberto, and Berggren (2012) conducted a review focusing explicitly on the use of group contingencies to address student disruptive and academic engagement behaviors in classroom settings. The authors also calculated a variety of effect size metrics to characterize the magnitude of treatment effects and synthesized effects across studies using multilevel models. Findings of this review again indicated that the research base underlying group contingencies was robust but also that—similar to the conclusions of Litow and Pumroy (1975)—there remained insufficient information to determine the contexts and students to whom the intervention could be expected to generalize.

Review Purpose and Research Questions

The current update serves two purposes. The first purpose is to ensure the inclusion of the most current research on group contingencies. Consideration of new research is potentially important for several reasons, including supporting or refuting current conceptualizations of evidential support and providing additional insight into the populations and conditions to which an intervention might generalize. The second purpose was to apply new methods for reviewing and meta-analyzing single-case research. Application of novel methods allows for the identification of strengths and weaknesses while also providing an opportunity for corroboration of the results from previous reviews based on earlier methods. With these purposes in mind, the present study pursues the following research questions:

Research Question 1: Does the inclusion of studies published since the earlier review provide additional evidence either in support of or in contrast to previous research?

Research Question 2: Does the inclusion of studies published since the earlier review provide additional insight into the conditions that might promote generalization?

Research Question 3: Do novel meta-analytic procedures produce similar results to those previously used with this body of research?

Research Question 4: Do selected moderator variables provide evidence of systematic variation in the magnitude of intervention effect?

Research Question 5: What are some considerations for further development of meta-analytic methods for single-case research?

Method

The methods for identifying and searching for relevant studies, coding descriptive and methodological information, and extracting data for analyses replicated those reported in Maggin et al. (2012). Because of this overlap, we provide only a brief summary of these parts of the method. Full operational details are available in an earlier, working version of this article (Maggin, Pustejovsky, & Johnson, 2017).

Study Identification Procedures

The research reports were identified using a three-step-process. First, candidate reports were located through an electronic database search of six educational and social science databases, including the *Educational Resources Information Center* (ERIC), *Proquest Dissertation Abstracts, PsychINFO, PubMed, Scopus*, and *Sociological Abstracts*. Keywords included *group contingency* and *contingency management* to retrieve the greatest number of returns. Second, the research team conducted an ancestral search of the reference lists of eligible studies. Third, the citation lists of relevant review articles published since the initial review were examined (e.g., Little et al., 2015). The search was completed on January 1, 2017.

Inclusion Criteria

The following six criteria were used to determine whether research reports were eligible for inclusion. First, the study had to be conducted no earlier than 1960. Second, the study had to be a single-case research investigation of a group contingency intervention. Group contingency interventions were defined as behavioral management systems in which reinforcement or punishment was delivered to a group of students based on the performance of one or more students in the class (Barrish, Saunders, & Wolf, 1969; Heering & Wilder, 2006). Third, the study the research included only students in kindergarten to the 12th grade. Fourth, the research had to be conducted during classroom instruction. Fifth, studies included students demonstrating behavioral problems, with or without disabilities; studies of students with cognitive or severe disabilities were excluded. Sixth, the dependent variable had to be a directly observable behavior related to problematic classroom performance such as social interactions, academic engagement, or disruptive behaviors. Figure S1 in the supplementary materials provides a schematic overview of the search process for the updated review, including the total number of studies and cases identified through the database search and the reasons for discarding ineligible studies.

Study Coding Procedures

The studies were reviewed to assess methodological rigor and evidential strength and to inventory essential characteristics of the research participants, settings, and intervention. Coders included two individuals with a doctoral degree and one individual with a master's degree in school psychology. Both doctoral-level coders had experience conducting systematic reviews of single-case research, carrying out primary single-case research, and teaching master's level courses on single-case design and applied behavior analysis. The master's level coder had taken classes in single-case design methods and behavior analysis prior to starting the coding process.

Evidence Evaluation

Design standards. The What Works Clearinghouse (WWC) procedures for evaluating single-case research (Kratochwill et al., 2013) were the primary method used to evaluate the rigor of included studies. The first step was to assess each study on the design criteria including (a) the systematic manipulation of the independent variable, (b) repeated collection of the dependent variable, (c) collection of interobserver agreement for at least 20% of sessions and an average of 80% agreement, (d) the presence of three opportunities to demonstrate the intervention effect, and (e) the presence of three or more data points per phase. Failure to meet any of these standards resulted in the study being classified as not meeting design standards. Moreover, cases with fewer than 3 points in any phase were classified as not meeting design standards, those with 3 or 4 points in any phase were classified as meeting design standards with reservations, and those with 5 or more data points in all phases were classified as meeting design standards.

Evidence standards. For cases meeting design standards with or without reservations, structured visual analysis was conducted to determine whether a functional relation was demonstrated. Specifically, the visual analyses were supported through the application of the visual analysis summary protocol described in Gast and Spriggs (2014). The summary protocol addresses a range of data properties commonly considered when conducting visual analyses, including information relating to both within- and between-condition data patterns as specified within the WWC evidence standards. Protocols were produced from the raw data using R syntax (Johnson, 2016). Each visual analyst used the information from the within- and between- phase assessments to independently determine whether there was a functional relationship. Visual analysts completed the analyses independently and came to a conclusion regarding the overall strength of the evidence using the WWC classifications of (a) meets evidence standards, (b) meets evidence standards with reservations, or (c) does not meet evidence standards.

Study characteristics coding. Following the application of the WWC evidence standards, the group of studies meeting evidence standards was reviewed along several dimensions, including information pertaining to (a) a series of methodological aspects that were not addressed within the evidence evaluation protocol but remain important for understanding the overall quality of evidence, (b) various characteristics of the research participants and settings in which the research took place, (c) various aspects of the independent variable and the manner in which it was operationalized and implemented within each study, and (d) features of the dependent variables within each study. Coding procedures were the same as those reported in Maggin et al. (2012). Interested readers are referred to an earlier version of this article (Maggin et al., 2017) for further details.

Meta-Analysis Procedures

The visual analyses conducted were supplemented with the application of meta-analysis to assess the magnitude and consistency of the intervention effects. Data from each study were analyzed using two recently developed between-case, standardized mean difference (BC-SMD, also known as "DHPS") effect size measures (Hedges, Pustejovsky, & Shadish, 2012, 2013; Pustejovsky, 2013), each of which is designed to produce an effect size index that is on the same scale as the SMD computed from a between-groups randomized experiment. The indices are premised on specific statistical models for the data from each single-case design and it is important to understand and evaluate the modeling assumptions upon which the effect size estimates are based. We therefore apply two distinct models—the "immediate effects model" and the "gradual effects model"—to examine the sensitivity of our findings under different assumptions.

Both of the effect sizes have technical limitations in that they require studies to have at least three participants within either a treatment reversal, multiple baseline, or multiple probe design. These requirements are necessary because the BC-SMD effect sizes involve between-participant variation in the outcome, which can only be estimated if the study includes multiple participants. Thus, the effect size estimates were computed for only a subset of the studies meeting the inclusion criteria outlined in the previous section. Both effect sizes were computed using the scdhlm package (Pustejovsky, 2016) for the R statistical computing environment.

Immediate effects model. BC-SMD effect sizes were first introduced by Hedges et al. (2012, 2013) for treatment reversal designs and across-participant multiple baseline/ multiple probe designs. The original version of the effect size is based on modeling assumptions that are appropriate for data patterns with stable baseline levels and abrupt level changes following the implementation of an intervention; we therefore describe it as the "immediate effects model." The modeling assumptions associated with this method include the following: (a) the dependent variable was measured on a continuous scale that is common across cases, (b) the baselines are stable and do not exhibit systematic time trends, (c) the intervention effect can be modeled as a change in the mean level of the outcome, (d) the intervention effect is homogeneous across cases, (e) the outcomes are normally distributed about case- and phase-specific mean levels, and (f) within each case, the outcomes follow a first-order auto-regressive process. The final assumption means that the model accounts for the possibility of autocorrelation in the measurements taken on each case. In light of these assumptions, the immediate effects model is best suited for data with stable data patterns across both baseline and intervention conditions.

Gradual effects model. Building on the initial development of the BC-SMD effect size index, Pustejovsky (2013) introduced a "gradual effects model" that relaxes two assumptions of the original model. First, the gradual effects model allows for the effects of the intervention to emerge gradually and to grow toward a new equilibrium level, rather than assuming that the full effect of intervention is immediate. This results in nonlinear time trends during intervention phases, as well as nonlinear trends during withdrawal phases of treatment reversal designs. Second, the gradual effects model allows for the full effect of intervention to vary across cases in a study, rather than assuming that the effect is homogeneous across cases. Compared with the immediate effects model, the more flexible assumptions requires more data (more cases, longer phases) to obtain accurate effect size estimates.

Meta-analysis method. The effect size estimates derived from the use of the immediate effects model and the gradual effects models were synthesized in separate meta-analyses. A subset of the studies examined the effects of group contingencies on multiple dependent variables in domains relevant to the purpose of this review, and so effect size estimates were computed for each dependent variable. A consequence of meta-analyzing multiple effect size drawn from the same study is that these estimates are not independent of each other. That is, effect size estimates drawn from the same participants—and using the same procedures—are likely to be more highly correlated than those based on data from other studies in the review. To account for the dependency among effect sizes based on the same set of participants, effect size estimates were synthesized using robust variance estimation (RVE) techniques, based on a correlated effects model (Hedges, Tipton, & Johnson, 2010). This technique accounts for the correlation among effect sizes within the same study, and is also robust to the possibility that the sampling variances of the effect sizes are not accurately estimated. Given the relatively small number of included samples, small-sample adjustments to RVE were used to maintain close-to-nominal error rates for hypothesis tests and confidence intervals (CIs; Tipton & Pustejovsky, 2015).

Heterogeneity and moderator analyses. Meta-analysis is concerned not only with questions of the average level of intervention effects, but also with the extent to which those effects vary across settings, participants, implementations of the intervention, and other factors. For the current study, we investigated variation in intervention effects in three ways. First, we used the estimated standard deviation of true effects (often called τ or "tau") and the I^2 statistic to quantify the degree of variability in the results across studies. The I^2 index provides an estimate of the proportion variability in effect size estimates that is associated with true heterogeneity of effects rather than sampling error alone (Higgins & Thompson, 2002). Second, we conducted moderator analyses to investigate whether the variability across studies could be explained by conceptually relevant variables. Variables relating to various samples, intervention, and methodological characteristics were tested using metaregression models, again using RVE techniques and small-sample corrections. Finally, the research team investigated whether the evidence ratings were associated with effect size magnitudes. For these analyses, the research team had to apply a decision rule regarding the evidence classification for studies with cases meeting and not meeting evidence standards. As such, the research team categorized studies based on the most prevalent evidence classification. For instance, if the categorization of most cases in a study was not meeting standards, all cases were classified as not meeting standards. It is important to note that, although we planned to examine all variables described in the coding scheme, reporting issues and the subsequent lack of variation precluded the examination of some moderators.

Procedures for extracting graphed data. To compute effect sizes, data from each eligible case identified for the review were extracted using the digitization program WebPlotDigitizer (Rohatgi, 2014). Data extracted from digitization software has been shown repeatedly to provide highly reliable and valid estimates from single-case research (Drevon, Fursa, & Malcolm, 2017; Flower, McKenna, & Upreti, 2016; Moeyaert, Maggin, & Verkullian, 2016).

Interobserver Agreement

The research team randomly selected 50% of the included papers and computed interobserver agreement on the design standards, evidence standards, and coding scheme items. Calculation of the agreement indices included comparing responses on each item of the coding scheme and visual analysis rating and dividing by the total number of opportunities for agreement. The research team randomly selected 50% of the papers to calculate interobserver agreement. Results indicated strong agreement for each domain with 97% agreement on design standard classifications and 93% agreement on evidence standard classifications. For the coding scheme domains, the coders agreed on 96% of the items in the sample and setting characteristics domain, 89% of the items in the independent variable features domain, 93% of the dependent variable features domain, and 87% of the items for methodological quality. In the event of discrepancies, the lead author made the final decision in consultation with the master's-level coder.

Results

WWC Design and Evidence Evaluation

WWC design standards. Figure S2 in the supplementary materials presents the results of applying the WWC design standards. For studies with the individual as the unit of analysis, there were 229 cases considered across both the initial and updated reviews. Of these, 28 cases met design standards without reservations, 65 cases met design standards with reservations, and 136 did not meet design standards. The updated review resulted in the identification of one additional study meeting design standards with reservations (Ling, Hawkins, & Weber, 2011). With this addition, 41% (n = 93) of the cases examining the effects of group contingency interventions at the student level were found to be eligible for the subsequent evidence evaluation. For studies with groups as the unit of analysis, 206 cases were eligible between the initial and updated review. Of these, 47 cases met design standards without reservations, 34 met design standards with reservations, and the remaining 125 did not meet design standards. As such, 39% (n = 81) of the group-level cases were eligible for the evidence review. The cases identified as part of the updated review were from 13 studies, bringing the total to 40 studies that contributed evidence in this review. Rodriguez and Anderson's (2014) study was included in the original review as a dissertation but was published in the interim. As such, we updated the citation and included it as part of the original review.

WWC evidence standards. The cases identified as meeting the WWC design standards with and without reservations were subjected to an evidence evaluation based on the principles of visual analysis. The metrics drawn from the Gast and Spriggs protocol were used to support the visual analysis process and inform determinations regarding the level of evidence provided by each case meeting design standards. Results of the visual analysis protocol developed by Gast and Spriggs (2014) are available in the supplementary materials. The results of the visual analyses correspond with the evidence evaluation in Figure S2. For the 93 cases with individuals as the unit of analysis that met design standards, 25% (*n* = 23) met evidence standards without reservations, 55% (*n* = 52) met evidence standards with reservations, and the remaining 20% (*n* = 18) did not meet evidence standards. The updated review resulted in the addition of one study and one case providing evidence with reservations, with no other additions found for the other evidence categories. For 81 cases with the classroom as the unit of analysis, 35% (*n* = 28) met evidence standards without reservations, 55% (*n* = 45) met evidence standards with reservations, and the remaining 10% (*n* = 8) did not meet evidence standards. The updated review resulted in the addition of 13 studies and 37 cases providing evidential support for group contingencies as a classroom-level intervention.

Methodological Characteristics

Table S1 in the supplementary materials presents the methodological characteristics of studies meeting the WWC evidence standards with or without reservations, from both the initial and current reviews. The studies identified as part of the updated review consistently addressed many of the methodological domains sampled. Methodological strengths across these studies included providing operational descriptions of the independent and dependent variables, using rigorous interobserver agreement procedures, and having adequate information and data stability for baseline phases. Moreover, the studies identified as part of the update represented a collective improvement in some areas over those included in the original review. Specifically, nearly all of the studies included an assessment of procedural fidelity and a majority reported on social validity data. Results from these measures generally indicated that intervention fidelity was high and that implementers found the procedures useful and easy to put into practice. Consistent with the methodological evaluation reported in the initial review, three areas were identified that might benefit from more attention including (a) reporting clear selection criteria for the students and classrooms, (b) providing additional information on the environmental conditions during baseline such as the particular management activities in place and a conceptual rationale for why current management practices are not working, and (c) the inclusion of generalization and maintenance phases to determine whether the intervention procedures can be applied to other settings or behaviors, or to evaluate the sustainability of the intervention without additional support.

Study and Case Characteristics

The studies, and their respective cases, meeting evidence standards with and without reservations were reviewed to determine the conditions under which the intervention could be expected to generalize. Table 1 provides an overview of study-level information for those studies providing evidential support with those included in the initial review and those identified as part of the update designated. The following sections provide an overview of the combined results of the initial review and the update. As with the original review (Maggin et al., 2012), the student characteristics for individual- and group-level analyses are reported separately to better understand the conditions for generalization for these different units of analysis. Because there was no procedural variability within studies, the other domains used to describe the operational procedures of group contingencies within the literature were reported at the study level. To distinguish between levels, we use *n* to denote the number of cases and *k* to denote the number of studies.

Individual characteristics. For studies examining the effects of group contingencies on individual-level performance, the average age of participants was 9.11 (*SD* = 2.46). Participants were drawn from first through 10th grades with a median grade level of fourth grade, a majority of students were male (*n* = 56, 75.42%), and the most common setting was general education (*n* = 48, 63.63%). In terms of the other demographic variables, there was considerable underreporting, with fewer than 60% of the studies examining the individual as the unit of analysis providing information on student ethnicity, academic achievement, disability, or primary language. As such, readers are cautioned to interpret these findings carefully and these characteristics are presented here as the proportion of cases reporting on these demographic domains rather than as an overall percentage of all cases. With these caveats, it was found that most students in these studies were White (*n* = 27, 60%) with moderate (*n* = 16, 53%) or low (*n* = 14, 47%) levels of academic achievement. Moreover, the most prevalent disabilities in studies with the individual as the unit of analysis were attention deficit disorders (*n* = 10, 37%) and emotional and behavioral disorders (EBD; *n* = 10, 37%). None of the studies reported the primary language of the students. In terms of the update, these findings are largely the same as those reported in the original Maggin et al. (2012) review.

Group-level characteristics. For studies examining the effects of group contingencies on group-level behaviors, the mean age of the classroom-level samples providing evidential support was 12.73 (*SD* = 2.90). One of the studies identified through the update provided an average age, although authors did universally report grade levels. The median grade of the classroom-level studies was third, with

Table 1. Key Characteristics of Studies Demonstrating Evidence for the Effectiveness of Group Contingency Interventions.

Study	Publication type	Unit of analysis	Research design	Dependent variable	Sample age range	Sample grade	Setting type	Contingency type	Reinforcer types
Studies identified as part of update									
Dart (2016)	Peer-reviewed journal article	Group	MBD	Engagement	N/R	Seventh–eighth	Gen Ed	Interdependent	Activity, edible, tangible
Donaldson, Vollmer, Krous, Downs, and Berard (2011)	Peer-reviewed journal article	Group	MBD	Behavioral	N/R	K	Gen Ed	Interdependent	Activity, edible, tangible
Ford (2015)	Thesis	Group	Reversal	Both	N/R	10th–11th	Gen Ed	Interdependent	Activity, edible, tangible
Herrera (2016)	Thesis	Group	MBD	Both	N/R	First–second	Gen Ed	Interdependent	Activity, edible, tangible
Lambert, Tingstrom, Sterling, Dufrene, and Lynne (2015)	Peer-reviewed journal article	Group	Reversal	Behavioral	N/R	Fourth–fifth	Gen Ed	Interdependent	Edible
Ling, Hawkins, and Weber (2011)	Peer-reviewed journal article	Individual	Reversal	Both	N/R	First	Gen Ed	Interdependent	Tangible
Lum (2015)	Thesis	Group	Reversal	Both	N/R	Ninth–10th	Gen Ed	Interdependent	Activity, edible, tangible
McHugh, Tingstrom, Radley, Barry, and Walker (2016)	Peer-reviewed journal article	Group	MBD	Behavioral	N/R	Second–third	Gen Ed	Interdependent	Activity, edible, tangible
Mitchell, Tingstrom, Dufrene, Ford, and Sterling (2015)	Peer-reviewed journal article	Group	Reversal	Behavioral	14–17	Ninth–10th	Gen Ed	Interdependent	Activity, edible, tangible
Radley, Dart, and O'Handley (2016)	Peer-reviewed journal article	Group	Reversal	Both	N/R	First	Gen Ed	Interdependent	Edible, tangible
Trevino-Mack, Kamps, and Wills (2015)	Peer-reviewed journal article	Group	Reversal	Engagement	N/R	10th	SPED	Interdependent	Edible
Wright and McCurdy (2012)	Peer-reviewed journal article	Group	Reversal	Both	N/R	K, fourth	Gen Ed	Interdependent	Tangible
Wright (2016)	Thesis	Group	MBD	Both	N/R	Ninth–10th	Gen Ed	Interdependent	Activity, edible
Studies identified as part of initial review									
Ascare and Axelrod (1973)	Peer-reviewed journal article	Group	Reversal	Engagement	N/R	Fifth–sixth	Gen Ed	Dependent	Activity
Barrish, Saunders, and Wolf (1969)	Peer-reviewed journal article	Group	Reversal	Behavioral	N/R	Fourth	Gen Ed	Interdependent	Activity, tangible
Christ and Christ (2006)	Peer-reviewed journal article	Group	MBD	Both	N/R	Ninth–10th	Gen Ed	Interdependent	Activity

(continued)

Table 1. (continued)

Study	Publication type	Unit of analysis	Research design	Dependent variable	Sample age range	Sample grade	Setting type	Contingency type	Reinforcer types
Cihak, Kirk, and Boon (2009)	Peer-reviewed journal article	Group	Reversal	Behavioral	N/R	Third	Gen Ed	Interdependent	Activity
Coleman (1970)	Peer-reviewed journal article	Individual	Reversal	Both	8–12	First, second, and fifth	Gen Ed	Dependent	Edible
Conklin (2010)	Dissertation	Group	Reversal embedded within MB	Engagement	N/R	K, second, and seventh	Gen Ed	Interdependent	Edible, tangible
Coogan, Kehle, Bray, and Chafouleas (2007)	Peer-reviewed journal article	Individual	Reversal	Behavioral	12	N/R	Gen Ed	Dependent / Interdependent	Edible, tangible, activity, work reduction
Crouch, Gresham, and Wright (1985)	Peer-reviewed journal article	Group	Reversal	Both	N/R	Third	Gen Ed	Interdependent	Edible, activity
Davies and Witte (2000)	Peer-reviewed journal article	Individual	Reversal	Behavioral	8–10	Third	Gen Ed	Interdependent	Activity, tangible
Grandy, Madsen, and Mersseman (1973)	Peer-reviewed journal article	Group	Reversal	Behavioral	N/R	N/R	Gen Ed	Interdependent	Activity
Greenwood, Hops, Delquardi, and Guild (1974)	Peer-reviewed journal article	Group	MB	Behavioral	N/R	First–third	Gen Ed	Interdependent	Activity
Hall et al. (1971)	Peer-reviewed journal article	Group	Reversal	Behavioral	N/R	First	Gen Ed	Interdependent	Activity
Hansen and Lignugaris-Kraft (2005)	Peer-reviewed journal article	Individual	Reversal	Behavioral	11–15	Seventh–ninth	SPED	Dependent	Edible, tangible, activity
Jones, Boon, Fore, and Bender (2008)	Peer-reviewed journal article	Individual	Reversal	Behavioral	12–13	Sixth	Resource	Dependent	Edibles, tangibles, activity, work reduction
Kelshaw-Levering, Sterling-Turner, Henry, and Skinner (2000)	Peer-reviewed journal article	Group	Reversal	Behavioral	N/R	Second	Gen Ed	Interdependent	Edibles, activity
Kiarie (2003)	Dissertation	Individual	Reversal	Behavioral	9–10	Fourth	Gen Ed	Interdependent	Tangible, edible, activity
McKissick, Hawkins, Lentz, Hailley, and McGuire (2010)	Peer-reviewed journal article	Group	MB	Both	N/R	Second	Gen Ed	Interdependent	Edible, tangible, activity
Rodriguez and Anderson (2014)	Peer-reviewed journal article	Group	MB	Behavioral	N/R	K	Resource	Interdependent	Activity, tangible

(continued)

Table 1. (continued)

Study	Publication type	Unit of analysis	Research design	Dependent variable	Sample age range	Sample grade	Setting type	Contingency type	Reinforcer types
Saigh and Umar (1983)	Peer-reviewed journal article	Group	Reversal	Behavioral	8	Second	Gen Ed	Interdependent	Activity, tangible
Salend and Lamb (1986)	Peer-reviewed journal article	Group	Reversal	Behavioral	11–13	First–third	Resource	Interdependent	Edible, tangible, activity
Salend and Gordon (1987)	Peer-reviewed journal article	Group	Reversal	Behavioral	8–10	Third–sixth	Resource	Interdependent	Activity
Salend, Reynolds, and Coyle (1989)	Peer-reviewed journal article	Group	Reversal	Behavioral	15–16	N/R	SPED	Interdependent	Edible, activity
Schmidt (2007)	Thesis	Individual	Reversal	Both	6	First	Gen Ed	Interdependent	Tangible
Switzer, Deal, and Bailey (1977)	Peer-reviewed journal article	Group	MB	Behavioral	N/R	Second	Gen Ed	Interdependent	Activity
Theodore, Bray, and Kehle (2001)	Peer-reviewed journal article	Individual	Reversal	Behavioral	N/R	N/R	SPED	Dependent / Interdependent	Edible, tangible, work reduction
Thorne (2005)	Dissertation	Individual	Reversal	Both	7–9	Second–third	Gen Ed	Interdependent	Edible, tangible, activity
Williamson, Campbell-Whatley, and Lo (2009)	Peer-reviewed journal article	Individual	Reversal	Engagement	N/R	10th	Resource	Dependent	Activity, work reduction

Note. MBD = multiple baseline design; N/R = not reported; SPED = special education.

classrooms from kindergarten through 10th grade included. These classroom samples tended to have a greater proportion of male students (M = 55.83%, SD = 23.38), with the most common setting being general education (n = 24, 86%). Similar to the individual-level findings, the reporting for other demographic variables occurred in fewer than 60% of the reports and therefore careful interpretation is warranted. For these analyses, the results represent the findings from those studies reporting on these outcomes and should not be interpreted as reflective of all cases. As such, the classroom samples reporting on these outcomes tended to include more White students (58.40%, SD = 20.38) and moderate levels of academic achievement (n = 20, 80%). Few studies provided information on the proportion of students with disabilities, although three studies focused primarily on students with EBD and four were comprised exclusively of students with learning disabilities. The primary language of the sample was not reported in any of the studies.

Independent variable features. Regarding intervention format, most of the studies used group contingencies that drew on interdependent procedures (k = 34, 85%). The primary implementer was most often the teacher (k = 35, 88%), with other studies using peers (k = 1, 3%), paraprofessionals (k = 1, 3%), and members of the research team (k = 3, 9%). Training was described in more than half of the studies (k = 23, 58%). Most of the group contingencies were implemented in conjunction with additional procedures (k =29, 73%), with the most common of these being the use of student feedback or self-monitoring procedures (k = 13) and modeling of the rules (k = 11). Rewards were delivered on a varying time schedule, with some studies providing immediate access (5 = 29, 73%) while the remaining studies indicated delaying delivery until a later time such as the end of the day. The group contingency interventions were implemented for an average of 47.95 min (SD = 56.30), with a range of 15 to 360 min. The study in which the intervention was implemented for 360 min was a substantial outlier, with the next closest time being 90 min. Regardless, the selection of time periods tended to correspond to a specific class period, such as math or language arts (k = 21, 53%) rather than to particular activities (k = 11, 28%), a specific length of time (k = 7, 18%), or the whole day (k = 1, 3%).

Dependent variable features. The dependent variables were predominantly related to disruptive behaviors (k = 22, 55%) rather than issues of academic engagement (k = 5, 15%), with the remaining studies including outcomes for both (k = 13, 33%). In addition, the majority of these operational definitions were defined to include less severe behavior (k = 37, 93%), with the remaining studies focusing on more intense behaviors such as aggression and verbal abuse. Data were most often collected using interval recording systems such as partial interval (k = 15, 38%) or momentary interval recording (k = 13, 33%), with other studies using event recording procedures (k = 11, 28%) or durational measures (k = 1, 3%). Most studies employed paper and pencil methods to collect the data (k = 35, 88%).

Meta-Analytic Results

Included cases. As noted previously, the techniques used to compute effect sizes could be applied only to studies that included three or more individual cases. Consequently, a number of studies had to be excluded from the meta-analytic sample. The cases providing evidence were from 40 studies that met design standards with or without reservations. Of these, 27 studies (68%) included a sufficient number of individual cases to compute BC-SMD effect sizes. Table S2 in the supplementary materials reports the specific studies included and excluded from the meta-analysis. Table S3 in the supplementary materials reports detailed results from the moderator analyses. The supplementary materials also include the machine-readable files containing raw data for each included study; effect size estimates, variances, and other associated information from each study, for both the immediate effects model and the gradual effects model; and computer code (written in R) for replicating all of the calculations and results.

Immediate effects model estimates. Table 2 includes the overall effect size estimates based on the immediate effects model, the estimates for each evidence classification, and an estimate for those studies that met evidence standards with and without reservations. Figure 1 presents a forest plot of these effect size estimates. Results are presented separately for disruptive behaviors and academic engagement. For the immediate effects model, the average effect size for disruptive behaviors was estimated as d = 1.95, 95% CI = [1.56, 2.34], $p < 0.001$. According to this estimate, the intervention led to average improvements of nearly a two-standard deviation over baseline levels. In addition, effects from individual studies were highly heterogeneous, with an estimated between-study SD of 0.74 (I^2 = 65%). Each level of evidence was associated with a progressively larger average effect size.

The average effect size for academic engagement outcomes was d = 1.80 95% CI = [1.36, 2.25]. These estimates indicate that intervention led to improvements in engagement of nearly 2.0 standard deviations over baseline levels. The effects were heterogeneous, with an estimated between-study SD of 0.62 (I^2 = 58%). The effect size estimates associated with each level of evidence were not linear; rather, those meeting with reservations were larger than those meeting without reservations. Both estimates, however, were larger than those from studies classified as not providing evidence.

Table 2. Average Effect Size Estimates, Overall and by WWC Evidence Classification, Based on Immediate Effects Model and Graduate Effects Model.

WWC evidence classification	Studies	Effects	Immediate effects model				Gradual effects model			
			Estimate	*SE*	*df*	tau	Estimate	*SE*	*df*	tau
Disruptive behavior										
Overall average	22	24	1.95	0.19	19.6	0.74	1.56	0.15	16.5	0.42
Does not meet	3	3	0.96	0.12	2.8	0.37	0.85	0.21	2.1	0.26
Meets with reservations	12	13	1.72	0.12	10.1		1.51	0.18	8.5	
Meets without reservations	8	8	2.66	0.30	5.4		2.11	0.16	4.2	
Meets with or without reservations	20	21	2.09	0.17	17.2	0.61	1.70	0.15	14.5	0.37
Academic engagement										
Overall average	14	14	1.80	0.20	11.9	0.62	1.88	0.23	12.5	0.72
Does not meet	3	3	1.15	0.17	2.0	0.55	1.06	0.03	2.0	0.56
Meets with reservations	9	9	2.09	0.27	7.0		2.23	0.28	7.5	
Meets without reservations	2	2	1.71	0.55	1.0		1.56	0.68	1.0	
Meets with or without reservations	11	11	2.05	0.24	9.0	0.61	2.15	0.25	9.5	0.66

Note. For estimated effects by WWC evidence classification, a single, pooled estimate of tau was calculated across all three design ratings. WWC = What Works Clearinghouse; *df* = degrees of freedom of the robust standard error; tau = estimated between-study standard deviation of true effect sizes.

The following moderators were tested in further metaregression models: student-level versus group-level unit of analysis, participant gender, participant grade level, general education versus special education setting, use of teacher training, dependent versus interdependent group contingency, length of intervention implementation, behavioral versus engagement outcomes, and measurement system for recording outcomes. The type of group contingency implemented moderated the results. None of the other moderators explained significant variability in the effect sizes.

Gradual effects model estimates. Table 2 also includes the overall effect size estimates, the estimates for each evidence classification, and an estimate for those studies that met reservations with and without reservations, all based on the gradual effects model. Figure 2 presents a forest plot of these effect size estimates. Results are presented separately for both disruptive behaviors and academic engagement. For disruptive behaviors, the average effect size across all studies and outcomes was d = 1.56, 95% CI = [1.25, 1.87], $p < .001$. These estimates indicate that the outcomes from the intervention phases represented approximately a 1.5 standard deviation improvement over baseline. The effects were found to be heterogeneous, with an estimated between-study *SD* of 0.42 (I^2 = 25%). The effect size estimates associated with each level of evidence were progressively larger, with those meeting without reservations larger than those meeting with reservations. Both estimates were also larger than those not providing evidence.

For academic engagement, the average effect size across all studies and outcomes was d = 1.88, 95% CI = [1.38, 2.38], $p < .001$. The effects were heterogeneous, with an estimated between-study *SD* of 0.72 (I^2 = 67%). The effect size estimates associated with each level of evidence were not linear with those meeting with reservations larger than those meeting without reservations. Both estimates, however, were larger than those not providing evidence.

None of the potential moderating characteristics explained a significant amount of variability in the effect size estimates based on the gradual effects model. Table S3 in the supplementary materials provides further details.

Immediate and gradual effects model comparison. For most studies, effect size estimates based on the gradual effects model were smaller than those produced by the immediate effects model. There are only a few instances where the discrepancies are large, indicating that these models generally performed similarly. Figure S3 in the supplemental materials presents a plot comparing the models.

Discussion

The purpose of this review was to provide an update to a previously published systematic review and meta-analysis of the group contingency literature. As noted in the introduction, it is essential that researchers expend the resources necessary to update evidence reviews to ensure dissemination of the most current information. As such, the present review was conducted to evaluate and synthesize recent group contingency research to determine if there was additional information that could be used to support its use in practice. Another goal of the review was to revisit the methods used to analyze and synthesize results from included studies, from both a visual and statistical standpoint. The visual- and

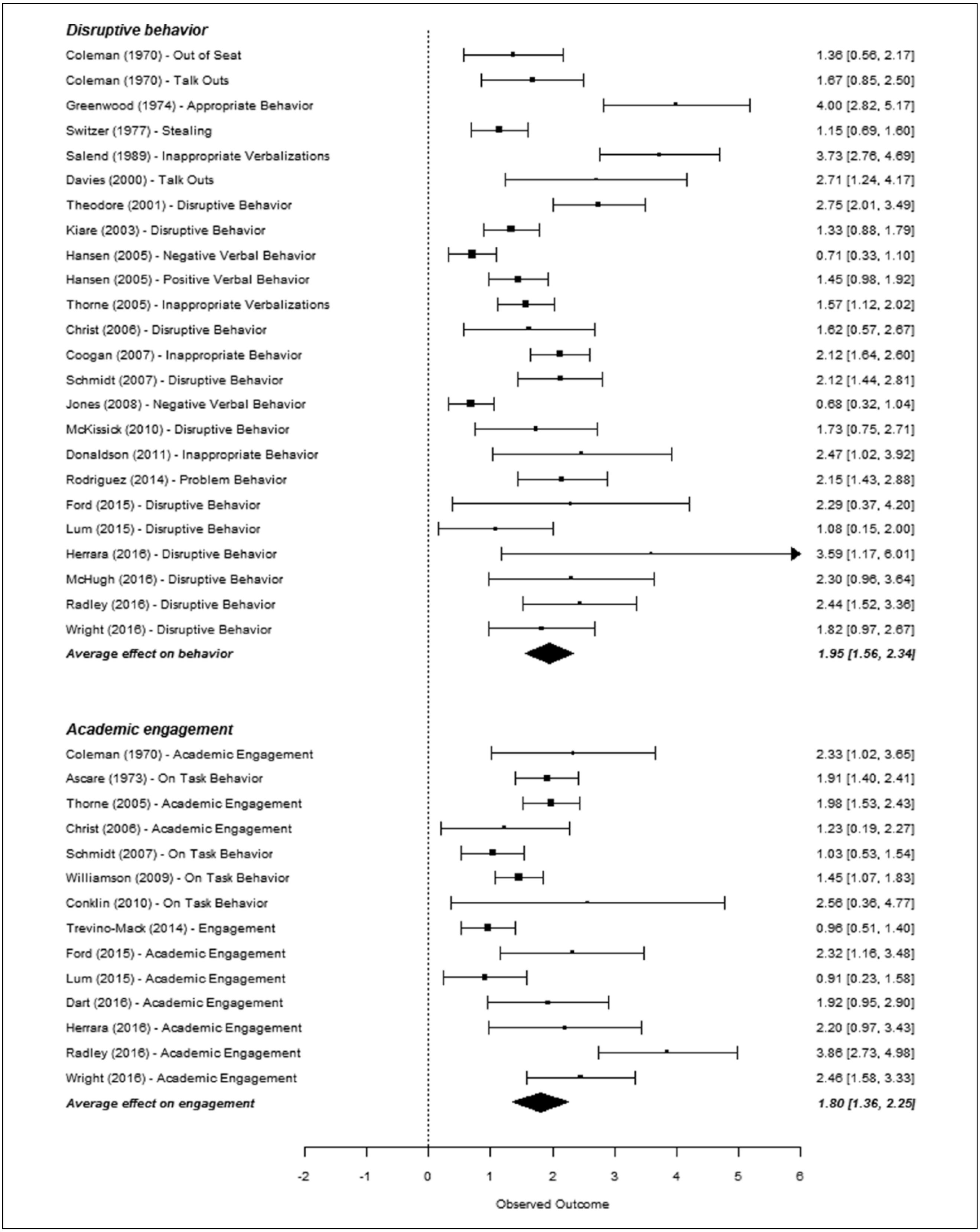

Figure 1. Forest plot of effect size estimates included in the meta-analysis based on the immediate effects model.

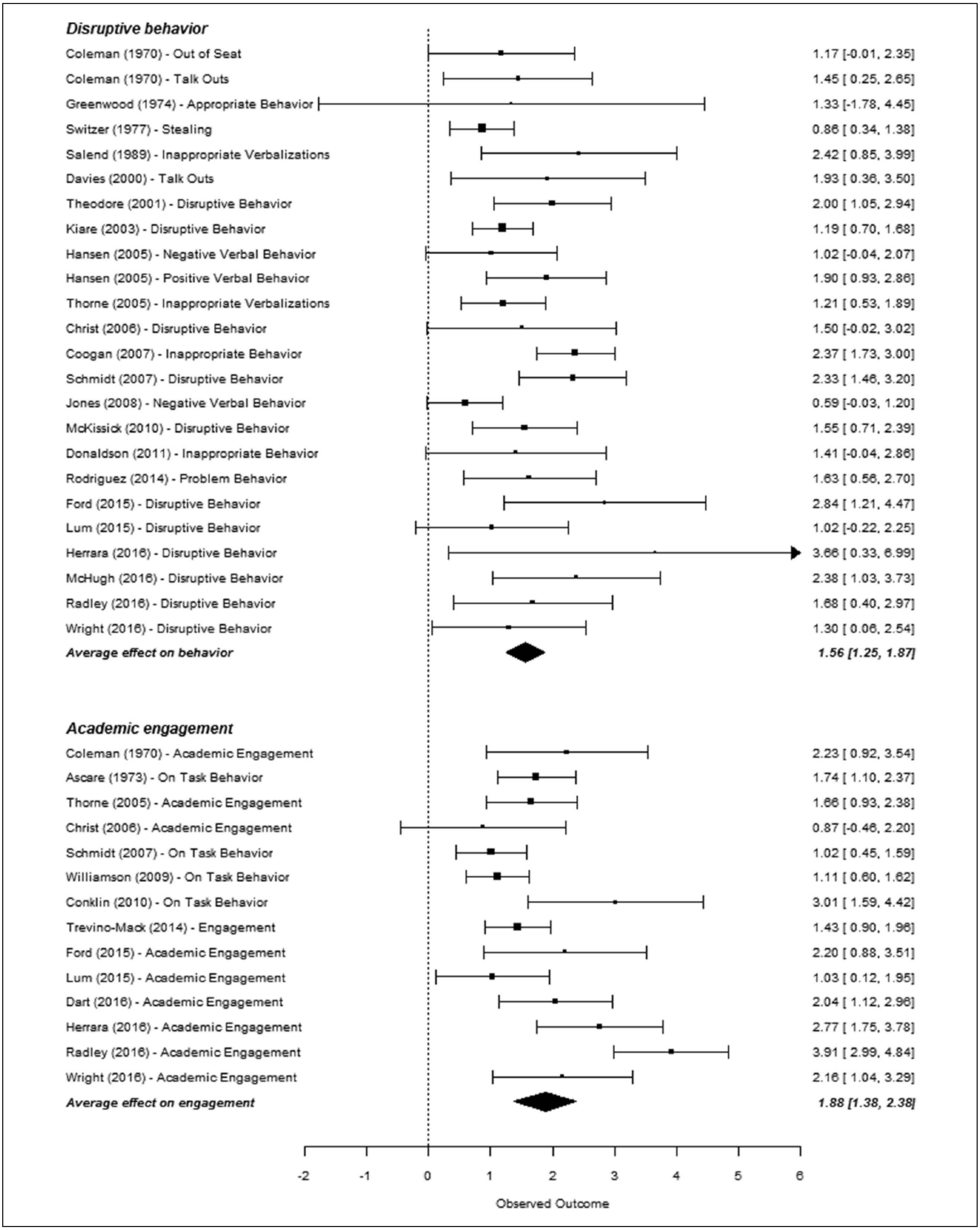

Figure 2. Forest plot of effect size estimates included in the meta-analysis based on the gradual effects model.

meta-analytic procedures used within the current article represent important extensions to those used previously and are advanced as critical updates in their own right. In the remainder, we summarize the results of this updated evidence review, provide some observations about the methodological advances used, and provide some recommendations for future systematic reviews of single-case research.

Updated Evidence Evaluation

The research on group contingency interventions appears to be robust with the identification of 40 total studies containing 148 unique cases providing evidential support for the intervention. These cases represented both student and classroom levels of behavior, with both meeting the WWC criteria for the number of cases, studies, and independent research teams needed to deem a practice as evidence based (Kratochwill et al., 2013). As such, there is empirical support for using group contingencies to target individual or group behavior. It is important, however, to consider sample and intervention characteristics of the included evidence in determining the students and contexts to which this intervention might apply (Maggin, 2015). Regarding generalization, the results are consistent with many of the observations made in the original review. For instance, the evidence base remains predominantly focused on students in late elementary and early middle school although recent research on group contingencies appears to be expanding to adolescents and high school students (e.g., Mitchell, Tingstrom, Dufrene, Ford, & Sterling, 2015; Trevino-Mack, Kamps, & Wills, 2015). A large majority of behaviors represented in these studies were less severe, such as calling out and out-of-seat behaviors, rather than aggressive behaviors. Furthermore, the average implementation time of group contingencies was less than an hour and most of the research took place in general education classrooms. Related to the latter point, there remains a dearth of research examining the use of group contingencies for students with disabilities or in special education classrooms. These findings suggest the need for more research to establish the effectiveness of group contingencies in special education settings and for students with disabilities.

The results of the present review also emphasize the need for additional research to understand the specific characteristics of those students who do and do not respond to group contingencies. Many demographic details were missing from the included reports, which presented difficulties for making strong recommendations. Perhaps more important than demographics, there was virtually no mention of functional characteristics of students in the included studies. Such a lack of information is striking, considering that, according to behavioral learning theory, function is essential for determining for whom the intervention is expected to work (Maggin, 2015). Results also suggest that group contingencies are often implemented with additional components, such as student self-monitoring and modeling the behavioral expectations, or perhaps varying expectations to address a range of functions. It is therefore possible—if not likely—that practitioners with a firm understanding of behavioral principles could extend these methods to a variety of behavioral functions (e.g., Maggin, Zurheide, Pickett, & Baillie, 2015). There remains a need for research to investigate such extensions through additional, rigorous experimental studies.

Methodological Observations

The current review extended previous research by employing novel data and visual analysis procedures to evaluate the available research base. In the following sections, some observations are provided on these issues.

The methodological area that has received the most attention since the publication of the original group contingency review has been single-case effect sizes (Shadish, 2014). Within the current update, we used two BC-SMD effect sizes based on somewhat different modeling assumptions. The most salient difference between these approaches is that the immediate effects model assumed that the data are stable and that the intervention effects are modeled as an overall level change, whereas the gradual effects model accounts for nonlinear data trends in the initial intervention phase and any subsequent phases. As many single-case researchers would attest (e.g., Gast & Spriggs, 2014), the latter model represents an important progression, given that many data patterns have perceptual trend. For the group contingency data, however, it appears that the influence of trend was generally small given that both effect size models produced similar results (see Figure S3 in the supplementary materials). These findings might signal to some that the models are essentially equivalent and that the more parsimonious approach would be preferred. To the contrary, we would argue that model selection should be informed both by visual inspection of the results and by underlying theory of behavior change. Consequently, we recommend that researchers attend to the statistical modeling assumptions that underpin effect size indices for single-case data to make informed decisions when selecting an effect size model that takes into account the data, the theorized intervention effect, and the research questions of interest.

The selection of an appropriate model for computing effect sizes is critical, yet there is also a need to acknowledge that the data requirements associated with the BC-SMD methods led to the exclusion of one third of the studies providing evidential support. If excluded studies differ in some systematic way from included studies, results of the meta-analysis would indicate bias. Addressing this lack of coverage will require the development of methods for including evidence from studies with fewer than the required number of cases. Incorporating external estimates of model parameters (such as the ICC, representing the proportion of variation in an outcome that is between cases) in estimating effect sizes represents one avenue. Another alternative would be to combine data across groups of studies that use common procedures for assessing the dependent variable, essentially creating "synthetic" studies with a sufficient number of cases to apply BC-SMD estimation methods.

The exclusion of some studies from the meta-analysis underscores the important role that visual analysis has within systematic evidence reviews of single-case research. That is, visual analysis is an important component of evaluating single-case research because it readily extends to all cases and studies within a review (Kratochwill, Levin, Horner, & Swoboda, 2014). For the current review, the WWC application standards were used to guide visual analysts (Kratochwill et al., 2013). Among the most prevalent criticisms of visual analysis, however, is that these methods can be subjective and, within the context of a systematic review where the graphs are often not readily available to readers, lack the transparency necessary to be fully evaluated (Horner, Swaminathan, Sugai, & Smolkowski, 2012). The current review took a step toward addressing this issue by applying a comprehensive visual analysis protocol advanced by Gast and Spriggs (2014) and making the results available to readers. By providing access to this information, it was our intention to provide readers with the opportunity to scrutinize and verify our results. We view this as a critical step toward developing more rigorous and transparent review methods for single-case research.

Recent scholarship on the use of statistical and visual analyses within systematic reviews and meta-analyses of single-case research has noted the importance of using these methods together (Kilgus, Riley-Tillman, & Kratochwill, 2016). In an effort to demonstrate the complementary role of visual and statistical analyses, the research team compared the average effect size estimates across the evidence classifications derived from the WWC standards. Results provided evidence that the average effect sizes for disruptive behavioral outcomes were larger for those studies meeting evidence standards versus those not meeting evidence standards. The correspondence between the evidence classifications and statistical analyses provide additional confidence in the classifications and represent one approach to use these methods together. However, the correspondence between some of the statistical analyses and evidence classifications were not fully consistent. For instance, the effect size estimates associated with academic engagement outcomes were larger for those studies meeting evidence standards with reservations than for those meeting without reservations.

On the surface, these findings are counterintuitive because cases with stronger classifications should logically have higher results. This is not necessarily the case, however, because visual analysts are concerned with a variety of other aspects aside from the magnitude of effect (Kilgus et al., 2016). That is, visual analysts must determine whether the conditions for establishing experimental control were met. This requires considering a variety of methodological aspects aside from the magnitude of treatment responses alone (Vannest & Ninci, 2015), such as the number of phase contrasts, the number of data points present within each phase, and interrater agreement indices, all of which contribute to making evidential judgments. Consistent with this view of experimental control, the WWC evidence classification system prompts the user to weigh a series of methodological characteristics in addition to the strength and consistency of the data patterns (Kratochwill et al., 2013). Whereas the effect size estimates represent the overall magnitude of effect regardless of research design issues, the visual analyst is more attuned with the range of characteristics required to deem the research as rigorous. The results reflect these different approaches in that some studies demonstrated larger effects despite having some methodological limitations. For example, a study might be downgraded to meeting evidence standards with reservations because the phases have three data points per phase even though there are large and consistent-level changes across baseline and intervention.

The association between methodological rigor and effect size magnitude has been an important issue in between-group experimental research (Moher et al., 1998). We believe that this question, and methods for separating assessments of methodological rigor from assessments of effect magnitude, requires much further investigation in the context of single-case research. Still, findings from the present investigation provide support for the recommendation of using statistical and visual analyses together within systematic reviews and meta-analyses that would aim to inform evidence-based practice.

Conclusion

The current review provided an update to a previously published systematic review and meta-analysis of the group contingency research. These updates related to both the identification of research conducted since the publication of the initial review and to the procedures used to synthesize and present results. As with the initial review (Maggin et al., 2012), findings provided strong support for the use of group contingencies in general education classrooms. The BC-SMD effect size estimates used in this review are on a different scale and are, therefore, difficult to compare directly. However, similar analyses in the original review led to similar conclusions. Moreover, there remains a need for researchers to continue to investigate issues related to determining for whom group contingencies are most likely to work. Providing this information to practitioners would allow for the development of more proactive adaptations to the basic group contingency protocol and could increase the efficiency with which students receive the necessary augmentative supports. As the field continues to develop this knowledge base, it appears that group contingencies warrant use in general education classrooms to address disruptive behaviors and promote increased academic engagement.

Acknowledgments

The authors would like to thank Dr. Will Shadish for his contributions on earlier statistical work for this article and Dr. Sandra M. Chafouleas for her contributions to the original review.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplementary Materials

Supplementary materials for this article are available on the journal's website and on the Open Science Framework at https://osf.io/8ygts/

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Ascare, D., & Axelrod, S. (1973). Use of a behavior modification procedure in four "open" classrooms. *Psychology in the Schools*, *10*, 243–248.

*Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, *2*, 119–124.

Briesch, A. M., Briesch, J. M., & Chafouleas, S. M. (2015). Investigating the usability of classroom management strategies among elementary schoolteachers. *Journal of Positive Behavior Interventions*, *17*(1), 5–14.

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

Chow, J. C., & Gilmour, A. F. (2016). Designing and implementing group contingencies in the classroom. *Teaching Exceptional Children*, *48*, 137–146.

*Christ, T. J., & Christ, J. A. (2006). Application of an interdependent group contingency mediated by an automated feedback device: An intervention across three high school classrooms. *School Psychology Review*, *35*, 78–90.

*Cihak, D. F., Kirk, E. R., & Boon, R. T. (2009). Effects of a classwide positive peer "tootling" to reduce the disruptive classroom behaviors of elementary students with and without disabilities. *Journal of Behavioral Education*, *18*, 278–286.

*Coleman, R. (1970). A conditioning technique applicable to elementary school classrooms. *Journal of Applied Behavior Analysis*, *3*, 293–297.

*Conklin, C. G. (2010). *The effects of class-wide function-related intervention teams (CW-FIT) on student's prosocial classroom behavior* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (3409240)

*Coogan, B., Kehle, T. J., Bray, M. A., & Chafouleas, S. M. (2007). Group contingencies, randomization of reinforcers, and criteria for reinforcement, self-monitoring, and peer feedback on reducing inappropriate classroom behavior. *School Psychology Quarterly*, *22*, 540–556.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson.

*Crouch, P. L., Gresham, F. M., & Wright, W. R. (1985). Interdependent and independent group contingencies with immediate and delayed reinforcement for controlling classroom behavior. *Journal of School Psychology*, *23*, 177–187.

*Dart, E. H., Radley, K. C., Battaglia, A. A., Dadakhodjaeva, K., Bates, K. E., & Wright, S. J. (2016). The classroom password: A class-wide intervention to increase academic engagement. *Psychology in the Schools*, *53*, 416–431.

*Davies, S., & Witte, R. (2000). Self-management and peer-monitoring within a group contingency to decrease uncontrolled verbalizations of children with attention-deficit/hyperactivity disorder. *Psychology in the Schools*, *37*, 135–147.

*Donaldson, J. M., Vollmer, T. R., Krous, T., Downs, S., & Berard, K. P. (2011). An evaluation of the good behavior game in kindergarten classrooms. *Journal of Applied Behavior Analysis*, *44*, 605–609.

Drevon, D., Fursa, S. R., & Malcolm, A. L. (2017). Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. *Behavior Modification*, *41*, 323–339.

Emmer, E. T., & Sabornie, E. J. (2015). Introduction to the section edition. In E. T. Emmer & E. J. Sabornie (Eds.), *Handbook on classroom management* (2nd ed.). New York, NY: Routledge.

Flower, A., McKenna, J. W., & Upreti, G. (2016). Validity and reliability of GraphClick and DataThief III for data extraction. *Behavior Modification*, *40*, 396–413.

*Ford, W. B. (2015). *Reducing disruptive behavior in high school: The good behavior game* (Unpublished master's thesis). University of Southern Mississippi, Hattiesburg, USA.

Gast, D. L., & Spriggs, A. D. (2014). General factors in measurement and evaluation. In D. L. Gast & J. R. Ledford (Eds.), *Single subject research methodology in behavioral sciences* (pp. 176–210). New York, NY: Routledge.

*Grandy, G. S., Madsen, C. H., & Mersseman, L. M. (1973). The effects of individual and interdependent group contingencies on inappropriate classroom behavior. *Psychology in the Schools*, *10*, 388–493.

*Greenwood, C. R., Hops, H., Delquardi, J., & Guild, J. (1974). Group contingencies for group consequences in classroom management: A further analysis. *Journal of Applied Behavior Analysis*, *7*, 413–425.

*Hall, R. V., Fox, R., Willard, D., Goldsmith, L., Emerson, M., Owen, M., . . . Porcia, E. (1971). The teacher as observer and experimenter in the modification of disrupting and talking-out behaviors. *Journal of Applied Behavior Analysis*, *4*, 141–149.

Hanley, G. P., & Tiger, J. H. (2011). Differential reinforcement procedures. In W. W. Fisher, C. C. Piazza & H. S. Roane (Eds.), *Handbook of applied behavior analysis* (pp. 229–249). New York, NY: Guildford Press.

*Hansen, S. D., & Lignugaris-Kraft, B. (2005). Effects of a dependent group contingency on the verbal interactions of middle school students with emotional disturbance. *Behavioral Disorders*, *30*, 170–184.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*, 224–239.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, *4*, 324–341.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65.

*Herrera, F. M. (2016). *The impact of a teacher-preferred group contingency with data-based decision making on class-wide behavior*. Tampa: University of Southern Florida.

Heering, P. W., & Wilder, D. A. (2006). The use of dependent group contingencies to increase on-task behavior in two general education classrooms. *Education and Treatment of Children*, *29*, 459–468.

Higgins, J., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558.

Hoglund, W. L., Klingle, K. E., & Hosan, N. E. (2015). Classroom risks and resources: Teacher burnout, classroom quality and children's adjustment in high needs elementary schools. *Journal of School Psychology*, *53*, 337–357.

Horner, R. H., Swaminathan, H., Sugai, G., & Smolkowski, K. (2012). Considerations for the systematic analysis and use of single-case research. *Education and Treatment of Children*, *35*, 269–290.

Johnson, A. H. (2016). *Functions for Gast and Spriggs' (2010) quantitative adjuncts to visual analysis* [R code]. Retrieved from https://osf.io/4cmfz/

*Jones, M., Boon, R. T., Fore, C., & Bender, W. N. (2008). Our mystery hero! A group contingency intervention for reducing verbally disrespectful behaviors. *Learning Disabilities*, *15*, 61–69.

*Kelshaw-Levering, K., Sterling-Turner, H. E., Henry, J. R., & Skinner, C. H. (2000). Randomized interdependent group contingencies: Group reinforcement with a twist. *Psychology in the Schools*, *37*, 523–533.

*Kiarie, M. (2003). *The differential effects of peer tutoring and peer tutoring with a group contingency on the spelling performance and disruptive behavior of fourth grade students in general education* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (3093664)

Kilgus, S. P., Riley-Tillman, T. C., & Kratochwill, T. R. (2016). Establishing interventions via a theory-driven single case design research cycle. *School Psychology Review*, *45*, 477–498.

Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, *86*, 643–680.

Kratochwill, T. R., Altschaefl, M. R., Bice-Urbach, B. J., & Kawa, J. M. (2015). Single-case experimental designs: Contributions to classroom management research. In E. T. Emmer & E. J. Sabornie (Eds.), *Handbook on classroom management* (2nd ed., pp. 538–559). New York, NY: Routledge.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, *34*(1), 26–38.

Kratochwill, T. R., Levin, J. R., Horner, R. H., & Swoboda, C. M. (2014). Visual analysis of single-case intervention research: Conceptual and methodological issues. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 91–125). Hillsdale, NJ: Lawrence Erlbaum.

*Lambert, A. M., Tingstrom, D. H., Sterling, H. E., Dufrene, B. A., & Lynne, S. (2015). Effects of tootling on classwide disruptive and appropriate behavior of upper-elementary students. *Behavior Modification*, *39*, 413–430.

*Ling, S., Hawkins, R. O., & Weber, D. (2011). Effects of a classwide interdependent group contingency designed to improve the behavior of an at-risk student. *Journal of Behavioral Education*, *20*, 103–116.

Litow, L., & Pumroy, D. K. (1975). A brief review of classroom group-oriented contingencies. *Journal of Applied Behavior Analysis*, *8*, 341–347.

Little, S. G., Akin-Little, A., & O'Neill, K. (2015). Group contingency interventions with children—1980-2010: A meta-analysis. *Behavior Modification*, *39*, 322–341.

*Lum, J. D. K. (2015). *The effects of tootling on disruptive and academic behaviors in high school* (Unpublished master's thesis). University of Southern Mississippi, Hattiesburg, USA.

Maggin, D. M. (2015). Considering generality in the systematic review and meta-analysis of single-case research: A response to Hitchcock et al. *Journal of Behavioral Education*, *24*, 470–482.

Maggin, D. M., Johnson, A. H., Chafouleas, S. M., Ruberto, L. M., & Berggren, M. (2012). A systematic evidence review of school-based group contingency interventions for students with challenging behavior. *Journal of School Psychology*, *50*, 625–654.

Maggin, D. M., Pustejovsky, J. E., & Johnson, A. H. (2017). *A meta-analysis of school-based group contingency interventions for students with challenging behavior: An update*. Retrieved from https://osf.io/hvuw5/

Maggin, D. M., Zurheide, J., Pickett, K. C., & Baillie, S. J. (2015). A systematic evidence review of the check-in/check-out program for reducing student challenging behaviors. *Journal of Positive Behavior Interventions*, *17*, 197–208.

*McHugh, M. B., Tingstrom, D. H., Radley, K. C., Barry, C. T., & Walker, K. M. (2016). Effects of tootling on classwide and individual disruptive and academically engaged behavior of lower-elementary students. *Behavioral Interventions*, *31*, 332–354.

*McKissick, C., Hawkins, R. O., Lentz, F. E., Hailley, J., & McGuire, S. (2010). Randomizing multiple contingency components to decrease disruptive behaviors and increase student engagement in an urban second-grade classroom. *Psychology in the Schools*, *47*, 944–959.

*Mitchell, R. R., Tingstrom, D. H., Dufrene, B. A., Ford, W. B., & Sterling, H. E. (2015). The effects of the good behavior game with general-education high school students. *School Psychology Review*, *44*, 191–207.

Moeyaert, M., Maggin, D., & Verkuilen, J. (2016). Reliability, validity, and usability of data extraction programs for single-case research designs. *Behavior Modification*, *40*, 874–900.

Moher, D., Jones, A., Cook, D. J., Jadad, A. R., Moher, M., Tugwell, P., & Klassen, T. P. (1998). Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *The Lancet*, *352*, 609–613.

Pas, E. T., Cash, A. H., O'Brennan, L., Debnam, K. J., & Bradshaw, C. P. (2015). Profiles of classroom behavior in high schools: Associations with teacher behavior management strategies and classroom composition. *Journal of School Psychology*, *53*, 137–148.

Pustejovsky, J. E. (2013). *Operationally comparable effect sizes for meta-analysis of single-case research*. Available from ProQuest Dissertations and Theses Global. (Order No. 3563832)

Pustejovsky, J. E. (2016). scdhlm: Estimating hierarchical linear models for single-case designs [R package version 0.3.*1]*. Retrieved from https://cran.r-project.org/package=scdhlm

Radley, K. C., Dart, E. H., & O'Handley, R. D. (2016). The quiet classroom game: A class-wide intervention to increase academic engagement and reduce disruptive behavior. *School Psychology Review*, *45*(1), 93–108.

*Rodriguez, B. J., & Anderson, C. M. (2014). Integrating a social behavior intervention during small group academic instruction using a total group criterion intervention. *Journal of Positive Behavior Interventions*, *16*, 234–245.

Rohatgi, A. (2014). *WebPlotDigitizer user manual version 3.4*. Retrieved from http://arohatgi.info/WebPlotDigitizer/userManual.pdf

*Saigh, P. A., & Umar, A. M. (1983). The effects of a good behavior game on the disruptive behavior of Sudanese elementary school students. *Journal of Applied Behavior Analysis*, *16*, 339–344.

*Salend, S. J., & Gordon, B. D. (1987). A group-oriented timeout ribbon procedure. *Behavioral Disorders*, *12*, 131–137.

*Salend, S. J., & Lamb, E. A. (1986). Effectiveness of a group-managed interdependent contingency system. *Learning Disability Quarterly*, *9*, 268–273.

*Salend, S. J., Reynolds, C. J., & Coyle, E. M. (1989). Individualizing the good behavior game across type and frequency of behavior with emotionally disturbed adolescence. *Behavior Modification*, *13*, 108–126.

*Schmidt, A. (2007). *The effects of a group contingency and individual behavior in an urban first-grade classroom* (Doctoral dissertation). Available from ProQuest Dissertations and Theses. (1443719)

Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology*, *52*, 109–122.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, *31*, 351–380.

Stage, S. A., & Quiroz, D. R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in public education settings. *School Psychology Review*, *26*, 333–368.

*Switzer, E. B., Deal, T. E., & Bailey, J. S. (1977). The reduction of stealing in second graders using a group contingency. *Journal of Applied Behavior Analysis*, *10*, 267–272.

*Theodore, L., Bray, M. A., & Kehle, T. J. (2001). Randomization of group contingencies and reinforcers to reduce classroom disruptive behavior. *Journal of School Psychology*, *29*, 267–277.

Theodore, L. A., Bray, M. A., Kehle, T. J., & Dioguardi, R. J. (2004). Contemporary review of group-oriented contingencies for disruptive behavior. *Journal of Applied School Psychology*, 20(1), 79–101.

*Thorne, S. (2005). *The effects of a group contingency intervention on academic engagement and problem behavior reduction in elementary school classrooms with at-risk students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (3191335)

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*, 604–634.

Trevino-Maack, S. I., Kamps, D., & Wills, H. (2015). A group contingency plus self-management intervention targeting at-risk secondary students' class-work and active engagement. *Remedial and Special Education*, *36*, 347–360.

Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling & Development*, *93*, 403–411.

*Williamson, B. D., Campbell-Whatley, G. D., & Lo, Y. (2009). Using a random dependent group contingency to increase on-task behaviors of high school students with high incidence disabilities. *Psychology in the Schools*, *46*, 1074–1083.

*Wright, R. A., & McCurdy, B. L. (2012). Class-wide positive behavior support and group contingencies: Examining a positive variation of the good behavior game. *Journal of Positive Behavior Interventions*, *14*, 173–180.

*Wright, S. J. (2016). *The effects of tootling combined with public posting in high school classrooms* (Unpublished master's thesis). University of Southern Mississippi, Hattiesburg, USA.