



2016

GROUP CONTINGENCY INTERVENTIONS FOR CLASSROOM BEHAVIOR MANAGEMENT: A SYSTEMATIC REVIEW OF THE LITERATURE 2011-2016

Danielle M. Crawford Duff

University of Kentucky, danielle.crawford@uky.edu

Digital Object Identifier: <https://doi.org/10.13023/ETD.2016.487>

[Click here to let us know how access to this document benefits you.](#)

Recommended Citation

Duff, Danielle M. Crawford, "GROUP CONTINGENCY INTERVENTIONS FOR CLASSROOM BEHAVIOR MANAGEMENT: A SYSTEMATIC REVIEW OF THE LITERATURE 2011-2016" (2016). *Theses and Dissertations--Early Childhood, Special Education, and Rehabilitation Counseling*. 32.

https://uknowledge.uky.edu/edsr_etds/32

This Master's Thesis is brought to you for free and open access by the Early Childhood, Special Education, and Rehabilitation Counseling at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Early Childhood, Special Education, and Rehabilitation Counseling by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@sv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Danielle M. Crawford Duff, Student

Dr. R. Allan Allday, Major Professor

Dr. Ralph Crystal, Director of Graduate Studies

GROUP CONTINGENCY INTERVENTIONS FOR CLASSROOM BEHAVIOR
MANAGEMENT: A SYSTEMATIC REVIEW OF THE LITERATURE 2011-2016

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in the
College of Education
at the University of Kentucky

By

Danielle M. Crawford Duff

Lexington, Kentucky

Director: Dr. R. Allan Allday, Associate Professor of Special Education

Lexington, Kentucky

2016

Copyright© Danielle M. Crawford Duff 2016

ABSTRACT OF THESIS

GROUP CONTINGENCY INTERVENTIONS FOR CLASSROOM BEHAVIOR MANAGEMENT: A SYSTEMATIC REVIEW OF THE LITERATURE 2011-2016

The purpose of this review was to evaluate the current literature base on the use of group contingency interventions to determine whether there is sufficient empirical evidence to recommend the practice for behavior management in K-12 classrooms. Studies were evaluated based on standards proposed by the What Works Clearinghouse. The results of the review indicate support for group contingencies as an evidence-based practice and highlight a need for increased experimental rigor and more detailed reporting to determine whether the interventions are effective and for which populations or settings they are most appropriate.

KEYWORDS: Group contingency, contingency management, behavior management

Danielle M. Crawford Duff

November 28, 2016

GROUP CONTINGENCY INTERVENTIONS FOR CLASSROOM BEHAVIOR
MANAGEMENT: A SYSTEMATIC REVIEW OF THE LITERATURE 2011-2016

By

Danielle M. Crawford Duff

Dr. R. Allan Allday

Director of Thesis

Dr. Ralph Crystal

Director of Graduate Studies

November 28, 2016

TABLE OF CONTENTS

LIST OF TABLES.....	iv
Section 1: Introduction	1
Problem Behavior	1
Evidence-Based Practice.....	3
Evaluating Group Contingencies	7
Section 2: Research Question	10
Section 3: Methods	11
Search Procedures	11
Inclusion criteria	11
Quality Indicators.....	12
Visual analysis	12
Evaluation of evidence base.....	13
Determining evidence-based practice	13
Descriptive characteristics	13
Interrater Agreement	14
Section 4: Results	20
Design Standards.....	20
Evidence Standards	20
Participants and Settings	21
Dependent Variables	21
Independent Variables.....	22
Student Outcomes	23
Determination of Evidence-Based Practice	24
Section 5: Discussion.....	26
Limitations	28
Implications for Future Research.....	28
Conclusion	30
References.....	31
Vita	37

LIST OF TABLES

Table 1, Evaluation of Studies Using What Works Clearinghouse Criteria.....	15
Table 2, Evaluation of Descriptive Characteristics of Studies Classified as Meets Standards and Meets Standards with Reservations.....	18

Section 1: Introduction

In recent years, federal and state legislation have placed an increased emphasis on accountability and teaching to high academic standards. Teachers and administrators must comply with a growing number of governmental mandates, such as the *Every Student Succeeds Act* (ESSA, 2015), which requires schools to measure and demonstrate academic progress while providing accommodations for all students to access the academic content. In addition, the *Individuals with Disabilities Education Improvement Act* (IDEA; 2004) mandated the accountability and equity of education for students with disabilities, which elevated the focus on the inclusion of individuals with disabilities into general education classrooms. These mandates have coupled learners with various levels of functioning together in the general education classroom. Often, children with disabilities exhibit higher rates of challenging behavior than their peers (Harrower & Dunlap, 2001), and these behaviors may interrupt teachers' ability to continue instruction.

Problem Behavior

Teachers and administrators have become more concerned about preventing disruptions to instruction (Sugai & Simonsen, 2012) because they may thwart the academic growth of students. Problem behavior in the classroom (e.g., talking out of turn, physical aggression toward property and others, inappropriate language) can disrupt instruction, and a teacher's inability to properly manage such behaviors can be detrimental for both the teachers and students. Problem behaviors in the classroom have been positively correlated to teacher stress (Clunies-Ross, Little, & Kienhuis, 2008) and burnout (Friedman, 2000; Kokkinos, 2007), and negatively affect student-teacher relationships and academic achievement (Horner et al., 2009; Sayal, Washbrook, &

Propper, 2015). Likewise, new teachers frequently transition into their careers and find themselves experiencing what has been termed “reality shock” as they are faced with problem behaviors and feel they are not adequately trained to manage the disruptions to instruction (Friedman, 2000).

Teachers who lack the expertise to address disruptive behavior have difficulty meeting the instructional demands within their classrooms, which also leads to poor student outcomes (Oliver & Reschly, 2010). Teacher education and training can play a significant role in developing the confidence and ability of teachers to effectively manage student problem behavior and minimize disruptions in their classrooms. However, multiple studies have demonstrated a deficit in teacher training where behavior management is concerned. In studies regarding higher-education teacher training programs, fewer than 50% of universities offered coursework solely focused on behavior management (Allday, Neilsen-Gatti, & Hudson, 2013; Oliver & Reschly, 2010). This is only compounded by the reports of teachers who feel unprepared and request additional training in classroom management (Stough, Montague, Landmark, & Williams-Diehm, 2015). White and Mason (2006) surveyed special education teachers and found that 83% of participants voiced a need for further training on how to properly manage behavior. These reports are particularly concerning because special education teachers are expected to be more knowledgeable and proficient in behavior management due to their responsibility for a population of students who typically require intensive behavior supports.

Due to the deficit in training, many educators use interventions that are *passed down* from other teachers or learned through trial-and-error experiences (Tillery, Varjas,

Meyers, & Collins, 2010). Clunies-Ross et al. (2008) also found that educators demonstrated the least amount of preference for consulting books to gain knowledge about behavior management strategies. Many teachers are now turning to the ever-expanding amount of digital content available via social media outlets (e.g., Pinterest, Facebook), wikis (e.g., WikiEducator, Wikipedia), online forums, and webpages to seek solutions and increase their understanding of topics within their profession (Hooks, 2015; Troutner, 2012). While web content allows individuals of shared professions to have means of contact for exchanging information and knowledge, it also has the potential to mislead users, which may leave some educators susceptible to misinformation regarding the use of content and strategies in their classrooms that may not be based on findings from rigorous studies in the special education and behavior management literature.

Evidence-Based Practice

Federal mandates have indicated the importance of using evidence-based practices for both academics and behavior management to foster progress and limit disruptions to instruction. Unfortunately, schools often fail to adopt such practices due to barriers such as teacher buy-in and lack of appropriate training (Pinkelman, McIntosh, Rasplica, Berg, & Strickland-Cohen, 2015). Teachers, administrators, and school districts are responsible for evaluating the quality and progress of both academic and behavioral interventions within the school setting (Wong et al., 2015). Given the increased demands on teachers to meet the academic and behavioral needs of all students, teachers need evidence-based behavior management strategies that are effective and practical to implement in the classroom. When implementing interventions in school settings, it has become increasingly imperative that the interventions are supported by research;

however, it was only in recent years that criteria for determining the quality of a study as a potential evidence-based practice was proposed for researchers and practitioners (Horner et al, 2005; Maggin, Johnson, Chafouleas, Ruberto, & Berggren, 2012).

Within education, there has been a shift toward utilizing and evaluating quantitative research to develop evidence-based practices (Horner & Kratochwill, 2012). One common experimental methodology, particularly in the field of special education, is single case research design (SCRD). SCR D is a quantitative research methodology in which all participants receive intervention and each participant serves as his or her own control (Gast & Ledford, 2014). The dependent variable (i.e., outcome/target behavior) is repeatedly measured in both the absence and presence of the independent variable (i.e., interventions). Unlike the more correlational or narrative descriptions of interventions of the past, SCR D seeks to determine if a functional relation exists between the independent and the dependent variables through measurement and systematic manipulation of the intervention across conditions.

While the term “single case” may suggest that the research techniques are only utilized with individuals, Kratochwill and Levin (2010) noted that researchers and practitioners have expanded the application of SCR D to dyads, small groups, classrooms, and institutions (e.g., schools and hospitals). In the field of education, SCR D is advantageous because educators are expected to effectively deliver and measure interventions with all students. However, all students do not respond to interventions in the same manner, and teachers may need to modify small aspects of their approach to yield greater gains in student progress. SCR D can be especially useful in this regard because it allows researchers and practitioners to measure the effects of interventions, the

effects of specific components of interventions, or to examine their effectiveness of interventions across multiple conditions. The evidence-base for a strategy is developed and strengthened through replication of the research with similar findings, which builds external validity by demonstrating similar effectiveness with different populations and/or learners in different locations. SCRD helps practitioners to select strategies based on those that have empirical support for their effectiveness (i.e., they are evidence-based), as well as derive with which populations of individuals and in what settings the interventions are most effective.

The federal government has recommended that educators and clinicians more widely adopt evidence-based strategies within their practice. Although SCRD contributed to the development of evidence-based practice, there has not been universally developed and accepted definitions of what translates to an “evidence-based practice”. The U.S. Department of Education established a Single-Case Design Panel (Kratochwill & Levin, 2010) to develop standards for SCRD and determine the level of empirical support required for an intervention to be recommended as “evidence based.” In addition, multiple researchers have proposed criteria for examining the quality of SCRD, with two of the most popular options being those developed by Horner et al. (2005) and What Works Clearinghouse (U.S. Department of Education, 2016).

Horner et al. (2005) proposed a list of components for defining and evaluating the quality of SCRD. The authors presented “quality indicators” for determining whether the methodological rigor of a study is “acceptable”. It was proposed that in order to be categorized as SCRD, research must include a description of participants and settings, dependent variable, independent variable, and baseline measures with detail that would

allow future researchers and practitioners to replicate procedures and select participants and settings that closely mirror those of the original study. In addition, the study must also control for threats to internal and external validity and include a report of social validity. Horner et al. (2005) believed that SCRD meeting the quality indicators could be categorized as evidence-based if: (a) the procedures, settings, and participants were defined in such a way that they could be replicated with fidelity, (b) the study demonstrated evidence that the procedures were carried out with fidelity, (c) threats from extraneous variables were controlled, (d) a clear functional relation existed between the independent and dependent variables, and (e) the intervention and its effects were replicated across a minimum of 20 participants in at least five single case studies, by at least three different researchers, in a minimum of three different locations.

With the goal of examining the quality of SCRD in areas like education and psychology, the WWC was established in 2002 by the Institute for Education Sciences (Wong et al., 2015). Kratochwill et al. (2013) reported that WWC arose from the increased “demands for accountability in education” and need for “identification of effective, evidence-based interventions” (p. 26). Based on criteria proposed by WWC for measuring the methodological rigor of SCRD, reviewers locate and examine studies, then classify each study as one of the following: (a) *Meets Standards*, (b) *Meets Standards with Reservations*, or (c) *Does Not Meet Standards*. During this process, the reviewers examine the documentation of interobserver agreement, the number of demonstrations of effect and data points per condition, and whether the independent variable was systematically manipulated (Kratochwill et al, 2013). If a meets design standards without or with reservations, data are analyzed using visual analysis. The extent to which data

represent a meaningful change in behaviors are categorized as follows: (a) *Strong Evidence*, (b) *Moderate Evidence*, or (c) *No Evidence* of a functional relation between the independent and dependent variables, which is based on a visual analysis of level, trend, and stability of the data, as well as overlap, the immediacy of the effect, and consistency of effect across conditions (Kratochwill et al, 2013).

As previously noted, federal mandates have increased requirements for evidence-based practices, while teacher knowledge and training has been shown to be lacking in managing potential behavior challenges (Oliver & Reschly, 2010). One likely contributor to educators' lack of interest in referencing literature when developing classroom management strategies is that many studies are conducted in highly controlled settings with resources (e.g., high ratio of staff) that are generally inaccessible or impractical for most classrooms. However, the recent focus on utilizing strategies that are grounded in research and empirical support has led to increased research in applied settings, which has also increased the overall social validity of the interventions that have been studied (Horner & Kratochwill, 2012).

Evaluating Group Contingencies

One practical approach for improving socially significant behavior in students is implementing a group contingency for *all* students in a classroom. Cooper, Heron, and Heward (2007) defined *group contingencies* as the delivery of a consequence, usually some form of a reward intended to serve as a reinforcer, that is accessible contingent on the behavior of a member, a designated portion, or all members of a group. The overarching goal of group contingencies is to promote adaptive behaviors and reduce problem behaviors. From a practitioner standpoint, group contingencies can be

advantageous because they minimize the number of staff needed to carry out the procedures and reduce the workload of adults who are responsible for managing the behavior of multiple learners at one time. For researchers and stakeholders, group contingencies are growing in popularity because they are simple for practitioners to implement, which increases the likelihood that teachers will be able to carry out the procedures effectively and with fidelity (Hulac & Benson, 2010).

The three most commonly described group contingencies are independent, dependent, and interdependent group contingencies (Cooper et al., 2007; Hulac & Benson, 2010). Independent group contingencies involve individualized contingencies, such as the teacher setting an expectation for the scores each student should receive (e.g., 90% or above) in order to receive a reinforcer. For example, a teacher may reward students with extended recess based upon the scores they receive on their spelling tests. Students who meet the goal will receive extra recess, and students who do not meet the goal will not be permitted to participate. Conversely, an interdependent group contingency would require all students to meet the goal for any member of the group to receive the reward. Lastly, a dependent group contingency is based on the performance of select members of the group. In this example, the entire group would receive access to the reward based on the performance of a teacher-selected individual(s).

Maggin et al. (2012) conducted a systematic review of group contingencies in school settings. The review examined studies from 1960 to 2012 based on WWC criteria. Based on the findings of 27 studies classified as *Meets Standards* or *Meets Standards with Reservations*, Maggin et al. (2012) concluded that sufficient data supported the categorization of group contingencies as evidence-based; however, it noted that there

were inconsistencies between the reported interventions and a general lack of documentation regarding participant demographics. The authors found that the inconsistencies and missing information from the reviewed studies prohibited them from reliably reporting on the types of individuals with whom group contingencies are most appropriate or effective. Maggin et al. (2012) reported multiple methodological weaknesses of studies examining group contingencies and called for researchers to expand on the literature base to provide greater support and detail for these interventions as an evidence-based practice.

Section 2: Research Question

The purpose of this review was to build upon the work of Maggin et al. (2012) and utilize WWC criteria to evaluate studies published within the last 5 years and determine whether recent publications contain sufficient descriptions of demographic variables. This review considered the following questions: (a) Do group contingency interventions employed in K-12 grade classrooms qualify as evidence-based strategies for managing problem behavior? and (b) With which behavior(s) and population(s) of students do group contingency interventions demonstrate empirical support for effective implementation?

Section 3: Methods

Search Procedures

The author reviewed experimental research to examine the empirical support for recommending group contingencies as an evidence-based practice for behavior management in school settings. The author used the search terms *group contingency* and *contingency management* (Maggin et al., 2012) in an electronic search of the following databases: *Academic Search Complete*, *Educational Research Information Center* (ERIC), *Psychology and Behavioral Sciences Collection*, *PsycINFO*, and *Sociological Collection*. Databases were selected based on overlap of publications searched by Maggin et al. (2012). The author also examined the reference list for each study included in the review and completed an ancestral search to identify additional articles not generated by the electronic databases. The titles and abstracts from the electronic and ancestral searches were scanned for inclusion criteria to determine the need for further examination.

Inclusion criteria. Studies included in this review met the following criteria: (a) targeted behavior management (i.e., decreasing problem behavior) or adaptive behavior (e.g., “on-task”) with dependent measures of disruptive behavior, (b) evaluated a group contingency intervention, (c) utilized a single case research design, (d) conducted the intervention in a classroom setting with a group of 5 or more students in kindergarten through 12th grade, (e) intervention occurred during typical classroom instruction (i.e., social activities such as recess and lunch were excluded), and (f) printed in English in a peer-reviewed journal between 2011 to 2016.

Quality Indicators

An electronic checklist (Table 1) was used to evaluate WWC criteria (U.S. Department of Education, 2016) for examining the rigor of experimental research. The checklist was used to record whether each study (a) systematically manipulated the independent variable, (b) collected interobserver agreement at least 20% of sessions, (c) agreement averaged at least 80% across all sessions, and (d) at least three opportunities to demonstrate an effect with the dependent variable being measured repeatedly over time (Kratochwill et al., 2013; Maggin et al., 2012). If these components were present, the number of data points per condition were also recorded. Studies were classified as (a) *Meets Standards* if the study met criteria and had five or more data points per phase, (b) *Meets Standards with Reservations* if there were 3-4 data points per phase, or (c) *Does Not Meet Standards* if there were less than 3 data points per phase or the study failed to meet criteria outlined in the checklist (Maggin et al., 2012).

Visual analysis. Studies classified as *Meets Standards* and *Meets Standards with Reservations* were evaluated using WWC (U.S. Department of Education, 2016) criteria for evidence standards of single case research design. Visual analyses assessed baseline and within-phase data patterns, which consisted of determining whether the data demonstrated (a) a dependent variable in need of change, (b) stability of data in each phase, (c) trend moving in a non-preferred direction, (d) change occurring between phases, (e) proportion of overlap between phases, and (e) similar patterns between similar phases (Maggin et al, 2012). Studies meeting these criteria were categorized as demonstrating a functional relation (Kratochwill et al., 2013) and the level of evidence was evaluated.

Evaluation of evidence base. Studies were classified as demonstrating *Strong Evidence*, *Moderate Evidence*, or *No Evidence* (Kratochwill et al., 2013). Studies labeled as providing *Strong Evidence* demonstrated at least three demonstrations of effect, each at a different time, with no non-effects. If a study demonstrated at least three demonstrations of effect at different times, but one non-effect (e.g., high variability of data within a phase, overlap between a control and intervention phase, inconsistent data patterns across similar phases), it was labeled as providing *Moderate Evidence*. Studies were classified as *No Evidence* if they did not provide at least three demonstrations of effect.

Determining evidence-based practice. The criteria used for determining whether group contingencies were an evidence-based practice were also based on WWC (Kratochwill et al., 2013) standards. To be recommended by WWC standards as an evidence-based practice, the strategy must be investigated by at least three different researchers, studied in three different geographical locations, and at least 20 subjects must have been included across those studies.

Descriptive characteristics. This review sought to determine the individual characteristics and conditions for which a group contingency intervention would be most applicable. WWC does not include indicators for examining descriptive characteristics, however, the determination of an evidence-based practice is based on the ability to replicate the effects of an intervention across multiple subjects in different geographical locations. For studies that were categorized as *Meets Standards* or *Meets Standards with Reservations* with “strong evidence or “moderate evidence,” quality indicators related to demographics, dependent variables, and setting were recorded (see Table 2) based on

quality indicators related to participant demographics and external validity as outlined by Horner et al. (2005). These characteristics included (a) participant age, gender, diagnoses, and diagnostic information, (b) type of group contingency, (c) target behavior(s), and (d) setting.

Interrater Agreement

The reviewer coded all the quality indicators (Table 1) and descriptive characteristics (Table 2) for examined studies. A second person independently coded 20% of studies after a 15-min training session to discuss data sheets which directly corresponded to Table 1 and Table 2. Studies were selected using the random function in Microsoft Excel. Interobserver agreement was calculated using the point-by-point method (Gast & Ledford, 2014). The number of agreements was divided by the total number of agreements plus disagreements, then the quotient was multiplied by 100 to obtain a percentage.

Table 1. Evaluation of Studies Using What Works Clearinghouse Criteria

Authors	Design	Systematic Manip. of IV	IOA 20% of sessions	IOA \geq 80%	≥ 3 Demonstr. of Effect	≥ 3 Data Points per Phase	≥ 5 Data Points per Phase	Design Standards	Evidence
Chafouleas et al. (2011)	Changing criterion	Y	Y	Y	N	N	N	--	--
Dart et al. (2016)	MB	Y	N	Y	Y	Y	Y	--	--
Denune et al. (2015)	ABCBC	Y	N	Y	Y	Y	Y	--	--
Donaldson et al. (2011)	MB	Y	N	Y	Y	Y	Y	--	--
Donaldson et al. (2015) 1	ABCDAD	Y	Y	Y	N	Y	N	--	--
Donaldson et al. (2015) 2	ABCD	Y	Y	Y	N	Y	N	--	--
Donaldson et al. (2015) 3	ABCD	Y	Y	Y	N	Y	N	--	--
Donaldson et al. (2015) 4	ABCD	Y	Y	Y	N	Y	N	--	--
Donaldson et al. (2015) 5	ABCD	Y	Y	Y	N	Y	N	--	--
Kamps et al. (2011) 1	ABAB	Y	N	Y	Y	Y	Y	--	--
Kamps et al. (2011) 2	ABAB	Y	N	Y	N	N	N	--	--
Kamps et al. (2011) 3	ABAB	Y	N	Y	N	N	N	--	--
Kamps et al. (2011) 4	ABA	Y	N	Y	N	N	N	--	--

Table 1 continued

Kamps et al. (2011) 5	ABAB	Y	N	Y	Y	Y	N	--	--
Kamps et al. (2011) 6	ABAB	Y	N	Y	N	N	N	--	--
Kamps et al. (2015) 1	ABCAC	Y	N	Y	N	N	N	--	--
Kamps et al. (2015) 2	ABCAC	Y	N	Y	Y	Y	N	--	--
Kamps et al. (2015) 3	ABCAC	Y	N	Y	N	N	N	--	--
Kamps et al. (2015) 4	ABCAC	Y	N	Y	N	N	N	--	--
Kleinman and Saigh (2011) 1	Changing criterion	Y	Y	Y	Y	Y	N	MSR	Moderate
Kowalewicz and Coffee (2014) 1	Changing criterion	Y	Y	Y	N	N	N	--	--
Kowalewicz and Coffee (2014) 2	Changing criterion	Y	Y	Y	N	N	N	--	--
Kowalewicz and Coffee (2014) 3	Changing criterion	Y	Y	Y	N	N	N	--	--
Kowalewicz and Coffee (2014) 4	Changing criterion	Y	Y	Y	N	N	N	--	--
Kowalewicz and Coffee (2014) 5	Changing criterion	Y	Y	Y	N	N	N	--	--
Kowalewicz and Coffee (2014) 6	Changing criterion	Y	Y	Y	N	N	N	--	--
Kowalewicz and Coffee (2014) 7	Changing criterion	Y	Y	Y	N	N	N	--	--

Table 1 continued

	Kowalewicz and Coffee (2014) 8	Changing criterion	Y	Y	Y	N	N	N	--	--
	Lambert et al. (2015) 1	ABAB	Y	Y	Y	Y	Y	N	--	--
	Lambert et al. (2015) 2	ABAB	Y	Y	Y	Y	N	N	MSR	Moderate
	Ling et al. (2011)	ABAB	Y	Y	Y	Y	Y	N	MSR	Moderate
	Mitchell et al. (2015) 1	ABAB	Y	Y	Y	Y	Y	Y	MSR	Moderate
	Mitchell et al. (2015) 2	AB	Y	Y	Y	N	Y	Y	--	--
	Mitchell et al. (2015) 3	ABAB	Y	Y	Y	Y	Y	Y	MSR	Moderate
17	Radley et al. (2016) 1	ABAB	Y	Y	Y	Y	Y	N	MSR	Moderate
	Radley et al. (2016) 2	ABAB	Y	Y	Y	Y	Y	Y	MS	Moderate
	Radley et al. (2016) 3	ABAB	Y	Y	Y	Y	Y	N	MSR	Moderate
	Wills et al. (2014)	MB	Y	N	Y	Y	Y	N	--	--

MB = multiple baseline; MS = meets standards; MSR = meets standards with reservations

Table 2. Evaluation of Descriptive Characteristics of Studies Classified as Meets Standards and Meets Standards with Reservations.

	Contingency	Target Behavior(s)	Age (yrs)	Grade	Gender	Diagnosis	Prerequisites	Setting
	INT	Vocal disruption; physical aggression; out-of-seat	15.39	9	15 males 11 females	--	Reports from teacher	General Education (History)
	INT	Disruptive behaviors	--	4	10 males 7 females	2 learning disability	SET rating 93%; referral from admin	General Education (Language Arts)
	INT	Off-task; engagement	8	1	1 male	--	Reports from teacher	General Education ("Carpet")
18	INT	Disruptive behavior	15.2	9 (18) 11 (2) 12 (1)	16 males 5 females	--	SET rating 79%; referral from admin; disruptive behavior 30% of intervals	General Education (Algebra)
	INT	Disruptive behavior	15.6	9 (5) 10 (11) 11 (7)	7 males 16 females	--	SET rating; 79%; referral from admin; disruptive behavior 30% of intervals	General Education (Spanish)

Table 2 continued

Radley et al. (2016) 1	INT	Disruptive behavior; engagement	--	1	9 males 14 females	3 individual behavior supports	Reports from teacher	General Education (Math)
Radley et al. (2016) 2	INT	Disruptive behavior; engagement	--	1	3 males 12 females	3 individual behavior supports	Reports from teacher	General Education (Language Arts)
Radley et al. (2016) 3	INT	Disruptive behavior; engagement	--	1	14 males 4 females	4 individual behavior supports	Reports from teacher	General Education

INT = interdependent; SET = School-Wide Evaluation Tool

Section 4: Results

Design Standards

A total of 38 studies met the inclusion criteria for this review. The studies were evaluated using WWC criteria, and the corresponding data can be found in Table 1. Of the 38 studies, 7 (18%) were categorized as *Meets Standards with Reservations* and 1 (3%) was rated as *Meets Standards*. Studies failed to “meet standards” due to an insufficient number of data points to meet criteria as outlined. The remaining 30 studies were classified as *Does Not Meet Standards* as a result of insufficient measures of interobserver agreement, lack of systematic manipulation of the independent variable, failure to display three demonstrations of effect, and/or an inadequate number of data points across conditions. Twenty studies (53%) failed to display three demonstrations of effect, which was one of the most common limitations of the reviewed studies. In addition, 14 studies (37%) were unsuccessful with measuring interobserver agreement during at least 20% of sessions, with eight of the 14 studies failing to meet both standards for interobserver agreement and adequate demonstrations of effect.

Evidence Standards

Based on a visual analysis of data reported by the studies classified as *Meets Standards with Reservations*, all studies demonstrated “moderate evidence” of a functional relation between the independent and dependent variables. The studies most frequently failed to demonstrate “strong evidence” due to limitations regarding unstable data immediately prior to phase change and lack of clear evidence of a functional relation.

Participants and Settings

The descriptive characteristics of the seven studies retained for review were also examined and reported in Table 2. A total of 144 students (75 males, 69 females) across eight classrooms participated in the studies. Four studies (50%) reported age of participants, which ranged from 8-15 years. In regard to ethnicity, there were a total 113 (78%) African American, 23 (16%) Hispanic, 4 (3%) Caucasian, 2 (1%) self-reported as African American and Caucasian, and 2 (1%) non-reporting participants. Six students were reported as receiving individualized behavior supports, although no specific diagnoses were noted for any of the participants. Ten students were reported as receiving individualized behavior supports and 2 students were diagnosed with a learning disability, although no specific diagnoses were noted for any of the participants.

All participants were students in classrooms that were referred for study participation based on teacher or administrator reports of difficulty managing problem behavior. Of the eight classrooms, there were four first grade, one fourth grade, one ninth grade classroom, in addition to two classrooms with students in grades 9-12. All studies were conducted in the general education setting during group instruction; descriptions of specific activities conducted during instruction were not provided in detail.

Dependent Variables

The dependent variables differed across the studies, although disruptive behavior (88% of studies) and “on-task” behavior or engagement (50% of studies) were most commonly measured. Disruptive behaviors included noise level (Radley, Dart, & O’Handley, 2016), being out of seat without permission (Mitchell, Tingstrom, Dufrene, Ford, & Sterling, 2015; Radley et al., 2016), engaging in physical aggression toward

others or property (Kleinman & Saigh, 2011; Lambert, Tingstrom, Sterling, Dufrene, & Lynne, 2015), and engaging in vocalizations that were unrelated to the classroom activity (Kleinman & Saigh, 2011; Lambert et al., 2015; Ling, Hawkins, & Weber, 2011). All studies provided observable and measurable definitions of target behaviors and provided examples that were directly related to the participants and settings.

To measure task engagement and disruptive behaviors, five studies used momentary time sampling at 10-s intervals (Lambert et al., 2015; Ling et al., 2011; Radley et al., 2016), two studies used partial interval 10-s (Mitchell et al., 2015), and one study employed partial interval recording at 30-s intervals (Kleinman & Saigh, 2011). One study employed changing criterion design and seven studies utilized ABAB withdrawal designs.

Independent Variables

All studies utilized an interdependent group contingency, although the framework differed. Four interventions involved earning smiley faces for appropriate behavior (Ling et al., 2012; Radley et al., 2016), three examined the Good Behavior Game (Kleinman & Saigh, 2011; Mitchell et al., 2015), and one focused on “tootling” (Lambert et al., 2015). All interventions involved a teacher-selected criterion for access to reinforcers. Selection of reinforcers was based on participant preference in three classrooms, parent suggestions for one classroom, and teacher-selected in four classrooms. Upon meeting criterion, two groups received edibles (e.g., cupcakes, pizza, chips, bite-sized candy), four received tangibles (i.e., stickers, beanie baby), and two groups received a choice from a provided menu of edibles, tangibles, and activities.

Student Outcomes

Kleinman and Saigh (2011) and Mitchell et al. (2015) implemented the Good Behavior Game with a classroom divided into two groups and a script provided for the teacher to announce the game as a team competition. As traditionally outlined, students who engaged in target deceleration behaviors were addressed in front of their peers and earned a point for their group. Reinforcers were delivered to the groups who had earned the least amount of points daily, and Kleinman and Saigh (2011) also incorporated a weekly reinforcer of higher magnitude than those received daily. Kleinman and Saigh (2011) found a decrease in disruptive behavior, particularly talking out of turn, which decreased by as much as 58% during intervention. Mitchell et al. (2015) found similar findings with a decrease in disruptive behavior of reportedly large effects based on Nonoverlap of All Pairs (NAP) scores of 1.00 for effect size.

Lambert et al. (2015) examined the effects of “tootling,” defined as a method involving students monitoring and reporting the prosocial behaviors of their peers, on the disruptive behavior of a fourth-grade classroom. Based on momentary time sampling data, it was found that students engaged in disruptive behavior during 27.3% of intervals in baseline, which decreased to 7.4% of intervals with the first introduction of the intervention. Levels of responding were similar across similar phases when the intervention was withdrawn and reintroduced. Overall, the authors reported a decrease in disruptive behavior and increase in appropriate behavior of moderate to strong effect size.

Ling et al. (2011) measured the effects of a group contingency related to earning smiley faces based on task engagement (i.e., orienting toward teacher or activity, responding to instructions) and off-task behaviors (i.e., motor activities unrelated to the

task, talking out of turn.) The classroom teacher provided an opportunity for the group to earn at least 3 smiley faces during their morning “carpet” instruction and provided vocal praise or a vocal reminder of appropriate behavior on an unspecified variable ratio schedule based on student behavior at the end of the interval. The largest impact was on off-task behaviors, which occurred during an average of 58% of intervals at baseline and decreased to an average of 25% of intervals during intervention.

Radley et al. (2016) also measured task engagement and disruptive behaviors, although the independent and dependent variables differed slightly. The independent variable utilized by Radley et al. (2016) rated student behavior with a smiley face or frowning face, and disruptive behaviors were not measured separately, but rather as a group of multiple topographies that included talking out of turn and being out-of-seat without permission. Radley et al. (2016) reported baseline levels of disruptive behavior ranging from an average of 23%-39.6% that decreased to an average range of 5.5%-13.5% following introduction of the intervention.

Determination of Evidence-Based Practice

The eight studies that demonstrated “moderate evidence” for the use of group contingencies for managing problem behavior were evaluated using the criteria for rating an evidence-based practice recommended by WWC. The criteria are as follows: (a) a minimum of five studies classified as *Meets Standards* or *Meets Standards with Reservations*, (b) the studies included at least 20 participants, and (c) the strategy was examined by at least three different researchers in three different geographical locations. Based on these criteria, interdependent group contingencies were utilized examined across three different locations (i.e., Midwestern United States, New York City, and

southeastern United States) and included a total of 144 students. Eight studies met standards for design and evidence, prohibiting group contingencies to be labeled as an evidence-based practice based on this review.

Section 5: Discussion

The purpose of this review was to expand upon the work of Maggin et al., (2012) and determine whether there is sufficient empirical support in recent literature to classify group contingency interventions as an evidence-based practice for managing disruptive behaviors in classrooms. Consistent with the findings of Maggin et al. (2012), this review found sufficient support for group contingencies to be recommended as evidence-based practice under the WWC guidelines, in addition to aligning with previous findings based on other components.

Similar to the findings of Maggin et al. (2012), the majority of reviewed studies were categorized as interdependent group contingencies. Additionally, the interventions appeared to be employed most frequently in general education classrooms of schools with low achievement and/or low socioeconomic status. In contrast, Maggin et al. (2012) found that group contingencies appeared to be used most often with Caucasian males in late elementary, while the evaluation of demographic characteristics in this review yielded a ratio of 52% males to 48% females, with only two participants reported as Caucasian. The eight studies retained for review focused heavily on minority populations in schools with a high ratio of students who receive free or reduced lunch. Radley et al. (2016) reported that the Good Behavior Game (GBG) has frequently been studied with students demonstrating low achievement and problem behavior. Given that GBG is a common interdependent group contingency, it is likely to have guided much of the research regarding group contingencies and disproportionately contributed to the evidence base of these interventions.

Although the findings of this review found similar results to those of Maggin et al. (2012), it is important to note that if combined with their review of the preceding 30 years of research, the results would continue to demonstrate sufficient evidence to recommend group contingencies as an evidence-based practice. The external validity would be strengthened by eight additional replications of effect in various geographical regions with different populations of individuals. Despite this contribution, the current review also examined whether research in recent years added to the literature by expanding on previously identified weaknesses, such as lack of detail regarding setting and selection criteria. It was found that many studies continued to display similar shortcomings.

Many of the reviewed studies demonstrated an absence of rigorous design quality. Studies failed to measure interobserver agreement across a sufficient number of sessions, did not display at least three demonstrations of effect, lacked an adequate number of data, and/or failed to demonstrate stable data in each phase. Furthermore, the research lacked fidelity measures and detailed descriptions of screening procedures, prerequisites, student achievement level, and the activities during which the interventions were employed. Maggin et al. (2012) also noted missing information that limited replication, which may indicate that this is an overall weakness of research with these types of interventions. Due to the focus on whole-group behavior in an ever-changing classroom setting, researchers may experience difficulty collecting such a high degree of information for each student and controlling for deviations to activities and procedures. For example, Kleinman and Saigh (2011) noted a change in teacher mid-year during their study, and researchers

reported difficulty with delivery of reinforcers as planned due to frequent deviations in classroom schedules and teacher preference.

Limitations

One limitation is related to the parameters of this review. Rather than overlapping with the review of Maggin et al. (2012), only studies conducted within the last five years were examined to determine whether research evolved to meet quality indicators as emphasis on evidence-based practice has increased. Additionally, the inclusion criteria for this review were narrowed so as not to include studies completed in fulfillment of credit for thesis or dissertation. Maggin et al. (2012) included a total of 5 (19%) theses or dissertation studies in their review, although indicators for determining sufficient rigor of a strategy specifically note that studies should be published in peer-reviewed journals (Horner et al., 2005). A direct replication of the procedures outlined by Maggin et al. (2012) was also restricted by to a lack of access to the digitizing software used for quantitative analyses.

Implications for Future Research

One limitation reported by Maggin et al. (2012) was the variability of procedures across reported interventions, which was also evident in the current review. For instance, although the three studies that *Met Standards with Reservations* implemented interdependent group contingencies, the criterion, reinforcers, and components of the interventions varied between the studies. Furthermore, many of the interventions involve multiple components, such as self-management, behavior-specific praise, and a more structured schedule of reinforcement that increases the density of reinforcement from baseline conditions. This hinders the ability to group the interventions together in order

to confidently draw conclusions regarding which interventions are most effective and for which populations in specific settings.

Data from the current review indicated that multiple studies lacked adequate experimental rigor, especially an insufficient number of data points and opportunities to demonstrate a basic effect, as well as detailed description (e.g., prerequisites, settings). As reported in this review, research continues to lack sufficient data (e.g., demonstrations of effect, number of data points per phase) and detailed description (e.g., prerequisites, settings). Thus, limiting recommendations that can be made to practitioners regarding group contingency interventions for managing problem behavior with specific populations of individuals and/or individuals within specific settings. Future research should adhere to quality standards of WWC to increase the experimental rigor of studies within this area.

It would also be beneficial for further research to consider specific interventions and their effects. As previously noted, there has been extensive research based on the GBG as a group contingency intervention, although other presented group contingency interventions have varying procedures and modifications. For example, Kleinman and Saigh (2011) adapted the GBG to focus on desirable behaviors rather than employing traditional procedures and drawing attention to undesirable behaviors. Additionally, many of the interventions contain multiple components founded in behavior principles that create more of a “packaged” intervention, rather than one particular independent variable. There appears to be little replication of specific interventions outside of the GBG, which will be required to even begin to disentangle and identify the most effective elements for different populations

Conclusion

Educators have both a legal and ethical obligation to employ evidence-based practices and foster positive outcomes for their students. Although previous studies have demonstrated positive effects with group contingencies and determined sufficient evidence to recommend the strategy as an evidence-based practice, the findings of this review suggest less conclusive results. This review focused on research conducted in within the last 5 years and did not find a sufficient number of rigorous studies to suggest that group contingency interventions employed in classrooms are an empirically supported practice. Given the ever-changing classroom dynamics, this indicates that the current classroom composition may be less responsive to previously successful interventions and highlights the need to continually evaluate their effects.

References

*Indicates articles included in review.

**Indicates articles retained for further analysis.

Allday, R. A., Neilsen-Gatti, S., & Hudson, T. M. (2013). Preparation for inclusion in teacher education pre-service curricula. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, *36*, 298-311.

*Chafouleas, S. M., Hagermoser Sanetti, L. M., Jaffery, R., & Fallon, L. M. (2012). An evaluation of classwide intervention package involving self-management and a group contingency on classroom behavior of middle school students. *Journal of Behavioral Education*, *21*, 34-57.

Clunies-Ross, P., Little, E., & Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behavior, *Educational Psychology*, *28*, 693-710.

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

*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.

*Denune, H., Hawkins, R., Donovan, L., McCoy, D., Hall, L., & Moeder, A. (2015). Combining self-monitoring and an interdependent group contingency to improve the behavior of sixth graders with EBD. *Psychology in the Schools*, *52*, 562-577.

- *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.
- *Donaldson, J. M., Wiskow, K. M., & Soto, P. L. (2015). Immediate and distal effects of the good behavior game. *Journal of Applied Behavior Analysis, 48*, 685-689.
- Every Student Succeeds Act of 2015, 114 U. S. C. of 2015. P. L. 114 -95.
- Friedman, I. A. (2000). Burnout in teachers: shattered dreams of impeccable professional performance. *Journal of Clinical Psychology, 56*, 595-606.
- Gast, D. L., & Ledford, J. R. (2014). *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed.). New York, NY: Routledge.
- Harrower, J., & Dunlap, G. (2001). Including children with autism in general education classrooms: A review of effective strategies. *Behavior Modification, 25*, 762-785.
- Hooks, M. (2015). Pinterest: A tool for lesson planning. *Mathematics Teacher, 108*, 466-468.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165-179.
- Horner, R. H., & Kratochwill, T. R., (2012). Synthesizing single-case research to identify evidence-based practices: Some brief reflections. *Journal of Behavioral Education, 21*, 266-272.
- Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial

- assessing school-wide positive behavior support in elementary schools. *Journal on Positive Behavior Interventions*, *11*, 133-144.
- Hulac, D. M., & Benson, N. (2010) The use of group contingencies for preventing and managing disruptive behaviors. *Intervention in School and Clinic*, *45*, 257-262.
- Individuals with Disabilities Education Improvement Act (2004). Public Law 108-446 (20 U.S.C. 1400 *et seq.*).
- Every Student Succeeds Act of 2015, 114 U. S. C. of 2015. P. L. 114 -95.
- *Kamps, D., Wills, H. P., Heitzman-Powell, L., Laylin, J., Szoke, C., Petrillo, T., & Culey, A. (2011). Class-wide function-related intervention teams: Effects of group contingency programs in urban classrooms. *Journal of Positive Behavior Interventions*, *13*, 154-167.
- *Kamps, D., Conklin, C., & Wills, H. (2015). Use of self-management with the CW-FIT group contingency program. *Education and Treatment of Children*, *38*, 1-32.
- **Kleinman, K. E., & Saigh, P. A. (2011). The effects of the hood behavior game on the conduct of regular education New York City high school students. *Behavior Modification*, *35*, 95-105.
- Kokkinos, C. M. (2007) Job stressors, personality and burnout in primary school teachers. *British Journal of Psychology*, *77*, 229-243.
- *Kowalewicz, E. V. & Coffee, G. (2014). Mystery motivator: A tier 1 classroom behavioral intervention. *School Psychology Quarterly*, *29*, 138-156.
- 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*, 26-38

- Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15*, 124-144.
- Kucucker, S., & Tekinarslan, I. C. (2015). Comparison of the self-concepts, social skills, problem behaviors, and loneliness of students with special needs in inclusive classrooms. *Educational Sciences: Theory & Practice, 15*, 1559-1573.
- **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.
- Maggin, D. M., Johnson, A. H., Chafouleas, S. M., Ruberto, L. M., & Berggren, M. (2012). A systematic review of school-based group contingency interventions for students with challenging behavior. *Journal of School Psychology, 50*, 625-654.
- **Mitchell, R. R., Tingstrom, D. H., Dufrene, B. A., & Ford, W. B. (2015). The effects of the good behavior agem with general-education high school students. *School Psychology Review, 44*, 191-207.
- Oliver, R. M., & Reschly, D. J. (2010). Special education teacher preparation in classroom management: Implications for students with emotional and behavioral disorders. *Behavioral Disorders, 35*, 189-199.
- Pinkelman, S. E., McIntosh, K., Rasplica, C. K., Berg, T., & Strickland-Cohen, M. K. (2015). Perceived enablers and barriers related to sustainability of school-wide

positive behavioral interventions and supports. *Behavioral Disorders*, 40, 171-183.

**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, 93-108.

Sayal, K., Washbrook, E., & Propper, C. (2015). Childhood behavior problems and academic outcomes in adolescence: Longitudinal population-based study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54, 360-368.

Stough, L. M., Montague, M. L., Landmark, L. J., & Williams-Diehm, K. (2015). Persistent classroom management training needs of experienced teachers. *Journal of the Scholarship of Teaching and Learning*, 15, 36-48.

Sugai, G. & Simonsen, B. (2012). *Positive behavior interventions and supports: History, defining features, and misconceptions*. Center for PBIS & Center for Positive Behavior Interventions and Supports.

Tillery, A. D., Varjas, K., Meyers, J., & Collins, A. S. (2010). General education teacher's perceptions of behavior management and intervention strategies. *Journal of Positive Behavior Interventions*, 12, 86-102.

Troutner, J. (2012). Cool tools, social media, and curriculum. *Teacher Librarian*, 9(4), 48-50. U.S. Census Bureau (2015). Regions and Divisions. Retrieved March 21, 2016 from http://www.census.gov/econ/census/help/geography/regions_and_divisions.html.

- U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, September). *What Works Clearinghouse: Procedures and Standards Handbook (Version 3.0)*. Retrieved from <http://whatworks.ed.gov>.
- White, M. & Mason, C. Y. (2006). Components of a successful mentoring program for beginning special education teachers: Perspectives from new teachers and mentors. *Teacher Education and Special Education, 29*, 191-201
- *Wills, H. P., Iwaszuk, W. M., Kamps, D., & Shumate, E. (2014). CW-FIT: Group contingency effects across the day. *Education and Treatment of Children, 37*, 191-210.
- Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., & Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders, 45*, 1951-1966.

Vita

Danielle M. Crawford Duff

University of Kentucky 2006-2010
Bachelor of Science in Psychology