

**Campbell Systematic Reviews**

2011:4

First published: 24 June, 2011

Last updated: June, 2011

# **Teacher classroom management practices: effects on disruptive or aggressive student behavior**

Regina M. Oliver, Joseph H. Wehby, Daniel J. Reschly



**THE CAMPBELL COLLABORATION**

# Colophon

---

<b>Title</b>	Teacher classroom management practices: Effects on disruptive or aggressive student behavior
<b>Institution</b>	The Campbell Collaboration
<b>Authors</b>	Oliver, Regina M. Wehby, Joseph H. Reschly, Daniel J.
<b>DOI</b>	10.4073/csr.2011.4
<b>No. of pages</b>	55
<b>Last updated</b>	June, 2011

---

<b>Citation</b>	Oliver R, Wehby J, Daniel J. Teacher classroom management practices: Effects on disruptive or aggressive student behavior. Campbell Systematic Reviews 2011.4 DOI: 10.4073/csr.2011.4
<b>Copyright</b>	© Oliver et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

---

## Keywords

---

<b>Contributions</b>	Regina M. Oliver, Joseph H. Wehby, and Daniel J. Reschly contributed to the writing and revising of this protocol. Statistical support was provided by Mark Lipsey, Director of the Peabody Research Institute at Vanderbilt University. Regina M. Oliver will be responsible for updating this review as additional evidence accumulates and as funding becomes available.
<b>Support/Funding</b>	Preparation of this manuscript was funded in part by the National Comprehensive Center for Teacher Quality (S283B050051). The opinions expressed in this article are those of the authors and do not necessarily reflect those of the funding agency.
<b>Potential Conflicts of Interest</b>	The authors have no vested interest in the outcomes of this review, nor any incentive to represent findings in a biased manner.

---

<b>Corresponding author</b>	Regina M. Oliver Peabody College, Department of Special Education Vanderbilt University 230 Appleton Place Nashville, Tennessee 37203-5721 USA E-mail: regina.m.oliver@vanderbilt.edu
-----------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

# Campbell Systematic Reviews

---

**Editors-in-Chief** Mark W. Lipsey, Vanderbilt University, USA  
Arild Bjørndal, Centre for Child and Adolescent Mental Health, Eastern and Southern Norway & Institute of Health and Society, University of Oslo, Norway

## **Editors**

*Crime and Justice* David B. Wilson, George Mason University, USA

*Education* Sandra Jo Wilson, Peabody Research Institute, Vanderbilt University, USA

*Social Welfare* William Turner, University of Bristol, UK  
Geraldine Macdonald, Queen's University, UK & Cochrane Developmental, Psychosocial and Learning Problems Group

**Managing Editor** Karianne Thune Hammerstrøm, The Campbell Collaboration

## **Editorial Board**

*Crime and Justice* David Weisburd, Hebrew University, Israel & George Mason University, USA  
Peter Grabosky, Australian National University, Australia

*Education* Paul Connolly, Queen's University, UK  
Gary W. Ritter, University of Arkansas, USA

*Social Welfare* Aron Shlonsky, University of Toronto, Canada  
Paul Montgomery, University of Oxford, UK

*Methods* Therese Pigott, Loyola University, USA  
Peter Tugwell, University of Ottawa, Canada

---

The Campbell Collaboration (C2) was founded on the principle that systematic reviews on the effects of interventions will inform and help improve policy and services. C2 offers editorial and methodological support to review authors throughout the process of producing a systematic review. A number of C2's editors, librarians, methodologists and external peer-reviewers contribute.

The Campbell Collaboration  
P.O. Box 7004 St. Olavs plass  
0130 Oslo, Norway  
[www.campbellcollaboration.org](http://www.campbellcollaboration.org)

---

# Table of contents

<b>TABLE OF CONTENTS</b>	<b>3</b>
<b>EXECUTIVE SUMMARY/ABSTRACT</b>	<b>4</b>
<b>1 BACKGROUND</b>	<b>6</b>
1.1 Classroom Management	7
1.2 Prior Research	8
1.3 Summary	14
<b>2 OBJECTIVES</b>	<b>15</b>
<b>3 METHODS</b>	<b>16</b>
3.1 Criteria for Inclusion and Exclusion of Studies in the Review	16
3.2 Search Strategy for Identification of Studies	17
3.3 Data Collection and Analysis	18
<b>4 RESULTS</b>	<b>23</b>
4.1 Description of Eligible Studies	23
4.2 Main Effects of Teachers' Classroom Management Practices	25
<b>5 DISCUSSION</b>	<b>35</b>
5.1 Limitations of the Review	37
<b>REFERENCES</b>	<b>39</b>
<b>APPENDIX A: CHARACTERISTICS OF INCLUDED STUDIES</b>	<b>44</b>
<b>APPENDIX B: CHARACTERISTICS OF EXCLUDED STUDIES</b>	<b>47</b>
<b>APPENDIX C: CODING MANUAL</b>	<b>51</b>

---

## Executive summary/Abstract

Disruptive behavior in schools has been a source of concern for school systems for several years. Indeed, the single most common request for assistance from teachers is related to behavior and classroom management (Rose & Gallup, 2005).

Classrooms with frequent disruptive behaviors have less academic engaged time, and the students in disruptive classrooms tend to have lower grades and do poorer on standardized tests (Shinn, Ramsey, Walker, Stieber, & O'Neill, 1987).

Furthermore, attempts to control disruptive behaviors cost considerable teacher time at the expense of academic instruction.

Effective classroom management focuses on preventive rather than reactive procedures and establishes a positive classroom environment in which the teacher focuses on students who behave appropriately (Lewis & Sugai, 1999). Rules and routines are powerful preventative components to classroom organization and management plans because they establish the behavioral context of the classroom by specifying what is expected, what will be reinforced, and what will be retaught if inappropriate behavior occurs (Colvin, Kame'enui, & Sugai, 1993). This prevents problem behavior by giving students specific, appropriate behaviors to engage in. Monitoring student behavior allows the teacher to acknowledge students who are engaging in appropriate behavior and prevent misbehavior from escalating (Colvin et al., 1993).

Research on classroom management has typically focused on the identification of individual practices that have some level of evidence to support their adoption within classrooms. These practices are then combined under the assumption that, if individual practices are effective, combining these practices into a package will be equally, if not more, effective. Textbooks are written and policies and guidelines are disseminated to school personnel based on these assumptions. Without research that examines classroom management as an efficient package of effective practices, a significant gap in our current knowledge base still exists. Understanding the components that make up the most effective and efficient classroom management system as well as identifying the effects teachers and administrators can expect from implementing effective classroom management strategies represent some of these gaps. A meta-analysis of classroom management which identifies more and less effective approaches to universal, whole-class, classroom management as a set of practices is needed to provide the field with clear research-based standards.

This review examines the effects of teachers' universal classroom management practices in reducing disruptive, aggressive, and inappropriate behaviors. The specific research questions addressed are: Do teacher's universal classroom management practices reduce problem behavior in classrooms with students in kindergarten through 12<sup>th</sup> grade? What components make up the most effective and efficient classroom management programs? Do differences in effectiveness exist between grade levels? Do differences in classroom management components exist between grade levels? Does treatment fidelity affect the outcomes observed? These questions were addressed through a systematic review of the classroom management literature and a meta-analysis of the effects of classroom management on disruptive or aggressive student behavior.

Twelve studies of universal classroom management programs were included in the review. The classroom-level mean effect size for the 12 programs was positive and statistically significant ( $d=.80$  with an  $ICC=.05$ ;  $d=.71$  with an  $ICC=.10$ ;  $p<.05$ ). Note that cluster adjustments were required due to differences in reporting measures between classroom level outcomes and individual student level outcomes. The resulting effect sizes index classroom-level differences and cannot be compared to the typical student-level effect sizes commonly reported in the literature. Due to a lack of power to detect heterogeneity and lack of information reported in the studies reviewed, only the first research question could be addressed.

Teacher's classroom management practices have a significant, positive effect on decreasing problem behavior in the classroom. Students in the treatment classrooms in all 12 studies located for the review showed less disruptive, inappropriate, and aggressive behavior in the classroom compared to untreated students in the control classrooms. The overall mean classroom effect size of either  $.80$  or  $.71$  indicates a positive effect that significantly impacts the classroom environment. To put our classroom-level mean effect sizes into a comparable format with the more typical effect sizes, we back-transformed our mean effect sizes using the original adjustment formulas (Hedges, 2007). Thus, the classroom-level mean effect sizes of  $.80$  and  $.71$  are roughly comparable to student level effect sizes of  $.18$  and  $.22$  for  $ICC=.05$  and  $ICC=.10$ , respectively. Teachers who use effective classroom management can expect to experience improvements in student behavior and improvements that establish the context for effective instructional practices to occur.

---

# 1 Background

Disruptive behavior in schools has been a source of concern for school systems for several years. Indeed, the single most common request for assistance from teachers is related to behavior and classroom management (Rose & Gallup, 2005).

Classrooms with frequent disruptive behaviors have less academic engaged time, and the students in disruptive classrooms tend to have lower grades and do poorer on standardized tests (Shinn et al., 1987). Furthermore, attempts to control disruptive behaviors cost considerable teacher time at the expense of academic instruction.

School discipline issues such as disruptive behavior and violence also have an increased effect on teacher stress and burnout (Smith & Smith, 2006). There is a significant body of research attesting to the fact that classroom organization and behavior management competencies significantly influence the persistence of new teachers in their teaching careers (Ingersoll & Smith, 2003). New teachers typically express concerns about effective means to handle disruptive behavior (Browers & Tomic, 2000). Teachers who have significant problems with behavior management and classroom discipline often report high levels of stress and symptoms of burnout and are frequently ineffective (Berliner, 1986; Browers & Tomic, 2000; Espin & Yell, 1994). The ability of teachers to organize classrooms and manage the behavior of their students is critical to achieving both positive educational outcomes for students and teacher retention.

Effective classroom management is also related to prevention efforts. Children's behavior is shaped by the social context of the environment during the developmental process (Kauffman, 2005). Many behavioral disorders begin with or are made worse through behavioral processes such as modeling, reinforcement, extinction, and punishment (Kauffman, 2005; Patterson, Reid, & Dishion, 1992). The classroom context plays a significant role in the emergence and persistence of aggressive behavior. Early intervention and treatment for students at-risk for emotional and behavioral disorders (EBD) is essential to prevent more serious behaviors from developing (Kauffman, 2005; Greer-Chase, Rhodes, & Kellam, 2002). The progression and malleability of maladapted behavior is affected by classroom management practices of teachers in the early grades (Greer-Chase et al., 2002). For example, classrooms with high levels of disruptive or aggressive behavior place children at-risk for more serious behavior problems and EBD. Research has

indicated that aggressive students in aggressive or disruptive classroom environments are more likely to be aggressive in later grades (Greer-Chase et al., 2002). Research-based approaches to classroom management are necessary to improve both academic and behavioral outcomes for students.

---

## **1.1 CLASSROOM MANAGEMENT**

---

A primary problem with determining research-based approaches to classroom management is establishing a definition. Classroom management has been defined broadly as any action a teacher takes to create an environment that supports and facilitates both academic and social-emotional learning (Evertson & Weinstein, 2006). Instructional procedures could also be considered classroom management by this definition; however, effective instruction alone is insufficient for establishing universal classroom management. Procedures that structure the classroom environment, encourage appropriate behavior, and reduce the occurrence of inappropriate behavior are necessary for strong classroom management (Evertson, Emmer, Sanford, & Clements, 1983). Instructional procedures, although equally important to the classroom environment, can be considered a separate set of procedures.

The components of effective classroom management are important in several ways. For example, focusing on preventive rather than reactive procedures establishes a positive classroom environment in which the teacher focuses on students who appropriately behave (Lewis & Sugai, 1999). Rules and routines are powerful preventative components to classroom organization and management plans because they establish a behavioral context for the classroom that includes what is expected, what will be reinforced, and what will be retaught if inappropriate behavior occurs (Colvin et al., 1993). This prevents problem behavior by giving students specific, appropriate behaviors to engage in. Monitoring student behavior allows the teacher to acknowledge students who are engaging in appropriate behavior and prevent misbehavior from escalating (Colvin et al., 1993).

One example of a whole-class classroom management approach is Classroom Organization and Management Program (Evertson et al., 1988). COMP is a professional development series developed by Carolyn Evertson and colleagues (1988) designed to create effective learning environments. The main components of COMP are: (1) organizing the classroom; (2) planning and teaching rules and procedures; (3) managing student work and improving student accountability; (4) maintaining good student behavior; (5) planning and organizing; (6) conducting instruction and maintaining momentum; and (7) getting the year off to a good start.

For the purpose of this review, universal or whole-class classroom management is defined only as: a collection of non-instructional classroom procedures implemented by teachers in classroom settings with all students for the purposes of teaching

prosocial behavior as well as preventing and reducing inappropriate behavior. This definition includes packaged interventions (e.g., COMP; Evertson et al., 1988) with multiple components (e.g., rules, classroom procedures, reinforcement, consequences) used as a comprehensive approach to universal classroom management. It also includes group contingencies such as the “Good Behavior Game” when used as a universal approach for classroom management. It does not include packaged social skills curricula used in isolation, as these are seen as a separate category already reviewed in the literature (e.g., Gresham, 1996).

---

## **1.2 PRIOR RESEARCH**

---

Although William Chandler Bagley wrote what may have been the first book on classroom management in 1907, systematic research on the topic did not begin until the 1950s (Brophy, 2006). The early research addressed teachers’ attitudes and concerns about classroom control. Studies were prominent in the 1950s and 1960s describing the leadership styles of teachers that were considered “better” classroom managers (e.g., Kounin, 1970; Ryans, 1952). With the influence of behavioral research in education came more specific behavioral methods (e.g., reinforcement and punishment) applied to classroom management (e.g., Hall, Panyan, Rabon, & Broden, 1968; Strain, Lambert, Kerr, Stagg, & Lenkner, 1983). Researchers also began identifying specific teacher behaviors and student-teacher interactions that promoted appropriate behavior and reduced inappropriate behavior (e.g., Anderson, Evertson, & Emmer, 1979; Shores, Jack, Gunter, Ellis, DeBriere, & Wehby, 1993).

Extensive theoretical and research bases exist for classroom management practices. In general, classroom management practices historically have been identified by observing effective teachers’ behavior, or combining behavioral approaches that have been established through research on effective behavior change procedures. Prior research falls into two broad categories: (1) observation studies used to identify how effective teachers organize and manage their classrooms; (2) experimental studies examining components of classroom management in isolation or in various combinations.

In studies of classroom management, typical behaviors that are targeted for intervention are disruptive, aggressive behaviors. Examples of these types of behaviors include noncompliance, verbal disruption, teasing others, being out of one’s seat, taking others’ property, damaging property, or attacking others; these are typically measured with observations or teacher reports (Kellam, Ling, Merisca, Brown, & Ialongo, 1998). Reductions in these types of individual student behaviors also reduce the overall classroom level of aggression. Identifying changes in student behavior is important for determining the effects of classroom management procedures. A review of this literature follows.

### 1.2.1 Observation Studies

Much of the early research on classroom management began with classroom observations to establish teacher behaviors that were observed in teachers considered highly effective. Effective teachers in these studies were defined as those who produced greater learning gains in their students or had classrooms with lower rates of disruptive student behavior and more on-task behavior (Anderson et al., 1979). By collecting narrative descriptions of effective teacher behavior, researchers were interested in identifying practices and behaviors across teachers that allowed them to make recommendations for effective classroom management.

Studies by Kounin reported in his oft-cited book (1970) were some of the earliest attempts to identify these practices. Kounin compared teachers' managerial behaviors in smoothly functioning classrooms with teachers from classrooms that had high rates of inattention and frequent disruptions. Based on observations of videotapes of teachers in both types of classrooms, Kounin identified a set of teacher behaviors that corresponded to highly managed classrooms. According to Kounin, effective classroom managers were aware of student behaviors and activities at all times in order to prevent small issues from escalating, a trait he termed "withitness" (p. 74). Effective classroom managers were also able to overlap more than one classroom task at a time in order to monitor student behavior and structure classroom activities that maintained high rates of student attention. These preventive strategies were not observed in classrooms with high disruptions and low student attention. Moreover, effective and ineffective teachers did not differ in how they responded to student misbehavior. The difference that set the effective classroom managers apart from the less effective classroom managers was in the preventive, organizational strategies used by the effective teachers. Kounin's work was the impetus for influential observational studies examining teachers' managerial practices during the 1970s and early 1980s.

Much of the discussion of effective classroom managers has been based on one year-long study by Anderson, Evertson, and Emmer conducted in the late 1970s. Researchers collected extensive narrative recordings of teacher behavior in 28 third grade classrooms over the course of an entire school year and analyzed trends in management styles of effective teachers (Anderson et al., 1979). In a preliminary report from this project (Anderson & Evertson, 1978), researchers identified one effective and one ineffective teacher based on student gains at the end of the school year. They then retrospectively compared those teachers' management practices from the beginning of the school year and found large differences in teacher behaviors between the effective teacher and the ineffective teacher. The effective teacher had better classroom management. On the first day of school, the better classroom manager had clear expectations about behavior and communicated them to students effectively. Classroom rules and routines were explicitly taught to students using examples and non-examples and students were acknowledged for appropriate behavior using behavior-specific praise. Likewise, the effective

classroom manager provided quick, prompt responses to inappropriate behavior before the behaviors escalated. The teacher was consistent with consequences to both appropriate and inappropriate behavior. Additionally, the better organized classroom teacher monitored student behavior and remained sensitive to the students' concerns and needs for information. The anecdotal information provided in this study supplied teachers with specific examples of what an effective classroom manager does and what poor classroom management looks like, specifically at the start of the year.

In the final analysis, Anderson, Evertson, and Emmer (1979) reported the results from the entire school year. Researchers found additional support for their initial findings from the preliminary analysis. The seven most effective teachers in the sample were compared with the seven least effective teachers. Again, teachers were considered effective based on the academic progress of students in their class. As was previously reported, the most effective teachers did not assume students would know the expectations of the classroom (Anderson et al., 1979). These teachers took an instructional approach to behavior and spent time teaching important discriminations between expected and unacceptable behavior. Additionally, effective teachers applied preventive procedures such as re-teaching the rules and routines of the classroom if there was a change to the typical routine or after long breaks such as Christmas break (Anderson et al. 1979). Transitions between activities were smooth and there were low levels of disruptive student behavior. Finally, the year-long analysis of observations in effective teachers' classrooms further supported the use of monitoring student behavior, consistent consequences, and behavior specific praise.

Although these studies provided a rich description of classroom management as evidenced by behaviors of effective classroom teachers, these data were correlational and therefore could not definitively determine if the differences in teacher behavior were responsible for student academic progress. Differences in teacher behavior were observed, but changes in student behavior were typically not documented due to the nature of the research. As more descriptive studies of classroom behavior were conducted, a series of experimental studies emerged examining the effects of teacher behavior on student behavior.

### **1.2.2 Experimental Studies**

Experimental studies have focused on a range of classroom management practices. These studies range from the manipulation of single practices of teacher behavior to broader based packages of practices including organization, structure, praise, and behavioral contingencies (e.g., Kelshaw-Levering, Sterling-Turner, Henry, & Skinner, 2000; Langland, Lewis-Palmer, & Sugai, 1998; Madsen, Becker, & Thomas, 1968). Most studies of this nature use single subject methodology to manipulate various teacher classroom management practices to establish functional relations with student behavior. Examples of these experimental studies follow.

### **1.2.2.1 Single Practice**

In one experimental study, an instructional approach to teaching rules in the classroom using lesson plans was shown to decrease inappropriate behavior (Langland et al., 1998). Teachers in the study designed lesson plans to teach classroom rules that incorporated the following: examples and non-examples of the rule; teaching examples; specific activities for students to practice the skill; and the use of precorrection, reminders, and praise after the lesson to facilitate fluency and generalization (Langland et al., 1998). Decreases in inappropriate behavior occurred when teachers taught classroom rules. Other studies have also been conducted examining single practices of classroom management such as classroom rules (Rosenberg 1986); structured classroom environments (Ahrentzen & Evans, 1984; Colvin, 2002); and reinforcement, praise, and consequences (Becker, Madsen, & Arnold, 1967; Conyers et al., 2004; Sutherland, Wehby, & Copeland, 2000).

### **1.2.2.2 Packages of Teacher Practices**

In addition, individual practices have been combined and determined by various researchers to be effective classroom management procedures. Packaged interventions using antecedent strategies (e.g., posting of rules, teacher movement, precision requests), reinforcement strategies (e.g., token economy, mystery motivator), and punishment strategies to respond to inappropriate behavior (e.g., response cost) have been used effectively to reduce disruptive behavior (Di Martini-Scully, Bray, & Kehle, 2000; Kehle, Bray, Theodore, Jenson, & Clark, 2000). This same classroom management package of strategies developed by Di Martini and others also has been used to decrease disruptive behavior for students with emotional and behavior disorders (Musser, Bray, Kehle, & Jenson, 2001).

Group contingencies, another example of a package of teacher practices used for class-wide behavior management, are well documented in research (e.g., Barrish, Saunders, & Wolf, 1969; Crouch, Gresham, & Wright, 1985; Darveaux, 1984; Fishbein & Wasik, 1981; Harris & Sherman, 1973; Kelshaw-Levering et al., 2000; Litow & Pomroy, 1975; Theodore, Bray, Kehle, & Jenson, 2001). Group contingency interventions apply contingent reinforcement to groups of students based on the behavior of one or more members of the group (Litow & Pomroy, 1975). A meta-analysis by Stage and Quiroz (1997) found group contingencies to have the greatest effect on reducing inappropriate behavior compared with other behavioral strategies examined in the review. The reason that group contingencies have been used for universal classroom management is due to the components of group contingencies which mirror important classroom management procedures. Rules are explicitly stated, and reinforcement and consequences delivered in the classroom are based on student behavior.

The most researched group contingency program is “The Good Behavior Game” based on the original study by Barrish et al. , (1969). In the original study, the authors implemented group consequences contingent on individual disruptive

behavior in the classroom through the use of a game. The game was easy to implement and did not require individualized plans. Winning the game was contingent on the behavior of each member of the team. Rules for the game were outlined ahead of time as were the rewards or reinforcement for winning. The teacher placed a mark on the board for any observed rule infraction from any team member (e.g., out-of-seat, talking-out). Any team with five or fewer marks won the game and the privilege of 30 extra minutes of free time at the end of the day. If a team did not win, they continued working during those 30 minutes. It was possible for both teams to win the game providing they met the established criteria. The researchers applied this approach of group contingencies for individual behavior in math period and then in reading. Results indicated a decrease in disruptive classroom behaviors by 84.3% over all baseline and intervention phases (Barrish et al., 1969).

### **1.2.3 A Systematic Review of Classroom Management Practices**

Recently a systematic best evidence review was conducted to identify evidence-based practices in classroom management to inform research and practice (Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). These researchers initially reviewed ten classroom management texts to identify typical practices described within texts and then systematically searched the research literature to identify experimental studies that examined these practices. The researchers used criteria for “evidence-based” similar to the What Works Clearinghouse criteria to evaluate the evidence of each practice (Simonsen et al., 2008). Results of the evaluation of 81 studies identified 20 general practices that met the criteria for evidence-based. These 20 general practices fell into five broad categories: (1) maximize structure and predictability; (2) post, teach, review, and provide feedback on expectations; (3) actively engage students in observable ways; (4) use a continuum of strategies to acknowledge appropriate behavior; and (5) use a continuum of strategies to respond to inappropriate behavior (Simonsen et al., 2008). A range of two to six practices were classified under each broad category and the empirical studies supporting each practice ranged from three to eight studies per practice. *Responding to inappropriate behavior* had the highest number of empirical studies while *maximizing structure and predictability* had the fewest (Simonsen et al., 2008). The results of this review were an important first step in identifying the evidence base for specific practices typically used in classroom management approaches.

Although that review was more closely aligned to classroom-based practices as opposed to school-based approaches, issues exist with how studies in the review were categorized as classroom management. The researchers identified eligible studies based on loose criteria for classroom management that included studies with as few as two students in classroom or non-classroom settings and practices including instructional management (Simonsen et al., 2008). For example, a study by Baker (1992) examined four different methods for correcting oral reading errors with one 6<sup>th</sup> grade participant. The treatment was provided in a one-on-one setting

and specifically addressed academic error correction rather than social behavior error correction. Another example is a study by White-Blackburn, Semb, and Semb (1977). In that study, the authors examined the effects of behavioral contracts on the disruptive behavior of four students in a general education classroom. This type of intervention is typically regarded as an example of a small group or secondary intervention rather than universal classroom management. The eligibility criteria used in the Simonsen review allowed inclusion of studies that did not evaluate whole-class, classroom-based management strategies. A practice used with as few as two students in a pull-out, small group setting is typically not considered universal classroom management. Moreover, because the purpose of that study was to identify evidence-based practices, an exhaustive review was not conducted by the researchers omitting potentially important studies in the review.

#### **1.2.4 Meta-Analyses of School-Based Programs**

The review by Simonsen and colleagues (2008) was an important first step in examining classroom management; however, more systematic approaches using meta-analysis are needed to determine the magnitude of the effects of classroom management. However, most prior research syntheses on interventions targeting antisocial behavior have examined school-based programs in general rather than classroom-based behavior management. A meta-analysis by D. Wilson, Gottfredson, and Najaka (2001) examined the effects of school-based prevention of crime, substance use, dropout, nonattendance, and other conduct problems. A wide variety of interventions were considered in that study, including individual counseling, behavior modification, and broader school procedures such as environmental changes or changes to instructional practice. The authors' analysis found differences in effects based on type of intervention, with cognitive-behavioral approaches showing larger effects than non-cognitive-behavioral counseling, social work, or other therapeutic interventions. Because the inclusion criteria were broad enough to cover any school-based intervention, it was beyond the scope of classroom-based, teacher implemented interventions.

S. Wilson, Lipsey, and Derzon (2003) extended the work by D. Wilson and colleagues (2001) on the effects of school-based intervention programs on aggressive behavior. S. Wilson and colleagues (2003) found similar effects for school-based prevention programs on problem behavior. However, nearly all of the studies were demonstration projects rather than routine practice programs implemented in typical school-based environments. Effective interventions need to be tested in typical school-based sites to further shore up their levels of evidence (U.S. Department of Education Institute of Education Sciences National Center for Educational Evaluation and Regional Assistance, 2003).

Another meta-analysis by S. Wilson and Lipsey (2006) examined school-based programs that targeted social information processing and found decreases in aggressive and disruptive behavior for students in treatment conditions. Based on

these reviews, school-based programs are an important part of prevention efforts; however, these reviews do not specifically address classroom management approaches. Classrooms are a primary context for prevention efforts within school systems. Students spend the majority of their day within the confines of the classroom; therefore determination of effective classroom behavioral management procedures is required.

---

### **1.3 SUMMARY**

---

Research on classroom management has typically focused on the identification of individual practices that have some level of evidence to support their adoption within classrooms. These practices are then combined with the assumption that if individual practices are effective, combining these practices into a package will be equally or more effective. Textbooks are written and policies and guidelines are disseminated to school personnel based on these assumptions. Without research that examines classroom management as an efficient package of effective practices, a significant gap in our current knowledge base still exists. Understanding what components make up the most effective and efficient classroom management system as well as identifying the effects teachers and administrators can expect from implementing effective classroom management strategies represent some of these gaps. A meta-analysis of classroom management which identifies more and less effective approaches to universal, whole-class, classroom management as a set of practices is needed to provide the field with clear research-based standards.

---

## 2 Objectives

Despite the large research base for strategies to increase appropriate behavior and prevent or decrease inappropriate behavior in the classroom, a systematic review of multi-component universal classroom management research is necessary to establish the effects of teachers' universal classroom management approaches. This review examines the effects of teachers' universal classroom management practices in reducing disruptive, aggressive, and inappropriate behaviors. The specific research questions addressed are: Do teacher's universal classroom management practices reduce problem behavior in classrooms with students in kindergarten through 12<sup>th</sup> grade? What components make up the most effective and efficient classroom management programs? Do differences in effectiveness exist between grade levels? Do differences in classroom management components exist between grade levels? Does treatment fidelity affect the outcomes observed? These questions were addressed through a systematic review and meta-analysis of the classroom management literature.

---

## 3 Methods

---

### 3.1 CRITERIA FOR INCLUSION AND EXCLUSION OF STUDIES IN THE REVIEW

---

#### 3.1.1 Interventions

Classroom management is defined as a collection of non-instructional classroom procedures implemented by teachers in classroom settings with all students for the purposes of teaching prosocial behavior and preventing and reducing inappropriate behavior. These procedures are considered universal because they are implemented with the entire class rather than with individual children or small groups requiring additional behavioral support. The classroom management practices reviewed were required to be actions performed by the classroom teacher in the context of the classroom, with the expectation that they would reduce problem behavior for the students in the classroom. Studies that delivered an intervention to the classroom teacher (e.g., teacher training in classroom management) needed to then have the teacher implement the strategies in the classroom to be included in this review. Studies involving universal school-wide strategies such as School-Wide Positive Behavior Support (Sugai & Horner, 2002) were not eligible because they did not address classrooms as the location for intervention. Additional definitional criteria included:

- 1) Interventions delivered universally to all subjects. Pull-out or small group interventions (e.g., small group social skills) were not eligible.
- 2) Interventions that began treatment outside of the classroom in a small group and then transferred it into the classroom were not eligible (e.g., guidance counsellor working with a small group of students outside of classroom and then working in students' classroom).
- 3) Additional treatment components (e.g., parent training) were allowed provided there was at least one outcome variable measuring treatment effects with students in the classroom.

#### 3.1.2 Setting and Subjects

Interventions were delivered universally to all school-aged subjects, K-12 or the equivalent formal schooling in countries with different grade structures than the U.S., in either general education or special education classrooms during school hours. Interventions in residential facilities or special schools (e.g. day treatment

facilities) were not eligible for inclusion. Studies from any country that met all other eligibility criteria were eligible.

### **3.1.3 Outcomes**

The study reported at least one outcome of problem student behavior in the context of the classroom as measured by the classroom teacher. Problem student behavior is broadly defined as any intentional behavior that is disruptive, defiant, or intended to harm or damage persons or property, and includes off task, inappropriate, disruptive or aggressive classroom behavior.

### **3.1.4 Study Design**

Studies were experimental or quasi-experimental designs with control groups. Control conditions could be “no treatment,” “treatment as usual,” or any other similar condition that served as contrast to the treatment condition and was not expected to produce change in the outcomes of interest. Conditions with academic interventions serving as the control were not included. Studies had to meet at least one of the following criteria:

- a. Participants were randomly assigned to treatment and control or comparison conditions;
- b. Participants in the treatment and control conditions were matched and the matching variables included a pretest for at least one qualifying outcome variable (see above) or the study statistically controlled for pretest differences using ANCOVA;
- c. If subjects were not randomly assigned or matched, the study needed to have both a pretest and a posttest on at least one qualifying outcome variable (see above) with sufficient statistical information to derive an effect size or to estimate group equivalence from statements of statistical significance. Posttest only non-equivalent comparisons (not randomized or matched) were not eligible.

---

## **3.2 SEARCH STRATEGY FOR IDENTIFICATION OF STUDIES**

---

Online database searches included ERIC, PsycINFO, Proquest Dissertations, and Proquest for the past 59 years (1950 until 2009). Given the vast research base for behavioral approaches in the classroom for individual students, keyword searches were purposefully narrow to restrict the search to identify whole-class classroom management practices. Keyword searches included the following terms and were restricted to studies with empirical outcomes:

- classroom management, classroom organization, classroom
- AND
- behavior, discipline
  - evaluation, experimental, outcomes, effects

Author searches were conducted in the above databases to identify potentially eligible articles by key researchers in the field of classroom management. Author searches included:

- Brophy, Jere
- Canter, Lee
- Evertson, Carolyn
- Kellam, Sheppard
- Kounin, Jacob
- van Lier, Pol

One of the above authors, Carolyn Evertson, was contacted once a list of potential studies was created. Dr. Evertson was selected based on the large number of studies obtained from her research group. Dr. Evertson was asked to review the identified studies and indicate if there were any known studies not included that might also have been eligible. No additional studies were identified by Dr. Evertson.

Prior meta-analyses on behavior management or reviews of classroom management were identified and the reference lists were searched. In addition, the citation lists for all studies identified through database searches were reviewed for potentially eligible studies. Searches of relevant websites were conducted to identify research that may not have been published in journals or indexed in the literature databases we searched. For example, the Classroom Organization and Management Program (COMP) website ([www.comp.org](http://www.comp.org)) was searched to identify potential studies. Hand searches of journals were conducted for 2008 and 2009 to identify potential studies that were not identified through other methods. Journals included were:

- *Behavior Disorders*
- *The Journal of Emotional and Behavioral Disorders*
- *The Journal of Educational Psychology*

Studies were obtained by downloading the PDF or Word document from the online journals or were photocopied from the journal source at the Peabody College, Vanderbilt University library. Other studies were obtained from a collection of early intervention research studies maintained by Mark Lipsey of the Peabody Research Institute at Vanderbilt University.

---

### **3.3 DATA COLLECTION AND ANALYSIS**

---

#### **3.3.1 Selection of Studies**

The online search produced 5,134 titles. Titles that clearly identified the report as single subject or identified other distinguishing characteristics which would exclude the study were omitted (e.g. a case study). Based on this procedure, 94 titles were retained for further screening. Abstracts for the 94 titles were reviewed to determine whether the reports should be obtained for a complete review. Abstracts that clearly did not meet the inclusion criteria were excluded. Abstracts that were either

questionable for inclusion or appeared potentially eligible were retained. Based on review of the abstracts, 18 reports were identified as potentially eligible and retrieved in their full-text form, of which five were retained for formal eligibility screening. Reports from prior meta-analyses were obtained from the Peabody Research Institute, of which five were retained for formal eligibility screening. Author searches and COMP research searches produced six additional journal articles and three research reports retained for formal eligibility screening. Hand searches produced no additional reports. A total of 18 articles or reports were thus retained for formal eligibility screening. One of the three identified research reports contained six individual studies eligible for screening bringing the study total from 19 to 24. Hereafter, the use of the word “study” will refer to independent samples used to calculate effect sizes, not necessarily a single published report or article.

All 24 studies were screened by the primary author using a detailed screening tool outlining the eligibility criteria. Each study was read completely and coded for each eligibility criterion. If a study did not pass the screening, the reason for exclusion was noted on the screening sheet. Several articles were authored by the same research group and may have contained the same sample across multiple articles; therefore, it was necessary to conduct a second stage of screening that involved cross verification of study samples to ensure independent samples. If a research study contained the same sample of participants from an earlier report already included in the selected studies, it was then excluded to ensure that studies were not counted twice. The first reported follow-up data were chosen to more closely align with the intervention duration of other studies. This decision was made because most studies were conducted over one year or less. A total of 13 studies were identified by the researcher for inclusion and 11 were excluded.

Screening reliability was conducted on 100% of the identified studies. The second screener was a doctoral student trained in meta-analysis and was blind to which studies had been included or excluded by the primary researcher. The second screener was trained in the inclusion criteria and was provided with the detailed description of inclusion criteria as well as the screening form to document whether each study passed or failed each of the criteria. The second screener also conducted a secondary screening procedure to ensure independent samples. The secondary screener identified 12 studies for inclusion and 12 studies for exclusion, producing 96% overall reliability across studies.

The single discrepancy was resolved through a discussion between the primary author and the reliability rater. The reliability screener screened out a study by Evertson and others (2000) based on concerns regarding assignment procedures and lack of evidence for procedures to control for pretest differences. Through careful discussion of the benefits of including this study compared to the methodological concerns if the study was included, both reviewers agreed that this study did not meet inclusion criteria and therefore was excluded from the analysis.

Of the remaining 11 excluded studies, five were excluded because the data were from samples reported in previous studies already selected for inclusion in this analysis; three were excluded because the treatment was not whole class or conducted in the classroom; two were excluded due to inappropriate dependent variables (e.g., child anxiety, teacher behavior); and finally, one study was excluded because the data necessary to compute an effect size were missing from the study. Detailed lists of included and excluded studies can be found in Appendices A and B.

### 3.3.2 Study Coding

Coding was conducted for each study included in the review based on a detailed coding protocol developed by the first author (see Appendix C). Effect sizes were calculated based on the available data in the study, most typically treatment and control means on post-test data with standard deviations. The standardized mean difference effect size statistic was used (Lipsey & Wilson, 2001) to code classroom management effects. In cases where treatment and control group means were not available, effect sizes were estimated based on the available data in the study (e.g., graph) using procedures described by Lipsey and Wilson (2001).

Coding reliability was performed on five randomly selected studies by a second trained coder. Point-by-point agreement was calculated on the 33 coded variables to obtain 84% overall agreement with a range of 0-100%. Discrepancies were handled through a meeting with the primary and secondary coder to obtain resolution. Problematic variables were reviewed with the second coder and determined to be an issue of definition specificity in the coding manual. Once the coding manual was revised, the problematic variables were re-coded by the second coder to reach 100% agreement.

### 3.3.3 Statistical Procedures

Standardized mean difference effect sizes were calculated on dependent variables that measured disruptive, inappropriate, or aggressive student behavior in the classroom. No studies included in the review reported prosocial student dependent variables. Due to the nature of the dependent variable, a positive outcome from treatment would mean decreases in the dependent measure, and therefore a negative effect size. When calculating the standardized mean difference effect size, the mean of the treatment group was subtracted from the mean of the control group in order to produce a positive effect size. The difference between the control group mean and treatment group mean was divided by the pooled standard deviation of the two groups, as follows:

$$ES_{sm} = \frac{\bar{X}_{g1} - \bar{X}_{g2}}{S_p}$$

Twenty-five effect sizes were calculated across studies, although only one effect size per study was included in the final analysis. Several studies provided additional outcome measures related to teacher behaviors or academic outcomes. Effect sizes were only calculated on problem behavior in the classroom, although outcomes on

prosocial behavior variables would have been calculated had they been reported in studies. Effect sizes from studies that reported more than one outcome variable related to inappropriate student behavior were averaged to create one independent effect size per study.

Two studies (Dolan et al., 1993; Ialongo, Werthamer, Kellam, Brown, Wang, & Lin, 1999) reported means and standard deviations for boys and girls separately rather than providing the means and standard deviations for the aggregate groups. The two subgroups were re-aggregated and the standardized mean difference effect sizes were calculated using the re-aggregated statistics. These same two studies also provided pretest scores on the outcome measures. Therefore, the posttest standardized mean difference effect sizes were adjusted for pretest differences by taking the difference between the pretest means and subtracting that from the difference between the posttest means and then dividing the result by the pooled standard deviation.

The effect size from one study (van Lier, Muthén, van der Sar, & Crijnen, 2004) was estimated based on the data supplied in tables and a graph.

One study (Hawkins, Von Cleve, & Catalano, 1991) required four separate calculations. First, the authors did not report the means and standard deviations for the treatment and control groups but did report the ANCOVA *F*-value for each outcome using free and reduced-price lunch status as a covariate adjustment. Second, the authors reported results for white females and black females separately. Third, all males were reported separately from females. Fourth, two separate behaviors were reported, externalizing and aggressive behavior. The first calculation involved extracting an effect size from the ANCOVA *F* results using the formulas in Lipsey and D. Wilson (2001). The next calculation involved averaging the covariate adjusted mean effect sizes for black and white females. Next, the male and female subgroups were re-aggregated. Finally, the mean of the aggressive behavior and externalizing behavior effect sizes was calculated and used in the final analysis.

### **3.3.3.1 Adjustment for Clustered Data**

An additional issue that required computational adjustments to the effect sizes was associated with the clustered or nested data provided in most of the studies. Some of the primary studies reported data at the classroom level (i.e., classroom level means and standard deviations) and some primary studies reported data at the individual student level (i.e., student level means and standard deviations); therefore, transformations were required to create equivalent effect sizes prior to analysis (Hedges, 2007). Deciding whether to adjust effect sizes based on individual data up to the classroom level or vice versa should be based on the dependent variable and how it was measured (Hedges, 2007). In the current analysis, the measurements taken at the individual student level could be adjusted to a classroom level variable defensibly while still maintaining the construct. However, it was less defensible to

consider measures of classroom levels as equivalent to individual student level data. Therefore, it was determined that the effect sizes based on individual student standard deviations would be adjusted up to the classroom level.

Adjusting effect sizes based on individual student data up to the classroom level required the use of an intraclass correlation (ICC) of behavioral measures and classroom outcomes. The Department of Education's What Works Clearinghouse ([ies.ed.gov/ncee/wwc/](http://ies.ed.gov/ncee/wwc/)) indicates an ICC = .10 as the convention. However, other researchers have found smaller values such as an ICC = .05 (e.g., Murray & Blitstein, 2003). Separate calculations using both ICCs were conducted as a sensitivity analysis. A total of five effect sizes from studies that did not report classroom level data (i.e., Dolan et al., 1993; Gottfredson et al., 1993; Hawkins et al., 1991; Ialongo et al., 1999; van Lier et al., 2004) were transformed by dividing the raw effect size by the square root of the intraclass correlation (ICC = .10 and ICC = .05).

$$ES_{cluster\ adj} = \frac{ES_{sm}}{\sqrt{.05}} \quad or \quad ES_{cluster\ adj} = \frac{ES_{sm}}{\sqrt{.10}}$$

The resulting effect size after this transformation is considered a classroom level effect size and not the standardized mean difference effect size typically reported in the research literature. This is an important distinction because these effect sizes are based on classroom level standard deviations which tend to be smaller than individual student standard deviations. Therefore, the resulting effect sizes will be larger. This classroom level effect size cannot be compared to the typical standard mean effect size.

Once effect sizes were adjusted to combine groups or outcome variables, 12 independent effect sizes were obtained, one per study. Subsequent to the effect size transformation to the classroom level, the standard error and variance of the effect sizes was adjusted for all 12 independent effect sizes using Hedges small sample correction (Hedges & Olkin, 1985). Because all effect sizes were transformed to the classroom level, the sample size used in this calculation was the number of treatment and control classrooms, not the number individual students. Finally, a weighted random effects analysis was conducted using the Comprehensive Meta-Analysis (CMA) software program. The random effects variance ( $V_o$ ) was based on maximum likelihood estimation using CMA.

---

## 4 Results

---

### 4.1 DESCRIPTION OF ELIGIBLE STUDIES

---

The 12 studies included in the systematic review had a range of characteristics (see Table 1). Most interventions in the studies were conducted in public school general education classrooms with students in K-12. When interventions were implemented across grades, the researchers did not break down the results by individual grade making it impossible to do an analysis by grade. One important distinction is that seven out of the 12 studies were from the same research group and assessed the efficacy the researcher’s program, Classroom Organization and Management Program (COMP; Evertson, 1988). Because COMP studies selected for inclusion represented 58% of total studies, an additional research question was added to examine whether COMP studies produced different outcomes compared to the other studies in the sample.

COMP is a professional development series developed by Carolyn Evertson and colleagues (1988) designed to create effective learning environments. The main components of COMP are: (1) organizing the classroom; (2) planning and teaching rules and procedures; (3) managing student work and improving student accountability; (4) maintaining good student behavior; (5) planning and organizing; (6) conducting instruction and maintaining momentum; and (7) getting the year off to a good start. COMP is the most highly researched classroom management packaged programs and has received validation from the U.S. Department of Education’s National Diffusion Network for effectiveness in decreasing disruptive classroom behavior, improving the classroom environment, and improving academic gains for students in COMP classrooms ([www.comp.org](http://www.comp.org)). Both instructional management techniques and behavioral management techniques are included in the intervention.

Another treatment used in three studies included in the review (i.e., Dolan et al., 1993; Ialongo et al., 1999; van Lier, Muthén, van der Sar, & Crijnen, 2004) was the “Good Behavior Game” (GBG; Barrish et al., 1969). Researchers used the GBG as a universal preventive treatment to reduce classroom levels of inappropriate behavior. Teachers in treatment classrooms outlined positively stated classroom rules and monitored students’ adherence to the rules. The criterion for winning the game was dependent on the behavior of each member of the team. Some form of response-cost

system was used in which cards or points would be removed from teams if a team member violated a classroom rule. Reinforcement was provided for teams that met the criteria. One study (c.f. Ialongo et al., 1999) incorporated curriculum enhancements and backup strategies for children who did not respond to the universal approaches in place. A family-school partnership intervention was an additional feature of the treatment package used in the Ialongo study (1999), although data for that portion of the treatment were not analyzed as part of this synthesis. Similarly, another study (c.f., Dolan et al., 1993) used an additional academic treatment condition, Mastery Learning, with some schools. Again, these data were not analyzed as part of this synthesis as they did not qualify as classroom management.

**TABLE 1: CHARACTERISTICS OF INCLUDED STUDIES (N=12)**

<b>Characteristic</b>	<b>N</b>	<b>%</b>	<b>Characteristic</b>	<b>N</b>	<b>%</b>
<b>Publication Year</b>			<b>Grades of Participants</b>		
1980s	2	17	K-12	1	8
1990s	9	75	K-6 (+resource)	8	67
2000s	1	8	K-9	1	8
			6-12	2	17
<b>Form of Publication</b>			<b>Location of Treatment</b>		
Published (peer review)	5	42	Regular classroom	8	67
Technical report	7	58	Both regular and special	4	33
<b>Country of Study</b>			<b>Treatment Agent</b>		
United States	11	92	Regular education teacher	8	67
Netherlands	1	8	Both regular and special	4	33
<b>Group Assignment</b>			<b>Duration of Treatment</b>		
Randomized (individual)	7	58	1-10 weeks	1	8
Randomized (group)	4	33	11-20 weeks	1	8
Nonrandomized	1	8	21-50 weeks	8	67
			>50 weeks	2	17
<b>Attrition</b>					
Not reported	2	14			
1-10%	10	71	<b>Focal Treatment Components</b>		

11-20%	2	14	Teacher training in COMP	7	58
			Good Behavior Game	2	17
<b>Sample Size (Tx + Control)</b>			Classroom-centered	1	8
Under 50	7	58	Multi-component	2	17
200-300	2	16			
400 and up	3	25	<b>Additional Treatment Components</b>		
			Parent training	2	17
<b>School Setting</b>			School structure changes	1	8
Public	10	84	Academic	1	8
Public and Private	2	16	None	8	67
<b>School Neighborhood</b>			<b>Treatment Program</b>		
Urban	1	8	Research project	5	42
Mix (urban, suburban, rural)	10	84	Demonstration project	7	48
Unknown	1	8			

Finally, the remaining two studies (c.f., Gottfredson, Gottfredson, & Hybl, 1993; Hawkins, Von Cleve, & Catalano, 1991) used multi-component treatments as part of their universal classroom management package. A continuum of treatments was provided for school, classroom, and individual students in the Gottfredson, Gottfredson, and Hybl study. Although Gottfredson and colleagues had multiple components, the researchers included a dependent measure of student behaviour directly from the classroom as was consistent with the inclusion criteria. Hawkins and colleagues trained teachers on proactive classroom management methods involving frequent use of encouragement and praise, as well as a social skills curriculum called Interpersonal Cognitive Problem Solving (ICPS; Spivack & Sure, 1982). ICPS teaches children to consider alternative solutions to interpersonal problems that they encounter. An interactive teaching component that required children to master content prior to progressing to more advanced work was also included in the classroom treatment package. A parent training component was an additional feature of the treatment package.

---

## 4.2 MAIN EFFECTS OF TEACHERS' CLASSROOM MANAGEMENT PRACTICES

---

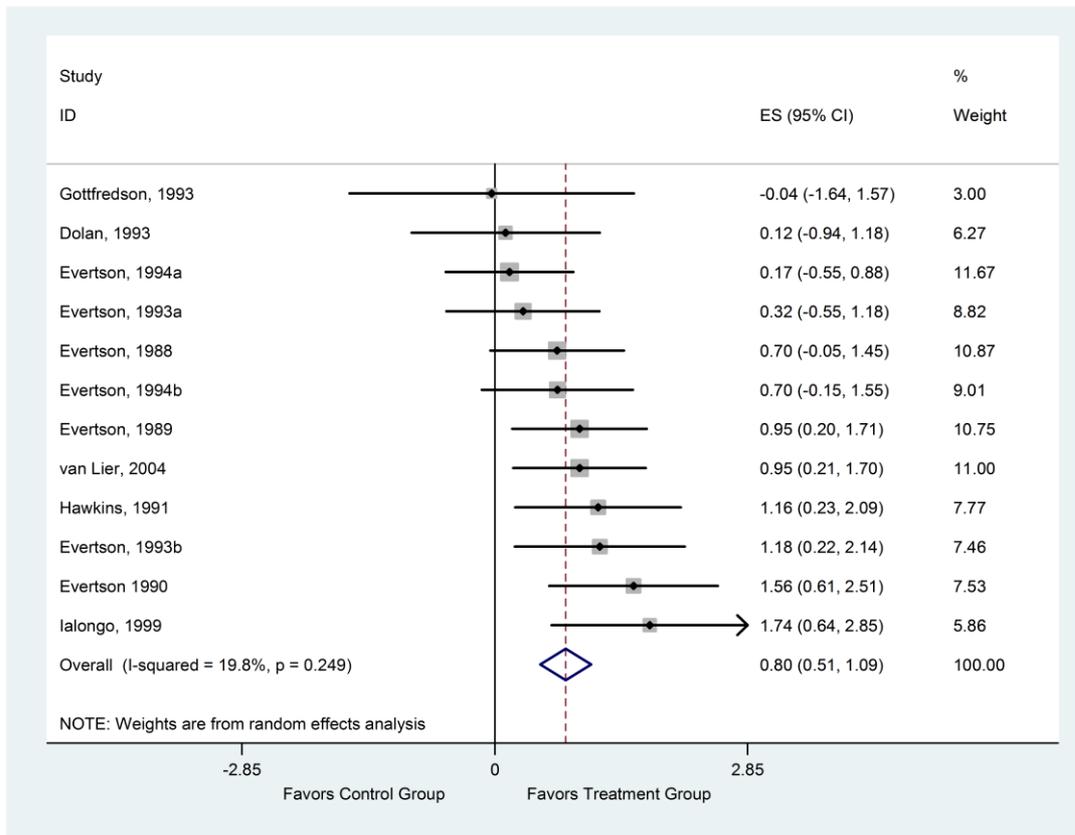
The primary research question, “Do teacher’s universal classroom management practices reduce problem behavior in classrooms with students in kindergarten through 12<sup>th</sup> grade?” was examined through a main effects analysis. A total of 25 effect sizes were obtained from the sample of 12 studies. A total of nine studies

reported two outcome measures: inappropriate behavior and disruptive behavior. Separate effect sizes were calculated for both outcome variables and then averaged to produce one effect size. The decision was made to average these effect sizes because inappropriate behavior and disruptive behavior could be considered the same construct of a broader category of “problem classroom behavior” and therefore they were not seen as distinct enough to warrant individual effect sizes. One study required several group calculations to obtain one grand mean effect size. These calculations produced 12 independent effect sizes, one per study.

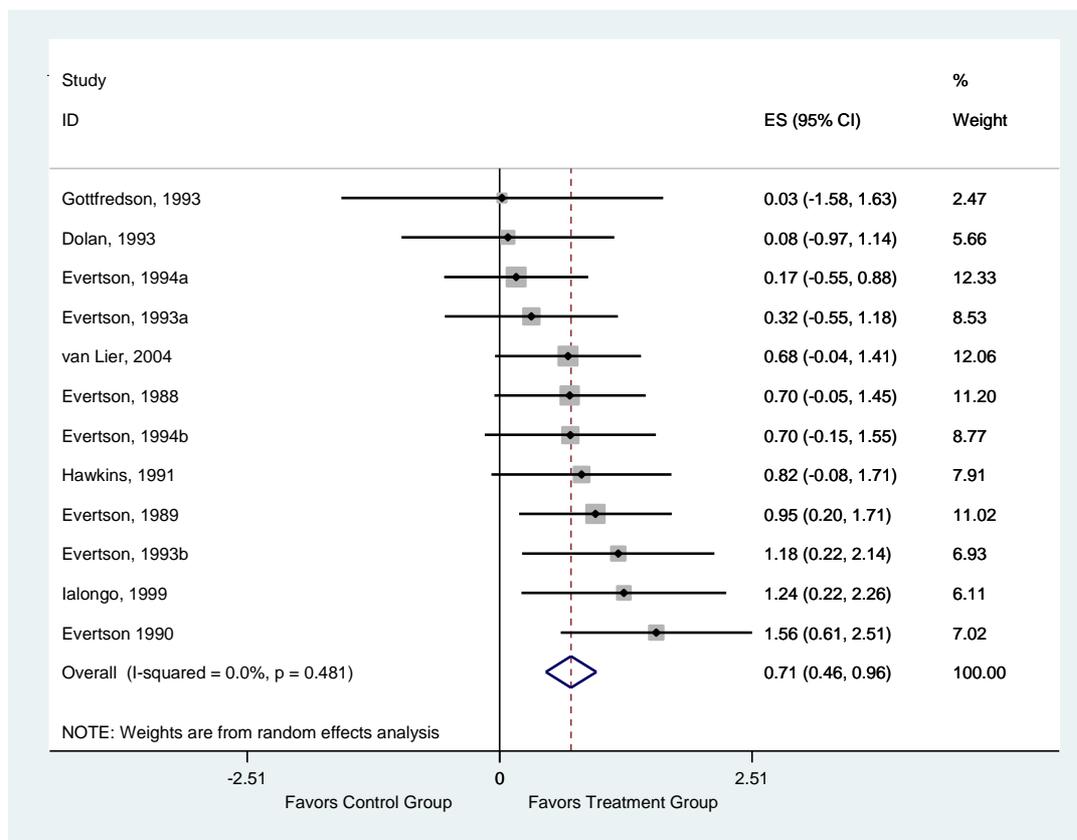
Only the most immediate posttest data were used to calculate effect sizes. That is, if a study was conducted over multiple years and follow-up data were collected, only the first follow-up measurement was used, typically after one school year. The reasoning behind this decision was that most studies in the sample reported data within one school year or less. The random effects analysis on the 12 effect sizes in the classroom management database produced a statistically significant mean classroom effect size of .80 (CI: 0.51-1.09;  $z = 5.44$ ,  $p < .05$ ) for ICC=.05 and a statistically significant mean classroom effect size of .71 (CI: 0.46-0.96,  $z = 5.54$ ,  $p < .05$ ) for ICC=.10 indicating that the participants in the classroom management intervention conditions exhibited significantly less problem classroom behavior after intervention. Recall that the effect sizes used here are based on classroom-level means and standard deviations and are not commensurate with the student-level effect sizes typical in educational research. To put our classroom-level mean effect sizes into a comparable format with the more typical effect sizes, we back-transformed our mean effect sizes using the original adjustment formulas (Hedges, 2007). Thus, the classroom-level mean effect sizes of .80 and .71 are roughly comparable to student level effect sizes of .18 and .22 for ICC=.05 and ICC=.10, respectively.

Figure 1 shows the forest plot of the effect sizes using ICC=.05 and Figure 2 shows the forest plot of the effect sizes using ICC=.10. The effect sizes from the 12 included studies ranged from -0.04 to 1.74 (ICC=.05) and from -.03 to 1.56 (ICC=.10) showing an overall positive effect for teachers’ classroom management practices. Additional analyses were conducted on the sample of effect sizes to determine if the sample was biased or if the sample was pulled from the same population of effect sizes.

**Figure 1.** Forest plot for the effects of classroom management on disruptive or aggressive student behavior ( $ICC = .05$ ). Heterogeneity Analysis:  $Q = 13.72$ ,  $df = 11$ ,  $p = .25$ , and  $I^2 = 19.83$ .



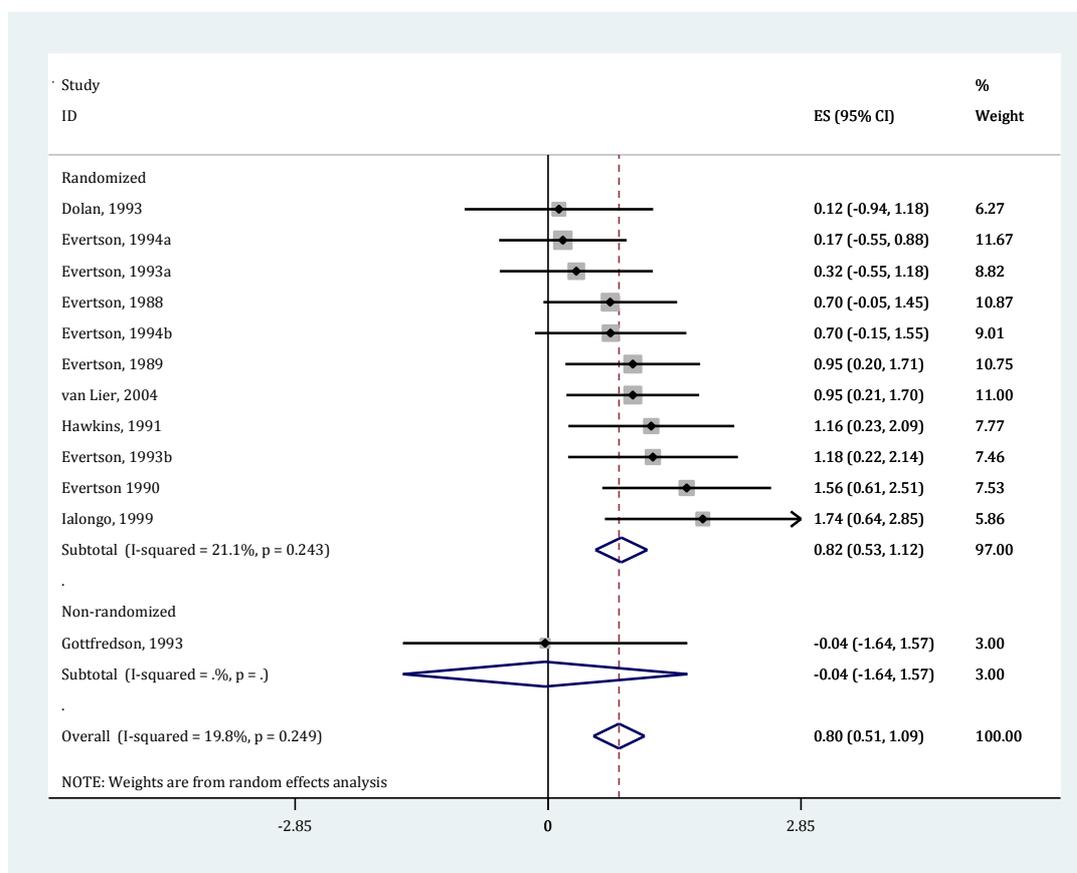
**Figure 2.** Forest plot for the effects of classroom management on disruptive or aggressive student behavior (ICC = .10). Heterogeneity Analysis:  $Q = 10.56$ ,  $df = 11$ ,  $p = .48$ , and  $I^2 = 0.00$ .



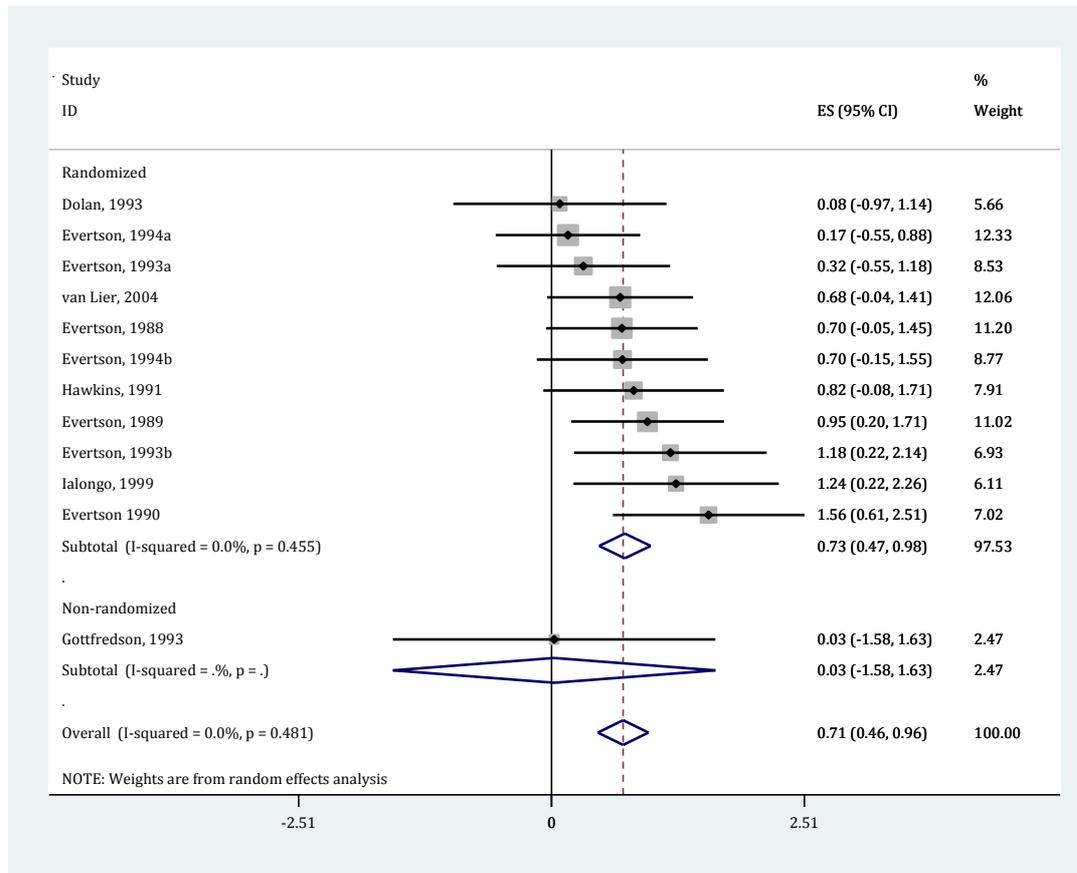
One study, Gottfredson et al., 1993, used a non-randomized design. An additional analysis was conducted to determine if removing this study from the analysis significantly changed the overall results. The random effects analysis on the 11 effect sizes in which random assignment was used produced a slightly larger statistically significant mean classroom effect size of .83 (CI: 0.53-1.12;  $z = 5.53$ ,  $p < .05$ ) for  $ICC = .05$  and a slightly larger statistically significant mean classroom effect size of .73 (CI: 0.48-0.98,  $z = 5.60$ ,  $p < .05$ ) for  $ICC = .10$  indicating that (1) the participants in the classroom management intervention conditions exhibited significantly less problem classroom behavior after intervention and (2) the differences in effect sizes when the Gottfredson et al. study was included did not significantly change the overall results. When these randomized study mean effect sizes are back-transformed using the original adjustment formulas (Hedges, 2007), the resulting classroom-level mean effect sizes of .83 and .73 can be roughly compared to student level effect sizes of .19 and .23 for  $ICC = .05$  and  $ICC = .10$ , respectively. These back-transformed effect sizes of only randomized studies are nearly identical to the back-transformed effect sizes with all studies were included in the analysis ( $ES_{sm} = .18$ ,  $ICC = .05$ ;  $ES_{sm} = .22$ ,  $ICC = .10$ )

Figure 3 shows the forest plot of the effect sizes using ICC=.05 and Figure 4 shows the forest plot of the effect sizes using ICC=.10 comparing the randomized and non-randomized studies. The effect sizes from the 11 randomized studies ranged from 0.12 to 1.74 (ICC=.05) and from .08 to 1.56 (ICC=.10) showing an overall positive effect for teachers' classroom management practices. Additional analyses were conducted on the sample of effect sizes to determine if the sample was biased or if the sample was pulled from the same population of effect sizes.

**Figure 3.** Forest plot for the effects of classroom management on disruptive or aggressive student behavior for randomized vs. non-randomized studies (ICC = .05). Heterogeneity Analysis:  $Q = 12.67$ ,  $df = 10$ ,  $p = .24$ , and  $I^2 = 21.1$ .



**Figure 4.** Forest plot for the effects of classroom management on disruptive or aggressive student behavior for randomized vs. non-randomized studies (ICC = .10). Heterogeneity Analysis:  $Q = 9.84$ ,  $df = 10$ ,  $p = .46$ , and  $I^2 = 0.00$ .



#### 4.2.1 Heterogeneity Analysis

The test for heterogeneity with all studies included was not statistically significant for ICC=.05 ( $Q = 13.72$ ,  $df = 11$ ,  $p = .25$ ) or for ICC=.10 ( $Q = 10.56$ ,  $df = 11$ ,  $p = .48$ ) indicating this sample of effect sizes are homogeneous and there may not be enough variability between studies to justify further analysis to examine potential moderators. The  $I^2$  values further support the conclusion ( $I^2 = 19.83$ , ICC=.05;  $I^2 = 0.00$ , ICC=.10) that the proportion of variance between studies is not large enough to suggest systematic differences between studies. We also performed heterogeneity analysis on the 11 randomized studies. As was the case when all studies were included, the test for heterogeneity was not statistically significant for ICC = .05 ( $Q = 12.67$ ,  $df = 10$ ,  $p = .24$ ) or for ICC = .10 ( $Q = 9.84$ ,  $df = 10$ ,  $p = .46$ ). Likewise the  $I^2$  values further support the lack of heterogeneity ( $I^2 = 21.1$ , ICC = .05;  $I^2 = 0.00$ , ICC = .10). Comparisons of both heterogeneity analyses indicate no difference in the interpretation that the sample of effect sizes are homogeneous. However, due to the small sample size there is little statistical power to detect heterogeneity. Moreover, classroom-level effect sizes have less power, further adding to the difficulty of detecting heterogeneity between studies. However, because one

program (COMP) was so prominent in the literature, it would be important to examine the average effect size for that program separately despite the lack of heterogeneity. The results for this analysis will be presented below.

### 4.2.2 Publication Bias Analysis

To determine if there is a potential for bias in the effect sizes due to unpublished small effects not being included in the analysis, funnel plots were produced. These are presented in Figure 5 (ICC=.05) and Figure 6 (ICC=.10). Visual analysis of the symmetry of studies on both sides of the vertical line that divides the funnel plot in half indicates there is a low risk of publication bias occurring in the current sample and the sample of studies are likely representative of most studies examining these outcomes. The lack of a study on the opposite side of the vertical line (e.g., larger Hedge's  $g$ ) for the study with low standard error may indicate a missing study but is insufficient to conclude publication bias.

Figure 5: Funnel Plot of Standard Error by the Hedges'  $g$  Effect Size for Problem Behavior (ICC=.05)

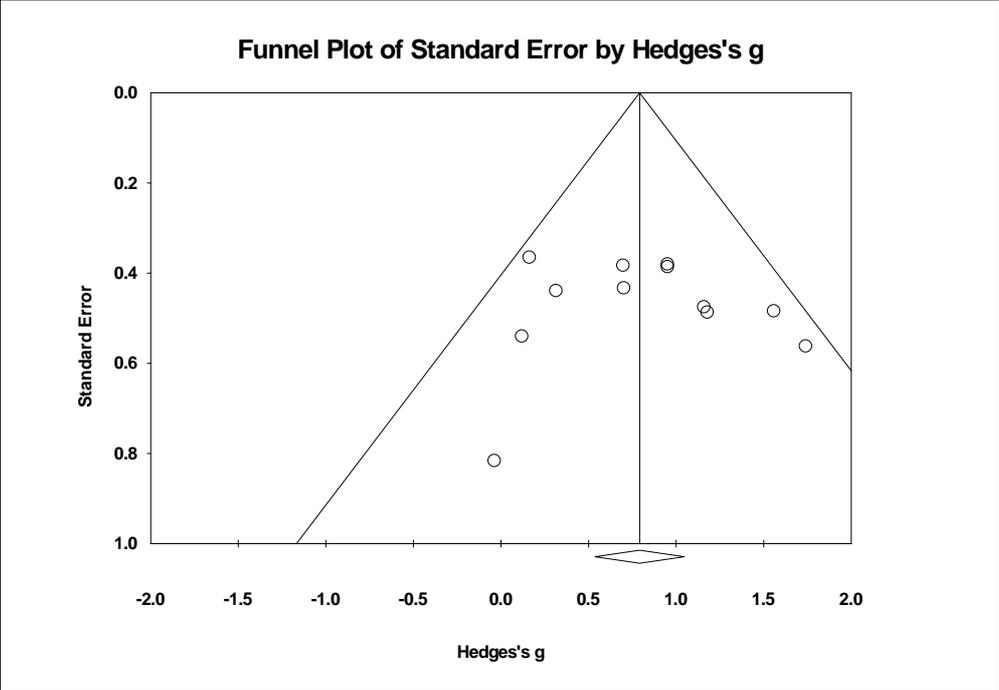
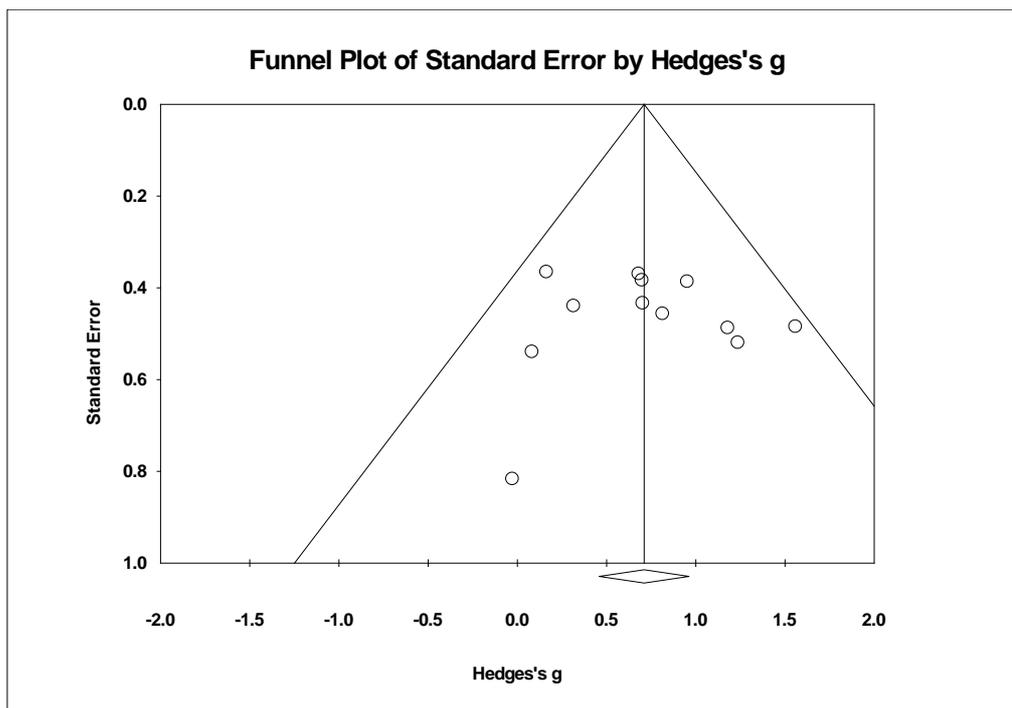


Figure 6: Funnel Plot of Standard Error by the Hedges' g Effect Size for Problem Behavior (ICC=.10)



#### 4.2.3 Moderator Analysis

Several research questions required the use of moderator analyses. These questions, however, could not be answered due to the small sample of studies, no evidence heterogeneity in the distribution of effect sizes to support such an analysis, and insufficient data reported in the studies themselves. The research questions that required a moderator analysis were: what components make up the most effective and efficient classroom management programs; do differences in effectiveness exist between grade levels; do differences in classroom management components exist between grade levels; and, does treatment fidelity affect the outcomes observed? Only the first research question related to treatment components could potentially be analyzed based on the data supplied in the studies. However, due to lack of statistical power to find heterogeneity, a moderator analysis was not supported. Studies did not report data by grade level and did not report sufficient treatment fidelity data to conduct a moderator analysis on these variables. Table 2 provides information regarding treatment fidelity data for each study in the meta-analysis. Further treatment of this issue is addressed in the discussion section.

**TABLE 2: TREATMENT FIDELITY REPORTED BY STUDY**

Study	Treatment Fidelity Measure	Outcome Data
Dolan et al. (1993)	None	None
Gottfredson, Gottfredson, & Hybl (1993)	Classroom environment survey <ul style="list-style-type: none"> <li>Proxy for classroom order, classroom order and organization, rule clarity, and teacher support</li> </ul> Teacher Survey <ul style="list-style-type: none"> <li>Global rating of program</li> <li>Team effectiveness rating</li> <li>Reports of level of implementation (e.g., 1 = never use, 5 = always use)</li> </ul> Effective School Battery <ul style="list-style-type: none"> <li>School-level data only</li> </ul>	Schools were grouped based on level of implementation (based on teacher-survey data) and grouped into high or medium level of implementation.  Provides somewhat detailed description of factors that potentially effected implementation at school level Reported means and standard deviations of teacher survey
Hawkins, Von Cleve, & Catalano (1991)	Teacher Self-Report checklist (completed weekly)	Used to provide additional support/training for teachers. No data reported.
Ialongo, Werthamer, Kellam, Brown, Wang, & Lin (1999)	Measured: <ul style="list-style-type: none"> <li>Measures of classroom setup</li> <li>Classroom observation</li> <li>Classroom visit record reviews</li> </ul>	5/9 classrooms identified as “high” implementation No specific data reported.
Van Lier, Muthén, van der Sar, & Crijnen (2004)	External school advisor evaluated whether the school implemented all phases of the GBG.	School level data not classroom level No specific data provided to determine categorizations.
Evertson (1988)	Teacher behaviors were measured as outcome data not as treatment fidelity	None
Evertson (1995) *6 studies	None	None

Based on the *post hoc* hypothesis that differences between COMP studies and other studies may exist, one moderator was selected for the analysis. It was hypothesized that treatment characteristics may moderate the magnitude of the effect sizes. Prior to conducting the moderator analysis, frequency distributions for the identified moderator variable were examined to determine if each cell had an appropriate amount of data. The independent variable “Treatment Characteristics” was coded into “COMP” or “GBG + Other”. This analysis examined the differences between the effects sizes for COMP and any other classroom management intervention used by researchers.

An inverse variance weighted analysis, similar to an ANOVA, was conducted to compare differences in mean effect sizes between COMP studies and the others. This

statistical analysis of differences between groups is tested using a  $Q_{between}$  instead of a t-test. The statistically significant mean effect size for studies categorized as “other” was  $ES=.88$  ( $p=.00$ ) using  $ICC=.05$  and  $ES=.66$  ( $p=.00$ ) using  $ICC=.10$ . COMP studies produced a statistically significant effect size  $ES=.75$  ( $p=.00$ ) (see Table 3). Based on the random effects analysis, differences between mean effect sizes were not statistically significant for either  $ICC=.05$  ( $Q_{between} = .38, df = 1, p = .54$ ) or  $ICC=.10$  ( $Q_{between} = .07, df = 1, p = .54$ ). Given the small number of studies and the fact that classroom-level effect sizes are used, there is little statistical power for detecting significant differences. It is difficult to determine if there are indeed no differences between studies using COMP and studies using other forms of classroom management or if there are differences between treatments, but there was not enough power to detect them.

**TABLE 3: RESULTS OF RANDOM EFFECTS WEIGHTED MODERATOR ANALYSIS FOR TREATMENT TYPE**

Treatment Characteristics	Mean Classroom ES	SE	-95% CI	+95% CI	z	p	$Q_{between}$
GBG + Other	.88 (ICC=.05)	.29	.22	.41	6.36	.00	.38
GBG + Other	.66 (ICC=.10)	.22	.23	1.10	3.01	.00	.07
COMP*	.75	.18	.40	1.10	4.23	.00	

\*All results were given at the classroom-level so no adjustments were performed for the COMP studies.

---

## 5 Discussion

Disruptive student behavior in the classroom is a major concern in school systems today. Students in classrooms with frequent disruptive behavior experience less academic engagement and lower academic outcomes (Shinn et al., 1987). Teachers who experience difficulty controlling classroom behavior have higher stress and burnout (Smith & Smith, 2006) and find it difficult to meet the instructional demands of the classroom (Emmer & Stough, 2001). Lack of effective classroom management may also worsen the progression of aggressive behavior for children in classrooms with higher levels of disruption (Greer-Chase et al., 2002). Effective approaches to managing the classroom environment are necessary to establish environments that support student behavior and the learning process as well as to reduce teacher stress and burnout. The purpose of this review was to examine the effects of teacher's universal classroom management practices to reduce disruptive, aggressive, or inappropriate behaviors of children in kindergarten through 12<sup>th</sup> grade.

Teacher's classroom management practices have a significant, positive effect on decreasing problem behavior in the classroom. Students in the treatment classrooms in all 12 studies located for the review showed less disruptive, inappropriate, and aggressive behavior in the classroom compared to untreated students in the control classrooms. The overall mean classroom effect size of either .71 or .80 indicates a positive effect that significantly impacts the classroom environment. Teachers who use effective classroom management can expect to experience improvements in student behavior and improvements that establish the context for effective instructional practices to occur.

Due to a lack of power, the homogeneity in the sample of effects sizes indicated no moderator variables could account for differences between studies. It is not possible to determine which treatment components contributed to the overall effects due to the small sample of studies included in the review. A leaner package such as GBG may be as effective as a more comprehensive package such as COMP. Without adequate treatment fidelity data, it is difficult to determine what level of fidelity is necessary to establish effective universal classroom management. Although fidelity data could not be analyzed in the current synthesis, one study did report outcomes based on level of implementation. The classroom level effect size in Gottfredson et al. (1993) calculated for all levels of implementation was -.04 (ICC=.05) and -.03

(ICC=.10). However, when only high implementation classrooms were analyzed, the classroom level effect size increased to .54 (ICC=.05) and .38 (ICC=.10). This disparity in classroom level effect sizes indicates the importance treatment fidelity may play in the magnitude of outcomes.

One important point to address is the lack of information about what type of management practices were occurring in the control classrooms. Presumably, teachers in control classrooms had some type of classroom management plan in place; otherwise, it is unlikely that they would be able to teach effectively. The implications for control conditions that have some level of classroom management may be that students in classrooms with highly structured classroom management practices such as those in the studies in this review demonstrate more appropriate behavior than students in typically managed classroom environments.

When studies employ a design with treatment and control conditions, presumably the control condition is “no treatment.” Effect sizes from these studies are interpreted based on “all or none” conditions. Classroom management studies, on the other hand, are comparisons of specific, structured, practices vs. current classroom management practices already in place. These are interpreted more accurately as “something different or more” vs. “current practice.” In light of this, an overall classroom mean effect size of .71 or .80 (.73 or .83 for randomized studies) may be more impressive given the fact that this effect was found over and above current classroom management practices vs. no classroom management. This difference can have a large impact on a classroom teacher struggling to meet the academic and social demands of the classroom.

Based on the current results and the extant literature on individual classroom management practices to reduce problem behavior, classroom organization and behavior management appears to be an effective classroom practice. But whose behavior is universal classroom management supporting? Studies on reducing problem behavior in schools frequently focus on changes in student behavior as the primary outcome measure of intervention effectiveness. While the ultimate goal may be to reduce problem behavior and increase prosocial behaviors, the fact remains that teacher behavior ultimately needs to change first to produce changes in student behavior. Classroom management, therefore, provides the structure to support teacher behavior and increase the success of classroom practices. Teacher proficiency with classroom management is necessary to structure successful environments that encourage appropriate student behavior. Adequate teacher preparation, therefore, is an important first step in providing content knowledge and opportunities to develop proficiency in classroom management (Oliver & Reschly, 2007).

What we know about teachers’ pre-service training and proficiency with classroom management indicates a less-than-preferable state of affairs. Investigations of

preservice teachers' perceptions on how prepared they are to effectively manage the classroom environment indicate that they frequently report feeling inadequately prepared and they receive little specific instruction in classroom management (Baker, 2005; Siebert, 2005). When there is coursework on classroom management in teacher preparation programs, it is perceived as too theoretical and broad or rarely in-depth enough to adequately prepare student teachers to handle significant antisocial behavior (Siebert, 2005). A recent evaluation of teacher preparation courses in one state supports teachers' perception of a lack of adequate pre-service training in classroom management. A review of 135 course syllabi indicated little to no special education teacher preparation in classroom organization and behavior management (Oliver & Reschly, 2011). Only 27% (N=7) of universities included in the review had an entire course devoted to classroom management. The remaining 73% (N=27) had classroom management content spread across several courses. If these data are representative of pre-service teacher training programs in other states, an apparent weakness in teacher preparation has been identified.

Inadequate competency in classroom management has detrimental effects on teachers challenged with handling disruptive behavior and meeting the instructional demands of the classroom (Emmer & Stough, 2001). Teachers who are unable to manage the classroom environment and have high rates of discipline problems and low rates of teacher responses to the problems are rated as ineffective by researchers. Poor classroom management is the main reason for being identified as ineffective (Berliner, 1986; Espin, & Yell, 1994). Teachers also report high levels of stress and burnout related to handling discipline issues (Browsers & Tomic, 2000; Smith & Smith, 2006). A clear understanding of the effectiveness of classroom management as a package of practices is necessary to establish teacher buy-in to implementing and sustaining classroom management. Identifying the most effective way to provide pre-service and in-service teachers with content knowledge as well as providing a system to support changes in teacher behavior are critical to improving the context of classroom environments and the persistence of teachers in the profession. Future research should address the issues posed in this review.

---

## **5.1 LIMITATIONS OF THE REVIEW**

---

Several limitations of this review can be identified. One such limitation is the noted lack of single subject studies included in the analysis. Much of the previous work in classrooms examining behavior management techniques was conducted using single subject methodology. This body of research provides important information about the functional relation between various classroom management practices and student behavior. However, given the concerns with current methods to analyze single subject data to calculate an effect size (Campbell, 2004; Parker et al., 2007), the authors consciously chose not to include these studies. The consequence of excluding single subject studies means the effect sizes of universal classroom management obtained in this study may be biased. However, obtaining effect sizes

from single subject studies that are not comparable to classroom effect sizes based on group studies was determined to not offer additional information to this study. Another limitation is that studies in which classroom management was examined but that did not include a measure of student behavior were not included in this review (e.g., Freiberg, Stein, & Huang, 1995). A meta-analysis with both academic and behavioral outcomes would be the ideal, however given the paucity of group studies in this area, it was not possible. Finally, due to the small sample of effect sizes a general lack of power affected the ability to do moderator analyses. Future research should address these issues.

---

## References

- Ahrentzen, S., & Evans, G. (1984). Distraction, privacy, and classroom design. *Environment and Behavior*, 16(4), 437-454.
- Anderson, L. M. & Evertson, C. M. (1978). *Classroom organization at the beginning of school: Two case studies*. Research and Development Center for Teacher Education: TX. [ERIC Document Reproduction Service No. ED 166 193].
- Anderson, L. M., Evertson, C. M., & Emmer, E. T. (1979). *Dimensions in classroom management derived from recent research*. Research and Development Center for Teacher Education: TX. [ERIC Document Reproduction Service No. ED 166 193].
- Baker, P. H. (2005). Managing student behavior: How ready are teachers to meet the challenge? *American Secondary Education*, 33, 51-64.
- Barrish, H., Saunders, M., & Wolf, 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.
- Becker, W.C., Madsen, C.H., & Arnold C. (1967). The contingent use of teacher attention and praise in reducing classroom behavior problems. *Journal of Special Education*, 1, 287-307.
- Berliner, D. C. (1986). In pursuit of the expert pedagogue. *Educational Researcher*, 15, 5-13.
- Brantlinger, E., & Danforth, S. (2006). Critical theory perspective on social class, race, gender, and classroom management. In C. Evertson & C. Weinstein (Eds.), *Handbook of Classroom Management: Research, practice, and contemporary issues* (pp. 157-179). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Brophy, J. E. (2006). History of research in classroom management. In C. Evertson & C. Weinstein (Eds.), *Handbook of Classroom Management: Research, practice, and contemporary issues* (pp. 17-43). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Browsers A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, 16, 239-253.
- Campbell, J. M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification*, 28, 234-246.

- Colvin, G., Kame'enui, E. J., & Sugai, G. (1993). Reconceptualizing behavior management and school-wide discipline in general education. *Education and Treatment of Children, 16*, 361-381.
- Conyers, C., Miltenberger, R., Maki, A., Barenz, R., Jurgens, M., Sailer, A., Haugen, M., & Kopp, B. (2004). A comparison of response cost and differential reinforcement of other behavior to reduce disruptive behavior in a preschool classroom. *Journal of Applied Behavior Analysis, 37*, 411-415.
- 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.
- Darveaux, D. (1984). The Good behavior game plus merit: Controlling disruptive behavior and improving student motivation. *School Psychology Review, 13*, 510-515.
- Di Martini-Scully, D., Bray, M. A., Kehle, T. J. (2000). A packaged intervention to reduce disruptive behaviors in general education students. *Psychology in the Schools, 37*, 149-156.
- Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*, 103-112.
- Espin, C. A., & Yell, M. L. (1994). Critical indicators of effective teaching for preservice teachers: Relationships between teaching behaviors and ratings of effectiveness. *Teacher Education and Special Education, 17*, 154-169.
- Evertson, C. M. et al. (1988). Improving elementary classroom management: A school-based training program for beginning the year. Peabody College, Vanderbilt University: TN [ERIC Document Reproduction Service No. ED 302 528].
- Evertson, C. M., Emmer, E. T., Sanford, J. P., & Clements, B. S. (1983). Improving classroom management: An experiment in elementary classrooms. *The Elementary School Journal, 84*, 172-188.
- Evertson C. M., & Weinstein, C. (2006). Classroom management as a field of inquiry. In C. M. Evertson & C. Weinstein (Eds.), *Handbook of Classroom Management: Research, practice, and contemporary issues* (pp. 3-15). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Fishbein, J. & Wasik, B. (1981). Effects of the good behavior game on disruptive library behavior. *Journal of Applied Behavior Analysis, 14*, 89-93.
- Freiberg, H. J., Stein, T. A., & Huang, S. (1995). Effects of classroom management intervention on student achievement in inner-city elementary schools. *Educational Research and Evaluation, 1*, 36-66.
- Greer-Chase, M., Rhodes, W.A., & Kellam, S.G. (2002). Why the prevention of aggressive disruptive behaviors in middle school must begin in elementary school. *The Clearing House, 75*(5), 242-245.

- Hall, R. V., Panyan, M., Rabon, D., & Broden, M. (1968). Instructing beginning teachers in reinforcement procedures which improve classroom control. *Journal of Applied Behavior Analysis, 1*, 315-322.
- Harris, V., & Sherman, J. (1973). Use and analysis of the "Good Behavior Game" to reduce disruptive classroom behavior. *Journal of Applied Behavior Analysis, 6*, 405-417.
- Hedges, L. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*, 341-370.
- Hedges, L., & Olkin, I. (1985). *Statistical models for meta-analysis*. New York: Academic Press.
- Ingersoll, R. M., & Smith, T. M. (2003). The wrong solution to the teacher shortage. *Educational Leadership, 60*, 30-33.
- Kauffman, J. M. (2005). *Characteristics of Emotional and Behavioral Disorders of Children and Youth* (8<sup>th</sup> ed.). Upper Saddle River, New Jersey: Pearson Education, Inc.
- Kehle, T. J., Bray, M. A., Theodore, L. A., Jenson, W. R., Clark, E. (2000). A multi-component intervention designed to reduce disruptive classroom behavior. *Psychology in the Schools, 37*, 475-481.
- Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., and Ialongo, N. (1998). The effect of level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology, 10*, 165-185.
- 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.
- Kounin, J. (1970). *Discipline and group management in classrooms*. New York: Holt, Rinehart & Winston.
- Langland, S., Lewis-Palmer, T., & Sugai, G. (1998). Teaching respect in the classroom: An instructional approach. *Journal of Behavioral Education, 8*, 245-262.
- Lewis, T. J., & Sugai, G. (1999). Effective behavior support: A systems approach to proactive schoolwide management. *Focus on Exceptional Children, 31*, 1-24.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage Publications.
- Litow, L., & Pomroy, D.K. (1975). A brief review of classroom group-oriented contingencies. *Journal of Applied Behavior Analysis, 8*, 341-347.
- Madsen, C. H., Jr., Becker, W. C., & Thomas, D. R. (1968). Rules, praise, and ignoring: Elements of elementary classroom control. *Journal of Applied Behavior Analysis, 1*, 139-150.
- Murray, D., & Blitstein, J. (2003). Methods to reduce the impact of intraclass correlation in group-randomized trials. *Evaluation Review, 27*, 79-103. doi: 10.1177/0193841X02239019

- Musser, E. H., Bray, M. A., Kehle, T. J., & Jenson, W. R. (2001). Reducing disruptive behaviors in students with serious emotional disturbance. *School Psychology Review, 30*, 294-304.
- Oliver, R. M., & Reschly, D. J. (2010). Teacher preparation in classroom management: Implications for students with emotional and behavioral disorders. *Behavioral Disorders, 35*, 188-199.
- Oliver, R. M., & Reschly, D. J. (December, 2007). *NCCTQ connections paper. Improving outcomes in general and special education: Teacher preparation and professional development in effective classroom management*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Parker, R. I., Brossart, D. F., Vannest, K. J., Long, J. R., De-Albe, R. G., Baugh, F. G., et al. (2005). Effect size in single case research: How large is large? *School Psychology Review, 34*, 116-132.
- Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial Boys*. Castalia: Eugene, OR.
- Rose, L. C., & Gallup, A. M. (2005). The 37<sup>th</sup> annual Phi Delta Kappa/Gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappan, 87*(1), 41-57.
- Rosenberg, M. S. (1986). Maximizing the effectiveness of structured classroom management programs: Implementing rule-review procedures with disruptive and distractible students. *Behavior Disorders, 11*, 239-248.
- Ryans, D. (1952). A study of criterion data—A factor analysis of teacher behaviors in the elementary school. *Educational and Psychological Measurement, 12*, 333-344.
- Shinn, M. R., Ramsey, E., Walker, H. M., Stieber, S., & O'Neill, R. E. (1987). Antisocial behavior in school settings: Initial differences in an at-risk and normal population. *The Journal of Special Education, 21*, 69-84.
- Shores, R. E., Jack, S. L., Gunter, P.L., Ellis, D. N., DeBriere, T. J., & Wehby, J. H. (1993). Classroom interactions of children with behavior disorders. *Journal of Emotional and Behavioral Disorders, 1*, 27-39.
- Siebert, C. J. (2005). Promoting preservice teacher's success in classroom management by leveraging a local union's resources: A professional development school initiative. *Education, 125*, 385-392.
- 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.
- Smith, D. L., & Smith, B. J. (2006). Perceptions of violence: The views of teachers who left urban schools. *The High School Journal, 89*, 34-42.
- Spivack, G., & Shure, M. (1982). Interpersonal cognitive problem solving and clinical theory. In B. B. Lahey, & A. E. Kazdin (Eds), *Advances in Child Psychology* (pp 322-372). New York: Plenum Press.
- Stage, S., & Quiroz, D. (1997). A meta-analysis of interventions to decrease disruptive behavior in the public education setting. *School Psychology Review, 26*, 333-368.

- Strain, P. S., Lambert, D. L., Kerr, M. M., Stagg, V., & Lenkner, D. A. (1983). Naturalistic assessment of children's compliance to teachers' requests and consequences for compliance. *Journal of Applied Behavior Analysis, 16*, 243-249.
- Sugai, G., & Horner, R.H. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child and Family Behavior Therapy, 24*, 23-50.
- Sutherland, K. S., Wehby, J. H. & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavior Disorders, 8*, 2-8.
- Theodore, L. A., Bray, M. A., Kehle, T. J., & Jenson, W. R. (2001). Randomization of group contingencies and reinforcers to reduce classroom disruptive behavior. *Journal of School Psychology, 39*, 267-277.
- U.S. Department of Education Institute of Education Sciences National Center for Educational Evaluation and Regional Assistance (2003). *Identifying and evaluating educational practices supported by rigorous research: A practical guide*. Princeton, NJ.
- White-Blackburn, G., Semb, S., & Semb, G. (1977). The effects of a good-behavior contract on the classroom behaviors of sixth-grade students. *Journal of Applied Behavior Analysis, 10*, 312.
- Wilson, D. B., Gottfredson, D. C., & Najaka, S. S. (2001). School-based prevention of problem behaviors: A meta-analysis. *Journal of Quantitative Criminology, 17*, 247-272.
- Wilson, S. J., & Lipsey, M. (2006). The effects of school-based information processing interventions on aggressive behavior: Part I universal programs. Center for Evaluation Research and Methodology. Nashville, TN: Vanderbilt University.
- Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology, 71*, 136-149.

## Appendix A: Characteristics of included studies

Study/Journal/Design	Participants	Grade/ Duration	Program Description	Outcome Measures
<b>Hawkins, Von Cleve, &amp; Catalano (1991)</b> (journal) <i>Random assignment</i>	Treatment ( $n = 11$ ) Control ( $n = 10$ )	Gr. 1-2 2 years	Multi-component: -Teacher training in classroom management, problem solving, and interactive teaching -Parent training	Student Aggressive, Externalizing Behavior CBCL TRF
<b>van Lier, Muthén, van der Sar, &amp; Crijnen (2004)</b> (journal) <i>Random assignment</i>	Treatment ( $n = 16$ ) Control ( $n = 15$ )	Gr. 1 baseline Gr. 2-3 2 years	Good Behavior Game	ADH, ODD, & Conduct Problem Behavior TRF/6-18—teacher rating PBSI—teacher rating
<b>Dolan et al., (1993)</b> (journal) <i>Random assignment</i>	Treatment ( $n = 8$ ) Control ( $n = 6$ )	Gr. 1 Fall to Spring	Good Behavior Game	Aggressive Behavior TOCA-R—Teacher rating

Study/Journal/Design	Participants	Grade/ Duration	Program Description	Outcome Measures
<b>Ialongo, Werthamer, Kellam, Brown, Wang, et al., (1999)</b> (journal) <i>Random assignment</i>	Treatment ( $n = 9$ ) Control ( $n = 9$ )	Gr. 1  Gr. 2 follow up	Multi-component -Classroom-centered	Aggressive Behavior TOCA-R—Teacher rating
<b>Gottfredson, Gottfredson, &amp; Hyble (1993)</b> (journal) <i>Nonequivalent control group</i>	Treatment ( $n = 6$ ) Comparison ( $n = 2$ )	Gr. 6-8 3 years (1 year baseline)	Teacher training on classroom management based on Evertson COMP	Student Behavior -Teacher ratings of disruptive, on-task
<b>Evertson (1988)</b> (research report) <i>Random individual matched</i>	Treatment ( $n = 14$ ) Control ( $n = 15$ )	Gr. K-9 4 months	COMP -Teacher training	Student Engagement -Observations and ratings of percent of students engaged
<b>* Evertson et al. (1988-1989)</b> (research report) <i>Random individual matched</i>	Treatment ( $n = 15$ ) Control ( $n = 15$ )	Gr. K-6/Res 4 months	COMP -Teacher training	Student Behavior Classroom Activity Record -Disruptive, Inappropriate
<b>* Evertson et al. (1989-1990)</b> (research report) <i>Random individual matched</i>	Treatment ( $n = 13$ ) Control ( $n = 10$ )	Gr. K-6/Res 4 months	COMP -Teacher training	Student Behavior Classroom Activity Record -Disruptive -Inappropriate -Off-task

Study/Journal/Design	Participants	Grade/ Duration	Program Description	Outcome Measures
* Evertson et al. (1992-1993) (research report) <i>Random individual matched</i>	Treatment ( $n = 10$ ) Control ( $n = 11$ )	Gr. 2-12 4 months	COMP -Teacher and mentor training	Student Behavior Classroom Activity Record -Disruptive -Inappropriate -On-task/off-task
*Evertson et al. (1992-1993) (research report) <i>Random individual matched</i>	Treatment ( $n = 13$ ) Control ( $n = 8$ )	Gr. K- 12/Rem 4 months	COMP -Teacher and mentor training	Student Behavior Classroom Activity Record -Disruptive -Inappropriate -On-task/off-task
* Evertson et al. (1993-1994) (research report) <i>Random individual matched</i>	Treatment ( $n = 15$ ) Control ( $n = 15$ )	Gr. K-5/Res 4 months	COMP -Teacher training	Student Behavior Classroom Activity Record -Disruptive -Inappropriate -On-task/off-task
* Evertson et al. (1993-1994) (research report) <i>Random individual matched</i>	Treatment ( $n = 13$ ) Control ( $n = 10$ )	Gr.6-12 4 months	COMP -Teacher training	Student Behavior Classroom Activity Record -Disruptive -Inappropriate -On-task/off-task

Note.  $n$  = Number; \* = Data obtained from a research report (Evertson et al., 1995) and not original published study; Gr. = Grade; Res. = Resource room; Rem = Remedial; COMP = Classroom Organization and Management Program; CBCL TRF = Child Behavior Check List Teacher Report Form (Achenbach, 1991); ADH = Attention Deficit Hyperactivity; ODD = Oppositional Defiant Disorder; PBSI = Problem Behavior at School Interview (Erasmus Medical Center, 2000); TOCA-R = Teacher Observation of Classroom Adaptation-Revised (Werthamer-Larsson, Kellam, & Wheeler, 1991).

## Appendix B: Characteristics of excluded studies

Study Citation	Grade/Ages	Treatment
Brown, J. H., Frankel, A., Birkimer, J. C., & Gamboa, A. M. (1976). The effects of a classroom management workshop on the reduction of children's problematic behaviors. <i>Corrective &amp; Social Psychiatry &amp; Journal of Behavior Technology, Methods &amp; Therapy</i> , 22, 39-41.	Grade 1-6	Teacher training
Reason for exclusion	Used individual behavior management instead of whole-class management.	
Emmer, E. T. and others (1983). Improving junior high classroom management. (Report NoSP-022-953). Austin, TX: Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED 234 021)	Grade 6-8	Teacher training
Reason for exclusion	Does not include student data sufficient to compute an effect size.	
Evertson, C. M., Emmer, E. T., Sanford, J. P., & Clements, B. S. (1983). Improving classroom management : An experiment in elementary school classrooms. <i>The Elementary School Journal</i> , 84, 172-188.	Grade 1-6	Teacher training

Study Citation	Grade/Ages	Treatment
Reason for exclusion	Does not provide student data of classroom behavior	
Evertson, C. M. & Smithey, M. W. (2000). Mentoring effects on protégés classroom practice: An experimental field study. <i>The Journal of Educational Research</i> , 93, 294-304.	Grade 1-12	Teacher mentoring workshop
Reason for exclusion	Group assignment was at mentor level but classroom teachers were implementers. No control for pretest but differences evident.	
Vuijk, P., van Lier, P. A. C., Huizink, A. C., Verhulst, F. C., & Crijnen, A. A. M. (2006). Prenatal smoking predicts non-responsiveness to an intervention targeting attention-deficit/hyperactivity symptoms in elementary schoolchildren. <i>Journal of Child Psychology and Psychiatry</i> , 47, 891-901.	Ages 7-11	Good Behavior Game
Reason for exclusion	Uses the same data as included study van Lier et al. (2004).	
Vuijk, P., van Lier, P. A. C., Crijnen, A. A. M., & Huizink, A. C. (2007). Testing sex-specific pathways from peer victimization to anxiety and depression in early adolescents through a randomized intervention trial. <i>Journal of Affective Disorders</i> , 100, 221-226.	Age 7	Good Behavior Game
Reason for exclusion	Dependent variable is anxiety and victimization.	
van Lier, P. A. C., Vuijk, P., & Crijnen, A. A. M. (2005). Understanding mechanisms of change in the development of antisocial behavior: The impact of a universal intervention. <i>Journal of Abnormal Child Psychology</i> , 33, 521-535.	Grade 6 follow-up	Good Behavior Game
Reason for exclusion	Follow-up data from included study van Lier et al., 2004.	

Study Citation	Grade/Ages	Treatment
Olexa, D. F., & Forman, S. G. (1984). Effects of social problem-solving training on classroom behavior of urban disadvantaged students. <i>Journal of School Psychology, 22</i> , 165-175.	Grade 4-5	Social problem-solving, response cost, social problem-solving with response cost
Reason for exclusion	Treatment is not provided in the subjects' primary classroom. Treatment is conducted in a small group pull-out.	
Kellam, S. G., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. <i>Journal of Child Psychology and Psychiatry, 35</i> , 259-281.	Grade 6 follow-up	Good Behavior Game
Reason for exclusion	Follow-up data from included study Dolan et al., 1993.	
Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior in middle school. <i>Development and Psychopathology, 10</i> , 165-185.	Grade 1-6 follow-up and re-analysis	Good Behavior Game
Reason for exclusion	Re-Analysis of data from included study Dolan et al., 1993 to determine effects of classroom level aggression.	
Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence. <i>Journal of Emotional and Behavioral Disorders, 9</i> , 146-160.	Grade 6 follow-up	Classroom-centered
Reason for exclusion	Follow-up data from included study Ialongo et al., 1999.	

Study Citation	Grade/Ages	Treatment
Besasil-Azrin, V., Azrin, N.H., & Armstrong, P. M. (1977). The student-oriented classroom: A method of improving student conduct and satisfaction. <i>Behavior Therapy</i> , 8, 193-204.	Grade 5	Multi-component "student-oriented"
Reason for exclusion	Treatment and control group were in same classroom.	

---

# Appendix C: Coding Manual

## Details of Study Coding Strategies

The following fields will be used to code and extract data from each article.

### A. Study Identifiers

- Study author(s) (lastname, first)
- Year of publication (four digits)
- Country in which study was conducted
  1. USA
  2. Canada
  3. Britain
  4. other English speaking
  5. other non-English speaking
  9. cannot tellother: \_\_\_\_\_
- Type of publication
  1. book
  2. journal article
  3. book chapter (in an edited book)
  4. thesis or dissertation
  5. technical report
  6. conference paper
  7. other
  9. cannot tell

### B. Study Context

- Study setting
  1. public school
  2. private school
  3. both
  9. cannot tell
- Study location
  1. urban
  2. suburban
  3. rural
  4. other: \_\_\_\_\_
  5. mix
  8. cannot tell

### C. Sample and Assignment procedures

Record the number of participants that participated in the treatment condition under “Observed N Tx” column and the number of participants that participated in

the control condition under the “Observed N Control” column. The observed number constitutes the number of participants that actually completed each condition after attrition. If the specific subgroup categories were not broken down by treatment and control condition, record ‘999’ under the columns for treatment and control and the total number under the “Observed Total” column.

	Observed N Tx	Observed N Control	Observed Total
<i>Total Sample:</i>			
<i>By subgroup: (code 999 if not identified specifically)</i>			
Gender			
Male			
Female			
Race/Ethnicity			
African American			
Latino			
White			
Asian/Pacific Islander			
Mixed			
Unknown			
Free/Reduced Lunch			
Other demographic factor:			

- Participants' mean Age
  1. \_\_\_\_\_ (e.g., 11.5)
  9. cannot tell
- Unit of assignment
  1. individual
  2. classroom
  3. school
  9. cannot tell
- Sampling procedure
  1. convenience sample
  2. random sample
  3. self-identified
  4. other: \_\_\_\_\_
  9. cannot tell
- Assignment procedure
  1. random assignment
  2. quasi-random
  3. non-equivalent
  4. matching on pretest only
  9. cannot tell
- How much attrition was evident in the study?
 

Original (prior to treatment) N = \_\_\_\_\_

Observed (completed treatment) N = \_\_\_\_\_

Percentage of attrition \_\_\_\_\_%

999 cannot tell

## D. Conditions

### *Control or Comparison Group*

- Characteristic identified in study
  1. treatment as usual
  2. other: \_\_\_\_\_

### *Characteristics of Focal Treatment*

- Describe the treatment based on the description from the study as closely as possible:
- Characteristics of focal treatment
  - \_\_\_ behavioral strategies (e.g., group contingency, positive reinforcement)
  - \_\_\_ cognitive strategies (e.g., problem solving)
  - \_\_\_ interpersonal/social skills (e.g., communication, refusal skills)
  - \_\_\_ a combination of strategies
  - \_\_\_ Good Behavior Game (GBG)
  - \_\_\_ Classroom Organization and Management Program (COMP)
- Location of treatment (i.e., where treatment is delivered)
  1. regular classroom
  2. special education classroom
  3. both
  9. cannot tell
- Type of program for treatment
  1. Research or demonstration project that involves a high level of involvement from the researcher(s).
  2. Evaluation of “real-world” or routine program (Practice/treatment that is initiated and implemented by school although researcher is involved with collecting data and evaluation.).
- Treatment agent (code general or special education teacher even if teachers were the “participants” of the study and initially trained by researcher)
  1. general education teacher
  2. special education teacher
  3. both
- Duration of treatment. Approximate or actual number of weeks. Divide days by 7; multiply months by 4.3; round to whole number. Code 999 if cannot tell. \_\_\_\_\_
- Additional treatments provided (identify all that apply and code yes=1 or no=0)
  1. parent training
  2. school structural changes
  3. medication
  4. counseling/therapy
  5. academic
  6. other: \_\_\_\_\_
- Was the treatment implemented with fidelity?
  1. Yes
  2. Maybe
  3. No
  4. can't tell

## E. Dependent Variables

At least one dependent variable *must* measure disruptive, inappropriate, or aggressive behavior **in** the classroom.

- The name of the measure as identified in the study should be written under each measure.
- The construct being measured should be identified for each measure (i.e., disruptive behavior, aggressive behavior, classroom level behavioral climate, inattentive, internalizing behavior, externalizing behavior).
- How the measure was administered should be identified for each dependent variable (i.e., parent report, teacher report, observation data, interview, standardized).
  1. parent report
  2. teacher report
  3. observation data
  4. standardized test
  5. interview
  9. cannot tell

		<u>Measure 1</u>	<u>Measure 2</u>	<u>Measure 3</u>	<u>Measure 4</u>	<u>Measure 5</u>	<u>Measure 6</u>
Name of Measure							
Construct being measured							
Control	Pre						
	Post						
	SD						
	N						
Tx	Pre						
	Post						
	SD						
	N						
ES (type)							
Exact P value							
Reliability Coefficient							
Type of Reliability Coefficient							
How was measure administered?							

Note: The treatment and control total N is also recorded in the “C. Sample and Assignment Procedures” section above