Summer 8-2014

Evaluating The Effectiveness of the Good Behavior Game with General Education High School Students Utilizing a Changing Criterion Component

Rachel Ritter Mitchell
University of Southern Mississippi

The University of Southern Mississippi

EVALUATING THE EFFECTIVENESS OF THE GOOD BEHAVIOR GAME WITH GENERAL EDUCATION HIGH SCHOOL STUDENTS UTILIZING A CHANGING CRITERION COMPONENT

by

Rachel Ritter Mitchell

Abstract of a Dissertation
Submitted to the Graduate School
of The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

August 2014

ABSTRACT

EVALUATING THE EFFECTIVENESS OF THE GOOD BEHAVIOR GAME WITH GENERAL EDUCATION HIGH SCHOOL STUDENTS UTILIZING A CHANGING CRITERION COMPONENT

by Rachel Ritter Mitchell

August 2014

The purpose of this study was to evaluate the effectiveness of the interdependent group contingency procedure known as the Good Behavior Game (GBG) with upper-level (i.e., 10th, 11th, and 12th grade) general education high school students utilizing a changing criterion design. The effectiveness of the GBG has been investigated with a variety of behaviors across many developmental levels; however, limited research has been done at the high school level. To date, only a few studies have examined the effectiveness of the GBG with a general education high school population, one with a single 9th grade classroom (Kleinman & Saigh, 2011) and one unpublished thesis utilizing an ABAB design across three high school classrooms, consisting mostly of 9th grade students (Mitchell, 2012). The present study adds to the literature base of the GBG by extending the versatility of the GBG to a broader age range of general education high school population with specific attention paid to older students as well as the criterion component of the GBG. These effects were evaluated across three classrooms, which all demonstrated decreases in disruptive behaviors during intervention phases (which were affected by the criterion level) as well as increases in academic engagement. Additionally, teachers found the GBG/TC to be acceptable for use in their classrooms.

These results support the use of a modified version of the GBG in high school classrooms.

The University of Southern Mississippi

EVALUATING THE EFFECTIVENESS OF THE GOOD BEHAVIOR GAME WITH GENERAL EDUCATION HIGH SCHOOL STUDENTS UTILIZING A CHANGING CRITERION COMPONENT

by

Rachel Ritter Mitchell

A Dissertation
Submitted to the Graduate School
of The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved:

Daniel Tingstrom
Committee Chair

Brad Dufrene

Keith Radley

Sara Jordan

Maureen Ryan
Dean of the Graduate School

August 2014

ACKNOWLEDGMENTS

First and foremost I would like to thank my committee chair, Dr. Daniel H. Tingstrom, “the detail man,” for his diligence, encouragement, and guidance throughout this project. I would also like to thank my committee members Drs. Brad A. Dufrene, Keith Radley, and Sara Jordan for their contributions to this project as well as my data-collector extraordinaires Emily Ness and Zachary LaBrot for their time and assistance. Lastly, I would like to thank my friends and family who have supported me throughout this process, especially my husband, Derek, for all of his unfailing patience and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF ILLUSTRATIONS

Figure

CHAPTER I

INTRODUCTION

Disruptive behaviors present a daily struggle for many educators and addressing them often leads to a decrease in instructional time, resulting in lowered academic performance by students (Canter, Paige, Roth, Romero, & Carroll, 2004). The techniques a teacher employs to handle disruption in his or her classroom are integral components of a consistent and effective learning environment for all students (Evertson & Emmer, 1982). Notably, effective classroom management is the ability of a teacher to handle inappropriate student behavior, manage events, and progress through instruction while maintaining the focus of the class (Kounin, 1970). Teachers mastering this skill result in maximized instruction and student attentiveness for their classrooms. Unfortunately, some teachers are not able to maintain well managed classrooms on their own and need additional support or training in various techniques to assist in achieving classroom management.

Furthermore, some researchers have found that many teachers report that they more frequently used corrective rather than preventative strategies for dealing with problem behavior and also that teachers did not use effective strategies in their classrooms or receive adequate support for handing challenging behavior (Reupert & Woodcock, 1970; Westling, 2010). Specifically, many teachers use a reactive method; that is, they wait until a student has engaged in inappropriate behavior to address it and then provide a consequence (e.g., office disciplinary referral), thereby taking class time from the group as a whole to focus on one student, resulting in lost instruction time (Milner & Tenore , 2010; Tillery, Varjas, Meyers, & Collins, 1992). Reactive methods

often take more time and do not actually prevent the behavior from occurring in the future, while preventative methods tend to eliminate the problem before it occurs. Procedures that focus on prevention of disruptive behaviors and maintaining appropriate behavior at the class-wide level are consistent with Positive Behavioral Intervention and Support systems (PBIS), a system popular in many school districts (Sayeski & Brown, 2011; Sugai et al., 2000).

PBIS involves the use of school-wide and individualized strategies that work to improve the social and learning outcomes of students while simultaneously preventing problematic behaviors in all students (Sugai et al., 2000). PBIS utilizes a three-tiered system of behavioral management, including a number of key features such as a prevention-focused continuum of support, proactive instructional approaches to teaching and improving social behaviors, conceptually sound and empirically validated practices, systems change to support effective practices, and data-based decision making (Sugai & Horner, 2002). The utilization of the tier system allows for a structured support for all students and provides more specialized or individualized support for students who are elevated throughout the tiers and are in need of additional services. The tiers work in a hierarchy, with Tier I offering universal supports for all students and focusing on primary prevention at the school and class-wide level. Students needing additional support will be elevated to Tier II which is used for smaller specialized groups of students who exhibit at-risk behaviors. Tier III is employed for students designated as having high-risk behaviors and utilizes a specialized intervention on the individual level.

One way to incorporate PBIS at the Tier I level and to incorporate group rather than individual consequences is by using group contingencies in the classroom. There

are several potentially problematic issues with focusing on individual behaviors in a classroom setting. One problem is that some students find attention (even social disapproval or reprimands) to be reinforcing. For these students, it is necessary to have additional contingencies in place for appropriate behavior in order to avoid disruption (Shaw & Simms, 2009). Three forms of group contingencies exist: dependent group contingences, independent group contingencies, and interdependent group contingencies (Litow & Pumroy, 1975). A dependent group contingency offers the same response to all group members, however, is contingent on only select members (e.g., offering the entire group a reward as long as the lowest score on a test is above a certain criterion) (Litow & Pumroy, 1975). An independent group contingency focuses more on an individual contingency, and although the same criterion applies to all group members, access to reinforcement is contingent on each individual's performance (Theodore, Bray, Kehle, & Dioguardi, 2004). A classroom with an independent group contingency might offer a reward to each student who scores a set grade or higher on a test. Similar to a dependent group contingency, an interdependent group contingency offers the entire group access to the same consequence but also requires collective behavior to access reinforcement (Tingstrom, Sterling-Turner, & Wilczynski, 2006). There are several variations of interdependent group contingencies with regard to classroom implementation; for example, every member of the group may be required to meet the set criterion, the group may be expected to achieve a certain average, or a certain percentage of group members may have to meet the criterion in order to gain access to reinforcement (Theodore et al., 2004).

Group contingencies may be superior to individual interventions for classroom management, are time efficient, and are effective across a broad range of populations (Theodore et al., 2004). Gresham and Gresham (1982) demonstrated success with group contingencies with 6- to 10-year-old students with mild intellectual disabilities in a self-contained classroom. In this study, the authors employed an A/B/C/D/A/B/C/D design to examine the success of all three group contingencies. All of the group contingencies improved classroom behavior. The dependent and interdependent group contingencies were more dramatic in their effects and both effectively decreased the disruptive behavior in the classroom, with the most dramatic decrease resulting during the interdependent contingency.

The Good Behavior Game (GBG) is a popular and successful procedure utilizing an interdependent group contingency. The GBG's easy implementation in classroom settings and success at lowering target behaviors has led to its popularity among teachers (Tingstrom et al., 2006). In the original study (Barrish, Saunders, & Wolf, 1969), a teacher divided her fourth-grade class into two teams. The teacher informed the students that they would play a game during a specific class period each day and if a student broke one of the rules, that student's team would receive a mark. The teacher displayed the marks on a blackboard for the class to see, and the team with the fewest marks won the game. For the game, a criterion of five marks was set. If both teams received no more than five marks, both teams won the game, and members of both teams were allowed access to the privileges. Target behaviors included out-of-seat and talking-out. During baseline, the median number of intervals with talking-out behavior was 96%, and out-of-seat was 82%. After implementing the game in the classrooms, talking-out decreased to

19% and out-of-seat to 9%. The drastic decreases in target behaviors during GBG phases help to establish a functional relationship between the implementation of the game and the decrease in occurrences of target behaviors. The GBG has since been identified as a well-established strategy for managing disruptive/aggressive behavior in the elementary school classroom setting (Franklin, Harris, & Allen-Meares, 2006).

The GBG has many advantages when implemented as a classroom management procedure. The game offers reinforcement for teams of students who do not exceed a criterion or exhibit the least amount of inappropriate behavior. Therefore, the GBG differentially reinforces low rate behaviors by rewarding students for exhibiting a low frequency of target inappropriate behavior during the class period (Litlow & Pumroy, 1975). In some cases, however, appropriate behavior is targeted, and differential reinforcement of incompatible behavior is used to reduce inappropriate behavior. In such cases the student is reinforced for the appropriate and incompatible, meaning he or she cannot exhibit the inappropriate behavior and the appropriate behavior simultaneously (Tingstrom et al., 2006).

Although the GBG is best known for its success in the classroom with inappropriate behaviors, since Barrish et al.'s original study, others have successfully adapted the GBG for use with a variety of behaviors in various settings. Numerous authors have created and applied modifications of the GBG and have been able to successfully address a variety of target behaviors including oral hygiene (Swain, Allard, & Holborn, 1982) and work productivity (Lutzker & White-Blackburn, 1979). Although there are numerous studies examining the effectiveness of the GBG with a wide array of participants, there is little literature on applications of the game with general education

high school students or that investigate the effect of the criterion component. Although modified versions of the GBG have shown versatility, yielding success with children as young as preschool and also with students in a general education sixth grade physical education class (Hunt, 2010; Patrick, Ward, & Crouch, 1998; Swiezy, Matson, & Box, 1992), the procedure has been extended to older students in regular education classrooms in only one published (Kleinman & Saigh, 2011) and one unpublished study (Mitchell, 2012). An additional component which has a limited research base regarding the GBG is the effect of the criterion component. While many studies have implemented and changed criteria across the study, few have investigated the effect of removing the criterion or including an "unknown/mystery" criterion component (McCurdy, Lannie, & Bardabas, 2009; Tingstrom, et al. 2006).

Though supportive of the effectiveness of the GBG in a high school setting, the Kleinman and Saigh (2011) study is limited in that it includes only one classroom. More studies, specifically ones with rigorous designs are needed to determine the effectiveness of the GBG at the high school level. The following section, organized by type of research design used to evaluate the intervention, includes a brief review of studies demonstrating the effectiveness of the GBG and its modifications.

Although the below review will be restricted to single-subject methodology and design, it is worth nothing that in addition to these reviewed studies, a number of studies on the GBG utilizing group designs exists (Tingstrom et al., 2006). While these studies do add to the literature base for effectiveness of the GBG, analysis of single-case design and group design utilize differing analyses. Larger sample numbers in group studies allow for extensive statistical analysis that may not be appropriate and/or meaningful for

single case designs, and therefore cannot be directly compared. Additionally, practitioners do not typically have the luxury of large groups of participants and must evaluate their interventions with one or a few participants at a time, hence the emphasis on single-subject methodology literature in the review below. As with studies employing single-case designs, the results of studies utilizing group designs suggest the GBG is effective in achieving desired outcomes (i.e., lowering inappropriate behaviors) across classrooms and schools. Designs used in these studies have included, but are not limited to, group repeated measures, group factorial designs, matched subjects designs, and group repeated measures longitudinal designs (Dolan et al., 1993; Kellam & Anthony, 1998; Kellam, Ling, Merisca, Brown, & Ialongo, 1998; Kellam, Rebok, Ialongo, & Mayer, 1994; Poduska et al., 2008).

Good Behavior Game

Simple A/B Phase Change Design

One of the simplest designs used to evaluate a study is the A/B design. This design consists of only two phases, the baseline and intervention phase. An A/B design is considered to be a weaker design than many of the others used because there are only two phases, making it impossible to account for any confounds or alternative explanations that may occur simultaneously with the phase change (Kazdin, 1982). While this design has limited experimental control, some researchers have used it to evaluate the effectiveness of the GBG.

One such study utilizing an A/B design is by Robertshaw and Hiebert (1973) in which the authors used a modified version of the GBG called the Good Astronaut Game with a first grade classroom. The authors used the game to increase attention-to-task

behavior in 24 students and collected independent data on one target student as well. Unlike the original study by Barrish et al. (1969), the students were divided into four teams. Rather than earning marks against the team, each time students completed a worksheet they earned a token for their team. Tokens were displayed on a board using spaceships, and the board was filled with other space themed pictures. At the end of the day, teams were able to move their spaceship according to the tokens earned, and the team with the most tokens was able to access reinforcers.

Data were collected during 30-minute seatwork periods. Data collectors recorded the target student's attentive behaviors and the number of worksheets completed by the target student as well as the entire class. Results indicated an increase in both attentiveness to work and number of worksheets completed; however, the use of an A/B design weakens the study.

Multiple Baseline Design

A popular and much stronger research design used to evaluate interventions and treatments is the Multiple Baseline design. This design involves collecting baseline data across various components (i.e., classrooms, behaviors, etc.). After collecting baseline data, the intervention is implemented in the first condition while the baseline data collection continues for the remaining conditions. Because the baseline data collection continues for remaining conditions, it is possible to say that any effects are not related to the timing of the implementation. After a clear treatment effect is evident, the treatment or intervention is implemented in the next condition, and so on. Though similar to a simpler A/B design, this design is considered much stronger because it demonstrates an

effect across multiple conditions and also because the "multiple baseline" component eliminates confounds related to time (Kazdin, 1982).

In a variation of the GBG, Maloney and Hopkins (1973) implemented a version of the GBG, called the Good Writing Game, which targeted improving compositional variables of stories written during a non-remedial summer school session with 14 students from fourth, fifth, and sixth grades. Each day the students had to exceed the set criteria in order to earn the reward. Also, if the students were not able to exceed the criteria, the students were still able to access the reward if the point difference between the two teams was less than 100 points.

The authors used a multiple baseline across parts of speech and sentence structure to improve three dependent variables: the number of different adjectives within written stories, the number of different action verbs within stories, and the number of different sentence beginnings. For the first phase of the Good Writing Game, winning was contingent on the number of different adjectives in the students' written stories while baseline data were collected for the other two target variables. During the second phase, the contingency only changed to only the number of different action verbs within the stories, and again all variables were tracked. During the third phase winning was contingent on the number of different adjectives, action verbs, and the number of different sentence beginnings. The results of the study indicate that the Good Writing Game was effective in improving the number of targeted parts of speech and sentence structure.

Johnson, Turner, and Konarski (1978) designed a study examining the effectiveness of the GBG in decreasing disruptive behavior and increasing academic

behavior in students who had been removed from regular classes due to low achievement-motivation. The authors implemented the game in a third and fourth grade transitional class. The authors targeted appropriate behavior, disruptive behavior, and teacher attention in the study. A multiple baseline design across settings, subjects, and time periods was used to evaluate the effectiveness of the GBG. The third grade class was divided into three teams, and the fourth grade class was divided into two teams based on the teachers' preference. The teachers also elected captains from each team to be responsible for counting the marks each day. Results from the game showed a dramatic decrease in disruptive behavior as well as teacher attention to disruptive behavior.

In a study using a multiple baseline across classrooms design, Swain et al. (1982) used a modified version of the GBG, called the Good Tooth Brushing Game, to improve dental hygiene with 22 first graders and 23 second graders. Oral hygiene was calculated using dissolving red tablets that colored to display any debris on the children's teeth; the Simplified Oral Hygiene Index (OHI-S) was used to calculate the amount of debris. Initially, all students were given a toothbrush, toothpaste, and red tablets and were given information on proper oral hygiene. During baseline, four students were randomly selected, without replacement, to have their teeth examined. Using the OHI-S, the amounts of the students' debris were determined. After baseline data were collected, the game was introduced to the students. Each day during the Good Tooth Brushing Game four children were, again, randomly selected to represent their team. Whichever team had the cleanest teeth would win for the day. Results suggest that the introduction and use of the game decreased the amount of debris and increased oral hygiene for the participants.

In yet another variation of the GBG, Patrick et al. (1998) utilized a multiple baseline design across fourth, fifth, and sixth grade classes of students targeting the occurrence of appropriate and inappropriate social behaviors and appropriate skill attempts during a volleyball unit in physical education classes. Like the original study, students were divided into teams; however, unlike the original study, students earned points based on displaying appropriate behaviors. The game also featured a response cost component in which the display of inappropriate behaviors resulted in the removal of points previously earned. The intervention showed an immediate increase in appropriate behavior and a decrease in inappropriate behavior; however, there was no change in correct or incorrect skill attempts by the students.

McCurdy et al. (2009) utilized a multiple baseline design across three lunch periods to evaluate another modified version of the GBG, called the Lunchroom Behavior Game (LBG), in the cafeteria of an urban elementary school. The study targeted disruptive behavior to include out-of-seat, play fighting, physical contact with force, throwing objects, and screaming. During the game, school staff developed expectations for the cafeteria, taught these expectations, and explained the LBG to their students one week prior to implementation. Teams for the LBG were comprised of entire classes. Staff members monitored the behaviors of the students and tallied marks on a large dry-erase poster. Mystery weekly criteria were set, and teams who did not exceed the criterion were announced over the intercom the following Monday. In addition, winners with the lowest points from each lunch period received rewards which included edible items, small tangibles, and certificates for movie time and class parties. Results of the

LBG were consistent with other studies of the GBG and resulted in clear and immediate decreases in disruptive behavior.

In one of only a few studies utilizing preschool students, Hunt (2010) used a nonconcurrent multiple baseline design across three preschool, Head Start classrooms to assess the effectiveness of the GBG in decreasing disruptive behavior. Participants included three target students as well as classroom peers. Hunt targeted inappropriate vocalizations, noncompliance, and aggression and used the original version of the GBG, as introduced by Barrish et al. (1969), introducing the game following baseline data collection. The GBG was effective in decreasing disruptive behavior across classrooms as well as with target students. According to Hunt, procedural integrity was quite low at times (67%) which required additional teacher training after implementation. Additionally, the criterion for the GBG was not based on baseline data and was, therefore, set too high at times, becoming unattainable for some students.

Hunt (2012) further examined the GBG's effectiveness within a preschool sample and evaluated the effects of the intervention on decreasing disruptive behaviors as well as increasing appropriate academic behaviors. Again Hunt utilized a multiple baseline design across three Headstart classrooms. Hunt's results indicated that the GBG successfully decreased disruptive behaviors and increased academic engagement within all three of the participating classrooms. Additionally, the GBG decreased disruptive behavior for three target students and increased academic engagement for two target students.

A/B/A/B Phase Change Design

Perhaps one of the most popular designs used in studies of the GBG (along with multiple baseline design or some combination of the two) is the ABAB phase change design (Tingstrom et al., 2006). This design involves the collection of baseline data (A phase), the intervention phase (B phase), a withdrawal phase (A phase) in which the intervention is removed, and then a re-implementation phase (B phase) in which the intervention is reintroduced (Kazdin, 1982). This design offers controls for internal validity because there are two separate implementations and therefore two separate opportunities for a treatment effect. Additionally, the inclusion of a baseline and withdrawal phase offer a prediction of what the behavior would have looked like had there been no implementation (Kazdin, 1982).

A study conducted by Medland and Stachnik (1972) partially replicated Barrish et al.'s (1969) investigation and also included systematic analysis of the GBG using a slightly modified version with a 5^{th} grade reading class. The effectiveness of the intervention in decreasing talking-out and out-of-seat behaviors was evaluated using an A/B/A/C/D/B phase change design which examined three separate interventions (B phase, C phase, and D phase). The game components consisted of rules, light indicators, and group consequences of extra recess or free time. Modifications to the game included the use of a light indicator, which provided feedback for when an error (e.g., inappropriate vocalization or being out of seat) was being made, in addition to the traditional marks utilized by Barrish et al. (1969). Additionally, students on the losing team, were allowed the opportunity, if necessary, to vote out a teammate for one day whom they believed to be a saboteur of the game.

The A/B/A/C/D/B design included six phases: $baseline_1$, $game_1$, $baseline_2$, rules, rules + lights, and $game_2$. Following implementation of the modified GBG ($game_1$), a withdrawal phase occurred ($baseline_2$) during which the game was not played, and the teacher utilized past classroom management procedures as she had during baseline. The following phase (rules) involved explaining only the rules with no game or related consequences. The next phase (rules + lights) involved explaining the rules as well as the light indicator feedback but did not involve the game or any of its consequences. For the final phase ($game_2$), the modified GBG was reintroduced to the classroom. Results indicated that both phases of the game produced meaningful change in all of the targeted behaviors. Results also indicated similar results in the rules + lights phase. A limitation of the study includes possible sequence or order effects, as well as possible carry-over effects, specifically regarding the rules + lights phase to the $game_2$ phase.

Bostow and Geiger (1976) conducted another early study using an A/B/A/B design in which they used the GBG with second graders to decrease out-of-seat behavior, talking-out behavior, lack of attention to assigned task, and bothering neighbors. The authors did not incorporate any modifications to the game. As in Barrish et al. (1969), the class was divided into two teams, and each time a student engaged in a target behavior he/she earned a mark against their team. Results showed a decrease in all targeted disruptive behavior.

In another unique take on the original study, Darch and Thorpe (1977) used a modified GBG called the Principal Game to increase on-task behavior. Unlike the traditional version of the GBG, the Principal Game targeted on-task behavior rather than off-task behavior. The study focused on the 10 students in a fourth grade classroom who

had been ruled “the most deviant” (p. 342) by their teacher. During the Principal Game, the class was divided into five teams based on the seating chart. During the game, students were given six opportunities to earn a point. In order to earn a point, the entire team had to be on-task when a timer sounded. Regardless of the performance of other teams, each team could win the game by earning five of the possible six points. The school principal delivered positive attention to the teams who won the Principal Game at the end of the period.

The effectiveness of the Principal Game was evaluated using an A/B/A/C/A design, a slight variation on the A/B/A/B design by including a second intervention (C phase). After collecting baseline data, the Principal Game was introduced during the second phase. Following the introduction of the game, a withdrawal phase was implemented. During the fourth phase, an independent contingency was put in place (C phase). During this time students earned consequences for individual behavior, and those behaviors were in no way impacted by the group. This phase was followed by an additional withdrawal phase in which all interventions were removed. Results indicated that the Principal Game increased on-task behavior above baseline levels as well as above levels during the independent contingency.

In a study by Lutzker and White-Blackburn (1979), the GBG was applied using four state hospital residents. The behaviors targeted for the residents included work output, on-task behavior, and staff attention. The game was evaluated using an A/B/A/C/A withdrawal design. The game was implemented during the second phase, and a feedback-only phase was implemented during the fourth phase. The game

produced higher levels of productivity and work output during the GBG phase compared to both baseline phases as well as the feedback-only phase.

In an investigation of the effectiveness of group contingencies, Gresham and Gresham (1982) implemented three separate programs during different phases with a self-contained classroom. The authors used an A/B/C/D/A/B/C/D design to compare the different forms of group contingency. During baseline, the classroom was conducted using the teacher's normal procedures. During the B phase, the GBG was implemented with no modifications to the Barrish et al.'s original study (1969). The following phase (c) utilized a dependent group contingency. At this time, the two most disruptive students were appointed as team captains. If the team captains exhibited five or more disruptive behaviors, the teams were not able to access reinforcers. During the D phase, the authors employed an independent contingency in which students competed independently for reinforcement. As long as a student received fewer than five marks, the student received the reinforcer (Gresham & Gresham, 1982).

The experimenters' use of an A/B/C/D/A/B/C/D design resulted in some limitations in the interpretation of results produced. Although the results showed that the interdependent and dependent contingency phases produced more desirable results for decreasing disruptive behavior, the design was not counterbalanced. Because the phases occurred in the same order, it is not possible to know whether order effects or carry-over effects contributed to the results. Furthermore, phase changes occurred as a function of time, further limiting the ability to make phase to phase comparisons.

An A/B/A/B design was also used to evaluate the GBG cross-culturally, as demonstrated by Saigh and Umar (1983) with second graders in Sudan, Africa. In this

study, the authors targeted aggression, seat leaving, and talking. As in the original study, students were divided into two teams, and each time a student violated any of the rules, the student's team received a mark (Barrish et al., 1969). The authors used victory tags, recess, and a sticker chart as rewards for the game. During the baseline phase, aggression occurred 8.5% of the intervals, talking occurred 12% of the intervals and seat leaving occurred 9.6% of intervals. After implementing the game, aggression decreased to 3.5%, talking to 4.7%, and seat leaving to 1.7% of intervals. Each additional phase change demonstrated that the game was associated with a decrease in all target behaviors. The results of this study demonstrate that the GBG can be used effectively to manage behavior across cultures.

One criticism of the GBG is that some view the distribution of marks as punishment for inappropriate behavior that does not teach or reinforce appropriate alternative behaviors. In an effort to address these limitations of the GBG, Darveaux (1984) modified the GBG by pairing it with a token reinforcement system in an effort to reinforce appropriate behaviors. This modification, the Good Behavior Game Plus Merit, was used both to reduce disruptive behavior and improve assignment completion. Although the game was used in a second grade classroom of 24 students, the target participants were two boys who were at high-risk for placement in a behaviorally impaired class. The class was divided into two teams, with one of the target students on each team. Like the original study conducted by Barrish et al. (1969), when students performed the targeted disruptive behavior their team received a mark; however, the game varied due to the merit portion of the modification. Students were awarded merit cards contingent on 75% accuracy of class work and participation. Once a team earned

five merits, they were able to remove one of the inappropriately earned marks against their team.

Both disruptive behavior as well as assignment completion were tracked. Using an A/B/A/B design, results from the study showed that during the GBG + Merit package phases, there was a dramatic decrease in disruptive behaviors and increase in assignment completion for the target students as well as an increase in assignment completion for the entire class compared to baseline phases. Following the removal of the intervention from the class, disruptive behavior and assignment completion returned to levels similar to those during baseline.

In another study utilizing special education students, Davies and White (2000) used a modified version of the GBG to target inappropriate vocalizations with four third grade students diagnosed with Attention Deficit Hyperactivity Disorder (ADHD) as well as their classroom peers. The study involved matching the four target students with ADHD with comparison peers without ADHD diagnoses. This modified version of the GBG incorporated a self-regulating component that involved a chart located in the middle of the classroom tables. These charts were divided into three sections: a green section, a blue section, and a red section, and each team had five Velcro dots on the charts. If a member of the team displayed one of the targeted inappropriate behaviors, he or she was to move a dot from the green section to the blue section. If the student did not move the dot appropriately, the teacher moved the dot into the red section. At the end of the game, each team needed at least one dot in the green section to win the reward. In addition to the group chart, each child had individual charts to track which students were responsible for each behavior. At the end of each day, students received teacher and peer feedback

on their performance. An A/B/A/B withdrawal design was used to assess the effects of the intervention. Results indicated that the target behavior decreased for all students. A major limitation was that mid-way through the study, the classes changed, three of the matched peers were moved to another classroom, and new peers were selected. Additionally, no treatment integrity data were collected.

In a study focusing on academic behaviors, Mudgal (2004) modified the GBG in a version called the Good Classwork Game (GCG). Participants for the GCG included three target students from three separate elementary-level classrooms, one in kindergarten, one in fourth, and one in fifth grade, as well as their classroom peers. For the GCG, the experimenter targeted three behaviors: work completion, work accuracy, and off-task behavior. Although Mudgal tracked all behaviors for data collection purposes, only the appropriate behaviors of work completion and work accuracy were addressed as part of the game and received consequences. Mudgal anticipated that off-task behavior would simultaneously decrease with the increase of appropriate behaviors. An A/B/A/B withdrawal design was used to demonstrate the effectiveness of the GCG. Although the GCG intervention was successful at increasing work completion and decreasing off-task behavior of the target students, there was no improvement in work accuracy (Mudgal, 2004).

Mudgal continued her examination of the Good Classwork Game in a follow-up investigation (Mudgal, 2006). In this study, she employed a cross-over phase change design to examine the effect of non-randomized compared to randomized game criteria on increasing math accuracy and completion. Mudgal's participants included four target students in third, fourth, and fifth grade, as well as their classroom peers. Each classroom

was randomly assigned to the A/B/C/B or A/C/B/C design. Non-randomized game criteria (criteria for work completion and accuracy based on baseline levels) were implemented during phase B, and randomized game criteria (based on different criteria established during the non-randomized phase) were implemented during phase C. In order for a team to win the "game" their criteria had to meet or exceed that which was set, depending on the phase. Both intervention phases appeared to be equally effective at increasing math work completion and accuracy.

Only one published study has utilized the GBG with a general education high school population (Kleinman & Saigh, 2011). Following the introduction of a new teacher into a ninth grade history class in a multi-ethnic New York City high school, Kleinman and Saigh implemented the GBG with the classroom in an effort to decrease disruptive behavior including talking or verbal disruption, aggression or physical disruption, and seat leaving. The class was comprised of 15 males and 11 females all with a mean age of 15.39 years. Of the 26 students, six were African American, 19 were Hispanic, and one did not report ethnicity. The GBG was modified in that it was not presented as a game, but instead as an opportunity for the students to earn prizes. Also, target behaviors were presented as "expectations" (p. 102) rather than rules.

The authors had the students complete a reinforcement preference questionnaire prior to treatment implementation to determine the daily and weekly prizes. The rewards included cost effective items limited to approximately $15 a week, such as candy and a pizza or cupcake party. The effects of the GBG were evaluated using an A/B/A/B withdrawal design. The authors initiated a week long adaptation period prior to baseline to limit reactivity caused by the observers. During this time the observers joined the

classroom and collected data on the presence or absence of behaviors; however, these data were not provided. During week two, the class was divided into two teams, and baseline data were collected. The GBG was implemented in the classroom during the third week. During the fourth week, a withdrawal phase began, during which the teams remained intact, and classroom expectations were read aloud each day. However, inappropriate behavior was handled in the teacher's traditional manner. The GBG was reintroduced during the fifth week in the same manner as it had been initially. The results of the study reflected an immediate decrease in the level of all disruptive behaviors from baseline following the game's introduction. Withdrawal of the game resulted in an increase in level of disruptive behaviors similar to those during baseline. The re-introduction of the game and follow-up phases resulted in levels consistent with the initial introduction of the game. These results provide initial support for the GBG as an effective tool for general education high school students.

Several limitations exist with the study. Phase changes were decided based on the beginning of a new week rather than the data. Though this did not end up causing any major issues, there was an increasing trend in the data prior to the withdrawal phase. Experimental control would have been improved by data-based decision making for phase changes. Additionally, because the classroom was comprised of all ninth grade students, it is not possible to say if the GBG is generalizable to older high school students or those of different educational or demographic backgrounds. Replication and additional studies are needed.

Additionally, Mitchell (2012) also evaluated the GBG in three general education high school classrooms. The purpose of the study was to decrease inappropriate

behavior, specifically, off-task, out-of-seat, and inappropriate vocalizations. Of the three classrooms, Classroom A was a Transitions to Algebra class and consisted of 21 students, 5 females and 16 males. Of those students, 20 were African American, and one self-identified as biracial (African American/Caucasian). One student was in the twelfth grade, two were in the eleventh grade, and eighteen were in the ninth grade. Classroom B was also a Transitions to Algebra classroom and consisted of 24 students. Of those students, 13 were females, and 11 were males. Classroom B consisted of 20 African American students, 1 Hispanic student, and 3 biracial students (African American/Caucasian). All students were in the ninth grade. Classroom C was a Spanish II classroom which consisted of 23 students, 16 females and 7 males; 20 of whom were African American students, one was Caucasian, one was Hispanic, and one was a biracial student (African American/Caucasian). Of those students, 7 were in the eleventh grade, 11 were in the tenth grade, and 5 were in the ninth grade.

Like Kleinman and Saigh (2011), Mitchell (2012) modified the GBG by presenting it as a teamwork competition rather than a game. Furthermore, students completed a preference questionnaire prior to treatment implementation to determine the rewards they would be willing to work for. The rewards included free or low-cost items such as candy, homework passes, or extra points for a test. The effects of the GBG were evaluated using an A/B/A/B withdrawal design across classrooms A and C, and an A/B design for classroom B. During baseline data collection, teachers conducted their classroom as usual. During treatment phases, the class was divided into two teams, and students were given marks when the teacher observed them engaging in targeted inappropriate behaviors or other behaviors the teacher felt to be unacceptable.

Results of the study suggest that the GBG was effective in decreasing inappropriate behaviors across all three classrooms. During baseline, inappropriate behaviors were observed an average of 67% of intervals for Classroom A, 74% of intervals for Classroom B, and 65% of intervals for Classroom C. An immediate decrease in level and trend of disruptive behaviors was evident following the implementation of the GBG in all three classrooms. During the initial implementation, disruptive behaviors averaged 30% of intervals for Classroom A, 35% of intervals for Classroom B, and 27% of intervals for Classroom C. To demonstrate experimental control, a withdrawal was employed in classrooms A and C. At this time inappropriate behaviors returned to near-baseline- levels with those behaviors occurring an average of 50% of intervals observed in Classroom A and 57% of intervals observed in Classroom C. Upon reimplentation of the game, observed inappropriate behaviors again decreased with the behaviors occurring an average of 26% of intervals for Classroom A and 27% for Classroom C.

In addition to evaluating the effect of the GBG on behavior, Mitchell (2012) examined the acceptability of the intervention by teachers and students in the participating classes. The teacher's acceptability of the GBG was assessed using a slightly modified *Intervention Rating Profile-15 (IRP-15;* Martens, Witt, Elliottt, & Darveaux, 1985). All three teachers rated the GBG as an acceptable intervention for use in their classrooms. Student acceptability was evaluated using a modified version of the Children's Intervention Rating Profile (CIRP; Witt & Elliottt, 1985). Students found the GBG to be mostly acceptable, with several indicating that they did not feel the game/competition was fair.

Some limitations of the study should be noted. The study is limited in its external validity in that the majority of participating students were 9th grade African American students. Furthermore, Classroom B was not subject to the same experimental control as the other two classrooms, and the author noted several issues regarding the teacher of that classroom (i.e., administrative concerns regarding curriculum).

Changing Criterion Design

The changing criterion design is another strong design that can be used to demonstrate the strength of an intervention, particularly one which affects the strength of the behavior over time (Kazdin, 1982). Like the A/B/A/B phase change design, a changing criterion design involves the collection of baseline data but also the use of subphases. During these subphases, a specific criterion is set for the participant, and as the participant demonstrates that he or she is able to achieve this criterion, the criterion is changed (i.e., either increased or decreased). Each new criterion represents a subphase. This design is strong because it demonstrates the continued tracking of the effect of the intervention on the behavior. If the behavior follows the criterion in a systematic way, then it is possible to say that the intervention is responsible for the change.

Using a multiple baseline including a changing criterion component, Harris and Sherman (1973) used the GBG across periods of math and English with 5th and 6th grade students. The multiple baseline occurred across the GBG with certain elements (i.e., traditional GBG, modified GBG without consequences, modified GBG without feedback, modified GBG without teams) to determine the most successful version of the intervention. Critieria changed through phase changes. The results from this study indicated that the most control was demonstrated during the traditional GBG phases and

that the student behavior oscillated around the criteria as it was changed. Results also demonstrated that if there was no team involved, the motivation for students to continue displaying appropriate behavior was diminished following the class passing the criterion.

Additionally, Hegerle, Kesecker, and Couch (1979), who were among the first to utilize the GBG with a special education population, used a changing criterion design to evaluate the GBG. The authors used a modified version of the GBG with a self-contained classroom, targeting out-of-seat behavior, talking-out behavior, and tattling (added four days after the implementation of the GBG). The authors provide no information regarding the age or grade of the students in the study. The experimenters divided the 22 students based on gender and also allowed certain students who earned a large number of marks to be omitted from counting against the team, a rule which was implemented seven days into implementation. The modified version also incorporated a token system, which allowed the winning team to place a star on a Victory Chart as well as a changing criterion component. Graphs were not included in the published study, and therefore it is not evident whether the behaviors followed the decreasing criterion. The authors reported that the results indicated that the implementation of the GBG was effective for reducing target behaviors and even though the criterion was eventually reduced from 25 marks to 2 marks, both teams always won the game.

A changing criterion within an A/B/A/B design was also used by Salend, Reynolds, and Coyle (1989) to decrease disruptive behaviors in emotionally disturbed high school students in a residential school. The authors used the GBG game in three separate classrooms at the residential school and focused on individualized student

behavioral concerns in the referred classrooms, specifically decreasing inappropriate vocalizations, cursing/negative comments, and touching/drumming.

Following baseline data collection, the GBG was introduced in all three classrooms. Classrooms were divided into teams based on the most frequent inappropriate behavior they displayed rather than seating chart or random selection (i.e., all students observed drumming were placed on Team 1, while students observed cursing were placed on Team 2). Unlike the traditional GBG, the teams were not in competition with each other, but rather needed to stay below the criterion in order to have access to the reward. Two of the classrooms had two teams, and one classroom had three teams. For the teams, criteria varied and were based on baseline levels. Once teams achieved criterion levels, the criterion was lowered. Following this initial implementation, a withdrawal phase was implemented. For all the classes this phase lasted for 5 days, at which point the GBG was re-implemented and monitored for an additional 9 days.

Results from this study suggest a functional relationship between the GBG and reductions in targeted disruptive behaviors. Specifically, both teams in all three classrooms showed immediate decreases in targeted behaviors following the implementation of the GBG. Additionally, data show that during nearly all observations, the teams stayed at or below the criterion for the day. Though several of the teams' performances appear to track the criterion, in some cases the behaviors occurred so infrequently from the beginning of the phase that they cannot be directly attributed to the changing criterion in place. During the withdrawal phase all teams immediately returned to near baseline levels. The reimplementation phase once again demonstrated an

immediate change in level and trend of the target behaviors. Again behaviors typically occurred at or below the criterion level.

Another study which included a changing criterion component within a multiple baseline design was the first study of the effectiveness of the GBG with preschool aged children (Swiezy et al., 1992). In this particular study, a variation of the GBG using a multiple baseline design-within subject pairs and across therapists with a changing criterion targeted compliance in four preschool children. Unlike the original version of the GBG used by Barrish et al. (1969), the game was applied in an analog setting, and the teams were composed of only two pairs of children rather than an entire classroom. In this modification, the game focused on rewarding the appropriate target behavior of compliance rather than focusing on inappropriate behavior and response cost procedures.

GBG sessions and observations occurred in a resource room three times a week for 15 minutes. Each of the pairs completed sessions separately from the other pairs. Therefore, each pair's performance and ability to win the game was independent of the other pair, and there was no competition between the two teams. An additional modification of the game included the use of a puppet called *Buddy Bear* as the session instructor. The authors targeted joint compliance from the pair of students. Each time the pair jointly complied with Buddy Bear's instruction, the pair was rewarded with a colored smiley face or dinosaur, which was then displayed on a felt board. For each session, a criterion was set. After reaching the criterion for two sessions, the criterion was increased. Unfortunately, the authors did not provide information regarding the specific criteria during the intervention phase in their results; therefore, there is no way of knowing whether or not the behaviors paralleled those changes in criteria. However, the

provided results indicated increased levels of compliance for both pairs during sessions. Although generalization occurred across therapists, there was no evidence of generalization across settings (Swiezy et al., 1992).

There are several limitations in the Swiezy et al. (1992) study. In terms of internal validity, no treatment integrity data were collected. Therefore, it is not possible to know whether or not changes in the dependent variable were directly related to the GBG. There were also threats to external validity. The sessions and observations were not conducted in a naturalistic setting. Instead of conducting the intervention in the students' classroom, the sessions were performed in a separate resource room. Therefore, it is unknown if the results of the study are generalizable to a classroom setting. Also, only a few students were used to form pairs rather than the entire classroom. This makes the generalizability of the intervention with larger groups unknown.

Purpose of the Present Investigation

Although there have been many studies supporting the effectiveness of the GBG as a classroom intervention, not all areas have been fully explored. Specifically, there is limited research with older students and the general education high school population. Introducing the GBG in a high school population may help to enable teachers to handle minor infractions in their classroom and avoid removing students from the academic setting for disciplinary action. To date, few studies exist (Kleinman & Saigh, 2011; Mitchell, 2012) in which the GBG has been used in general education high school classrooms. Additional studies are necessary to confirm whether or not the GBG is truly an appropriate and effective intervention for high school students across both upper and lower grade levels. Through the use of a stronger A/B/A/B design with a changing

criterion component across multiple classrooms, the present study will utilize a modified GBG, referred to as a "Teamwork Competition" (TC), in general education high school classrooms, and will target teacher-referred disruptive behavior to determine the relationship between the GBG and behavioral improvement. Furthermore, through the use of the changing criterion component, the present study will investigate how the criterion component will affect the level of disruptive behavior, specifically when it is high, low, and unknown (i.e., a "mystery" criterion). The "mystery" criterion component has not been previously incorporated in the GBG with general education high school students.

The present study aims to investigate the following research questions:

1. Will the Good Behavior Game/Teamwork Competition (TC) effectively decrease disruptive behavior of upper level (10^{th}, 11^{th}, and 12^{th} grades) general education high school students in the classroom?
2. What effect, if any, will the changing criterion component have on levels of disruptive behavior of the class in the Good Behavior Game/Teamwork Competition?
3. Will the Good Behavior Game/Teamwork Competition increase academic engagement?
4. What effect will the use of a "Mystery Criterion" have on target behaviors?
5. Is the Good Behavior Game/Teamwork Competition acceptable to teachers for use in a general education high school classroom?

CHAPTER II

METHODOLOGY

Participants and Setting

Participants included three high school classrooms and their teachers at a high school located in a southeastern state referred for participation by administrative referral for classroom disruptive behavior. Approximately 800 students were enrolled in the school which consisted of students in the ninth through twelfth grades, with 89% of the school's population eligible for free or reduced lunch. At the time of the study, the school had been implementing PBIS for three years. The school received a 95% on the School-wide Evaluation Tool (SET), a widely recognized and empirically supported School-wide PBIS evaluation tool which was utilized to assess the level of PBIS implementation for the school (Sugai, Lewis-Palmer, Todd, & Horner, 2005). It is important to note that the GBG was implemented as a Tier I level intervention (meaning all students in referred classrooms benefitted from its implementation), and therefore, it is reasonable to assume that a classroom utilizing this intervention without a solid PBIS system in place in the school may see similar results.

Prior to screening classrooms, teacher consent was obtained by the primary investigator for inclusion (See Appendix A for Teacher Consent Form). Because of the nature of the study, teachers could decline to participate in the study or withdraw at any time. Over the course of the study, a total of six classroom teachers were referred to participate in the study. One teacher declined to participate, one classroom did not screen in, and one teacher chose to withdraw from the study prior to intervention implementation. Teachers who did not agree to participate were referred for other

services outside the context of the study. Teachers who agreed to participate were briefly interviewed and asked to list and describe behaviors they felt were most troublesome in their classroom. These primary concerns of the teacher, along with the behaviors identified by the school administrator and the author, were used to operationally define the targeted behaviors for each of the classrooms. Teachers and all students in the classroom served as participants and had their behavior monitored by trained observers. All teachers served as interventionists for the study. Because no individual student data were collected, only teacher consent was obtained for classroom participation. Details of general classroom demographics and intervention agents are provided below. All materials and procedures used in the study were approved by The University Southern Mississippi Institutional Review Board (IRB; See Appendix B).

Teacher A was a 32-year-old Caucasian female who had taught for three years; this was her first year teaching at the present school. She was enrolled in classes for her master's degree at the time of the study. Classroom A was a tenth grade World History class and consisted of 25 students. Of those students, 9 were females, and 16 were males. The classroom students' self-reported race consisted of 22 African American students and three biracial students (African American/Caucasian). Teacher A also indicated that four students in the classroom were inclusion students (i.e., not self-contained) and had Individualized Educational Plans (IEPs).

The teacher for Classroom B changed midway through the study, Teacher B1 and Teacher B2. Teacher B1 was a 26-year-old African American female who had taught for four years, all at the present school. She had a Master's of Business Administration degree. Due to a teacher in a "high stakes," state-wide subject area testing classroom

taking maternity leave, approximately two months after the study began, Teacher B1 moved to that class, and a permanent substitute, Teacher B2, began teaching Classroom B. Because this was toward the end of the first intervention phase, but prior to the withdrawal phase, the primary investigator determined the change would not significantly impact the conclusions that could be drawn from the data. Furthermore, the primary investigator felt this change could potentially further expand the external validity of the intervention. Teacher B2 was a 42-year-old Caucasian male who had taught for nine years and had earned his bachelor's degree. Serving as the permanent substitute in Classroom B was his first time teaching at the present school. Classroom B was a twelfth grade English class with 20 students enrolled. Of those students, 9 were females, and 11 were males. The classroom students' self-reported race consisted of 17 African American students, one Hispanic student, and two biracial students (African American/Caucasian). Teacher B1 indicated that two students in the classroom were inclusion students (i.e., not self-contained) and had IEPs.

Teacher C was a 28-year-old Caucasian male who had taught for seven years and was in his second year teaching at the present school. He completed his master's degree while the study was in process. Classroom C was an eleventh grade U.S History class and consisted of 18 students. Unlike the subjects taught in the other participating classrooms, U.S. History is a state-wide subject area testing subject. Of those students, 6 were females, and 12 were males. Teacher C's students' self-reported race consisted of 15 African American students, one Caucasian student, and two biracial students (African American/Caucasian). Teacher C indicated that three students in the classroom were inclusion students (i.e., not self-contained) and had IEPs.

Materials

A teacher script (Appendix C), board for displaying team names/points, and teacher-approved rewards were all used during the implementation of the Good Behavior Game (GBG)/Teamwork Competition (TC). Utilizing a prewritten script allowed for consistent introduction of the game to each of the classes. This also ensured that the teachers presented the appropriate rules and game procedures to their classes. The classroom board was necessary to allow students to track their performance throughout the game. A requirement of the board placement was that it was easily viewed from all seats and easily accessible to the teacher. In addition to displaying the team names, the board served as the tracker for the teacher to mark points for disruptive behavior. After baseline data collection, students were asked to list items they would like to earn from these requested items, and the teacher and primary investigator developed a list of rewards to present the winning teams. Rewards for all classrooms were of low or no monetary value (i.e., tangibles, edibles, PBIS tickets, homework passes, access to free time, etc.).

Acceptability

After all classroom data collection was completed, the acceptability of the intervention was assessed from the participating teachers. The teachers' acceptability of the GBG/TC was assessed using the *Intervention Rating Profile-15 (IRP-15;* Martens, Witt, Elliottt, & Darveaux, 1985; Appendix D). The *IRP-15* is a questionnaire containing 15 questions regarding the intervention using a Likert scale (1 = strongly disagree; 6 = strongly agree). Scores can range from 15 to 90, and the higher a score, the higher the acceptability of the intervention. Von Brock and Elliott (1987) recognize scores of 52.50

and higher as indicating acceptability and also report that the IRP-15 has a Cronbach's Alpha of .98. Teachers completed the *IRP-15* following the final day of data collection. The IRP-15 was slightly modified, changing tense of some items from present to past tense as well as the use of word substitutions including changing the word *child/children* to *student(s)* to accommodate high school aged participants. Minor modifications to such scales have not been found to affect technical adequacy (Freer & Watson, 1999).

Dependent Variables, Observation Procedures, and Data Collection

Four dependent variables (three primaries and one secondary) were measured for each classroom. The specific targeted behaviors were developed according to teacher referral; because all three teachers reported similar target behaviors, the same behaviors were tracked across all three classrooms. The primary targeted behaviors included inappropriate vocalizations, out-of-seat behavior, and inappropriate touching; academic engagement was also tracked as a secondary dependent variable. Each behavior was operationally defined to ensure interobserver agreement. Inappropriate vocalizations were defined as any audible verbalization made without teacher permission such as speaking, yelling, humming, singing, and/or whispering excluding involuntary verbalizations such as sneezes. Out-of-seat behavior was defined as the student's buttocks breaking contact with his or her seat for more than three seconds without teacher permission. Inappropriate touching was defined as the student touching another student at any point during the interval including hitting, pushing, poking, hugging, rubbing, flicking, or any other contact between two students that was not pertinent to the assignment. Academic engagement was defined as the student being actively involved or attending to (e.g. looking at or writing on) independent seatwork, teacher instruction, designated classroom

activities, and/or engaging in task related (permissible) vocalizations with teachers and/or peers.

The primary investigator and trained graduate students served as observers for the three classrooms. Observations lasted 20 minutes and used 10-second intervals for observations. Coding sheets had the 20-minute observation broken into 10-second intervals (Appendix E). Observers used partial interval recording for coding disruptive behavior (inappropriate vocalization, out-of-seat, and inappropriate touching), and momentary time sampling for academic engagement. An audio recording (heard via headphones) cued the changing of each interval. For each interval, the observer looked at the interval's target student (for each interval the target student alternated across the entire classroom, until every student had been observed, and then the observer started over with the first student until the 20 minute observation was complete) and coded for that interval. When coding disruptive behavior, observers utilized partial interval recording; therefore, any occurrence of a target behavior in an interval resulted in that behavior being recorded as "present" during the interval. The occurrence of academic engagement was coded utilizing a momentary time sampling procedure. This meant that during the interval, an additional tone on the recording halfway through the interval cued the academic engagement recording. When the tone sounded, the observer marked if the student met the requirement for academic engagement. When the recording cued a new interval, the observer looked to the next student and recorded whether that interval's target student was exhibiting any of the target inappropriate behaviors at any time during the interval or was academically engaged at the moment of the interval's tone. Observations occurred between two and four times a week and began no less than five

minutes after the start of the period to allow the students and teacher an opportunity to transition between classes. Although data were not collected everyday or for the entire period, the TC was in place, and permanent products were used (i.e., team names/marks on board, teacher tracking winners, decrease in rewards, etc.) to ensure that the intervention remained in place.

The data collected for all of the students each day were collapsed into the percent of intervals for each targeted behavior. At the end of each observation, the total percentage of intervals with each target behavior was graphed as a representation of the classroom as a whole. All disruptive target behaviors (inappropriate vocalization, out-of-seat, and inappropriate touching) were also collapsed and graphed as the total percentage of intervals with disruptive behavior. Academic engagement was graphed separately from the disruptive behaviors.

Procedures

Screening

Following administrative referral for the study and teacher consent, screening observations were conducted. During this observation, teachers were instructed to utilize their standard classroom management procedures and to deal with appropriate and inappropriate behavior in their typical manner. At this time, all students were monitored in alternating intervals for any of the targeted behaviors. If any of the targeted behaviors occurred in 30% or more of the intervals, the classroom screened in to the study. Data collection for the screening process was identical to the method used during data collection for baseline and intervention phases. Therefore, the screening observation served as the first baseline point for classrooms that screened in. Of the referred teachers

who consented to participate in the study, one did not screen in (only 11% of intervals contained disruptive behavior). One teacher did not consent to be included in the study and was referred for other services outside of the context of the study. During the screening observation, Classroom A displayed 44% of intervals with disruptive behavior, Classroom B displayed 36% of intervals with disruptive behavior, and Classroom C displayed 35% of intervals with disruptive behavior.

Baseline

After screening into the study, baseline data collection continued for each class. At this time, the teacher continued to conduct his or her classroom as usual with no additional contingencies in place for targeted behaviors. Teachers reported that there were no classroom interventions in place during this time, and this was confirmed by anecdotal observations (i.e., observers noted the absence of any intervention materials in the classroom). Data were collected for the four targeted behaviors, with data collectors observing all students in the classroom during alternating intervals. The data for the students were then collapsed to represent the classroom as a whole and were graphed daily.

Teacher training

Following the collection of baseline data, the teachers were trained on the Teamwork Competition (TC) procedures. Training occurred during each teacher's planning period after baseline data collection and prior to implementation. Training consisted of introducing the script and explaining each of the TC components, modeling the script and steps of the TC for the teacher, and watching the teacher rehearse each of the steps. Teachers were required to practice the steps of the intervention with 100%

integrity prior to implementation with their classrooms. The teacher received feedback following the training session and following each data collection session to ensure that integrity was maintained throughout the study. Feedback included information about any steps a teacher did not complete during a session as well as praise for teachers who implemented the TC with high/perfect integrity. If a teacher did not complete at least 80% of the steps possible, the teacher would have been retrained (see Appendix F and G for checklist). No teacher received less than 80% integrity; therefore, no teacher required retraining.

Preference Assessment

Following the collection of baseline data, but prior to implementation, teachers administered a brief preference assessment questionnaire to their students (See Appendix H). At this time students were able to suggest rewards that they would like to earn and would be willing to work for. Students' responses included edibles (i.e., candy, Cheetos, donuts, pizza), homework passes, bathroom passes (which were otherwise limited to the teacher's discretion), bonus points/grades, school supplies (i.e., pencils, erasers, notebooks), and money. All three teachers approved the use of edibles, homework passes, bathroom passes, bonus points, and school supplies as daily rewards for the students. All teacher approved rewards were placed in a box, and winning students were allowed to choose from all of the options on any day. All items of monetary value were provided by the author; however, cost of these items was relatively low (approximately $5-10 per week).

Good Behavior Game/Teamwork Competition

Following baseline data, teacher training, and student preference assessment, the teacher introduced the TC to the classroom. At this time, the teacher used the script and followed the steps modeled and rehearsed during prior training. After explaining the rules and expectations of the game, the teacher divided the students into two teams based on the seating chart. Because the teachers knew about the development of teams prior to data collection, any adjustments to balance the most disruptive students across the two teams was to happen prior to baseline data collection. Only one teacher (teacher B1) elected to move students, and this was limited to 3 students. All other teachers felt the classrooms were balanced and did not opt to move any students. The teachers were instructed to give the teams generic names (e.g., the red and blue team). Drawing on the school's colors, Classroom A and B had the Gold and Maroon teams, and Classroom C (a U.S. History class) had the Red and Blue teams. The team names were written on the classroom whiteboard and were easily visible by all students and accessible by the teachers. Each time the teacher observed one or more students on a team engaging in one or more inappropriate behaviors, the teacher placed a mark under the team of the offending student and provided feedback to students when necessary (i.e., pointing out or verbalizing names of specific team members). Each day that observations occurred, the observer(s) also completed a teacher integrity checklist.

Following baseline data collection, a criterion was set for each class based on a private frequency count kept by the teachers during the baseline phase. Frequency counts were kept by the teacher during baseline to demonstrate what the number of marks would be if the game was in place. The mean number of marks of the classroom was divided by

two to account for two separate teams. The criterion was then set to 10% below the divided mean number of marks in each classroom so that rewards could be reasonably attained. The initial criteria (C1) were set at 16 marks for Classroom A, 15 marks for Classroom B, and 10 marks for Classroom C. If both teams met this criterion, meaning both teams had fewer marks than the pre-set criterion, both teams had access to the rewards at the end of the class period. If neither team scored fewer marks than the set criterion, then the team with the fewest marks won the reward. Therefore, both teams had an opportunity to win the game every day; however, at least one team always won.

Design and Data Analyses

The effects of the GBG on reducing disruptive behavior were evaluated using an A/B/A/B withdrawal design, containing a changing criterion component in "subphases" within each intervention phase. Phase changes occurred based on the trend and stability of the targeted disruptive behaviors. Baseline data were collected for five sessions in Classroom A, six sessions in Classroom B, and four sessions in Classroom C. Treatment effects were analyzed visually for level, trend, and variability. Additionally, Percentage of Non-Overlapping data (PND) was calculated. PND involves identifying the lowest data points in the baseline or withdrawal phase and then determining if any data points during the intervention phases exceed this number (Scruggs, Mastropieri, & Casto, 1987). A percentage is then calculated for these points. PND can range from 0% – 100% and is said to reflect unreliable treatment if the percentage falls below 50%; PND between 50% and 70% are said to display questionable effectiveness. PND are said to be fairly effective if they fall between 70% and 90%, and those above 90% are considered to be highly effective (Scruggs et al., 1987).

Following the implementation of the TC and establishing a clear, stable treatment effect (based on level, trend, and variability), the criterion was decreased approximately 50 percent (criterion was rounded to the nearest whole number) to eight marks for Classroom A after five sessions, eight marks for Classroom B after four sessions, and five marks for Classroom C after five sessions. After a clear and stable treatment effect was achieved, the criterion was once again lowered approximately 50 percent to four marks after four sessions in Classroom A, four marks after three sessions in Classroom B, and one mark after five sessions in Classroom C. After a clear and stable effect was evident during this third "subphase," a withdrawal phase was implemented.

During the withdrawal phase, the TC and all of the associated components (i.e., team names, rewards, etc.) were removed, and the teachers were asked to return to their typical classroom management procedures. Following the withdrawal phases, the TC was put back in place for a re-implementation phase and three subsequent subphases.

During the re-implementation treatment phase, one of the subphases was set to a "mystery criterion" in which the criteria varied between the lowest and highest prior criteria. Students were not told what the criterion was until the end of the period. Again treatment effects were analyzed visually for level, trend, and variability. In this final re-implementation phase, this variation (i.e., mystery criterion) was implemented to further evaluate the importance of a consistent and obtainable goal. During this phase, three criteria were implemented: C4, C5 and C6. The re-introduction of the intervention at C4 included a low, but obtainable criterion. During C5 the criterion was introduced as a "mystery criterion," and therefore, the class was not told what the criterion was for both teams to be able to earn a reward. However, at least one team earned the reward every

day. Criteria during the mystery subphase were comprised of all numbers between the lowest and highest criterion levels that were demonstrated in that class (i.e., if a class had a subphase with a 15 mark criterion, a six mark criterion, and a three mark criterion during the first implementation phase, the range of the mystery criteria would fall between 15 and three marks). The class criterion was drawn from a jar at the end of the period to determine the criterion. During the final subphase, C6, the criterion was once again set to the lowest criterion of the previous five subphases.

Treatment Integrity

During each day of data collection, the observers completed an integrity checklist (See Appendix F & G). This checklist served to ensure that treatment integrity remained high throughout the course of the intervention implementation phase. The checklist, adapted from Hunt (2010), included questions regarding the adherence to the game and awarding of prizes to the game winner(s). Integrity was also evaluated during non-intervention sessions to ensure that the teachers were not utilizing intervention components during baseline/withdrawal (Appendix F). Teachers received feedback each day of data collection regarding any steps missed on the integrity checklist following data collection. Teacher integrity was calculated by dividing the number of steps carried out by the number of total steps possible and was then multiplied by 100 to create a percentage. Teachers were informed that anytime he or she failed to complete 80% of the steps, he/she would be retrained. However, at no point during the study did any teacher fall below 80%.

All three teachers maintained high integrity throughout the entire study. During baseline and withdrawal phases, all three teachers maintained 100% integrity.

Specifically, during these phases, none of the teachers mentioned or utilized rewards or teams. Integrity for Teacher A averaged 92% (range = 80% - 100%) overall, 90% (range = 80% - 100%) during initial implementation, and 92% (range = 80% - 100%) during re-implementation. Teacher integrity for teacher B averaged 90% (range = 80% - 100%) overall, 87% (range = 80% - 100%) during initial implementation, and 88% (range = 80% - 100%) during re-implementation. Teacher integrity for teacher C averaged 88% (range = 80% - 100%) overall, 83% (range = 80% - 100%) during initial implementation, and 84% (range = 80% - 100%) during re-implementation.

Observer Training

Trained graduate students served as observers. Prior to data collection, observers were trained on the exact behavioral definitions of each behavior. The operational definitions were also included on data collection sheets to ensure full access to definitions at all times (Appendix E). Observers were trained by practicing data collection in a classroom(s) until 90% or higher agreement for each of the targeted behaviors with the primary investigator was obtained. At this point, observers were allowed to enter classrooms independently to observe and collect data.

Interobserver Agreement

Two observers collected data during at least 25% of the sessions for each phase and subphase for all classrooms. Interobserver agreement (IOA) for disruptive behavior as well as for teacher integrity checklists remained above 90% throughout the entire study. Had IOA fallen below 90% at any point the observer would have been retrained. However, at no point during the study did IOA fall below 90%. IOA was calculated separately for each of the four behaviors as well as for the teacher integrity checklist.

IOA was calculated by adding the total number of agreements for occurrence and for nonoccurrence of the behavior between the two observers and dividing that number by the total number of intervals then multiplying by 100.

For Classroom A, IOA was collected for 40% of baseline sessions, 31% of initial treatment sessions, 25% of withdrawal sessions, and 35% of re-implementation sessions. Total target IOA for Classroom A averaged 94% (range = 92% - 100%). IOA for individual target behaviors for Classroom A was 93% (range = 91% - 97%) for inappropriate vocalizations, 99% (range = 98% - 100%) for out-of-seat behavior, 96% (range = 94% - 100%) for inappropriate touching, and 94% (range = 92% - 99%) for academic engagement.

For Classroom B, IOA was collected for 50% of baseline sessions, 40% of initial treatment sessions, 33% of withdrawal sessions, and 41% of re-implementation sessions. Total target IOA for Classroom B averaged 93% (range = 90% - 95%). IOA for individual target behaviors for Classroom B was 91% (range = 90% - 93%) for inappropriate vocalizations, 99% (range = 98% - 100%) for out-of-seat behavior, 96% (range = 94% - 100%) for inappropriate touching, and 95% (range = 92% - 97%) for academic engagement.

For Classroom C, IOA was collected for 50% of baseline sessions, 38% of initial treatment sessions, 40% of withdrawal sessions, and 30% of re-implementation sessions. Total target IOA for Classroom C averaged 93% (range = 91% - 99%). IOA for individual target behaviors for Classroom C was 94% (range = 90% - 97%) for inappropriate vocalizations, 99% (range = 98% - 100%) for out-of-seat behavior, 97%

(range = 95% - 100%) for inappropriate touching, and 98% (range = 94% - 99%) for academic engagement.

IOA for teacher integrity was also collected and calculated by adding the total number of agreements regarding the completion/incompletion of TC steps among observers and dividing that number by the total number of agreements possible. That number was then multiplied by 100. Teacher Integrity was collected during all data collection sessions; therefore, the percentage of IOA sessions for procedural collection was identical to that of IOA (range = 25% - 50% of phases). There was 100% agreement for all teacher integrity observations.

Cohen's Kappa Coefficient

Cohen's kappa coefficient is a statistical measure of inter-rater agreement. Cohen's Kappa is considered a more robust and conservative measure of agreement as it accounts for chance agreement (Watkins & Pacheco, 2000). Cohen's kappa coefficient was calculated for disruptive behavior as well as academic engagement. After determining a kappa coefficient, the value can be interpreted as poor agreement (values equal or less than 0.40), moderate agreement (values falling between 0.40 - 0.60), good agreement (values falling from 0.60 - 0.75), and very good agreement (values of .75 and above) (Fleiss, Levin, & Cho Paik, 2003; Watkins & Pacheco, 2000).

Kappa was calculated for disruptive behaviors and academic engagement for each of the three classrooms. The mean kappa for Classroom A for disruptive behavior was 0.634, which can be interpreted as a good agreement. Academic Engagement for Classroom A was 0.860, revealing a very good agreement between observers. For Classroom B, an average of 0.521 for disruptive behavior was calculated, indicating a

moderate agreement. Kappa values for academic engagement in Classroom B averaged 0.575, also indicating a moderate agreement across observers. Mean kappa for Classroom C for disruptive behavior was 0.506, demonstrating a moderate agreement across observers. Mean kappa for academic engagement was 0.609, demonstrating a good agreement across observers. Overall, Kappa scores revealed moderate to good agreement for disruptive behavior among observers across all classrooms and moderate to very good agreement for academic engagement.

CHAPTER III

RESULTS

Effects on Disruptive Behavior and Academic Engagement

Figure 1 includes the percentage of intervals in which total disruptive behavior (collapsed across individual behaviors) as well as academic engagement occurred across phases in each classroom. During baseline, students in Classroom A (top panel) displayed disruptive behaviors during an average of 43% (range = 30% - 51%) of intervals. Academic engagement was observed during 44% (range = 30% - 58%) of intervals during baseline. There was an immediate reduction in targeted disruptive behavior following the implementation of the TC. Disruptive behaviors decreased to an average of 10% (range = 4% - 17%) of intervals during the entire first TC phase, and academic engagement increased to an average of 72% (range = 50% – 83%) of intervals. The initial intervention criterion (C1) was set approximately 10% below the baseline average (to 16 marks). During this first criterion the disruptive behavior averaged 15% (range = 10% - 17%) of the intervals. The criterion was then lowered (C2) by approximately 50% of the prior subphase (to 8 marks), and the disruptive behavior average decreased to 9% (range = 5% - 16%) of intervals. During the final initial implementation subphase (C3), the criterion was lowered 50%, (to 4 marks), and the disruptive behavior average dropped to 6% (range = 4% - 7%) of intervals.

During withdrawal disruptive behaviors increased to an average of 48% (range = 43% - 50%) of intervals, and academic engagement decreased to an average of 38% (range = 31% - 43%). Following this withdrawal phase, the TC was re-implemented, and occurrence of disruptive behaviors again decreased.

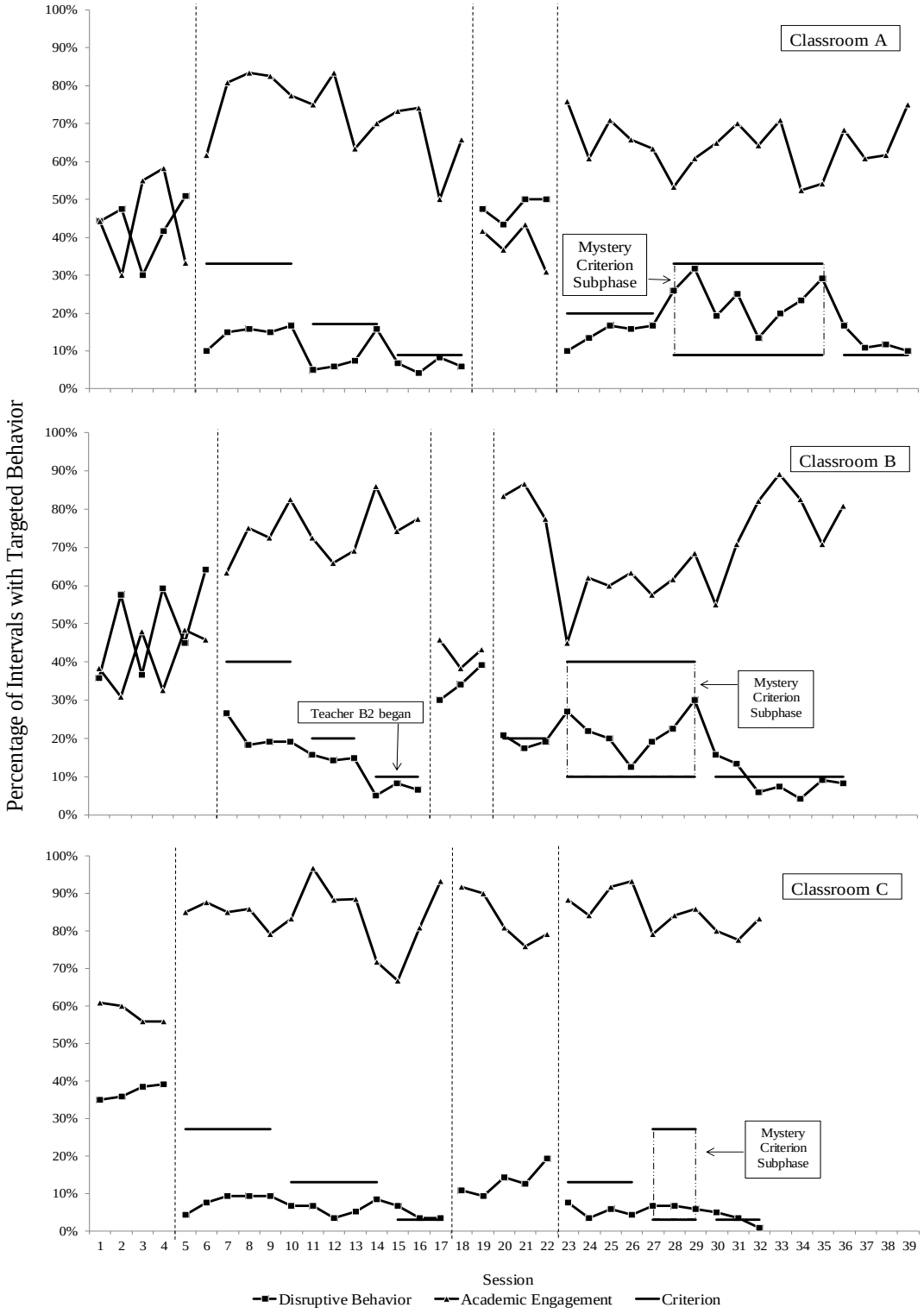

Figure 1. Percentage of Intervals of Combined Disruptive Behaviors and Academic Engagement Across Phases for Classrooms A, B, and C.

During the initial reimplementation subphase (C4), the criterion was set to 10 marks, and the students' disruptive behavior decreased to an average of 15% (range = 10% - 17%) of intervals. For the second subphase (C5) in the reimplementation phase, a "mystery criterion" was introduced. In this subphase, the students only knew that the criterion for the day would fall somewhere between four and 16 marks. During this subphase disruptive behavior averaged 23% (range = 13% - 32%) of intervals. The observers noted that the disruptive behavior for the class was much more variable than during past intervention phases. The criterion was lowered to four marks for the final subphase (C6). The average of the disruptive behavior was 12% (range = 10% - 17%) of intervals during this subphase. For Classroom A, there was only one overlapping datum point between intervention phases and non-intervention phases; therefore, the percent of non-overlapping data points (PND) was 96.7%. Academic Engagement varied somewhat but remained relatively high during all intervention subphases: C1 was 77% (range = 62% - 83%) of intervals, C2 was 73% (range = 63% - 83%) of intervals, C3 was 66% (range = 50% - 74%) of intervals, C4 was 67% (range = 61% - 76%) of intervals, C5 was 61% (range = 53% - 71%) of intervals, and C6 was 66% (range = 61% - 75%) of intervals.

In addition to tracking overall collapsed disruptive behaviors, individual behaviors were tracked and graphed and can be seen in Figure 2 and Table 1. In all three classrooms, inappropriate vocalizations clearly accounted for the vast majority of disruptive behavior across phases compared to out-of-seat and inappropriate touching.

During baseline, students in Classroom B (middle panel, Figure 1) displayed disruptive behaviors during an average of 50% (range = 36% - 64%) of intervals.

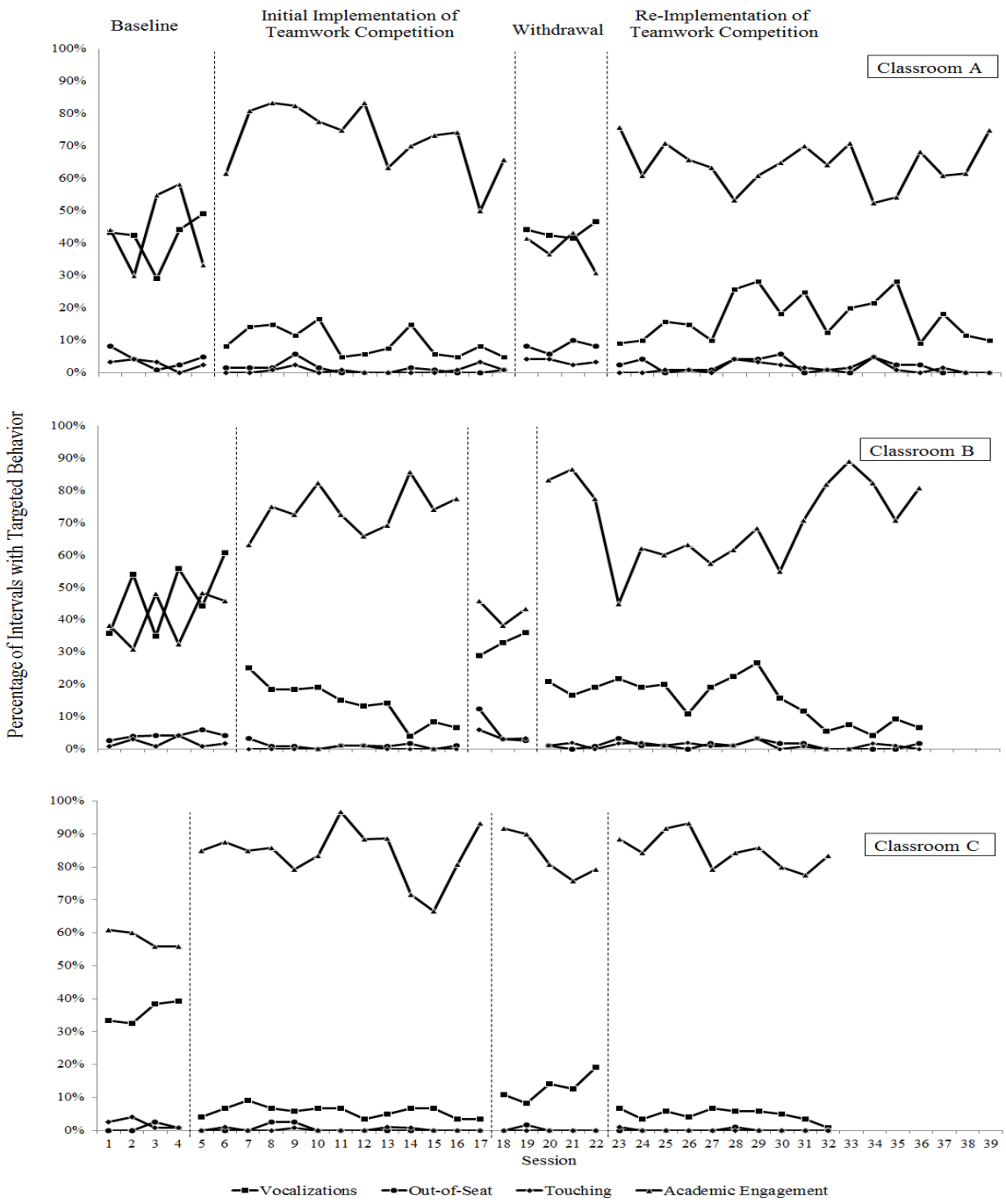

Figure 2. Percentage of Intervals of Inappropriate Vocalizations, Out-of-Seat, Inappropriate Touching, and Academic Engagement Across Phases for Classrooms A, B, and C.

Table 1

Mean Percent of Total and Individual Disruptive Behaviors

Behavior	**Baseline**	**TC**	**Withdrawal**	**TC**
Classroom A:				
Vocalizations	42%	9%	44%	17%
Out-of-Seat	4%	1%	8%	2%
Touching	3%	1%	4%	1%
Total DB	**43%**	**10%**	**48%**	**18%**
Engagement	44%	72%	38%	64%
Classroom B:				
Vocalizations	48%	14%	33%	15%
Out-of-Seat	4%	1%	6%	1%
Touching	2%	<1%	4%	1%
Total DB	**50%**	**15%**	**34%**	**16%**
Engagement	41%	74%	43%	70%
Classroom C:				
Vocalizations	36%	6%	13%	5%
Out-of-Seat	1%	<1%	<1%	<1%
Touching	2%	<1%	0%	<1%
Total DB	**37%**	**6%**	**13%**	**5%**
Engagement	58%	84%	84%	85%

Academic engagement was observed during 41% (range = 31% - 48%) of intervals during baseline. As with Classroom A, there was an immediate reduction in targeted disruptive behavior following the implementation of the TC. Overall, disruptive behaviors decreased to an average of 15% (range = 5% - 27%) of intervals during the entire first TC phase, and academic engagement increased to an average of 74% (range = 63% – 86%) of intervals. The first criterion (C1) was set at approximately 10% below the baseline average (to 15 marks). During C1 the disruptive behavior averaged 21% (range = 19% - 27%) of intervals. The criterion was lowered by approximately 50% (to 8 marks) during C2, and the disruptive behavior average decreased to 15% (range = 14% - 16%) of intervals. During C3 the criterion was lowered 50% further (to 4 marks), and the disruptive behavior average dropped to 7% (range = 5% - 8%) of intervals. It is important to note that during C3 the original teacher, Teacher B1, left the class, and the permanent substitute, Teacher B2, took over and continued to utilize the TC.

Approximately one week after Teacher B2 took over, a withdrawal phase occurred in which the intervention was removed from the classroom for a short time. During this phase disruptive behaviors increased to an average of 34% (range = 30% - 39%) of intervals. Following this phase, Teacher B2 re-implemented the TC and the occurrence of disruptive behaviors again decreased. During C4 the criterion was set to 8 marks and the students' disruptive behavior decreased to an average of 19% (range = 18% - 21%) of intervals. For C5 a "mystery criterion" was introduced. Criteria each day fell between four and 15 marks. During this subphase disruptive behavior averaged 22% (range = 13% - 30%). Again the observers noted that the disruptive behavior for the class was more variable than past intervention subphases. C6 was lowered to 4 marks. The

average of the disruptive behavior was 9% (range = 4% - 16%) of intervals in this subphase.

For Classroom B there was only one overlapping datum point between intervention phases and non-intervention phases; therefore, the PND was 96.3%. Academic Engagement varied but remained higher during all intervention subphases: C1 was 73% (range = 63% - 83%) of intervals, C2 was 69% (range = 66% - 73%) of intervals, C3 was 79% (range = 74% - 86%) of intervals, C4 was 83% (range = 78% - 87%) of intervals, C5 was 60% (range = 45% - 68%) of intervals, and C6 was 76% (range = 55% - 89%) of intervals.

During baseline, students in Classroom C (bottom panel, Figure 1) displayed disruptive behaviors during an average of 37% (range = 35% - 39%) of intervals. Academic engagement was observed during 58% (range = 56% - 61%) of intervals during baseline. As with Classrooms A and B, there was an immediate reduction in targeted disruptive behavior following the implementation of the TC. Overall, disruptive behaviors decreased to an average of 6% (range = 3% - 9%) of intervals during the entire first TC phase, and academic engagement increased to an average of 84% (range = 67% – 97%) of intervals. C1 was set approximately 10% below the baseline average to 10 marks. During this first criterion the disruptive behavior averaged 8% (range = 4% - 9%). The criterion was lowered by approximately 50% during C2 (to 5 marks), and the disruptive behavior average decreased to 6% (range = 3% - 8%) of intervals. Because the marks in the class were so low, the criterion was lowered approximately 80% from the previous subphase during C3 (to 1 mark), and the disruptive behavior average dropped to 4% (range = 3% - 7%) of intervals.

During the withdrawal phase, the percentage of intervals containing disruptive behaviors increased slightly to an average of 13% (range = 9% - 19%). Following this phase, the TC was re-implemented, and the occurrence of disruptive behaviors again decreased. During C4 the criterion was set at 5 marks, and the students' disruptive behavior decreased to an average of 5% (range = 4% - 8%) of intervals. For C5 a "mystery criterion" was introduced. Criteria each day fell between one and 10 marks. During this subphase disruptive behavior averaged 6% (range = 6% - 7%) of intervals. Unlike with Classrooms A and B, for Classroom C, the mystery criterion phase was the most stable. For C6 the criterion was lowered to one mark. The average of the disruptive behavior was 3% (range = 1% - 5%) of intervals during this subphase.

For Classroom C there were three overlapping data points between intervention phases and non-intervention phases; therefore the PND was 86.9%. Academic Engagement varied but remained higher during all intervention subphases: percentage of intervals for C1 was 85% (range = 79% - 88%), C2 was 86% (range = 72% - 97%), C3 was 80% (range = 67% - 93%), C4 was 89% (range = 84% - 93%), C5 was 83% (range = 79% - 86%), and C6 was 80% (range = 76% - 83%).

Teacher Acceptability

Following data collection the teachers of each classroom completed the modified *IRP-15*. Teacher A rated the GBG/TC as an 84, Teacher B1 responded with a rating of 76, Teacher B2 with a rating of 74, and Teacher C endorsed a rating of 60. Teacher A and C did not agree that the intervention was consistent with ones they had used in the past. Teacher C also slightly disagreed that he liked the procedures of the intervention or that he would be willing to use the intervention again in the classroom setting. Otherwise, all

three teachers agreed (to some extent) with all other statements on the *IRP-15*. Additionally, Teacher A noted that she was implementing the GBG/TC in class periods outside of her class referred for inclusion in the study and noted she was pleased with the results in those other classrooms.

CHAPTER IV

DISCUSSION

Although the effectiveness of the Good Behavior Game (GBG)/Teamwork Competition (TC) on managing a variety of behaviors has been investigated across many developmental levels, the high school population has not been adequately investigated. Studies have found the GBG to be effective across ages ranging from pre-school to adulthood. However, the present study is only one of three to utilize the procedure with a general education high school population (Kleinman & Saigh, 2011; Mitchell 2012). One reason for the smaller research base of implementation of the GBG/TC intervention in high school classrooms is that many (e.g., teachers, researchers) believe the procedure may be developmentally inappropriate. However, components of the GBG/TC are based on sound behavioral principles (i.e., differential reinforcement of low rates of behavior, clear expectations, feedback, and monitoring) which are not limited by age; therefore, the intervention may in fact be effective and developmentally appropriate for this age group with some modifications.

Research Questions

Research Question 1

The results of the present study are consistent with those of Kleinman and Saigh (2011) and Mitchell (2012) indicating that a modified version of the GBG can be effectively employed with high school students to decrease disruptive behavior. The data reflect clear and immediate decreases in disruptive behavior for all classrooms during intervention phases. Additionally, there was little to no overlap of the data for disruptive behavior during intervention phases with baseline and withdrawal phases. As in previous

literature, these data suggest that the GBG/TC can, in fact, effectively decrease the disruptive behavior of general education high school students, even those in upper grades (i.e., 12th grade), thereby affirming Research Question 1.

Research Question 2

In addition to evaluating the effect of the GBG/TC, the present study sought to examine the criterion component of the intervention and to determine if its level differentially influenced the level of disruptive behavior. In most cases the behavior of the students was far lower than the initial criterion; however, with the progression of the changing criterion, disruptive behavior levels typically tracked the criterion and were stable and below the level of the criterion, thereby affirming Research Question 2 and demonstrating that gradually making the criterion more stringent can produce additional decreases in disruptive behavior. One exception to this pattern was the mystery criterion subphase during re-implementation. While Classroom C responded well to this phase, Classrooms A and B showed greater variability in day-to-day disruptive behavior during this subphase. Observers noted that several students in both classes verbally requested that the teacher tell them in advance what the criterion would be.

Research Question 3

The effect of the GBG/TC on academic engagement was also investigated. In all three classrooms, intervals in which students were academically engaged increased during intervention phases, affirming Research Question 3. In two of the three classrooms (Classrooms A and B), there was little to no overlap of the data for academic engagement during intervention phases with baseline and withdrawal phases. There was, however, greater variability in academic engagement data than in disruptive behavior.

One reason for this variability is that the students typically finished their work much faster when the intervention was in place. Observers noted that during several sessions in Classroom A, students were observed quietly playing games or reading unrelated materials after finishing their classwork. Though the students were not engaged in disruptive behavior, they were also not academically engaged. Teacher A expressed excitement that the students were getting through their work faster but did not want that class to get too far ahead of the other classes due to lesson planning. She eventually implemented the GBG/TC in several of her other classes.

Research Question 4

With regards to the three classrooms, the mystery criterion seemed to be less favorable than a preset criterion. Observers noted that several students voiced displeasure with the use of this component and stated "just tell us what we need to get" when the teacher reiterated that the criterion would be revealed at the end of class. This concept is interesting when considering one of the basic principles of PBIS and many other behavioral systems recommend setting clear, consistent expectations for students. By implementing an unknown criterion, the students were not given the clarity of expectations of the other subphases. With the exception of Classroom C which remained consistently low throughout all intervention phases, the mystery criterion subphase lead to extremely variable data although it remained lower than baseline and withdrawal levels.

Research Question 5

Additionally, the present study sought to examine the acceptability of the GBG/TC by the teachers implementing the intervention. Although some teachers may

have reservations about the use of a technique originally designed to target disruptive behaviors in elementary aged students, the basic principles of the game still served to decrease the targeted behaviors of older students. However, because an intervention works or is feasible does not deem it acceptable or developmentally appropriate. Based on the results of a modified IRP-15, all three teachers felt that the GBG/TC was an overall acceptable intervention for use in their classrooms. Furthermore, two of the three teachers in the study noted that they employed the intervention in additional classes and reported satisfaction with those students' behavioral improvements.

Limitations

The present study supports the effectiveness and acceptability of the GBG with a general education high school population; however, some limitations should be noted. The study was conducted in a naturalistic setting (general education high school classrooms); therefore, there were many variables beyond the control of the study (i.e., variations in schedule, whole group/small group/independent desk work). Some variations in the school schedules included assemblies (e.g., pep rallies) and school-wide testing days which may have impacted attendance or class period length. Although none of these variations were present during phase changes, these variations may have affected the stability of some data.

Another possible limitation of the study includes the use of tangible rewards. The most popular reward included tangibles (i.e., candy or pencils) which are of a low monetary value. It is important to note that although this cost was low for the purpose of the current study, over the course of months or with multiple classrooms, costs could add up to a value that is not feasible or affordable to some teachers. Again, it is also important

to note that though the primary investigator purchased all tangible rewards for the referred classrooms, Teacher A elected to utilize the GBG/TC in other classrooms at her own expense and noted that she felt it was feasible for her.

Some potential limitations to external validity are also worth noting. The classrooms had different responses to the mystery criterion. It is unclear what caused these differences. However, with the exception of Classroom C, this subphase was the least stable with regard to disruptive behavior. Additionally, the majority of students included in the study were African American and of low SES. Therefore, it is not certain whether the game would be effective with students of differing backgrounds, higher SES, or higher grades (i.e., all/mostly eleventh or twelfth graders). Additionally, because of the nature of the study, all teachers who participated were voluntary and given the option to withdraw at any point (as one did prior to implementation). It is unknown what effects would have been rendered had a teacher been *forced* to implement such an intervention by school administrators.

Future Research

Although this study highlighted upper level high school students and the criterion component, another area that may be worth investigating is the type of rewards utilized in the GBG/TC. Although not monitored formally, teachers and observers noted that the students tended to prefer the edible rewards (candy) to the free rewards (i.e., extra points, homework passes) they requested during the preference assessment. Limiting the amount of a certain reward (e.g., only getting candy once a week) or only offering some of the preferred rewards may alter the results. Furthermore, as previously noted the students utilized in the present study were of low SES. An area of further investigation may be to

see whether these types of rewards or the GBG/TC in general, would be effective with students of higher SES who might readily be able to purchase the tangibles on their own. That is, due to an abolishing operation of already having greater access to tangibles, other types of rewards may be necessary for the intervention to be effective with higher SES students. Although the rewards for this population may need to be altered, the basic structure of the game would still be intact and would not change. It is also not known whether dividing the class into teams is necessary at the secondary level. If the GBG is effective without the team component, it might further streamline teachers' monitoring and record-keeping, and, thus, time involvement with the procedure.

Implications

The present study contributes to the current literature in a number a ways. Consistent with previous studies, this study serves as evidence that group contingencies may be viable techniques for managing inappropriate behaviors and increase in academic engagement in high school classrooms of all levels, and that the GBG/TC is a viable option to incorporate within a PBIS system at the secondary level. High school teachers, administrators, and school psychologists are often searching for appropriate Tier I and Tier II management procedures at the high school level. The present study demonstrates that the GBG/TC may serve as an effective procedure to use with high school students at either a Tier I or Tier II level and that altering the criterion can help to reduce disruptive behaviors further. In addition to demonstrating the effectiveness of the GBG/TC, the study also demonstrated that in addition to being an effective method for managing disruptive behavior, the GBG/TC is also deemed acceptable by high school teachers who participated in the study.

APPENDIX A

TEACHER CONSENT FORM

Dear Teacher,

I am a doctoral student in the School Psychology Program at The University of Southern Mississippi working under the guidance of Dr. Daniel Tingstrom. As part of my dissertation, I am researching the effectiveness of a classroom-based intervention, the Good Behavior Game (GBG), a procedure used to decrease disruptive behavior. Your classroom has been referred for class wide disruptive behavior, which the intervention aims to address, therefore we hope you will participate in the study.

If you agree to participate in this study, we will ask you to perform several tasks. First, prior to the implementation of the GBG, you will be asked to complete a consultation session with me to obtain information regarding your students' behavioral concerns. Following this consultation, a screening procedure will be conducted to verify your classroom's capacity for participation. If your classroom qualifies for participation, I will conduct a training session to explain and practice the steps of the intervention with you prior to implementation. The GBG is an intervention in which two teams compete to obtain the fewest amount of check marks for disruptive classroom behavior. Low numbers of disruptive behavior will enable students on the winning team to gain access to rewards. In order to participate in the study, your classroom must demonstrate disruptive behavior in at least 30% of the observation intervals at the time of the screening session in order qualify for the study. If the classroom does not qualify for participation other services will be made available to you.

Throughout the study, classroom observations will be conducted multiple times a week by myself or another trained graduate student from the USM School Psychology program. The study will consist of two phases. Following the initial screening observation, data will be collected on the targeted disruptive behavior. At this time, you will conduct class as normal without the implementation of the GBG. During the second phase, the GBG will be implemented in the classroom. The game will consist of dividing the students into two teams and marking points against a team each time a team member performs and inappropriate behavior or breaks a classroom rule. At the end of each day the team with the least marks against them, or both teams if neither exceeds a pre-set criterion, will win the game and will earn access to an approved reinforcer.

Following each day of observations, you will be provided with feedback on the game implementation. At the end of the study, you and your students will be asked to complete a questionnaire to assess your satisfaction with the GBG.Agreeing to participate in this study may offer several benefits for you and your students. By participating in this study you will be trained on the implementation of a new intervention technique that can be used with other students. An additional benefit is the expected decrease in inappropriate behaviors and the increased appropriate behaviors by your students.

Students' behavior will be monitored to ensure undesired effects (e.g., increase in inappropriate behaviors) do not happen. Should we observe any unanticipated effects on your students' behavior, modifications or discontinuation of the intervention will occur, and your students will be provided with other appropriate services. There appear to be very few risks for either you or your students participating in this study. The greatest discomfort for you may be related to implementing a new procedure in the classroom. To reduce discomfort, I and/or other trained graduate students will provide training, materials, and will be available to answer any questions you may have. Your students should not experience any discomfort from the implementation of the recommended intervention.

All interviews, observations, and other information obtained during this study will be kept strictly confidential. Your name, students' names, and other identifying information will not be disclosed to any person not connected with this study. Results from this research project may be shared at professional conferences or published in scholarly journals; however, all identifying information will be removed from publications and/or presentations. Your participation in this study is entirely voluntarily. In addition, you may withdraw from this study at any time without penalty, prejudice, or loss of benefits. Further services, if needed, may be provided outside the scope of this study. Whereas no assurance can be made concerning results that may be obtained (as results from investigational studies cannot be predicted) the researcher will take every precaution consistent with the best scientific practice.

If you agree to participate, please read, sign, and return the following page. Please keep this letter for your records. If you have any questions about this study, please contact Rachel R. Mitchell at (228.327.2005; Rachel.Mitchell@eagles.usm.edu) or Dr. Daniel Tingstrom (601.266.4594; Daniel.Tingstrom@usm.edu). This project and this consent form have been reviewed by the Human Subjects Protection Review Committee at USM, which ensures that research projects involving human subjects follow federal regulations. Any questions or concerns about rights as a research subject should be directed to the Institutional Review Board Office, The University of Southern Mississippi, Box 5147, Hattiesburg, MS 39406-5147, (601) 266-6820.

Sincerely,

_____

Rachel R. Mitchell, M.A.
School Psychologist in Training

THIS SECTION TO BE COMPLETED BY TEACHER

Please Read and Sign the Following:

I have read the above documentation and consent to participate in this project. I have had the purpose and procedures of this study explained to me and have had the opportunity to ask questions. I am voluntarily signing this form to participate under the conditions stated. I have also received a copy of this consent. I understand that I will be asked to implement a classroom-based intervention called the Good Behavior Game, and observations will be conducted in the classroom on the students' behavior. In order to do so, I will be required to complete a consultation session, to implement the intervention, and to complete a structured questionnaire to assess my satisfaction with the intervention. In addition, I will be trained on all of the intervention procedures by the primary experimenter. I further understand that all data collected in this study will be confidential and that my name and the students' names will not be associated with any data collected. I understand that I may withdraw my consent for participation at any time without penalty, prejudice, or loss of privilege.

_____ _____

Signature of Teacher *Date*

_____

Signature of Witness

APPENDIX B

IRB APPROVAL

THE UNIVERSITY OF
SOUTHERN MISSISSIPPI

INSTITUTIONAL REVIEW BOARD
118 College Drive #5147 | Hattiesburg, MS 39406-0001
Phone: 601.266.6820 | Fax: 601.266.4377 | www.usm.edu/irb

NOTICE OF COMMITTEE ACTION

The project has been reviewed by The University of Southern Mississippi Institutional Review Board in accordance with Federal Drug Administration regulations (21 CFR 26, 111), Department of Health and Human Services (45 CFR Part 46), and university guidelines to ensure adherence to the following criteria:

- The risks to subjects are minimized.
- The risks to subjects are reasonable in relation to the anticipated benefits.
- The selection of subjects is equitable.
- Informed consent is adequate and appropriately documented.
- Where appropriate, the research plan makes adequate provisions for monitoring the data collected to ensure the safety of the subjects.
- Where appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of all data.
- Appropriate additional safeguards have been included to protect vulnerable subjects.
- Any unanticipated, serious, or continuing problems encountered regarding risks to subjects must be reported immediately, but not later than 10 days following the event. This should be reported to the IRB Office via the "Adverse Effect Report Form".
- If approved, the maximum period of approval is limited to twelve months. Projects that exceed this period must submit an application for renewal or continuation.

PROTOCOL NUMBER: **12092502**
PROJECT TITLE: **Evaluating the Effectiveness of the Good Behavior Game with General Education High School Students Utilizing a Changing Criterion Component**
PROJECT TYPE: **Dissertation**
RESEARCHER/S: **Rachel R. Mitchell**
COLLEGE/DIVISION: **College of Education & Psychology**
DEPARTMENT: **School Psychology**
FUNDING AGENCY: **N/A**
IRB COMMITTEE ACTION: **Expedited Review Approval**
PERIOD OF PROJECT APPROVAL: **10/02/2012 to 10/01/2013**

Lawrence A. Hosman, Ph.D.
Institutional Review Board Chair

APPENDIX C

TEACHER SCRIPT

1) Introduction of the Teamwork Competition
    - Inform students that there will now be a team competition each day during the set class period. At this time, students are expected to follow all of the classroom rules.

2) State and demonstrate class expectations
    - Remind the class of each classroom rule. If the target behaviors are not a part of the classroom rules, those should also be explained. The teacher should demonstrate the expected appropriate behaviors for the class to see.

3) Explain Competition procedures and divide the class into teams
    - Divide the students into two teams and write the names on the board. Explain the kinds of behaviors that will result in marks against teams. Teams will be informed of the reinforcers that will be awarded to winning teams and that both teams may be able to earn the rewards by staying under the set criterion. On days in which there is no criterion, only one team will have the opportunity to win.

4) Following the introduction to the class, the Teamwork Competition will immediately begin
    - Disruptive behaviors and those that go against the classroom rules will immediately begin earning marks for teams.

5) End the competition and award the winning team(s)
    - At the end of the game each day the points will be tallied and the winner(s) announced

APPENDIX D

INTERVENTION RATING PROFILE-15 (IRP-15)/MODIFIED VERSION

Please respond to each of the following statements thinking about the intervention implemented. Please then circle the number associated with your response. Be sure to answer all statements.

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
This was an acceptable intervention for the problem behavior(s).	1	2	3	4	5	6
Most teachers would find this intervention appropriate for behavior problems in addition to the ones described.	1	2	3	4	5	6
This intervention proved effective in helping to change the problem behavior(s) of the classroom.	1	2	3	4	5	6
I would suggest the use of this intervention to other teachers.	1	2	3	4	5	6
The classroom behavior problem was severe enough to warrant the use of this intervention.	1	2	3	4	5	6
Most teachers would find this procedure suitable for the problem behavior(s) described.	1	2	3	4	5	6
I would be willing to use the intervention again in the classroom setting.	1	2	3	4	5	6
The intervention did *not* result in negative side effects for the students.	1	2	3	4	5	6
This intervention would be appropriate for a variety of students.	1	2	3	4	5	6
This intervention was consistent with those I have used in the classroom setting before.	1	2	3	4	5	6
This intervention was a fair way to handle problem behavior in the classroom.	1	2	3	4	5	6
This intervention was reasonable for the problem behavior(s) described.	1	2	3	4	5	6
I liked the procedures used in this intervention.	1	2	3	4	5	6
The intervention was a good way to handle the behavior problem(s).	1	2	3	4	5	6
Overall, this intervention was beneficial.	1	2	3	4	5	6

Taken and adapted from, Martens, B.K., Witt, J.C., Elliott, S.N. & Darveaux, D. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research and Practice, 16*, 191-198

APPENDIX E

OBSERVATION SHEET

Teacher name: _____ Date: _____ Observer name: _____

Interval	IV	OOS	IT	AE		Interval	IV	OOS	IT	AE
1.1						11.1				
1.2						11.2				
1.3						11.3				
1.4						11.4				
1.5						11.5				
1.6						11.6				
2.1						12.1				
2.2						12.2				
2.3						12.3				
2.4						12.4				
2.5						12.5				
2.6						12.6				
3.1						13.1				
3.2						13.2				
3.3						13.3				
3.4						13.4				
3.5						13.5				
3.6						13.6				
4.1						14.1				
4.2						14.2				
4.3						14.3				
4.4						14.4				
4.5						14.5				
4.6						14.6				
5.1						15.1				
5.2						15.2				
5.3						15.3				
5.4						15.4				
5.5						15.5				
5.6						15.6				
6.1						16.1				
6.2						16.2				
6.3						16.3				
6.4						16.4				
6.5						16.5				
6.6						16.6				
7.1						17.1				
7.2						17.2				
7.3						17.3				
7.4						17.4				
7.5						17.5				
7.6						17.6				
8.1						18.1				
8.2						18.2				
8.3						18.3				
8.4						18.4				
8.5						18.5				
8.6						18.6				
9.1						19.1				
9.2						19.2				
9.3						19.3				
9.4						19.4				
9.5						19.5				
9.6						19.6				
10.1						20.1				
10.2						20.2				
10.3						20.3				
10.4						20.4				
10.5						20.5				
10.6						20.6				

Occurrence of IV = ___/120= ___%
Occurrence of OOS = ___/120= ___%
Occurrence of IT = ___/120= ___%

Occurrence of Total DB =___/120= ___%
Occurrence of AE =___/120= ___%

Inappropriate vocalizations was defined as any voluntary audible verbalization made without teacher permission such as speaking, yelling, humming, singing, and/or whispering.
Out-of-seat behavior was defined as the student's buttocks breaking contact with his or her seat for more than three seconds without teacher permission.
Inappropriate touching was defined as the student touching another student at any point during the interval including hitting, pushing, poking, hugging, rubbing, flicking, or any other contact between two students that is not pertinent to the assignment.
Academic engagement was defined as the student being actively involved or attending to (e.g. looking at or writing on) independent seatwork, teacher instruction, designated classroom activities, and/or engaging in task related (permissible) vocalizations with teachers and/or peers

APPENDIX F

BASELINE/WITHDRAWAL PROCEDURAL INTEGRITY CHECKLIST

Teacher Name:_____ Date:_____

Observer:_____

Training Steps	✓
Class not divided into teams	
Teams not mentioned	
Rewards not mentioned	
Rewards not given	
Steps Completed	
Steps Possible	

Percentage of steps completed: _____

APPENDIX G

INTERVENTION PROCEDURAL INTEGRITY CHECKLIST

Teacher Name:_____ Date:_____ Observer:_____

Training Steps	**Yes**	**No**
1. Teacher announces/reminds students of the game/rules		
2. Rules are displayed so that students can see them		
3. Students are divided into teams (visual inspection)		
4. Team names are displayed on board where they can be seen by students		
5. Teacher reminds students of the daily criterion for both teams to win, or that there is a mystery criterion.		
6. Displays/References the appropriate pre-set criterion		
7. Teacher identifies/records disruptive behavior as marks on the board against teams		
8. Teacher announces when the game is over		
9. Teacher announces winning team(s)		
10. Teacher allows winning team to access reward		
Steps Completed		
Steps Possible		

Percentage of steps completed: _____

Teacher requires retraining: Yes No

*Taken and adapted from, Hunt, B. M. (2010). *The good behavior game with a preschool population* (Unpublished master's thesis). The University of Southern Mississippi, Hattiesburg, MS.

APPENDIX H

STUDENT PREFERENCE ASSESSMENT

In the space below, please list at least three rewards you would like to earn from (teacher name) in (class name) that would make you work harder and behave better in class. List as many things as you would be willing to "work" to earn (for example free time, extra points on a test, candy, pencils, etc.).

1)

2)

3)

Other:

REFERENCES

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

Bostow, D., & Geiger, O. G. (1976). Good behavior game: A replication and systematic analysis with a second grade class. *School Applications of Learning Theory, 8,* 18-27.

Canter, A. S., Paige, L. Z., Roth, M. D., Romero, I., & Carroll, S. A. (Eds.). (2004). Helping children at home and school II: Handouts for families and educators. Bethesda, MD: National Association of School Psychologists.

Darch, C. B. & Thorpe, H. W. (1977). The principal game: A group consequence procedure to increase classroom on-task behavior. *Psychology in the Schools, 14,* 341-347.

Darveaux, D. X. (1984). The good behavior game plus merit: Controlling disruptive behavior and improving student motivation. *School Psychology Review, 13,* 510-514.

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

Dolan, L. J., Kellam, S. G., Brown, C. H., Werthamer-Larson, L., Rebok, G. W., Mayer, L. S., et al. (1993). The short-term impact of two classroom-based preventive interventions on aggressive and shy behaviors and poor achievement. *Journal of Applied Developmental Psychology, 14,* 317-345.

Evertson, C. M., & Emmer, E. T. (1982). Effective management at the beginning of the school year in junior high classes. *Journal of Educational Psychology, 74(4),* 485-498.

Fleiss, J. L., Levin, B., & Cho Paik, M. (2003). *Statistical methods for rates & proportions*. Hoboken, NJ: John Wiley & Sons, Inc.

Franklin, C., Haris, M. B., & Allen-Meares, P. (Eds.). (2006). *The school services sourcebook: A guide for school-based professionals.* New York, NY: Oxford University Press.

Freer P., & Watson, T. S. (1999). A comparison of parent and teacher acceptability ratings of behavioral and conjoint behavioral consultation. *School Psychology Review*, *28,* 672.

Gresham, F. M., & Gresham, G. N. (1982). Interdependent, dependent, and independent group contingencies for controlling disruptive behavior. *Journal of Special Education, 16,* 101-110.

Harris, V. W., & Sherman, J. A. (1973). Use and analysis of the "good behavior game" to reduce disruptive classroom behavior. *Journal of Applied Behavior Analysis, 6*, 405-417.

Hegerle, D. R., Kesecker, M. P., & Couch, J. V. (1979). A behavior game for the reduction of inappropriate classroom behaviors. *School Psychology Digest*, 8, 339-343.

Hunt, B. M. (2010). *The good behavior game with a preschool population* (Unpublished master's thesis). The University of Southern Mississippi, Hattiesburg, MS.

Hunt, B. M. (2012). *Using the good behavior game to decrease disruptive behavior while increasing academic engagement*. (Unpublished doctoral dissertation). The University of Southern Mississippi, Hattiesburg, MS.

Johnson, M. R., Turner, P. F., & Konarski, E. A. (1978). The "good behavior game": A systematic replication in two unruly transition classrooms. *Education and Treatment of Children, 1*, 25-33.

Kazdin, A. E. (1982). *Single-case research design: Methods for clinical and applied settings*. New York, NY: Oxford University Press.

Kellam, S. G., & Anthony, J. C. (1998). Targeting early antecedents to prevent tobacco smoking: Findings from an epidemiologically based randomized field trial. *American Journal of Public Health, 88*, 1490-1495.

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & 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.

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. *Journal of Child Psychology and Psychiatry, 35*, 259-281.

Kleinman, K. E., & Saigh, P. A. (2011). The effects of the good behavior game on the conduct of regular education New York City high school students. *Behavior Modification, 35*, 95-105.

Kounin, J. (1970). *Discipline and group management in classrooms*. New York, NY: Holt, Rinehart, & Winston.

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

Lutzker, J. R., & White-Blackburn, G. (1979). The good productivity game: Increasing work performance in a rehabilitation setting. *Journal of Applied Behavior Analysis, 12*, 488.

Maloney, K. B., & Hopkins, B. L. (1973). The modification of sentence structure and its relationship to subjective judgments of creativity in writing. *Journal of Applied Behavior Analysis, 6,* 425-433.

Martens, B.K., Witt, J.C., Elliott, S.N., & Darveaux, D. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology*: *Research and Practice, 16*, 191-198.

McCurdy, B. L., Lannie, A. L., & Bardabas, E. (2009). Reducing disruptive behavior in an urban school cafeteria: An extension of the good behavior game. *Journal of School Psychology, 47*, 39-54.

Medland, M. B., & Stachnik, T. J. (1972). Good-behavior game: A replication and systematic analysis. *Journal of Applied Behavior Analysis, 5*, 45-51.

Milner, H. R., & Tenore, F. B. (2010). Classroom management in diverse classrooms. *Urban Education*, 5, 560-603.

Mitchell, R. R. (2012). *The effects of the good behavior game with general education high school students* (Unpublished master's thesis). The University of Southern Mississippi, Hattiesburg, MS.

Mudgal, D. (2004). *Good classwork game: Efficacy of an interdependent group contingency on mathematics seatwork*. (Unpublished master's thesis). The University of Southern Mississippi, Hattiesburg, MS.

Mudgal, D. (2006). *Randomizing criteria within an interdependent group contingency: Effects on math seatwork*. (Unpublished dissertation). The University of Southern Mississippi, Hattiesburg, MS.

Patrick, C. A., Ward, P., & Crouch, D. W. (1998). Effects of holding students accountable for social behaviors during volleyball games in elementary physical education. *Journal of Teaching in Physical Education, 17*, 143-156.

Poduska, J.M., Kellam, S.G., Wang, W., Hendricks, B.C., Ialongo, N.S., & Toyinbo, P. (2008). Impact of the good behavior game, a universal classroom-based behavior intervention, on young adult service use for problems with emotions, behavior, or drugs or alcohol. *Drug and Alcohol Dependence, 95,* S29-S44.

Reupert, A., & Woodcock, S. (1970). Success and near misses: Pre-service teachers' use, confidence and success in various classroom management strategies. *Teaching and Teacher Education, 6,* 1261-1268.

Robertshaw, C. S. & Hiebert, H. D. (1973). The astronaut game: A group contingency applied to a first grade classroom. *School Applications of Learning Theory, 6,* 28-33.

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

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

Sayeski, K. L., & Brown, M. R. (2011). Developing a classroom management plan using a tiered approach. *TEACHING Exceptional Children, 1*, 8-17

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

Shaw, R., & Simms, T. (2009). Reducing attention-maintained behavior through the use of positive punishment, differential reinforcement of low rates, and response marking. *Behavioral Interventions, 24,* 249-263.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., Scott, T., Liaupsin, C., Sailor, W., Turnbull, A. P., Turnbull III, H. R., Wickham, D., Wilcox, B., & Ruef, M. (2000). Applying positive behavioral assessment in schools. *Journal of Positive Behavior Interventions, 2,* 131-143.

Sugai, G., & Horner, R. H. (2002). Introduction to the special series on positive behavior support in schools. *Journal of Emotional and Behavioral Disorders, 10*, 130-135.

Sugai, G., Lewis-Palmer, T., Todd, A. & Horner, R. (2005). School-wide Evaluation Tool (version 2.1). Educational and Community Supports, University of Oregon.

Swain, J. J., Allard, G. B., & Holborn, S. W. (1982). The good toothbrushing game: A school-based dental hygiene program for increasing the toothbrushing effectiveness of children. *Journal of Applied Behavior Analysis, 15*, 171-176.

Swiezy, N. B., Matson, J. L., & Box, P. (1992). The good behavior game: A token reinforcement system for preschoolers. *Child and Family Behavior Therapy, 14,* 21-32.

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

Tillery, A. D., Varjas, K., Meyers, J., & Collins, A. S. (1992). General education teachers' perceptions of behavior management and intervention strategies. *Journal of Positive Behavior Interventions, 12*, 86-102.

Tingstrom, D. H., Sterling-Turner, H. E., & Wilczynski, S. M. (2006). The good behavior game: 1969-2002. *Behavior Modification, 30,* 225-253.

Von Brock, M. B., & Elliottt, S. N. (1987). Influence of treatment effectiveness information on the acceptability of classroom interventions. *Journal of School Psychology, 25*, 131–144.

Watkins, M. W., & Pacheco, M. (2000). Interobserver agreement in behavioral research: Importance and calculation. *Journal of Behavioral Education, 10*(4), 205-212.

Westling, D. L. (2010). Teachers and challenging behavior: knowledge, views, and practices. *Remedial & Special Education, 31*, 48-63.

Witt, J. C., & Elliot, S. N. (1985). Acceptability of classroom intervention strategies. In T. R. Krotochwill (Ed.), *Advances in School Psychology, 4,* 251-288